Structural and Functional Characterization of Plant and Fungal LOV Proteins

Jameela Lokhandwala
Southern Methodist University, jlokhandwala@smu.edu

Follow this and additional works at: https://scholar.smu.edu/hum_sci_chemistry_etds

Part of the Biochemistry, Biophysics, and Structural Biology Commons

Recommended Citation
https://scholar.smu.edu/hum_sci_chemistry_etds/5

This Thesis is brought to you for free and open access by the Chemistry at SMU Scholar. It has been accepted for inclusion in Chemistry Theses and Dissertations by an authorized administrator of SMU Scholar. For more information, please visit http://digitalrepository.smu.edu.
STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF PLANT AND FUNGAL LOV PROTEINS

Approved by:

_____________________________
Dr. Brian D. Zoltowski
Professor of Chemistry

_____________________________
Dr. Nicolay Tsarevsky
Professor of Chemistry

_____________________________
Dr. Patty Wisian Neilson
Professor of Chemistry

_____________________________
Dr. Peng Tao
Professor of Chemistry

_____________________________
Dr. Steven B Vik
Professor of Biology
STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF PLANT AND FUNGAL LOV PROTEINS

A Dissertation Presented to the Graduate Faculty of

Dedman College
Southern Methodist University

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

with a
Major in Chemistry

by

Jameela Lokhandwala

Bachelor’s in Biotechnology, Jeppiaar Engineering College, Chennai, India

May 19, 2018
The plant blue light photoreceptor proteins Flavin binding Kelch F-box protein-1 (FKF1) and LOV Kelch protein-2 (LKP2) play an important role in maintaining circadian rhythmicity and photo-periodic flowering responses. These proteins along with light sensitive Light Oxygen Voltage (LOV) domain contain F-box and Kelch repeat domains. Therefore, these proteins control blue light mediated degradation of various protein targets to regulate circadian rhythmicity, growth, and photo-periodic flowering. Blue-light excitation of these proteins leads to formation of a cysteinyl-flavin adduct, which then decays depending upon various factors. Therefore, spectral studies were conducted to measure the kinetics of light–dark recovery. Size Exclusion Chromatography (SEC) was performed to determine its oligomeric state. Attempts to crystallize these proteins to understand the signal transduction landscape were unsuccessful. Although crystal structure information of Zeitlupe a protein belonging to the same family provided by Ashutosh Pudasaini gave useful insights about potential signaling mechanisms that could apply to FKF1 and LKP2. These results collectively aid to develop a model of LOV domain function
We also studied ENV1 a blue light photoreceptor protein from the industrially important fungus *Trichoderma reseei*. ENV1 is responsible for blue-light mediated regulation of gene transcription, growth, reproduction and stress responses along with another blue light photoreceptor Blue light receptor-1 (BLR1). Previous studies of a homologous protein VVD that regulates circadian rhythmicity in fungus *Neurospora crassa* provides useful insight into blue-light dependent signaling mechanisms. Biophysical characterization of ENV1 using techniques such as size-exclusion chromatography, multi-angle light scattering, and X-Ray crystallography shows that it controls gene transcription through competitive hetero- and homo-dimer formation. Further, it was observed that ENV1 can respond to oxidative stress conditions in addition to blue light sensitivity. Integration of blue light signaling, and oxidative stress responses were confirmed by *in-vivo* studies. Thus, ENV1 has evolved from its homologous protein VVD to incorporate oxidative stress sensing. Sequence alignment results indicate that it can also serve as a model to understand physiology of pathogenic fungus that affect crops. Thus, key findings of this research will aid in understanding signal transduction pathways and development of plant and fungal variants with desirable traits. Further, since these proteins are sensitive to blue light, they can be utilized for opto-genetic applications.
# TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

LIST OF ABBREVIATIONS xiv

CHEMICAL STRUCTURES xvi

ACKNOWLEDGEMENTS xviii

CHAPTER 1
PAS DOMAIN STRUCTURE AND FUNCTION 1

1.1 Introduction 1

1.2 PAS nomenclature 2

1.3 PAS structure 2

1.4 PAS dimerization 4

1.5 PAS domain diversity 6

1.5.1 PAS domains in prokaryotes 7

1.5.2 PAS domains in eukaryotes 9

1.6 PAS classification based on various cofactors 10

1.6.1 Heme binding PAS domains 10

1.6.2 Flavin binding PAS domains 11

1.6.3 Carboxylic acid binding PAS domains 12

1.6.4 Divalent metal binding PAS domains 13

1.7 PAS-like proteins 13

1.8 PAS classification based on structure 14

1.8.1 Isolated PAS domains 15

1.8.1.1 Photoactive yellow protein 15
1.8.2. Single PAS domains

1.8.2.1 Ytva signaling

1.8.3. Multiple PAS domains

1.8.3.1 Phototropin signaling

1.9 Specific research goals

REFERENCES

CHAPTER 2
PLANT CIRCADIAN CLOCK AND CHARACTERIZATION OF LOV PROTEINS LKP2 AND FKF1

2.1 Circadian clock

2.2 Vertebrate clock

2.3 Plant circadian clock

2.3.1 ZTL family and plant circadian clock

2.3.2 Photochemical characterization of plant LOV domains

2.3.3 Photo-cycle kinetics of FKF1 and LKP2

2.3.4 Photo-cycle kinetics of rate altering variants

2.4 Material and methods

2.4.1 Cloning and protein purification of LKP2 and FKF1 proteins

2.4.3 UV-Vis absorbance spectroscopy and kinetics

2.4.4 Base catalysis studies

REFERENCES

CHAPTER 3
F-BOX PROTEINS AND CHARACTERIZATION OF PLANT F-BOX PROTEINS FKF1 AND LKP2

3.1 Introduction to the F-box protein family

3.2 Mechanism of protein degradation by SCF complexes

3.3 F-box and substrate recruitment
3.4 F-box dimerization  75
3.5 Structural characterization of LKP2 and FKF1  77
3.6 Materials and methods  84
  3.6.1 Cloning and protein purification of LKP2 and FKF1 proteins  84
  3.6.2 Size-exclusion chromatography:  85
  3.6.3 Determination of dissociation constants (K_d)  86
REFERENCES  88

CHAPTER 4
STRUCTURAL CHARACTERIZATION OF LOV PHOTORECEPTOR ENV  93

  4.1 Introduction to the fungal circadian clock  93
  4.2 Crystal structure of ENV1  99
    4.2.1 Crystallographic dimer interface  103
    4.2.2 ENV-64 dark state dimer resembles VVD light state dimer  105
  4.3 Size Exclusion Chromatography of ENV1  108
  4.4 In-vivo significance of disulfide bond formation  113
  4.5 A Competitive dimer model of LOV function in filamentous fungi is modified to sense oxidative stress:  115
  4.6 Cys96 Is conserved in closely related sordariomycetes  118
  4.7 Materials and methods  121
    4.7.1 Cloning and purification  121
    4.7.2 Structural analysis  122
    4.7.3 Size-exclusion chromatography  122
    4.7.4 Spectroscopy and kinetics  123
    4.7.5 Disulfide reduction potential  123
    4.7.6 Strains and cultivation conditions  124
    4.7.7 Construction of ENV1-C96T complementation strains in T. reesei  125
    4.7.8 Isolation of total RNA and qRT-PCR  125
CHAPTER 5
CHARACTERIZATION OF ENV1 AND VVD MUTANTS

5.1 Solution characterization of ENV1 S99A mutant
5.2 Characterization of ENV1 V98C mutant
5.3 Importance of HI loop in signaling
5.4 A72S mutation blocks VVD dimerization
5.5 Materials and methods
   5.5.1 Cloning and purification
   5.5.2 Size exclusion chromatography
   5.5.3 Determination of dissociation constants (K_d)
   5.5.4 Spectroscopy and kinetics
   5.5.5 Structural analysis
REFERENCES

CHAPTER 6
ENV1 PHOTOCHEMISTRY AND OSMOTIC STRESS SIGNALING

6.1 ENV Photochemistry
6.2 Photocycle kinetics of ENVT101I
6.3 Computational analysis for water occupancy at FMN active site
6.4 Effect of T101I mutation in-vivo
6.5 T101 is involved in tolerance of osmotic stress in T. reesei
6.6 Materials and methods
   6.6.1 Cloning and protein purification
   6.6.2 Spectroscopy and kinetics
   6.6.3 Molecular dynamics simulations
   6.6.4 Strains
6.6.5 Stress response 172
6.6.6 Construction of ENV1-T101I complementation strains in T. reesei 173

REFERENCES 174

APPENDIX

A.1 Peroxide production in LOV proteins 179
A.2 Photochemical Characterization of ENV and VVD mutants 182
LIST OF TABLES

Table 1: Kinetic recovery of rate altering variants of FKF1: 54
Table 2: Kinetic recovery of rate altering variants of LKP2: 57
Table 3: Dissociation constants of ZTL family: 82
Table 4: Kinetics of thermal reversion for ENV1 constructs and variants at 296 K. 160
LIST OF FIGURES

Figure 1: Schematic diagram of a PAS domain fold: 2

Figure 2: PAS classification based on structure 15

Figure 4: Sequence Alignment of phototropin LOV2 domain with other LOV proteins from plant and fungal species: 20

Figure 3: FMN binding in phototropin LOV2 domain: 20

Figure 5: Three components of a circadian clock: 33

Figure 6: Parameters of circadian rhythm: 37

Figure 7: ZTL family of proteins: 40

Figure 8: Plant circadian clock: 42

Figure 9: Mechanism of adduct formation. 46

Figure 10: Photo-cycle kinetics of FKF1, ZTL, and LKP2: 49

Figure 9: Mechanism of adduct formation. 50

Figure 11: Photo-cycle kinetics of rate altering variants of FKF1: 53

Figure 12: Base Catalysis Studies of LKP2 WT and LKP2 L153Q: 56

Figure 14: SCF E3 ubiquitin ligase: 73

Figure 15: Size exclusion studies of dimer disruption mutants: 81

Figure 16: Competitive dimer formation: 95

Figure 17: Dimer Formation: 96
Figure 18: LOV domain: 99
Figure 19: ENV1 Photocycle and crystal structure: 101
Figure 20: Crystal structure of ENV1: 102
Figure 21: N-terminal forms important dimer contacts: 104
Figure 22: Comparative Studies of VVD ENV dimer: 105
Figure 23: Sequence alignment of ENV1 and VVD: 107
Figure 24: Comparison of the ENV-64 crystal structure with VVD: 107
Figure 25: Size exclusion chromatography of ENV1 FL and N-terminal truncated constructs: 109
Figure 26: Cys96 mediates disulfide-linked dimer formation 110
Figure 27: Calculation of Midpoint Reduction potential of disulfide-linked dimer formation: 112
Figure 28: C96T mutants are able to complement growth defects in high light and darkness: 114
Figure 29: C96T mutant is incapable of rescuing growth defects in low light: 115
Figure 30: Model of divergent signaling in Trichoderma ressei: 117
Figure 31: Sequence alignment of selected Sordariomycetes: 119
Figure 32: Evolutionary Relationships of ENV1 Homologs in Sordariomycetes: 120
Figure 40: Superposition of VVD WT dark and VVD A72S: 142
Figure 43: Kinetics and spectra of ENV1 proteins: 155
Figure 44: Eyring plots and T101I kinetics: 159
Figure 45: Imidazole catalysis of ENV1: 159
Figure 46: Evolutionary relationships of fungal ENV1 homologs. 160
Figure 47: Wild type ENVOY active site: 163
Figure 48: Hyphal extension under different condition: 168
Figure 49: Peroxide production in LOV proteins: 181
Figure 50: Photocycle kinetics of ENV and VVD mutants: 183
LIST OF ABBREVIATIONS

AHR (Aryl hydrocarbon receptor)
ARNT (Aryl hydrocarbon receptor nuclear translocator)
ASK2 (Arabidopsis Skp1 like protein 2)
bHLH (Basic Helix loop helix)
BLR1 (Blue light receptor-1)
BLR2 (Blue light receptor-2)
BLUF (Blue light using FAD)
BMAL (Brain and Muscle ARNT like-1)
βTrCP (Beta transducin repeat containing protein)
CACHE (Calcium channels and Chemotaxis receptors)
CCA1 (Circadian clock associated protein-1)
CDF1 (Cyclic DOF Factor1)
CDK1 (Cyclin dependent kinase-1)
CKS1 (Cyclin dependent kinase regulatory subunit-1)
CLOCK (Circadian Locomotor Output Cycles Kaput)
CO (Constans)
CRY (Cryptochrome)
dCLK (Drosophila CLOCK)
dCYC (Drosophila Cycle)
DTT (Dithiothreitol)
ENV1 (Envoy)
FAD (Flavin adenine dinucleotide)
FBXL (F-box Leucine rich repeat protein)
FBXW (F-box WD repeat domain)
FBXO (F-box only protein)
FBA (F-box Associated domain)
FKF1 (Flavin binding kelch repeat F-box protein-1)
FMN (Flavin mononucleotide)
FRQ (Frequency)
FRH (Frequency interacting RNA helicase)
FT (Flowering Time gene)
GI (Gigantea)
HIFα (Hypoxia interacting factor-α)
hPASK (Human PAS kinase)
HTH (Helix turn helix)
H-NOXA (Heme-Nitric oxide/Oxygen binding A)
IRP2 (Iron regulatory protein-2)
IkBα (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha)
LHY (Late Elongated Hypocotyl)
LKP2 (LOV domain containing kelch repeat protein-2)
LOV (Light Oxygen Voltage)
LRR (Leucine rich repeats)
NifL (Nitrogen fixing gene L)
NPH1 (nuclear factor kappa-light-chain-enhancer of activated B cells)
PER (Period)
PAS (Period Aryl hydrocarbon receptor nuclear translocator Single minded)
PDB (Protein Data Bank)
PCR (Polymerase Chain Reaction)
PRR5 (Pseudo response regulator 5)
PYP (Photo-active yellow protein)
PhyA (Phytochrome A)
PIF (Phytochrome Interacting Factor-3)
RHT (Retinal Hypothalamic tract)
Rbx1 (Ring box protein-1)
SCF (Skp1-cullin-Rbx1)
SCN (Super chiasmatic nucleus)
SEC (Size Exclusion Chromatography)
Skp1 (S-phase kinase associated protein-1)
TIM (Timeless)
TOC-1 (Timing of CAB expression -1)
UIM (Ubiquitin Interacting motif)
WC-1 (White Collar-1)
WC-2 (White Collar-2)
ZT (Zeitgeber)
ZTL (Zeitlupe)
CHEMICAL STRUCTURES

Flavin Mononucleotide (FMN)

Flavin Adenine Dinucleotide (FAD)
ACKNOWLEDGEMENTS

I feel my words are not enough to express my gratitude to my research advisor Prof Brian D. Zoltowski. I am extremely grateful to Dr. Zoltowski for his constant support and guidance throughout the graduate program and helping me become a confident and independent researcher. I would like to extend special thanks to my colleagues Aditi Nagar and Ashutosh Pudasaini for their companionship and support. I would also like to thank Ashutosh Pudasaini for ZTL kinetic and structure data. Also, I would like to thank my family for their constant support and encouragement.

I want to extend my gratitude to the committee members Prof. Patty Wisian-Neilson, Prof. Peng Tao, Prof. Nicolay V. Tsarevsky, and Prof. Steven Vik for taking time out of their busy schedule to review this dissertation and for their valuable advice. I would also like to thank SMU chemistry department, and several funding sources such as Herman Frasch Foundation (Grant #: 739-HF12), NIGMS (Grant #: GM-103485) and Dissertation Fellowship. This work is based upon research conducted at the Cornell High Energy Synchrotron Source (CHESS), which is supported by the National Science Foundation and the National Institutes of Health/National Institute of General Medical Sciences under NSF award DMR-1332208, using the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award GM-103485 from the National Institute of General Medical Sciences, National Institutes of Health.
CHAPTER 1:

PAS DOMAIN STRUCTURE AND FUNCTION

1.1 Introduction

The PAS family of proteins are ubiquitous and are found in all forms of life. They are involved in regulating myriad of physiological responses, for example, regulation of enzyme activity, maintaining circadian rhythms, phototactic and chemotactic responses in bacteria, etc. (1) The structure of PAS proteins is such that it allows binding of different co-factors and small molecules which makes them sensitive to various internal and external stimuli. Most commonly, PAS proteins are sensitive to light, cellular oxygen concentration and changes in cellular redox potential (1). A subclass of PAS proteins binds to a flavin chromophore and is sensitive to blue light. For instance, VVD a blue light photoreceptor from fungus Neurospora crassa, employs a flavin chromophore to modulate the circadian clock in response to blue-light. Other, PAS proteins regulate physiological responses to chemical stimuli upon binding of small metal ions, C3 and C4 carbon molecules and fatty acids (1,2). Therefore, PAS proteins either already have a bound co-factor that make them sensitive to a certain stimulus or binding of a small molecule to the protein can elicit a response. Although there are some PAS proteins which are not involved in binding of small molecules but, are involved in regulating a physiological response through interactions with other PAS-sensors.
Intrinsic to PAS domains is the ability to act as protein: protein interaction module upon signal perception (3). Hence, PAS domains are primary sensory and signal transducing proteins present in almost all living organisms. Their ability to respond to various environmental stimuli to promote protein: protein interactions helps organisms cope with environmental changes.

1.2 PAS nomenclature

The PAS family of proteins are named after the proteins, in which they were first discovered, PERIOD a circadian clock protein from the fly, the vertebrate protein- Aryl Hydrocarbon Nuclear Translocator (ARNT) and fly developmental regulator protein-Single-minded (SIM) hence Per-Arnt-Sim or PAS (2). It was believed to be a conserved family of proteins sharing sequence similarity. Since then, PAS domain containing proteins have been identified in all kingdoms of life such as bacteria, fungi, algae, plants, and mammals. They are primary sensory and signal transducing proteins utilized by organisms in different physiological responses in accordance with internal and external stimuli.

1.3 PAS structure

![Schematic diagram of a PAS domain fold](image)

**Figure 1: Schematic diagram of a PAS domain fold:** PAS domains consists of five antiparallel β-strands (green arrows) and four α-helices (blue cylinders). Series of α-helices intervene the β strands. The diagram shows the organization of β-sheets and α helices in PAS domain. In general, five distinct β-strands and a variable number of α-helices make up the PAS domain with N/C-terminal extensions that often connect to functional effector domains.
PAS domains are about 100 amino acid residues long and have a mixed α/β fold. All PAS domains whose structures have been deposited in Protein Data Bank (PDB) share the same overall fold (2,3). The PAS fold comprises a central antiparallel β-sheet with five strands Aβ, Bβ, Gβ, Hβ and Iβ which is traversed with α helices denoted as Cα, Dα, Eα, and Fα. The strands of the β sheet are in the topological order B-A-I-H-G or 2-1-5-4-3 (4). Some PAS domains, such as the first PAS domain structure obtained from Photoactive Yellow Protein contain a six-stranded central β-sheet (5). The central β-sheets has two faces the A and B face. The A-face consists of β-strands and α-helices which create a hydrophobic pocket for co-factor binding. The B-face is hydrophobic in nature and is involved in protein: protein interactions (5,6). Many PAS domains contain either N-terminal (N-cap) or C-terminal (C-cap) extensions, which interact with the hydrophobic B-face of the central β-sheet. The central β sheet is highly conserved in the structure in most PAS domains, whereas the structure and number of α helices connected to the central β sheet varies. PAS proteins share structural similarity but not sequence similarity. Therefore, to identify PAS proteins in different organisms from the protein database the structure of the PAS fold is used instead of the sequence.

There are two types of PAS proteins: (i) Modular PAS proteins are covalently linked to an effector domain. There are a wide variety of effector domains such as histidine kinases, serine/threonine kinases, guanylate cyclases, phosphodiesterase’s, transcription factors, ion channels, and chemotaxis proteins (7). Therefore, the PAS domain functions as a sensor to regulate the activity of these effector domains. Most commonly PAS proteins are linked to N-terminal of their effector domains. (ii) PAS proteins have either N-terminal or C-terminal extensions to the
PAS core. This type of PAS proteins upon stimuli perception, interact with other proteins, or undergo a change in its oligomerization state to bring about the function of the protein (1,2,8).

1.4 PAS dimerization

As we have already discussed, PAS proteins are sensory, signal transducing and protein:protein interaction modules that help organisms regulate a variety of responses. One of the most critical aspects of PAS signaling is the formation of homo and heterodimers to regulate a response. There are different modes of dimer association observed in PAS proteins which will be discussed in this section. Most commonly the dimer interface involves the hydrophobic B face of the central β scaffold of the PAS fold. N-terminal and C-terminal extensions could provide additional stability to the core PAS fold. The hydrophobic nature of the β scaffold makes it a natural location for intra and inter protein interactions. Therefore, many PAS proteins dimerize utilizing the β scaffold. The dimer interaction could be parallel or antiparallel in orientation. Below we will discuss some examples of PAS dimers.

PAS: PAS homodimers were observed in bacterial stress sensor protein YtvA in both inactive and active states (9,10,11). Light activation leads to rotation of two monomers relative to each other which leads to activation of the attached sulfate transporter and antisigma factor antagonist (STAS) domain. The activation of STAS domain by light leads to positive regulation of general stress transcription factor σB. YtvA forms a parallel head to head dimer utilizing the hydrophobic interactions between the two β sheets (10). The hydrophobic contacts are made between residues located on β strands of Aβ, Bβ, Hβ, and Iβ.

Multiple PAS domain containing protein aryl hydrocarbon receptor nuclear translocator (ARNT) present in mammals is involved in regulating a variety of biological responses by forming hetero-dimers. ARNT contains two PAS domains and a basic helix-loop-helix region
and it forms heterodimeric complexes with other bHLH-PAS proteins to act as a transcription activator complex to initiate gene transcription (12,13). Under low oxygen or hypoxic conditions, ARNT forms a heterodimer with various isoforms of bHLH-PAS protein hypoxia inducible factor alpha (HIF-α) to initiate gene transcription of various proteins to aid cell survival under hypoxic conditions (1,12,14). NMR spectroscopic characterization of the heterodimeric complex of ARNT and HIF 2-α revealed the involvement of β scaffold as the dimer interface. The PAS-B of ARNT forms a complex with PAS-B of HIF 2α by associating their β sheets in an antiparallel fashion (14). In addition ARNT also forms a hetero-dimer with aryl hydrocarbon receptor (AHR) and single-minded (SIM) proteins to regulate diverse biological responses. Therefore, β-scaffold interactions are involved in the formation of homo as well as hetero dimeric complexes.

A variant dimer interface involving the FG loop connecting the Fa and Gβ was exemplified by the heterodimer of LuxP-LuxQ proteins from marine bacterium Vibrio harveyi (15). LuxP is a periplasmic membrane binding protein associated with a PAS protein LuxQ. LuxQ has two PAS domains in tandem, a histidine kinase and a response regulator domain. Ligand binding to LuxP regulates kinase/phosphatase activity of LuxQ. LuxP associates with LuxQ even in absence of ligand, the hetero-dimer interface between a non-PAS and a PAS protein involves N-terminal of LuxP which inserts in a cleft between two PAS domains of LuxQ and forms hydrogen bonds with amino acid residues on FG loop of proximal PAS domain and HI loop of distal PAS domain (15). The FG loop of the proximal PAS domain appears to be disordered in the crystal structure. The FG loop is involved in undergoing critical conformational changes upon ligand binding in bacterial oxygen sensor FixL and human PAS kinase (hPASk) whereas its involved in forming essential dimer interactions in LuxP-LuxQ pathway (15,16,17).
The circadian clock protein PERIOD (PER) from the fruit fly *drosophila melanogaster* forms a heterodimer with a non-PAS protein TIMELESS (TIM) to regulate gene transcription and maintenance of circadian rhythmicity (18). The crystal structure of two PAS domains in PER form a homo-dimer, the residues involved in the formation of homo-dimer are also relevant for hetero-dimer contacts. Further, homo-dimer of PER is also biologically crucial for its function. The PER dimer consists of contacts between Aα helix of PAS-B against the Cβ, Dβ, Eβ, and Cα of PAS-A domain. Further, a C-terminal helix Fα from PAS-B runs across the β scaffold of PAS-A. Thus, the crystal structure of PER showed the involvement of C-terminal and α-helical elements of PAS fold in dimer formation. The dimeric interface observed in bacterial oxygen sensor proteins of FixL, EcDosH and NifL utilize N-terminal helices to form a dimer. The two N-terminal helices from each monomer dock on each other with hydrophobic interactions. The N-terminal helix may also hydrogen bond to some residues on the β scaffold.

Thus, we have observed that dimerization in PAS proteins involves various interfaces such as the β scaffold, N-terminal/C-terminal extensions, and the FG loop region. Some PAS proteins remain dimeric in the inactive and the active state, in such proteins signal perception causes rotation of two monomers relative to each other which then relays the signal to attached output domain by allostery. In contrast, in some PAS proteins signal perception leads to dimerization with specific partners such as that seen in ARNT and HIF to carry out a particular function. Therefore, homo and hetero dimerization between PAS and non-PAS proteins upon signal perception is the primary signaling event to bring about the purpose of the protein.

### 1.5 PAS domain diversity

To date, more than 21,000 proteins containing PAS domains have been discovered across several species of bacteria (81%), archaea (6%) and eukaryote (13%) (2,19). PAS domains are
primarily sensory, and signal transducing proteins and regulate diverse physiological responses. They are found in serine/threonine kinases, histidine kinases, photoreceptors, chemoreceptors for taxis and tropism, circadian clock proteins, voltage activated ion channels, regulators of response to hypoxia and embryological development of the central nervous system (7). Below we will discuss about PAS proteins and their functions in diverse organisms.

1.5.1 PAS domains in prokaryotes

Prokaryotes are microscopic single-celled organisms found in various environmental niches. Survival of a single celled organism depends mainly upon its adaptation to changes in the physicochemical makeup of its environment (2). Therefore, prokaryotes express a variety of sensory and signal transduction proteins among which the PAS domains are widely utilized. Most PAS domains in members of bacteria and archaea are a part of a two-component regulatory system such as those found in histidine kinase sensors (16). The two-component system consists of a histidine kinase sensor and a cognate response regulator. An N-terminal PAS domain senses stimulus directly or indirectly and transmits the signal to the C-terminal (20,21,22). The C-terminal module contains a conserved histidine which is phosphorylated upon stimuli reception. The phosphoryl moiety is then transferred from the sensor histidine to a conserved aspartate in the receiver module of the response regulator. The response regulator gets activated upon phosphorylation and either interact with DNA or another signaling protein. Mostly, response regulator is a transcriptional activator.

FixL-FixJ pathway in bradyrhizobium japonicum and other related bacteria regulate the expression of nitrogen fixation genes (17). FixL is an input PAS domain which is an oxygen sensor. Oxygen dissociation from the input FixL domain induces a structural change and increases autophosphorylation of the conserved histidine (23). FixL then catalyzes the transfer of a
phosphoryl moiety from the histidine to an aspartate residue in the receiver module of the response regulator FixJ. Phosphorylated FixJ then activates the transcription of genes involved in nitrogen fixation (23).

PAS domains also act as signal transducing proteins such as in the case of Aerotaxis transducer protein Aer in Escherichia coli which, regulates aerotaxis responses. Aerotaxis responses refer to the navigation of bacteria to environments where the concentration of oxygen is optimal for growth. The Aer protein has a PAS domain at its N-terminal which is non-covalently bound to a FAD cofactor (24). The FAD cofactor does not directly sense oxygen but senses redox changes in the electron transport system (24). In this regard, Aer belongs to the class of PAS transducers instead of sensors. The predicted structure of Aer protein provides clues to signal transduction mechanism. A central hydrophobic sequence anchors the PAS domain to the cell membrane. Aer forms a dimer in-vivo. The C-terminal of Aer protein contains an HCD domain present in all chemotaxis transducers. The HCD domain in the presence of CheW regulates the histidine kinase activity of CheA (25). Detection of change in redox potential by N-terminal PAS domain causes the CheA protein to phosphorylate the CheY response regulator (26). Upon phosphorylation, CheY protein binds to the FliM protein on the flagellar motor which reverses the direction of the motor from anti-clockwise to clockwise and changes the direction of motion (26). Therefore, the prokaryotic genome encodes for many PAS domain-containing proteins which regulate different physiological responses. Some of the physiological responses controlled by PAS domains include chemotaxis, sporulation, nodulation, virulence and maintaining energy metabolism.
1.5.2 PAS domains in eukaryotes

Eukaryotes have membrane bound organelles and are more complex than prokaryotes. Their genomes encode 13% of all PAS proteins identified. They are mainly involved in regulating light responses, maintaining circadian rhythms, regulating voltage-ion channels, hypoxia responses and regulation of central nervous system development (8,14). In this thesis we will primarily focus on the role of PAS proteins involved in photoreception in plants.

In plants, PAS domains play a role in regulating conformational changes in photoreceptors, and in regulating protein: protein interactions with various classes of plant photoreceptors sensitive to all wavelengths of light. For instance, phytochromes are red light sensitive proteins found in plants and are responsible for regulation of various processes such as seed germination, seedling de- etiolation, shade avoidance and flowering (27). They bind to a bilin cofactor which is sensitive to red and far-red light. They are typically composed of a PAS:GAF:PHY motif, where the N-terminal PAS domain helps cradle the bilin cofactor. Cryptochromes are blue light photoreceptor proteins found in plants and animals and they are involved in maintaining circadian clocks (28). They primarily achieve signal transduction through blue-light regulated formation of protein: protein interactions, often involving diverse members of the PAS family. The third class of plant photoreceptors are the Light Oxygen Voltage domain or LOV proteins. LOV proteins form a PAS subclass and are mainly found in bacteria, plants, and fungi. In this thesis, we will discuss in detail eukaryotic LOV proteins FKF1, LKP2, VVD, and ENV1.

FKF1 and LKP2 are blue light photo-receptor proteins found in plants and are involved in maintaining circadian responses and other processes such as flowering (29). They are modular proteins containing LOV domain at the N-terminal and F-box and Kelch repeat domain at the C-terminal. The LOV domain binds to a flavin cofactor and is sensitive to blue light. The Kelch
repeat domain is involved in protein: protein interaction while F-box targets protein for degradation (29). Therefore, these proteins are involved in light mediated targeted protein degradation in all planta.

In fungi, VVD is an isolated LOV protein from fungus *Neurospora crassa* and it regulates light mediated gene transcription and circadian rhythmicity along with other proteins. We were interested in the VVD homolog ENV from fungus *Trichoderma reseei* (30,31). In addition to light mediated gene-regulation, ENV also adapted to incorporate oxidative and osmotic stress sensing. It has evolved from its homologue VVD to sense additional stimuli. In this thesis, we will discuss in detail about differences in signaling mechanisms of ENV and VVD and the basis of adaptation to additional stimuli.

**1.6 PAS classification based on various cofactors**

PAS domains can bind to chemically diverse small molecules. The structure of PAS domains is very flexible to allow co-factor binding (32). Although, there are few PAS domains that do not bind to a co-factor and function without a co-factor. The most common co-factors that bind to PAS proteins are heme, flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), hydroxycinnamic acid (4-HCA), C3-C4 carboxylic acids, and divalent metal cation (2).

**1.6.1 Heme binding PAS domains**

Heme is a co-factor consisting of an Fe\(^{2+}\) ion coordinated at the center of a porphyrin ring. Heme is also found in other important biological molecules such as hemoglobin, myoglobin, cytochrome, catalase and heme peroxidases (33,34). FixL, as we have discussed above, is a sensor histidine kinase which senses oxygen with a heme-b cofactor (17). Another related PAS domain direct oxygen sensor, DosP from *Escherichia coli*, also binds oxygen to heme b cofactor to regulate the activity of phosphodiesterase enzymes. FixL and DosP have a common heme ligation site via
a conserved histidine residue (33). Notably, interactions between heme and PAS domain can vary as some PAS domains also bind to c-type heme cofactors. The chemoreceptor proteins from Desulfovibrio vulgaris and Geobacter sulfurreducens bind to a c-type heme cofactor via a conserved bi-cysteine ligation site. The heme cofactor is bound at the surface of a PAS domain on its opposite face by a helical motif from the second PAS domain. Although the binding mechanisms may differ, all heme binding PAS domains serve as oxygen sensors.

1.6.2 Flavin binding PAS domains

A subtype of PAS domains binds to flavin co-factors they are known as Light Oxygen Voltage domains or LOV proteins. In this thesis, we will be mainly discussing function and signaling mechanisms of some plant and fungal LOV domains. The bound flavin co-factor in LOV proteins senses internal and external stimuli and translates the signal to bring about the function of the protein. Upon stimuli perception, the flavin cofactor changes its chemical state which affects the hydrogen bonding nature within the protein leading to intra and inter-protein interactions. Most LOV proteins are blue light sensors. The flavin cofactor gets reduced upon blue light absorption and forms a covalent bond with a conserved cysteine residue.

In 2001, ultrafast time-resolved spectroscopy studies on Avena sativa (Oat) LOV2 domain revealed the mechanisms of light-induced spectral changes and transitions in LOV proteins. In its dark-adapted state LOV2 domain exhibited absorption peaks at 450 nm with shoulder peaks at 425 and 475 nm representing the vibronic states. Blue light absorption bleached the peak at 450 nm and revealed a transient 660 nm peak within 30 ns, indicative of the triplet excited state. The triplet state disappeared at a time-scale of 4 µs resulting in a 390 nm absorbing species (LOV390). LOV390 represents the covalent cysteine-flavin adduct. Thus, the cysteine-flavin
adduct can be characterized by a characteristic peak around 390nm. The adduct state then decays back to its ground state. Of particular interest, is the fact that the rate of adduct decay varies in different LOV proteins. The kinetics of adduct decay can be followed spectrometrically by observing the rise in free flavin peak at 450 nm. Further, the protonation of flavin at the N5 position by light causes changes in the hydrogen bonding network within the protein. There is a conserved Gln residue in most LOV proteins which hydrogen bonds to the N5 position of flavin co-factor in the ground state. Upon light activation and N5 protonation, the Gln residue flips and hydrogen bonds to its neighboring residue which leads to changes in the hydrogen bonding network. The changes in the hydrogen bonding network affects the protein structure and either causes it to homo-dimerize or interact with neighboring domains or proteins to bring about the function. For example, Phototropin is a LOV protein found in plants responsible for phototropic responses. Phototropin consists of two LOV domains at its N-terminal and a serine/threonine kinase at its C-terminal. Light activation of LOV domain at its N-terminal leads to autophosphorylation and activation of serine/threonine kinase at its C-terminal. EL222, a LOV protein from marine bacterium *Erythrobacter litoralis*, is a transcription factor protein. It consists, of a LOV domain at its N-terminal and a helix-turn-helix HTH domain at its C-terminal (35). In its dark state, the β sheet of LOV domain inhibits the HTH domain, upon light activation the inhibitory interaction between LOV: HTH domain is released and the protein dimerizes allowing DNA binding (35).

1.6.3 Carboxylic acid binding PAS domains

Carboxylic acid binding PAS domains are utilized for various purposes. They serve as a carbon energy source during aerobic metabolism, and they can also be utilized as electron acceptors during anaerobic metabolism. PAS domains get activated upon binding of carboxylic acids and further regulate the gene expression of proteins required for their uptake and metabolism.
DcuS is an extracellular PAS domain from *Escherichia coli* it binds to C4-dicarboxylates and upon binding it activates the kinase domain (36). The kinase domain further phosphorylates the response regulator known as DcuR which binds to promoter regions of genes required for C4-dicarboxylate metabolism. CitA a citrate binding protein from *Klebsiella pneumoniae* is also a sensor histidine kinase which regulates gene expression of proteins needed for citrate metabolism upon citrate binding (37). The PAS structure allows binding of various carboxylic acids using different backbone and side-chain interactions.

### 1.6.4 Divalent metal binding PAS domains

PhoQ is a PAS domain present in bacteria which binds to divalent metal ions such as Mg$^{2+}$, Ca$^{2+}$, and Ni$^{2+}$. PhoQ consists of an acidic cluster of residues at the surface of the PAS domain which allows binding of divalent metal ions. PhoQ is a transmembrane protein kinase which regulates various responses such as Mg$^{2+}$ transport, stress responses, lipopolysaccharide remodeling (38). In pathogenic bacteria, PhoQ is involved in playing a role in bacterial virulence.

### 1.7 PAS-like proteins

As discussed above we have observed that PAS proteins have the ability to bind to various types of co-factors. Although, the ability to bind to small molecules and initiate signaling is not confined to PAS proteins there are numerous other protein families collectively called as PAS-like proteins which have similar architecture to PAS proteins with some variations. The protein families that have similar structure to PAS fold include SH2, profilin, CACHE and GAF domains (39).

There are some variations observed even within PAS proteins. Usually, a PAS fold consists of five antiparallel $\beta$ strands and four $\alpha$ helices. However, there are some PAS proteins such as PYP and H-NOXA that have an extended $\beta$ scaffold with 6$\beta$ strands (41). The GAF protein family
also have an extended β scaffold. The GAF family of proteins are also ubiquitous and are found across various kingdoms of life (39). They bind to cyclic nucleotides such as cyclic-GMP and initiate signaling, or they are involved in regulating the activity of phosphodiesterase enzyme (40,41). Since, GAF proteins do not bind to typical PAS ligands, it led to their identification as a unique family, although in many protein families PAS and GAF domains are employed interchangeably.

CACHE domains are primarily found in bacterial chemoreceptors. They are extracellular receptors capable of binding to various ligands to initiate signaling (42). They have transmembrane helices which anchor them to the membrane. Earlier extracellular PAS domains were a part of PAS family. Later, they were identified as a part of CACHE domain family which are homologous to PAS proteins (42). These PAS-like proteins share a similar general fold but differ in their functionalities and types ligands they bind. Therefore, they were categorized as a separate family. Considering their similarities in protein fold and signaling mechanisms, they could have evolved from the same evolutionary superfamily.

1.8 PAS classification based on structure

PAS proteins can be classified into three types based on their structure. The first type of PAS proteins are those proteins which only consists of an isolated PAS domain not attached to any other domains. For example, Photoactive yellow protein from bacterial species and VVD/ ENV1 from fungal species. The second type of PAS proteins are modular proteins containing a single PAS domain attached to one or more than one effector domains such as ser/thr kinase, histidine kinase, guanylate cyclase, transcription factor etc. for example, YtvA a bacterial stress sensor protein, LKP2/FKF1/ZTL circadian clock proteins from plants. The third type of PAS proteins are modular proteins containing more than one PAS domains attached to effector domains they are
also called as Tandem PAS domains for example, Phototropin, a blue light photoreceptor, from plants contains two LOV domains at its N-terminal, FixL, a bacterial oxygen sensor. In this chapter we will be discussing signaling mechanisms of photoactive PAS proteins under each category to

![Figure 2: PAS classification based on structure](image)

1.8.1 Isolated PAS domains

1.8.1.1 Photoactive yellow protein

Photoactive yellow protein (PYP) from purple photosynthetic bacteria *Halorhodospira halophila* is the first photo-sensor protein to be structurally characterized (5). It also serves as a structural prototype for PAS domain-containing proteins due to its sequence homology (5). Computational analysis was used to predict the structure of human ARNT protein based upon the structural data from PYP. Thus, the use of structural data of PYP for structure prediction of other PAS proteins further validated the application of PYP structure as a prototype for PAS proteins (5).

Photoactive yellow protein is a short protein containing only the photo-sensory PAS domain with a short N-terminal extension. It binds to a 4-hydroxycinnamyl chromophore which undergoes trans-cis isomerization upon light reception (18). The crystal structure of PYP revealed
a mixed α/β fold consisting of six β strands and five α helices. The structural components of PYP can be described as follows: N-cap consisting of helices α1 and α2, PAS core is composed of β strands β1, β2 and β3 traversed by helices α3 and α4, a helical connector (α5) and a β-scaffold consisting of β strands β4, β5 and β6 (5).

The chromophore is a planar, six-membered ring of 4-hydroxycinnamyl anion forming a covalent bond with the protein in its inactive state. The thiol of Cys69 forms a thioester linkage with the chromophore (43). The chromophore binding pocket consists of charged residues from α3 helix and the adjacent loop required for maintaining hydrogen bonding contacts to the chromophore. The phenolic O4 group of chromophore forms hydrogen bonds with Tyr 42 and Glu46. The Thr50 side chain forms hydrogen bonding contacts with Tyr42 and Glu46. Additionally, the positively charged group of Arg52 packs against the aromatic ring of the chromophore. Arg52 forms a shield between chromophore and solvent (43). Upon light reception, the phenolic oxygen gets protonated, and the chromophore isomerizes to its cis conformation. This leads to changes in hydrogen bonding arrangement within the chromophore binding pocket. Eventually, the Arg52 side chain flips exposing the chromophore to the solvent. The signal is then translated to the N-terminal confirmed by solution studies such as NMR and SAXS (44,45). The downstream signaling partner of PYP is yet unknown, but so far it has been shown that light signal is translated from chromophore binding pocket to the N-terminal.

1.8.2. Single PAS domains

1.8.2.1 Ytva signaling

Ytva is a blue light photoreceptor protein from bacterium Bacillus subtilis. It regulates the organism’s reaction to stress conditions (11). Ytva contains an N-terminal LOV domain which is
connected to a C-terminal STAS (Sulfate Transporter Anti-Sigma factor agonist) domain by a Jα helical connector (9). The solution, as well as crystallographic studies, show that the protein forms a dimer in both light and dark state. The crystallographic data revealed the formation of a head to head dimer burying the hydrophobic residues of the β scaffold of the LOV core (10). The position of Jα helix relative to LOV core is very different from that observed in phototropin-LOV2 domains. In phototropin, LOV2 domain Jα helix is packed against the β scaffold whereas in Ytva the Jα helix is rotated away from the LOV core. The Jα helix forms intermolecular contacts with other Jα helices from symmetry-related molecules within the crystal. Light activation leads to the formation of cysteinyl-flavin adduct and rotation of two monomers by 4-5° relative to each other. The rotation of two monomers resembles a scissor-like motion. Although YtvA has a typical LOV domain fold, its quaternary structure is different from other LOV domains. YtvA also has more hydrophobic residues on its β scaffold as compared to other LOV proteins and an extended Hβ and Iβ strands. The rotation of two monomers relative to each other may lead to activation of the C-terminal STAS domain (10). Although, there is high sequence similarity between YtvA and phototropin LOV domains the signaling mechanisms of YtvA are more like bacterial oxygen sensor FixL. Nevertheless, LOV proteins may share high sequence similarity and general fold, but how the stimuli signal is transduced through the structure and effector domains may differ. Therefore, the structural elements and amino acid composition of PAS proteins are adapted to undergo variations in signaling mechanisms.

1.8.3. Multiple PAS domains

1.8.3.1 Phototropin signaling

Phototropin is a blue light photoreceptor protein present in various species of plants. It controls multiple blue light responses such as phototropism, stomatal opening, light-mediated
chloroplast relocation (51). The primary role of phototropin proteins is to regulate photosynthesis at appropriate light intensities to maximize energy production while avoiding radiation damage by affecting chloroplast relocation. Phototropin1 proteins respond from low to moderate levels of blue light to localize the chloroplast at the periclinal surface for photosynthesis (52,53). Whereas, phototropin2 proteins react to high levels of blue light and relocalize the chloroplasts away from the surface to avoid radiation damage (52,53). Thus, phot-1 and phot-2 together regulate photosynthesis at appropriate light intensities. Phototropins are membrane-associated ser/thr kinases which undergo autophosphorylation upon light activation (54).

The phototropin domain architecture contains two LOV domains referred as LOV1 and LOV2 at its N-terminal and a C-terminal ser/thr kinase domain. Light absorption by LOV1 and LOV2 regulates kinase activity (50). Both LOV1 and LOV2 non-covalently bind to an FMN molecule (54). Both LOV1 and LOV2 undergo a reversible photocycle shown by the blue shift of three bands observed on spectrometer upon light activation and its recovery upon incubation in dark. The blue shift is due to the formation of a cysteinyflavin adduct. Mutation of active site cysteine allowed flavin binding but abolished the photocycle. The conformational changes occurring upon light activation are more enhanced in LOV2 and are required for kinase activation. Mutation of the active site cysteine in LOV1 did not affect kinase activity although, mutation of active site cysteine in LOV2 altered kinase activation (50). This showed that LOV2 is mainly involved in kinase activation upon light reception whereas LOV1 is responsible for accentuating the response and plays a role in LOV: LOV dimer formation (55).

The LOV1 and LOV2 domain of phototropin-1 have been structurally characterized in various plant species. The phototropin-1 are also first LOV domain proteins to be structurally characterized. The crystal structure of the LOV2 domain of phototropin from chimeric fern
photoreceptor phy3 revealed residues essential for flavin binding and confirmed the formation of a covalent adduct between flavin co-factor and a conserved cysteine upon light activation (47). The LOV2 structure revealed a typical PAS fold consisting of antiparallel β-sheet with five strands Aβ, Bβ, Gβ, Hβ and Iβ which is traversed with α helices denoted as Cα, Dα, Eα, and Fα (47). The strands of the β sheet are in the topological order B-A-I-H-G or 2-1-5-4-3.

A single molecule of FMN is non-covalently bound between the β sheet and α helices. There are various amino acid residues responsible for making non-covalent interactions with the FMN molecule to allow stable binding of the co-factor (47). The majority of flavin contacts are made by the amino acid residues on the β strands of C, D, and E and the Aα helix. The polar contacts are made on the pyrimidine side of isoalloxazine ring and non-polar contacts are made on the dimethyl benzene moiety. The ribityl chain of FMN terminates on the surface of LOV2, and the phosphate moiety of FMN forms salt bridges with guanidium groups of Arg967 and Arg983 on αa and αc. Side chain carbonyl atoms of Asn965 and Gln970 from αa helix form hydrogen bonding interactions with hydroxyl groups of the ribityl chain. The side chain carbonyl atoms of Asn998 from βc forms hydrogen bonds with N3 of FMN and amide side chain of Asn998 forms hydrogen bonds with O2 of FMN. The amide nitrogen of Asn1008 from βd and Gln1029 from βe form hydrogen bonds with O4 of FMN. Val979 of αc makes van der Waals contact with 4′ hydroxyl of ribityl moiety and Iso982 contacts C1 of FMN.
Figure 3: FMN binding in phototropin LOV2 domain: FMN is bound to the LOV domain with various hydrogen bonding, Van der Waals and hydrophobic interactions. Some of the critical hydrogen bonding interactions as discussed above are shown in this figure. The FMN molecule is held between the β scaffold and α helices. Hydrogen bonding interactions are made between N3, O2 and O4 of FMN and amide or carboxylic side chains of amino acid residues Gln 1029, Asn 1008, Asn 998 from the β strands C, D and E. The phosphate group of FMN forms a salt bridge with positive guanidium groups of Arg967 and Arg983.

Figure 4: Sequence Alignment of phototropin LOV2 domain with other LOV proteins from plant and fungal species: The amino acid residues involved in interactions with FMN in LOV2 domain of phototropin are conserved in LOV domains of FKF1, ZTL, LKP2 from plant Arabidopsis thaliana and ENV1 from fungus Trichoderma reseei whose signaling mechanisms we will discuss in this dissertation.
Residues V986, L1012, F1025, T934, and V932 form the hydrophobic pocket around the dimethylbenzene moiety. Some of the amino acid residues essential for flavin binding in phototropin1 are also conserved in other LOV proteins such as ZTL, FKF1, and WC-1. The most conserved stretch present in most LOV domains is the consensus sequence NCRFLQ, which constitutes the $3_{10}$ helix ($\alpha'A$) in phy3 LOV2.

The proximity of Cys966 to C4a carbon of FMN suggested the formation of a covalent adduct between $\gamma$ of Cys966 and C4a carbon of FMN in its light state. Absorption of photons by the isoalloxazine ring of FMN causes redistribution of charges on the ring; therefore, the basicity of N1 decreases and basicity of N5 increases which could abstract a proton from Cys966 and caused the thiolate ion of Cys966 to attack the C4a carbon to form a covalent adduct. Indeed, the light state crystal structure of this protein confirmed the formation of the cysteinyi-flavin adduct (48). The proximity of $\gamma$ of Cys966 from $\alpha'A$ to C4a carbon and puckering of isoalloxazine ring validated the electron density to be of a covalent adduct.

Further, the light state structure revealed the hydrogen bonding rearrangement occurring within the flavin binding pocket upon light excitation. The amide side chain of Gln 1029 is hydrogen bonded to O4 atom of FMN and carbonyl side chain is hydrogen bonded to Ser930 in dark state. Upon N5 protonation, the carbonyl side chain of Gln1029 breaks the hydrogen bond with hydroxyl of Ser930 and forms a new hydrogen bond with the N5th position of flavin (48,49). The amide side chain of Gln1029 breaks the hydrogen bond with O4 and now forms a hydrogen bond.

Thus, the dark and light crystal structure of phy3LOV2 provided an understanding as to how LOV domain structure binds to flavin cofactor and changes occurring within flavin binding.
pocket upon light activation. The conservation of active site cysteine and residues interacting with flavin in other LOV proteins suggest a common mechanism of light activation and co-factor binding in other LOV proteins.

An understanding of how the formation of cysteinylation-flavin adduct leads to structural changes and kinase activation has been explained by NMR and crystallographic studies of phototropins from other plant species. NMR studies of oat phototropin LOV2 domain in light and dark state gave insight into structural changes that could lead to kinase activation. NMR studies were done on a more extended protein construct (404-560) that contained a 40-amino acid C-terminal extension to the PAS core. Dark state NMR studies revealed that the C-terminal extension forms an amphipathic helix called as Jα which is packed against the β sheet surface. Specifically, the Jα helix is packed against solvent exposed β strands of Gβ, Hβ, and Iβ burying the hydrophobic surfaces. Therefore, the solvent-exposed hydrophobic surface of the β sheet of PAS core is protected by the Jα helix. Light state NMR studies of oat phototropin revealed destabilization and movement of Jα helix away from the β sheet core upon light activation. Further experimentation confirmed the movement of Jα helix away from the β sheet.

Thus, it was proposed that light activation induces changes in the flavin binding pocket of LOV2 domain further leading to structural changes at the C-terminal which leads to kinase activation. Crystallographic studies of light and dark oat phototropin LOV2 construct 404-546 also confirmed the movement of Jα helix and provided information about hydrogen bonding rearrangement required to bring about the structural response (56). The crystal structure in addition to C-terminal Jα helix also revealed an N-terminal extension to PAS core consisting of a helix-turn-helix. The contacts between β sheet and Jα helix are primarily made between hydrophobic amino acid residues of Iβ strand and Jα helix. Light activation leads to cysteinyln-
flavin adduct formation and protonation of the N5th position of flavin. N5th protonation causes Gln 513 on Iβ strand to break a hydrogen bond with O4 of FMN and form a hydrogen bond with the N5th position of FMN. Light also causes breaking of hydrogen bond between Asn414 of Aβ and Asp515 of Iβ. This leads to the formation of new hydrogen bond between Gln513 and Asn414. These hydrogen bonding rearrangements displace Gln513 outward from the flavin pocket and towards the N-terminal helix turn helix. The rupture of a hydrogen bond between Asn414 and Asp515 leads to displacement of Jα helix. Therefore, the formation of cysteinyl-flavin adduct on light activation causes local structural changes in the flavin binding pocket which is then propagated to N and C-terminal extensions to bring about the function of the protein. Thus, phototropin provides a model system for LOV domain function where the light signal is transduced from flavin binding pocket to other regions of the protein such as N and C-terminal extensions to bring about the function of the protein.

1.9 Specific research goals

In the above review, we observed the ubiquitous nature of PAS proteins and its versatility in binding different cofactors and regulating various responses. In this thesis, we will outline the photochemistry and structural characterization of an isolated PAS protein ENV from fungus *Trichoderma reseei* and modular circadian clock proteins FKF1/LKP2 from *Arabidopsis thaliana*. These proteins belong to the LOV domain family, a subclass of PAS superfamily. Phototropin proteins from plants were first LOV domain proteins to be characterized and provide useful insights about co-factor binding and signal transduction mechanisms in this family. Based on knowledge obtained from previous literature and combination of various biophysical techniques such as UV-spectrometry, size-exclusion chromatography, and X-Ray crystallography, signal transduction mechanisms in these proteins will be elucidated. Further, structural characterization
of these proteins will provide knowledge of signaling mechanisms and interaction surfaces which will help in designing mutations to affect the function in an economically advantageous manner. The modular nature of LOV proteins allows them to be used as tools for optogenetics. The various adduct lifetimes can be obtained by designing mutations that alter the photochemistry of LOV proteins. Therefore, photochemical characterization of these proteins can provide information about tunability of LOV proteins to be used in the field of optogenetics.
REFERENCES


CHAPTER 2: PLANT CIRCADIAN CLOCK AND CHARACTERIZATION OF LOV PROTEINS LKP2 AND FKF1

2.1 Circadian clock

Earth rotates about its axis every 24 hours. Due to this rotation, a certain region of earth is either facing towards or away from the sun resulting in day and night cycles. Many organisms have evolved an endogenous mechanism to predict the time of day, known as the circadian clock (1,2). Circadian clocks help organisms to conduct certain activities according to the 24 hrs cycle to ensure balance and fitness. For example, sleep-wake cycles, eating patterns and hormonal secretion in humans, leaf movements, seasonal flowering and photosynthesis in plants all demonstrate circadian rhythmicity (1,2,3). Although the circadian clock is entrained by external stimuli such as light and temperature, circadian rhythms and activities can be maintained under conditions of constant darkness or light. Therefore, these rhythms are endogenously generated and are self-sustaining (2,4). The first circadian activity in plants was observed in 4th century B.C by a ship captain Androsthenes, he observed diurnal leaf movements of a tamarind tree (5). In 1729, a French scientist Jean-Jacques d’Ortous de Mairan observed 24 hr pattern in leaf movements of plant Mimosa pudica which was even maintained in constant darkness (6).

Therefore, there are specific criteria, by which circadian rhythms are defined. Firstly, circadian rhythms are endogenous and are observed under constant conditions. The second criteria is that the circadian clocks are entrainable by external stimuli such as light and adapt to change
in their surroundings, for example when we travel across various time zones we experience jet-lag, but eventually, our bodies get entrained to the new time zone (3,7). Lastly, circadian clocks are temperature compensated, where they can maintain rhythmicity across wide ranges of temperatures.

The circadian machinery consists of three components (i) The input pathway comprising of photoreceptors and other proteins which entrain the circadian oscillator with environmental cues such as light, (ii) The circadian oscillator which is a central component regulating various biological responses to a 24 hrs cycle. (iii) The output pathway regulated by the circadian oscillator to bring about the circadian response (4).

**Figure 5: Three components of a circadian clock:** Environmental light is detected by photoreceptor that associate signaling to an oscillator protein that controls the expression of genes that regulate the physiological responses. The variable activity of the photoreceptor and the nuclear oscillator protein can result in variable clock output as shown in the A, B, and C gene expression profiles.
At the core molecular level, the circadian oscillator consists of transcription-translation feedback loops. The core oscillator is a transcriptional activator/repressor that activates or represses expression of circadian genes depending upon the input signal. The activity of the oscillator itself is controlled by various positive and negative feedback loops (4,8,9).

In mammals, understanding of the generation of circadian rhythms at molecular level was first facilitated by the discovery of the period gene (PER) in fruit fly Drosophila melanogaster (10). The PER gene and its product the PER protein both showed a circadian oscillation in its expression (3,10). The circadian rhythms can be studied by monitoring the gene expression throughout the course of the day. If the expression pattern of a gene resembles a wave peaking at a particular time of the day and reducing at a particular time, this indicates that the gene is either involved or regulated by the circadian machinery. Several physiological activities in Drosophila occur at a specific time of the day. For example, adult flies emerge from the pupa at dawn when it is cool and moist. This process is known as eclosion. Study of circadian rhythms in the fly was greatly facilitated by studying their eclosion behavior under different circumstances. Pupae kept under constant darkness also timed their eclosion at dawn by anticipating the dawn time similar to pupae maintained under 12 hr light/ 12 hr dark. This shows the presence of an internal pacemaker.

Further forward genetics enabled revealing the molecular elements involved in maintaining the circadian machinery in flies and was recently awarded the Nobel Prize in the year of 2015. Two clock genes dClock (dClk) and cycle (cyc), whose orthologues are also found in mammals, encode for proteins containing the PAS domain and a DNA binding bHLH domain (11). These two proteins dClk and Cyc heterodimerize via PAS domains and bind to E-box
elements of various genes to activate their transcription. They also activate the transcription of PER and Timeless (TIM) genes (11). The two genes per and tim are actively transcribed during the day, therefore, proteins PER and TIM heterodimerize in the cytoplasm during the day and later in the evening the heterodimer complex translocates to the nucleus to inhibit its own transcription. The PER and TIM hetero-dimer inhibits the dClk/cyc complex, thereby inhibiting its own transcription. The PER and TIM proteins are degraded before dawn, relieving the repression of dClk/cyc complex, thus starting a new cycle of transcription and translation. Therefore, these positive (dClk/cyc) and negative (PER/TIM) feedback loops maintain circadian rhythmicity in flies (12).

2.2 Vertebrate clock

In humans and other mammals, the core biological clock resides in the superchiasmatic nuclei (SCN) in the brain located at the base of the hypothalamus (13). Research in the field of the mammalian clock revealed that the biological clock is cell autonomous and present in peripheral tissues. The master clock in the SCN regulates the peripheral clocks and distant cells by signaling through hypothalamic and non-hypothalamic neuronal circuits (13). The SCN thus communicates with other tissues in the body to release specific hormones or to perform a specific activity at a particular time of a day to maintain homeostasis. For example, food intake, activity, body temperature, blood pressure, plasma adrenaline and urinary excretion are higher in day-time than night. The information about the time of the day is facilitated by photic entrainment of biological clock in SCN by the light entering through the retina of the eye. There is a direct pathway between retina and hypothalamus through the retina hypothalamic tract (RHT). The mammalian clock oscillator consists of two PAS domain-containing transcription factor proteins CLOCK (Circadian Locomotor Output Cycles Kaput) and BMAL1 (Brain and Muscle ARNT-Like) (8,14). CLOCK
and BMAL1 heterodimerize and activate the transcription of various clock-controlled genes (8). They also activate the transcription of PER and Cryptochromes (CRYs) genes. The PER/CRY heterodimers inhibit the BMAL1/CLOCK complex similar to how PER/TIM impede the dClk/cyc in drosophila, forming the negative feedback loop (14).

In mammals, CRYs incorporate the metabolic state of a cell as a cue for tissue-specific function. CRYs regulates gluconeogenesis and thus plays a role in glucose homeostasis (15). CRYs prevent gluconeogenesis, a process of synthesizing glucose from a non-carbohydrate source in hepatic cells, during early morning hours or during the feeding hours. During night-time, CRY levels are low to allow gluconeogenesis to occur at the fasting hours (15). Therefore, irregularities in circadian machinery can cause blood glucose levels to rise abnormally, a symptom common in type-II diabetes. Hirota and co-workers reported that CRYs can be stabilized with small molecule activators that prevent its proteasomal degradation (16). High levels of CRYs, in turn, decrease the rate of gluconeogenesis and reduce blood glucose levels. Thus, circadian clock research and understanding of its components is essential to develop treatments against metabolic diseases such as type-II diabetes and much more.

2.3 Plant circadian clock

Due to the sedentary nature of plants, their growth, development and metabolic processes need to be well-coordinated with the environmental day/night cycles (17). The plant's circadian machinery plays a significant role in its survival and fitness under varying environmental conditions (18). The viability of plants decreases when their endogenous clock is dysfunctional. One of the major response regulated by the clock in plants is photoperiodism (19). Photoperiodism refers to detection of changes in duration of days and nights and adapting physiological responses to the seasonal changes in the environment. Photoperiodism controls seasonal flowering in plants.
which ensures that plants flower during appropriate season (20). Plants at lower latitudes tend to flower during short-days to avoid extreme summer heat, whereas plants in more temperate climates flower during long days before the onset of winter. Additionally, it is important to match the timing of flowering with the life cycle of important pollinators such as bees, to guarantee reproductive success of the species (21). There are plenty of physiological responses governed by the circadian clock such as the leaf movement, stomatal opening, hypocotyl elongation, cotyledon movement, etc.

Figure 6: Parameters of circadian rhythm: The phase or amplitude of the circadian clock is defined as the maximum level of a circadian output (e.g., protein or mRNA), whereas the period is defined as the time between two phases of the circadian response (protein or mRNA levels).

Further, enzymes and proteins responsible for photosynthesis show high expression during dawn as compared to dusk to achieve maximum utilization of sunlight during the day. Experimentally, circadian rhythms can be studied by monitoring the gene expression or mRNA levels throughout the course of the day (4). Genes showing a cyclic expression exhibiting a peak in its expression at certain time of the day and a dip in its expression at a certain time point are said to be under circadian control. Circadian oscillations are defined by specific terms such as phase which refers to the time where the peak in expression of a particular gene occurs (time-point of maximum amplitude), period refers to the time period between two phases or the time period between two
troughs (time point of minimum amplitude) (4). Phase is often defined in zeitgeber time (ZT). Zeitgeber is a German term meaning time-giver; any external cue giving information about the time of the day is a zeitgeber (4). The onset of light is an important zeitgeber and dawn is defined as ZT0. Thus, circadian rhythms are defined by 24 hours between ZT0 to ZT24.

Previous studies have shown that 6% or more than 8000 genes are under circadian control in plants (4). These genes include the genes implicated in light harvesting reaction of photosynthesis, genes encoding photo-receptor proteins of phytochrome B (phy B), Cryptochrome 1 (CRY1), Cryptochrome 2 (CRY2), and phototropin (NPH1), genes implicated in carbon, sulfur and nitrogen metabolism, genes implicated in photoperiodic flowering and many other processes (22). The plant circadian oscillations are maintained by several genes intercepted by feedback loops (23).

The plant circadian oscillator consists of two myb transcription factors CIRCADIAN CLOCK ASSOCIATED1 (CCA1) and the LATE ELONGATED HYPOCHOTYL (LHY) and a pseudo-response regulator protein TIMING OF CAB EXPRESSION (TOC1) which maintain healthy circadian rhythm in plants (17,24,25). The expression of CCA1 and LHY itself is rhythmic and reaches peak levels at dawn (25). CCA1 and LHY then bind to specific sites on promoter regions of various genes to induce their transcription. Plants overexpressing CCA1 and LHY show arrhythmic behavior while plants lacking CCA1 and LHY show period lengthening, thus, proving that CCA1 and LHY are core clock components and mediators of circadian rhythms. CCA1 and LHY act independently to induce gene transcription (25).

During morning, CCA1 and LHY bind to evening elements on the promoter of TOC1 and repress its transcription. At dusk, as the levels of CCA1 and LHY reduce TOC1 starts expressing and the transcript levels peak at dusk. The maximum transcript levels of CCA1/LHY and TOC1
are 12 hours out of phase. TOC1, in turn, positively regulates the expression of CCA1 and LHY. Thus, CCA1 and LHY inhibits its own transcription by inhibition of TOC1 in the morning forming an autoregulatory feedback loop. CCA1 and LHY positively regulate pseudo-response regulators (PRRs) PRR5, PRR7, and PRR9 which are also members of the TOC1 family. PRR5/PRR7/PRR9 in turn, negatively regulate CCA1 and LHY, forming a second negative feedback loop (26).

Light input to the plant circadian clock is regulated by various photoreceptor proteins capable of sensing from far red to blue light (22). Phytochromes are sensitive to red light and provide early light signals to the clock. Previous research and experiments have hypothesized that at dawn phytochrome interacts with phytochrome interacting factor 3 (PIF3), a bHLH transcription factor and initiates expression of CCA1 (22,27). The exact mechanism how photoreceptors entrain the clock is not known yet. By contrast, plants lacking PhyA exhibit a long period phenotype when grown under dim red light. Similarly, plants lacking PhyB exhibit a long period phenotype under high fluence red-light and plants lacking CRY1 show a long period phenotype under blue light. These results suggested that PhyA mediates input of low red-light intensity to the clock and PhyB mediates input of high red-intensity light to clock while CRY1 mediates blue light input to the clock (22). Further, over-expression of these photo-receptor proteins leads to short-period phenotype exhibiting the significance of these proteins in entraining the biological clock. Additionally, blue light photo-receptor proteins from the ZTL family play a significant role in plant circadian clock and mediating photoperiodic responses (28,29). The role of the ZTL family in plant circadian clock will be discussed in the following section.
2.3.1 ZTL family and plant circadian clock

The Zeitlupe (ZTL) family consists of three proteins: Zeitlupe (ZTL), Flavin Binding Kelch Repeat-1 (FKF1), and LOV Kelch Protein-2 (LKP2). These proteins, along with a photo-reactive LOV domain, also possess F-box and Kelch domains. The presence of F-box and Kelch domain suggest that these proteins regulate protein stability. F-box proteins are components of the SKP–Cullin–Rbx–F-box (SCF) complex, indicating the formation of SCF E3 ubiquitin ligases. Therefore, the presence of F-box domain in the ZTL family of proteins suggests that they function as light-regulated E3 ubiquitin ligases that target proteins for degradation (30,31). Previous studies have shown that ZTL protein plays a role in plant circadian clock by regulating the stability of TOC1 protein (32,33). The transcript levels of ZTL do not show circadian oscillation but the protein level peaks at dusk. During the day, the LOV domain of ZTL protein interacts with another protein called Gigantea (GI); this interaction stabilizes ZTL (34). Upon light:dark transition during dusk, the ZTL:GI complex dissociates, and ZTL then is involved in TOC1 protein degradation.

**Figure 7: ZTL family of proteins:** Schematic representation of domains present in ZTL family of proteins.
during night. ZTL is also involved in degradation of PRR5 (31) Studies have also shown that the F-box domain of ZTL interacts with *Arabidopsis* SKP1 like protein2 (ASK2) which indicates that ZTL is a part of SCF complex and is involved in ubiquitin-mediated protein degradation (30,31). Thus, ZTL plays an essential role by regulating the stability of a critical component of plant oscillator. Plants lacking *ztl* show period lengthening. Since three members of the ZTL family show 80% amino acid similarity, they may target the same proteins for degradation. Plants lacking *fkf1* show minor circadian clock defects but a double mutant lacking *ztl* and *fkf1* shows period lengthening higher than the single deletion mutant of *ztl* itself (35). A triple mutant lacking *ztl* *fkf1* and *lkp2* shows enhanced circadian clock defects (35).

Additionally, FKF1 and LKP2 also interact with TOC1 and PRR5 (28). Studies have shown that TOC1 and PRR5 proteins are more stable in *ztl* *fkf1* and *ztl* *fkf1* *lkp2* deletion mutants as compared with a single deletion mutant *ztl*. This indicates that ZTL, FKF1, and LKP2 together regulate TOC1 and PRR5 degradation and are major contributors to determining the period of circadian oscillation (28,31,35). FKF1 also plays a vital role in regulating photo-periodic flowering. Arabidopsis is a facultative long-day plant which flowers during longer photo-periods. During long days, FKF1 regulates the gene transcription and protein stability of CONSTANS (CO) which is a crucial regulator of flowering (36,37,38). CO protein binds to the promoter region Flowering Time (FT) gene and induces its expression which in turn encodes a florigen that induces flowering.
During daytime, Cyclic Dof factor1 (CDF1) binds to the promoter regions of CO and FT and represses its transcription. During afternoon of long-days, the phase of expression of GI and FKF1 coincide and, whereas in short-days, FKF1 is transcribed three hours later than GI. The coincidence of FKF1 and GI expression is important to induce CO transcription during the light period (38,39,40). Light activates FKF1 and GI complex formation that degrades CDF1 on the promoter region of CO and FT. Degradation of CDF1 by FKF1:GI complex during afternoon of long-days enables CO transcription by the end of the day. CO transcription during the light period...
enhances FT transcription leading to induction of flowering in long days. During short-days since the expression of FKF1 and GI does not coincide and FKF1: GI complex formation does not occur therefore, CO repression by CDF1 is not relieved. Thus, FKF1 along with GI regulates photoperiodic flowering in long-day plants (40). Studies have shown that FKF1 not only regulates CO gene transcription but also interacts with CO protein during the day and stabilizes CO. Thus, the ZTL family of proteins play a significant role in regulating circadian clock and photo-periodic responses (17,28,29,35, 38,41,42).

The mechanism of operation of plant circadian machinery differs from mammalian circadian clocks. Although, both plant and mammalian clocks involve feedback mechanisms controlling transcription/translational activity, the proteins involved differ in mechanism of action. For example, CCA1 and LHY are two main plant myb transcription factors that act separately on their gene targets, not as dimers, whereas in mammalian circadian machinery the two main transcription factors BMAL1 AND CLOCK members of PAS family form heterodimer complexes to act on their gene targets. Further, in plants, various photoreactive PAS proteins sense different wavelengths of light and regulate the light input to the plant circadian machinery. Also, the plant circadian machinery is complicated due to presence of various intertwined feedback mechanisms and redundant proteins that have similar function. Below we will discuss photochemical characterization of photoreactive PAS proteins belonging to the ZTL family.

2.3.2 Photochemical characterization of plant LOV domains

As mentioned earlier, proteins belonging to the LOV domain family form a sub-class of PAS domains that bind to a flavin co-factor. The flavin moiety present in LOV domains could either be flavin mononucleotide (FMN) or flavin dinucleotide (FAD) that are sensitive to blue light (400 nm-500 nm) in the visible region of the electromagnetic spectrum (43). The ZTL family of
proteins bind to FMN as a chromophore. Photoreception by flavin changes the chemical state of the co-factor that initiates signaling by altering the protein structure (44). The initial steps occurring following the photo-excitation were studied in detail in LOV2 domains of phototropin using ultra-fast time-resolved spectroscopy measuring intermediates on a time-scale of a nanosecond (45,46).

Upon blue light absorption, a covalent adduct is formed between the flavin moiety and a conserved cysteine residue present in all LOV proteins (43). This adduct state signifies the active signaling state of the protein. Blue light excitation of the flavin moiety initiates a series of electron transfer reactions, which results in protonation of the N5 position of the flavin ring and formation of a covalent adduct at the C4a position. The mechanism of adduct formation is shown (Figure 9).

The protonation of N5 position leads to changes in the hydrogen bonding network of the protein starting from the flavin binding pocket to the other regions of the protein (47,48,49,50). These changes in hydrogen bonding network serve to transmit the signal by causing structural changes affecting inter-protein or intra-protein interactions. The adduct state then decays back to its ground state.

The ground state of a LOV protein has two major peaks at 360 nm and 450 nm with the values differing by 1-10 nm in different LOV proteins. The peak at 450 nm is characteristic of a free flavin, and it also has two shoulder peaks at 425 nm and 475nm representing the vibronic states. The cysteine-flavin adduct formation bleaches the characteristic free flavin peak at 450 nm and forms a single peak at 390 nm (46). Mutation of the active site cysteine residue in LOV domains abolishes adduct formation and formation of peak at 390 nm suggesting that spectral changes correspond to adduct formation (51). Of particular interest to the manuscript, is the rate
at which the adduct decays back to its ground state varies in different LOV proteins. Some proteins undergo extremely fast decay in seconds like phototropins, whereas, some proteins remain in the adduct state for longer times requiring hours or days to return to its ground state (43,52). Previous studies utilizing techniques such as proton inventory and solvent isotope effect have shown that the rate-limiting step during the course of adduct-decay is N5 deprotonation (52). To calculate the solvent isotope effect, the rate of adduct decay is measured in buffer containing H$_2$O and D$_2$O. The rate of adduct decay decreases in presence of buffer containing D$_2$O as it takes time to abstract a heavier deuterium atom than a hydrogen atom from N5. Thus, such experiments showed that N5 deprotonation is the rate-limiting step.
Figure 9: Mechanism of adduct formation. A) Blue light excitation of flavin results in electron transfer from a nearby cysteine to the N5 position, which then abstracts a proton from the cysteine residue, which creates a free radical at C4 position of flavin and at the sulphur of cysteine which then undergo radical recombination to form a cysteine-flavin adduct. B) Light and dark state spectra of an LOV domain.
2.3.3 *Photo-cycle kinetics of FKF1 and LKP2*

We measured the rate of adduct decay in FKF1 and LKP2 using a UV-Vis spectrophotometer. Photo-cycle studies were performed on a protein construct just containing the LOV domain FKF1 28-174 and LKP2 16-165. The F-box and Kelch domains were excluded due to poor expression of full-length proteins in bacterial systems. The light state of the protein was generated by using a 450 nm (Z-bolt Sapphire) laser or broad-spectrum white flood light (EKE 21V 150W, Phillips). The generation of the light-state was confirmed by bleaching the dark-state peak at 450nm and formation of a single peak at 390 nm. After generation of the light-state, UV-Vis spectra were measured at regular intervals for several hours until the dark state peak at 450 nm fully recovered. The rate of adduct decay was calculated by obtaining the absorbance values at three specific wavelengths, 450 nm, 478 nm, and 600 nm at specified time intervals. The absorbance at 450 nm and 478 nm represents the dark state peak. The absorbance at 600 nm serves as a background or control as it does not change in the light- or dark-state. The absorbance values at 450nm and 478nm were then plotted against time. Using a mono-exponential fit function in Origin Lab software, the time constants at 450 nm and 478 nm were calculated. Finally, the average time constants were calculated from at least three biological replicates. The rate constants were calculated from time constants using a first-order kinetics rate equation. Figure 10A-C show light to dark recovery plots with mono-exponential fitting for both 450 nm and 478 nm wavelengths for LKP2, FKF1, and ZTL LOV domains. The photo-cycle kinetics data for ZTL was obtained from Ashutosh Pudasaini (33).

LKP2 and FKF1 have extremely long photo-cycles and are difficult to measure on an experimental timescale. Therefore, a small base like imidazole was used that accelerated adduct scission by N5 proton abstraction. The photo-cycle kinetics of FKF1 and LKP2 was measured with
different concentrations of imidazole, in triplicates, and the rate constants were calculated as discussed earlier. The photo-cycle of FKF1 and LKP2 showed a linear dependence to imidazole concentration, i.e., the rate of adduct decay increased with increasing concentrations of imidazole. The lifetime of the adduct state with no imidazole was extrapolated by plotting rate constants against corresponding concentration of imidazole.
Figure 10: Photo-cycle kinetics of FKF1, ZTL, and LKP2: Mono-exponential fitting of light to dark state recovery at 450 nm (black circles/line) and 478 nm (red circles /line) of A) ZTL, B) LKP2, and C) ZTL. The FKF1 and LKP2 experiments were base catalyzed at 300 mM imidazole and 100 mM imidazole respectively, to enhance the recovery rates. D) Linear plots of rate constants v/s imidazole concentration of LKP2 (D) and FKF1 (E). F) Summary of rate constants (k) and time constants (τ) of light to dark state recovery in ZTL, FKF1, and LKP2. The time constants (τ) of light to dark state recovery vary from few hours to several days in the ZTL family. Both FKF1 and LKP2 have extremely large time constants.
The use of a small base like imidazole also helps predict solvent accessibility of a protein (45,52). Proteins with high solvent accessibility show greater imidazole dependence. Although belonging to same family, ZTL has an extremely fast rate of adduct scission of about ($\tau=1.4-1.7$ hrs) as compared to LKP2 ($\tau=136-138$ hrs) and FKF1 ($\tau=110-112$ hrs). The lifetime of the adduct state may also signify the biological function of the protein. Plant photoreceptors FKF1 and LKP2 have prolonged decay rates. It has been shown that light-dark recovery rates dictate the population of active state molecules of protein present in the plants (24). FKF1 that has an extremely slow light-dark kinetics can have a high population of protein in its active state throughout the day, thereby enabling FKF1 to remain active and degrade the repressors of flowering during long days and induce flowering.

**2.3.4 Photo-cycle kinetics of rate altering variants**

Previous studies in the fungal photoreceptor protein VVD have shown that the rate of adduct decay can be modulated by substitution of active site residues (52). Substitutions made near the active site can affect the conformation of the conserved cysteine or alter the electronic state of flavin and can also alter the solvent accessibility to the active site thereby can either accelerate or decrease the rate of adduct scission (47,48,52). In VVD substitution of an isoleucine residue within the flavin active site region to a valine (I74V) residue increased the rate of adduct decay by 25-fold (52). Crystal structures revealed that the isoleucine residue projects a methyl group within the van der Waals contact to the active site cysteine residue. The methyl group sterically stabilizes the adduct state of the protein. Substitution to a valine residue removes the methyl group within the active site region and accelerates adduct decay by steric destabilization. To evaluate the effect of equivalent sites in FKF1, we examined the FKF1 sequence and created mutants to attempt to accelerate FKF1 kinetics.
Sequence comparison shows that FKF1 also has a corresponding isoleucine residue within the similar region. Substitution of this residue with a valine (I57V) increased the rate of adduct decay by 7-fold. Whereas, substitution of this residue to a threonine (I57T) increased the rate of adduct decay by 9-fold. Threonine is a polar residue, and it may also increase the solvent accessibility to the active site in addition to steric destabilization of the cysteine-flavin adduct.

Although these substitutions accelerated FKF1 kinetics as expected, these mutants still exhibited adduct decay rates substantially slower than ZTL, VVD, and other LOV proteins. To further accelerate the adduct decay rate of FKF1, we examined other residues known to affect LOV kinetics in other proteins. Previous studies on VVD showed that there is another isoleucine residue within the active site Ile85 which also projects a methyl group towards the active site cysteine. Mutation of this residue to valine I85V increased the rate of adduct scission by 25-fold as compared to wildtype VVD (52). Moreover, a double variant I74V:I85V increased the rate of adduct scission by 600-fold (52). The residue Ile85 is conserved in LOV domain family except FKF1 that has a leucine residue at the corresponding position (Leu68). A double variant I57T:L68T in FKF1 increased the rate of adduct scission dramatically by 2200-fold. The half-life of the adduct is approximately three mins. In this regard, the double mutant can convert FKF1 from one of the slowest LOV photocycles known, to a photocycle comparable to fast cycling LOV proteins (e.g., phototropins). This dramatic decrease in adduct lifetime can be attributed to steric destabilization of the cysteine-flavin adduct and an increase in solvent accessibility due to the presence of two polar threonine residues within the active site.

The most conserved stretch observed in all LOV domain family across various taxa is the consensus sequence of GXNCRFLQ that also contains the active site cysteine responsible for forming cysteine-flavin adduct (70). The X residue can be any amino acid and varies across the
LOV domain family. Upon sequence comparison of ZTL family of proteins, it was observed that ZTL has a consensus sequence of GGNCRFLQ whereas LKP2 and FKF1 have GRNCRFLQ. The presence of glycine (Gly80) instead of arginine in LKP2 and FKF1 within the flavin active site may contribute to fast adduct-state decay in ZTL (33). Glycine only has a single hydrogen atom as its side chain as compared to arginine which has a large side chain. Therefore, the presence of glycine can make the active site more solvent accessible and can also affect the conformation of the active site cysteine. Mutation of the corresponding arginine to glycine in FKF1 on I57V background I5V:R89G increased the rate of adduct decay by 28-fold. The half-life of the double variant I57V:R89G is $\tau=4$ hrs which is only approximately 3-fold higher than ZTL.

Thus, we can conclude that evolutionary selection of a Val at position 48 and Gly at position 80 (I57 and R89 in FKF1 respectively) account for the faster photocycle kinetics observed in ZTL that are required for ZTL function in the evening.
Figure 11: Photo-cycle kinetics of rate altering variants of FKF1: Linear plot of rate constant of adduct decay v/s imidazole concentration for FKF1 WT (28-174) (A) FKF1 I57V (B) FKF1 I57T L68T (C) Error bars are depicted as ±10%, reflecting errors in sample construction. D) Comparative analysis of imidazole dependence of FKF1 WT (black) FKF1 I57V (red) FKF1 I57T L68T (blue). Mutant FKF1 I57V shows only a slight increase (1.3 fold) in imidazole dependence compared to the wild-type FKF1 whereas mutant FKF1 I57T L68T shows an increase of 5144-fold. E) Imidazole dependence comparison of FKF1 WT and FKF1 I57V is shown here as not apparent in Fig 11D. F) Sequence alignment of ZTL family of proteins and VVD.
Table 1: Kinetic recovery of rate altering variants of FKF1: The table shows the rate of adduct-decay and half-life of various FKF1 mutants. The rate constant for FKF1 WT was calculated by base-catalysis studies. The rate constants and half-life are averaged over three sets of experiments.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Rate Constant, k (s⁻¹)</th>
<th>Time constant (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FKF1 28-174 WT</td>
<td>2.5*10⁻⁶</td>
<td>6600</td>
</tr>
<tr>
<td>FKF1 28-174 I57V</td>
<td>1.913*10⁻⁵</td>
<td>900</td>
</tr>
<tr>
<td>FKF1 28-174 I57T</td>
<td>2.3*10⁻⁵</td>
<td>740</td>
</tr>
<tr>
<td>FKF1 28-174 I57T</td>
<td>5.6 *10⁻³</td>
<td>3</td>
</tr>
<tr>
<td>FKF1 28-174 I57V</td>
<td>6.8*10⁻³</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The rate of adduct-decay in LKP2 is even slower than FKF1. Interestingly LKP2 has a leucine residue at position 153, which corresponds to a conserved glutamine residue in ZTL, FKF1, VVD and other LOV domain proteins. In VVD, Gln182 undergoes a 180° rotation as the N5 position of flavin gets protonated upon light activation and is located within the LOV active site. The rotation of glutamine is critical to initiate conformational changes and signaling upon light activation in VVD and other LOV proteins. In fact, a Gln at this position is considered required for LOV function, and mutations at this site abrogate function in all known LOV proteins.
Therefore, Leu153 in LKP2 is unusual and indicates a possible divergent mechanism of LKP2-like proteins.

To examine the effect of Leu153 on LKP2 photokinetics, a L153Q point mutation was prepared. Examination of the L153Q photocycle indicated that mutation to glutamine increased the rate of adduct decay by 14-fold (Figure 12B and Table2). The half-life of the mutant L153Q was found to be about 10 hrs (Table 2). Mutation of a hydrophobic residue (Leu) to a polar residue (Gln) may increase the solvent accessibility. Indeed, mutant L153Q exhibited a greater imidazole dependence compared to the wild-type LKP2; the results indicated a 10-fold increase in imidazole dependence (Figure 12D).
Figure 12: Base Catalysis Studies of LKP2 WT and LKP2 L153Q: A) Dark state (black) and Light state (red) spectra of LKP2. Light state spectra (red) has a peak at 372 nm corresponding to the adduct state. The dark state spectrum (black) has a peak at 442 nm corresponding to $S_0$-$S_1$ transition and two vibrational bands at 417 nm and 467 nm. B) Imidazole catalyzed Light-Dark Recovery Kinetics of LKP2 16-165 exhibited linear dependence of adduct decay to various concentrations of imidazole. Error bars are depicted as $\pm$ 10%, reflecting errors in sample construction. C) Imidazole catalyzed Light-Dark Recovery Kinetics of mutant LKP2 L153Q 16-165 also exhibited linear dependence of adduct decay to various concentrations of imidazole. D) Comparison of imidazole dependence of wild-type LKP2 (black) and Mutant LKP2 L153Q (red). Mutant LKP2 L153Q shows a 10-fold increase in imidazole dependence compared to Wild-type LKP2. E) Protein sequence comparison of ZTL, LKP2, and FKF1 and VVD indicates that LKP2 has a leucine instead of a conserved glutamine.
Further, examination of the LKP2 sequence revealed the presence of an arginine residue in the conserved sequence GRCRNFLQ whereas ZTL has a glycine instead of arginine. As discussed earlier, the fast rate of adduct decay in ZTL is due to presence of this glycine (G80) residue at this position. Therefore, arginine 80 (R80) in LKP2 was mutated to a glycine to see if it affected the rate of adduct-decay. The mutation was made on a construct similar to ZTL-LOV construct on which kinetic studies were performed, and it also contained a mutation of L153Q (29-164 L153Q R80G). The rate of adduct decay in this double variant was very similar to ZTL. Thus the presence of a leucine residue (L153) and arginine (R80) in LKP2 dictate divergence from ZTL-like photo-cycle kinetics.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Rate Constant, k (s(^{-1}))</th>
<th>Time constant (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LKP2 16-165 WT</td>
<td>2.04 \times 10^{-6}</td>
<td>8160</td>
</tr>
<tr>
<td>LKP2 16-165 L153Q</td>
<td>3.6 \times 10^{-5}</td>
<td>480</td>
</tr>
<tr>
<td>LKP2 29-164 L153Q R80G</td>
<td>1.26 \times 10^{-4}</td>
<td>132</td>
</tr>
</tbody>
</table>

**Table 2: Kinetic recovery of rate altering variants of LKP2:** The table shows the rate of adduct-decay and half-life of some LKP2 mutants. The rate constant for LKP2 WT was calculated by base-catalysis studies. The rate constants and half-life are averaged over three sets of experiments.

In summary, the rate of adduct decay rate of a photo-receptor protein can be modulated by substitution of active site residues. These tunable photo-cycle lifetimes give a better control to design of optogenetic tools. FKF1 is currently being utilized as an optogenetic tool to control gene expression in mammalian cell cultures. Therefore, tunable photo-cycle kinetics makes these photoreceptors a promising tool for optogenetics and also, they are blue-light sensitive, which is
within the visible light spectrum harmless to living organisms. Further understanding of its signaling mechanisms and identifying its degradation targets will provide useful information to improve crop productivity. In order to gain knowledge of its signaling mechanisms, efforts were taken to structurally characterize these proteins which will be discussed in the next chapter.
2.4 Material and methods

2.4.1 Cloning and protein purification of LKP2 and FKF1 proteins

The nucleotides encoding LKP2 and FKF1 were obtained from The *Arabidopsis* Information Resource (Tair). LOV Domain region of LKP2 and FKF1 were cloned into 6His, and pGST parallel vector using NcoI, SpeI, Bam-hI and HindIII cut sites respectively. Point mutations I57V, I57T, L68T, R89G, I160R, and L153Q were introduced into FKF1 28-174 and LKP2 16-165 constructs using Quick-change protocol (Strata-gene).

All 6His constructs of LKP2 were expressed in *Escherichia coli* BL21DE3 cells, and pGST constructs of FKF1 were expressed in *Escherichia coli* JM109 cells. Cells were grown at 37 °C until an OD at 600 nm of approximately 0.6 was reached. The temperature was then decreased to 18°C for 40 min to assure thermal equilibrium. At 18 °C 0.3 mM isopropylthiogalactoside (RPI) was added to initiate protein expression. After 22 hrs the bacterial pellets were harvested and stored in a buffer containing 50 mM Hepes, 100 mM NaCl, 10% Glycerol pH 8.0.

For protein purification, bacterial pellets were thawed and sonicated for four cycles. After sonication, the soluble cell lysate was fractionated by centrifugation at 18,000 rpm for 1 hr. All 6-His constructs of LKP2 were purifed by Ni-NTA affinity chromatography. After binding the columns were cleaned with buffer containing 50 mM Hepes, 100 mM NaCl, 20 mM imidazole, 10 % glycerol pH 8.0 to wash away non-specifically bound proteins. The bound protein was then eluted from the column with buffer containing 50 mM Hepes, 100 mM NaCl, 500mM imidazole, 10 % glycerol pH 8.0. The eluted protein was treated with 2 mg/ml of TEV protease to cleave the 6His tag. The protein was then incubated with TEV for 2 hrs at 22 °C or overnight at 4 °C. An additional round of Ni-NTA chromatography was carried out to remove 6His-TEV.
The protein was further purified using a Superdex S200 size exclusion column (GE Lifesciences) equilibrated with 50 mM Hepes, 100 mM NaCl, 10% glycerol pH 8.0. The corresponding protein fractions from size exclusion column was collected and verified using SDS-PAGE and UV-vis spectroscopic methods. The pure protein was then concentrated to desired concentration to carry out further experiments.

All constructs of FKF1 in pGST vector were purified using glutathione resin (QIAGEN) at 22 °C. After binding protein was treated on column with 2 mg of TEV protease per milliliter of resin overnight at 4 °C. Proteins were then washed from the column with buffer. An additional round of Ni-NTA chromatography was conducted to remove 6His-TEV. The proteins was further purified using Superdex S200 size exclusion column as mentioned.

2.4.2 Determination of protein concentration

The concentration of protein was determined using a UV-spectrometer. The absorbance at 380nm was measured. The protein concentration was calculated using beer-lamberts law.

Beer’s Law:

\[ A_{380} = \varepsilon_{380}bc \quad (1) \]

where \( A_{380} \) is the UV-Vis absorbance value at 380 nm (Light state peak), the extinction coefficient at 380 nm (\( \varepsilon_{380} \)) is 8500 mole\(^{-1}\) cm\(^{-1}\), \( b \) is path length (cm), and \( c \) is the concentration (M). The extinction coefficient at 380 nm was calculated using the known value at 450 nm from previous studies (\( \varepsilon_{450} = 12,500.0 \) mole\(^{-1}\) cm\(^{-1}\)). Keeping the protein concentration (\( c \)) and path length (\( b \)) constant we can estimate \( \varepsilon_{380} \) by using the following equation:

\[ \frac{A_{380}}{A_{450}} = \frac{\varepsilon_{380}}{\varepsilon_{450}} \quad (2) \]
2.4.3 UV-Vis absorbance spectroscopy and kinetics

UV–visible absorbance spectroscopy for all constructs and mutants of LKP2 and FKF1 were conducted on Agilent 8453 spectrophotometer. Photo-physical properties of all constructs were verified to show LOV type chemistry exhibiting a single characteristic peak around 380-400 nm for C4a adduct and two broad absorption peaks for the ground state. The ground state spectra consisted of a free flavin peak at 450 nm and two vibrational bands at 425 nm and 475 nm. The LKP2 protein samples were concentrated down to 100µM, and FKF1 protein samples were concentrated to 30 µM for measurement of light-dark recovery rates. Protein samples were exposed to a broad-spectrum white floodlight source (150 W) while being incubated on ice to populate the light state. Kinetic measurements of light-dark recovery rates were determined using a spectrometer where light adapted protein's spectra was obtained at regular time intervals in a 1.0 cm path length cuvette. The absorbance values at 450 and 478 nm representing the dark state was noted down for each spectral measurement. All values for 450 and 478 nm were corrected for deviations in the baseline by subtracting the absorbance at 600 nm. The net absorbance values at 450nm and 478nm were plotted against time, using a scatter plot in Origin software. Data were fit using mono- and biexponential equations as required to extract kinetic parameters.

\[ y = A_{450}e^{-\frac{t}{t_{450}}} + y_0 \]  

\[ y = A_{478}e^{-\frac{t}{t_{478}}} + y_0 \]
Where \( A_{450} \) and \( A_{478} \) is the amplitude of absorption of protein at 450 nm and 478 nm wavelengths respectively. Whereas, \( t_{450} \) and \( t_{478} \) are time constants of exponential fits for absorbance at 450 nm and 478 nm respectively. To obtain the rate constant for exponential recovery of the dark state (or decay of light state), the average thermal time constant (\( \tau \)) was used as shown.

\[
\tau = \frac{t_{450} + t_{478}}{2} \quad (5)
\]

\[
k = \frac{1}{\tau} \quad (6)
\]

All time constants were reported as 1/ \( k_{\text{adduct scission}} \) that are averaged between the values obtained at 450 and 478 nm.

2.4.4 Base catalysis studies

Light-dark recovery rates were studied with help of a small base imidazole to expedite the conversion rates. Imidazole acts as a base to catalyze adduct scission by abstraction of N5. Stock solutions containing 800 mM imidazole in 100 mM NaCl, 50 mM Hepes (pH 8.0), and 10 % glycerol were prepared. Light-dark recovery rates were measured at various concentrations of imidazole ranging from 100 mM-600 mM imidazole. Each data set at a particular concentration of imidazole was repeated thrice. The protein concentration was maintained constant for base-catalysis studies.
REFERENCES


CHAPTER 3:

F-BOX PROTEINS AND CHARACTERIZATION OF PLANT F-BOX PROTEINS FKF1 AND LKP2

3.1 Introduction to the F-box protein family

As already discussed in the previous chapter, the ZTL family of proteins play a significant role in maintaining circadian rhythm in plants. These proteins also belong to F-box family as they contain an F-box domain that mediates protein degradation. F-box proteins are a large group of proteins found in eukaryotes and are mainly involved in recruiting substrates for degradation. F-box domain was first identified in 1995 as a homologous region in mammalian proteins Cdc4, β-TrCP, Met30, Scon2, and MD6 (2). Later in 1996, they were recognized as a motif for protein:protein interaction. They were first characterized as a component of SKP1-cullin-F-box ubiquitin ligase complexes (SCF) (1,2). The SCF complex regulates ubiquitin-mediated protein degradation. The function of F-box protein is to interact and recruit protein substrates to the SCF complex to undergo ubiquitin-mediated degradation. Usually, F-box domain is about 50 amino acid long and is present at the amino terminal of the protein; the carboxy-terminal consists of domains that can interact with protein substrates. The C-terminal domain of F-box proteins in humans is comprised of leucine-rich regions (LRR) or WD repeats (3). These regions are responsible for interacting with protein substrates while the F-box domain interacts with the SCF complex. Thus, F-box proteins form a platform for attaching protein substrates to SCF complex for protein degradation.
There are 11 F-box proteins found in yeast *Saccharomyces cerevisiae*, about 326 predicted in nematode worm *Caenorhabditis elegans*, 22 in *Drosophila*, at least 38 in humans and about 100 or more in *Arabidopsis thaliana* (4). It was estimated that higher than 80% of proteins undergo ubiquitin-mediated degradation (5). F-box domains are attached to wide variety C-terminal domains that interact with protein substrates such as proline-rich regions, zinc fingers, cyclin domains, leucine zippers, and ring fingers. F-box proteins are classified according to the C-terminal domain attached to the F-box domain for interacting with protein substrates. In humans, there are three families of F-box proteins, the FBXL family contains an F-box domain and an LRR region at its C-terminal, the FBXW family contains an F-box domain and a WD domain at its C-terminal, the FBXO family includes an F-box and either any other domain or no domain at its C-terminal (5,6). Proteins belonging to the FBXL and FBXW family are conserved in *C.elegans* and yeast, whereas only some of the proteins of the FBXO family are conserved in yeast.

In plants, there are two F-box families, the FBK family that contain an F-box domain and a kelch repeat domain at the C-terminal and the FBA family that contain F-box domain and an F-box associated domain at its C-terminal (1,2). The ZTL family of proteins belong to the FBK family that contains kelch repeats for protein substrate interaction (7). The presence of kelch repeats at the C-terminal is more specific to plant F-box proteins (1,7). Both kelch repeats and WD-40 domain form β-propeller structures (8,9). β propeller structures are specialized in protein: protein interaction and current models predict the ZTL family functions as obligate dimers.
Protein degradation is an important post-translational regulatory mechanism that helps maintain cellular homeostasis by degrading misfolded proteins. Also, protein degradation is required for certain cellular processes to occur in a time of day specific manner. As mentioned earlier the SCF E3 ubiquitin ligase complex plays a major role in cellular protein degradation by ubiquitination. (9,10,11). Ubiquitin is a small protein of about 8-9 kDa that is mostly present in all eukaryotic tissues (15,16). The process of ubiquitination requires three steps of activation, conjugation, and ligation. The ubiquitin activation takes place in two steps where E1 ubiquitin-activating enzyme adenylates one ubiquitin molecule utilizing ATP (12,13,14). The adenylated ubiquitin then forms a thioester linkage at cysteine residue of the E1 enzyme. During, conjugation, the ubiquitin molecule is transferred from E1 to the cysteine residue of E2 ubiquitin-conjugating enzyme via trans-thioesterification reaction (13). Finally, E3 ubiquitin ligase transfers the ubiquitin molecule from E2 to the lysine residue of the protein substrate to be marked for degradation. The E3 ubiquitin ligase is a multiprotein complex also known as the Skp1-Cullin-F-box (SCF)

3.2 Mechanism of protein degradation by SCF complexes

Figure 13: Schematic representation of F-box attached to various C-terminal domains:
The F-box is connected to different types of C-terminal domains that attach to specific protein substrates to target for protein degradation.
complex). The E3 ligase is composed of four proteins, S phase kinase-associated protein 1 (Skp1), a Cullin 1 (CUL1), RING-box protein-1 (Rbx-1) and an F-box protein. The CUL1 is a scaffold protein that brings other proteins in the complex together (17,18). The carboxyl-terminal of CUL1 attaches to the Rbx1 protein, and amino terminus binds to the Skp1 protein. The F-box protein attaches to the protein substrate to be degraded through its C-terminal domain. The F-box domain at the N-terminal of F-box protein attaches to the Skp1 protein to create a link to CUL1(17). Due to a certain stimulus, the F-box protein attaches to the protein substrate and recruits it to the SCF scaffold via Skp1 (17). The Rbx1 protein binds to the E2 ubiquitin-conjugating enzyme and directly transfers the ubiquitin molecule from E2 to a lysine residue on the protein substrate. After the first ubiquitin has been transferred to the lysine residue of protein substrate, the second ubiquitin molecule attaches to the first ubiquitin molecule, and subsequently, other ubiquitin molecules attach to the previous ones forming a chain of ubiquitin molecules on the lysine residue of the protein substrate (20,21). At least four ubiquitin molecules should be attached to the substrate for it to undergo degradation. After ubiquitination, the ubiquitin-interacting motif (UIM) on the proteasome recognizes the protein substrate for protein degradation (19). The proteasome is a large cylindrical structure that contains proteolytic enzymes called proteases that degrade proteins to smaller peptides of 8-9 amino acids (11).
The C-terminal domain of F-box proteins is responsible for binding to protein substrates. F-box proteins are attached to a variety of C-terminal domains that provide specificity to different protein substrates. F-box proteins recruit substrates for degradation in response to a specific stimulus. There are short defined regions on the substrate protein called as degradation motifs or degrons that interact with the F-box proteins (22). In most cases phosphorylation of these degradation motifs or degrons trigger binding to F-box proteins for degradation. β transducin repeat-containing protein (βTrCP) an F-box protein from humans binds to a consensus degron Asp-Ser-Gly-Xaa-Xaa-Ser (Xaa refers to any amino acid) on its substrates after both the serine residues are phosphorylated (23,24). There are certain variations to phosphorylation mediated substrate recruitment. Sometimes a degron can be phosphorylated by two different kinase enzymes, For example, cyclin-dependent kinase-2 (CDK2) and glycogen synthase kinase 3

Figure 14: SCF E3 ubiquitin ligase: Schematic representation of ubiquitination of substrate protein by the SCF complex.

3.3 F-box and substrate recruitment

The C-terminal domain of F-box proteins is responsible for binding to protein substrates. F-box proteins are attached to a variety of C-terminal domains that provide specificity to different protein substrates. F-box proteins recruit substrates for degradation in response to a specific stimulus. There are short defined regions on the substrate protein called as degradation motifs or degrons that interact with the F-box proteins (22). In most cases phosphorylation of these degradation motifs or degrons trigger binding to F-box proteins for degradation. β transducin repeat-containing protein (βTrCP) an F-box protein from humans binds to a consensus degron Asp-Ser-Gly-Xaa-Xaa-Ser (Xaa refers to any amino acid) on its substrates after both the serine residues are phosphorylated (23,24). There are certain variations to phosphorylation mediated substrate recruitment. Sometimes a degron can be phosphorylated by two different kinase enzymes, For example, cyclin-dependent kinase-2 (CDK2) and glycogen synthase kinase 3
(GSK3) both phosphorylate different residues on Cyclin E degron (25). Degron phosphorylation sometimes requires priming phosphorylations that are nearby to the degron but not at the substrate-F-box interaction surface.

In addition to phosphorylation, some F-box proteins require accessory proteins or cofactors for substrate recognition. Human F-box protein FBXL requires an accessory protein cyclin-dependent kinase regulatory subunit 1 (CKS1) for binding to phosphorylated CDK inhibitor protein p27 (26,27). FBXO4 also uses a chaperone, αβ crystalline, for substrate recruitment (28,29). There are some cases where modification of degron motif by phosphorylation is not required for substrate recruitment to F-box. Cyclin F an F-box protein from humans recognizes an Arg-Xaa-Ile/leu motif on its substrates and targets them for degradation. Since Arg-Xaa-Ile/leu degron is independent of any modification, the substrates containing this motif would be degraded constitutively unless the presence of additional regulatory mechanisms (30,31). Cyclin F can recognize and access the Arg-Xaa-Ile/leu degron on its substrate ribonucleotide reductase M2 enzyme (RRM2) only when its phosphorylated by Cyclin-dependent kinase at Thr33. Thr33 is not a part of the degron, but its phosphorylation enables access of Arg-Xaa-Ile/leu to Cyclin F (30,31). Substrate recognition by F-box protein is also affected by subcellular-localization of protein substrate and F-box protein. RRM2 is a cytoplasmic protein whereas cyclin F is a nuclear protein; therefore, nuclear import of RRM2 is required for its degradation.

Notably, substrate recruitment and ubiquitination by F-box proteins are sometimes controlled by sensory domains attached to the F-box domain that senses environmental cues and regulates the activity of the F-box protein. For example, FBXL5 controls degradation of iron regulatory protein 2 (IRP2) in the presence of high iron levels (32,33). IRP2 regulates the expression of proteins required for iron metabolism. In the presence of high iron levels, iron non-
covalently binds to the haemerythrin domain in FBXL5 and stabilizes it (34,35,36). In the presence of low iron levels, haemerythrin domain cannot bind iron causing FBXL5 unfolding and degradation (35,36). Therefore, high iron levels control FBXL5 stability and subsequent degradation of the IRP2 protein. Similarly, plant F-box proteins LKP2, ZTL, and FKF1 have an LOV domain attached to the F-box domain that senses blue light from the environment and controls light-mediated degradation activity of these proteins (37). It has been shown that ZTL and FKF1 form a complex with Gigantea (GI) during the day which stabilizes these proteins (38). Also, studies in ZTL has shown that during the day ZTL: GI interaction stabilizes both proteins and upon day to night transition, ZTL dissociates from GI and targets TOC1 for degradation (39). Therefore, light perception by LOV domain leads to stabilization of these proteins similar to FBXL5 where iron attachment to haemerythrin domain stabilizes them.

So far, we have seen phosphorylation of degrons as a stimulus for substrate binding, but there are also other types of covalent modifications that regulate substrate recognition, FBXO2 and FBX06 bind to glycosylated substrates via F-box associated domains (FBA) (40,41). Lastly, some F-box proteins bind to a substrate in a domain dependent manner where they recognize a particular domain independent of any modification. FBXL3 binds to cryptochrome 1 and cryptochrome 2 in a modification independent manner and mediates its degradation (42,43)

### 3.4 F-box dimerization

Dimerization has been observed in some F-box proteins that may or may not regulate substrate binding and ubiquitylation. F-box dimers also facilitate the formation of the SCF complex dimer. In F-box proteins the F-box domain is responsible for forming contacts with the Skp1 protein and C-terminal is responsible for substrate recruitment. Therefore, dimeric contacts are mediated either by N-terminal domain preceding the F-box or the linker region connecting the
F-box and C-terminal domain. The latter case is exemplified by a human F-box protein FBX4. FBX4 dimerization regulates ubiquitination of substrate Telomere regulator Pin 2 (44). The crystal structure of Fbx4-Skp1 complex showed that it formed an antiparallel dimer where the linker region of one molecule formed dimeric contacts with the C-terminal domain of the other molecule (44). The C-terminal domains where on opposite sides of the dimer therefore substrate recruitment would not be hindered by dimer formation. Further ubiquitination assays showed that Pin2 ubiquitination was inefficient when a F-box construct not capable of dimerization was used. The FBX4 construct capable of dimerization efficiently added ubiquitin conjugates to Pin2. Therefore, FBX4 dimerization is essential for ubiquitination of Pin2.

βTrCP1 and βTrCP2 are F-box proteins that regulate degradation of nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor alpha (Ikβα) by homodimer formation (45). Phosphorylation of Ikβα stimulates its degradation by βTrCP1 and βTrCP2 (45). Ikβα binds to nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kβ) and inhibits its activity. Degradation of Ikβα allows NF-kβ to translocate to the nucleus and induce transcription of variety of genes involved in cell survival and immune responses. The N-terminal domain of βTrCP1 and βTrCP2 preceding the F-box domain also known as the D-domain is responsible for forming dimeric contacts. The homodimer of βTrCP1 as well as βTrCP2 and not the heterodimer is involved in ubiquitination of Ikβα. Although, hetero-dimeric complexes of βTrCP1 and βTrCP2 form they are not involved in degradation of Ikβα. Thus, homodimer formation utilizing the N-terminal domain preceding the F-box domain in βTrCP1 and βTrCP2 regulates degradation of its substrate.

ZTL/LKP2 and FKF1 contain an LOV domain preceding the F-box domain. As discussed previously, LOV proteins act as protein: protein interaction modules and also function as homo
and heterodimers. Therefore, since LOV proteins tend to dimerize and previous literature showed that the N-terminal domain preceding the F-box domain forms dimeric contacts in some proteins, it is hypothesized that the ZTL family of proteins mediate dimerization via LOV domain and this homo or heterodimers regulate substrate ubiquitination and degradation.

3.5 Structural characterization of LKP2 and FKF1

The ZTL family of proteins contain LOV domain, F-box domain, and a Kelch domain. We were not able to express full-length proteins in bacterial cultures due to instability and misfolding therefore, we focused on shorter protein constructs containing just the LOV domain with N/C terminal extensions flanking the LOV core. Unfortunately, the shorter LOV constructs FKF1 28-174 and LKP2 16-165 demonstrated long lifetimes of adduct decay rendering them not suitable for crystallization studies. Attempts to crystallize these constructs yielded no results. Therefore, structural efforts focused on rate altering variants of FKF1 and LKP2 with fast rates of adduct decay.

We attempted to crystallize FKF1 28-174 I57V that has a higher rate of adduct decay than FKF1 WT. The protein FKF1 28-174 I57V was grown in bacterial cultures and purified by affinity chromatography. The purified protein was used in crystallization attempts utilizing a series of crystallography reagents and buffers from Hampton research. Hanging-drop method of crystallization was used where an equal ratio of protein and the crystal screening reagent were mixed on a transparent glass coverslip and sealed facing a reservoir of the same screening agent. Vapor diffusion from the mixture of protein and screening reagent and the screening reagents in the reservoir promotes phase separation and crystal growth. The purified FKF1 I57V protein was mixed with more than 100 individual screens and incubated in the dark, at 22 °C. We observed FKF1 I57V crystals in about 2-3 months of incubation under a particular-screen condition. More
crystal trays were set by modifying the screen conditions further to obtain better quality crystals. These dark grown crystals were sent for X-ray diffraction. Unfortunately, these crystals diffracted very poorly, and so structural information could not be obtained from these crystals. The fact that these crystals start to appear on a timeframe of 2-3 months suggested that the protein undergoes degradation to a stable construct and then crystallizes. An SDS PAGE of an obtained crystal confirmed protein degradation. We then attempted to crystallize the double variant that has an extremely fast adduct decay of 3 mins FKF1 28-174 I57T: L68T. Crystals appeared in an about a month in the same screen condition as FKF1 28-174 I57V. Protein degradation was also found by SDS-PAGE in this double variant. We also attempted to crystallize FKF1 28-174 I57T L68T by adding proteolytic enzymes such as chymotrypsin to aid in protein degradation. Crystals appeared in two days by addition of chymotrypsin but even these crystals diffracted poorly.

To improve crystallization, we attempted to clone and crystallize the degradation product for which crystals readily formed. Toward this aim, we made shorter length constructs that could be stable for crystallization such as FKF1 37-174 I57T L68T, FKF1 41-171 I57T L68T, and FKF1 41-178 I57T L68T. Crystals were observed in two of these shorter constructs, but none of them diffracted to required resolution. Attempts to crystallize LKP2 16-165 and LKP2 16-165 L153Q also failed. The intrinsic disorder in these proteins makes them extremely hard to crystallize. Nevertheless, crystal structure information for ZTL LOV domain provided by Ashutosh Pudasaini (Graduate student, Zoltowski lab, SMU) can provide useful insights about FKF1 and LKP2 structure (46).

PAS/ LOV domains are known for protein: protein interactions and forming homo and heterodimers. Size-exclusion chromatography (SEC) studies of ZTL family of proteins revealed that they exist as dimers in solution. The crystal structure of ZTL-LOV domain consists of a
tetramer within the asymmetric unit (48). Due to the dimeric nature of ZTL in solution, it is possible that the tetrameric architecture is the result of crystal packing and an artifact of crystal growth. The ZTL-LOV structure contains a characteristic protein fold of Aβ-Bβ-Cα-Dα-Eα-Fα-Gβ-Hβ-Iβ, a standard feature of PAS/LOV domain.

Upon crystal structure examination two modes of dimerization were observed in ZTL-LOV domain (46). These dimers have two different types of interface with different orientation of the PAS domains: 1) a parallel helical dimer interface and 2) an antiparallel β-sheet dimer interface (46). In order to understand the type of dimer interface that exists in solution, mutations that disrupt the dimer interactions in each type of interface were made and tested for dimer formation. The parallel helical dimer interface is primarily stabilized by electrostatic interaction between an arginine (R95) and the phosphate side chain of FMN co-factor. Introduction of an R95A point mutation should disrupt dimerization, as the methyl side chain of alanine could not form the electrostatic interaction with the FMN-phosphate tail. The SEC elution profile of ZTL-LOV R95A is dimeric with little to no change in apparent molecular weight. This suggests that the α-helical dimer does not exist in solution.

The anti-parallel β-sheet dimer interface is formed by direct contacts between Aβ, Iβ and Hβ sheets of the PAS domains of ZTL. The F47, I151 and I153 residues from each subunit make the core of the dimer interface. The dimer interface mainly consists of hydrophobic interactions comprised of pi(π)-pi(π) stacking interactions between the side chains of F47 pairs, and the dispersion interactions formed by the tetrad of I151 (a pair) and I153 (a pair) side chains are the prominent stabilizing interactions that hold the dimer interface together. Mutation of I153 to an arginine disrupted dimer formation as SEC elution profile of the mutant suggested a monomer.
Thus, the introduction of a polar residue disrupts the hydrophobic interactions at the interface core. Therefore, the antiparallel β-sheet is the dimeric interface that exists in solution in ZTL.

In order, to analyze if antiparallel β-sheet dimer interface is relevant for LKP2 and FKF1, amino acid residues corresponding to I153 in ZTL were mutated to arginine in LKP2 and FKF1. Indeed, the SEC studies of FKF1 28-174 I160R and LKP2 I150R exhibited dimer disruption. Therefore, the anti-parallel β-sheet dimer interface is common dimer interface in the ZTL family of proteins. Further, we ran LKP2 16-165 and FKF1 28-174 on an S200 analytical column at different concentrations in the light as well as dark state and calculated the dimer dissociation constant ($K_d$).
Figure 15: Size exclusion studies of dimer disruption mutants: A) The mutant FKF1 I160R eluted as a monomer (~ 80µM red) as compared to FKF1 WT which eluted as a dimer (~ 80µM black). The second peak in the mutant FKF1 I160R elution (red) is due to protein degradation. B) The mutant LKP2 I150R eluted as a monomer (~ 80µM red) in comparison to LKP2 WT (~ 80µM black). The small curve in the elution (red) may be due to the presence of other impurities. Also, the mutant proteins expressed poorly and were not stable. C) Elution profile of ZTL (red), FKF1 (black), LKP2 (blue) on a S200 column. ZTL forms a high-affinity dimer as compared to FKF1 and LKP2. D) LKP2 WT was run on an s200 analytical column at different concentrations 100 µM (black), 50 µM (red), 20 µM (blue), 10 µM (magenta) and 5 µM (green) to calculate dimer dissociation constant (K_d). The elution profile exhibited concentration dependent dimer formation. E) FKF1 WT was run on a s200 analytical column at different concentrations 100 µM (black), 50 µM (magenta), 20 µM (blue), 10 µM (green) and 5 µM (red) to calculate dimer dissociation constant (K_d). The calculation of dissociation constants is discussed in section.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Dissociation constant $K_d$ (µM)</th>
<th>Light/Dark state</th>
</tr>
</thead>
<tbody>
<tr>
<td>LKP2 16-165</td>
<td>12.8</td>
<td>Light</td>
</tr>
<tr>
<td>LKP2 16-165</td>
<td>13.5</td>
<td>Dark</td>
</tr>
<tr>
<td>FKF1 28-174</td>
<td>0.7</td>
<td>Light</td>
</tr>
<tr>
<td>ZTL 29-165 G80R</td>
<td>0.22</td>
<td>Light</td>
</tr>
</tbody>
</table>

**Table 3: Dissociation constants of ZTL family:** The dissociation constants ($K_d$) were calculated by using a method from previous literature (49). Ashutosh Pudasaini provided the data of ZTL-G80R. Due to low expression of ZTL WT, the experiment was done with ZTL G80R which has similar elution profile as ZTL-WT.

Thus, we have demonstrated that the LOV domains of LKP2 and FKF1 form a dimer under *in-vitro* conditions. The dimer interface of FKF1 and LKP2 LOV domains in solution is similar to ZTL and involves antiparallel β sheet interface. Further, the dimer dissociation constants indicate that ZTL forms high-affinity dimer similar to FKF1. In contrast, LKP2 forms a very weak dimer indicated by large dissociation constants. Currently, dimer formation has not been shown in full-length proteins or *in-vivo* due to the instability of full-length proteins.

Studies on ZTL protein as discussed in the previous chapter have shown that it interacts with Arabidopsis Skp1 like protein 2 which indicates it's a part of the SCF complex. Therefore, from previous studies of ZTL and F-box protein in mammals, it could be hypothesized that the F-box domain of FKF1 and LKP2 form a complex with ASK2 to undergo SCF complex-mediated degradation of their protein substrates. The kelch repeats in these proteins may be involved in
substrate recruitment for targeting proteins for degradation. The presence of LOV domain at the N-terminal preceding the F-box domain may implicate that these proteins require dimer formation utilizing the LOV domain for substrate ubiquitination and degradation. Also, it has been shown that ZTL and FKF1 form a complex with GI during the day. The FKF1:GI complex also interacts with CDF1 and mediates its degradation during the afternoon of long-days to control photoperiodic flowering. Therefore, FKF1 mediated degradation of CDF1 may follow one of two mechanisms i) GI acts as an accessory protein that mediates substrate recruitment of CDF1 to FKF1 to regulate its degradation. ii) GI acts as an accessory protein to mediate substrate recruitment to FKF1 but dissociates after substrate recruitment and FKF1 forms a dimer mediated by its LOV domain for substrate ubiquitination and degradation. In case of ZTL, ZTL:GI complex is not observed during the night therefore ZTL may or may not function as a dimer for substrate ubiquitination and degradation. Further studies investigating interaction of LKP2 and FKF1 with ASK2 is required to confirm that they undergo SCF mediated protein degradation and also if these proteins require dimer formation for substrate ubiquitination and degradation. Also, light and dark state crystal structure information will provide knowledge about how light regulates signal transduction in these proteins.
3.6 Materials and methods

3.6.1 Cloning and protein purification of LKP2 and FKF1 proteins

The nucleotides encoding LKP2 and FKF1 were obtained from The Arabidopsis Information Resource (Tair). LOV Domain region of LKP2 and FKF1 were cloned into 6His and pGST parallel vector using NcoI, SpeI, BamHI and HindIII cut sites respectively. Point mutations I57V, I57T, L68T, R89G, I160R and L153Q were introduced into FKF1 28-174 and LKP2 16-165 constructs using Quick-change protocol (Strata-gene).

All 6His constructs of LKP2 were expressed in Escherichia coli BL21DE3 cells, and pGST constructs of FKF1 were expressed in Escherichia coli JM109 cells. Cells were grown at 37 °C until an OD at 600 nm of approximately 0.6 was reached. The temperature was then decreased to 18 °C for 40 min to assure thermal equilibrium. At 18 °C 0.3 mM isopropylthiogalactoside (RPI) was added to initiate protein expression. After 22 hrs the bacterial pellets were harvested and stored in a buffer containing 50 mM Hepes, 100 mM NaCl, 10% glycerol pH 8.0.

For protein purification, bacterial pellets were thawed and sonicated for four cycles. After sonication, the soluble cell lysate was fractionated by centrifugation at 18,000 rpm for 1 hr. All 6- His constructs of LKP2 were purified by Ni-NTA affinity chromatography. After binding the columns were cleaned with buffer containing 50 mM Hepes, 100 mM NaCl, 20 mM imidazole, 10% glycerol pH 8.0 to wash away non-specifically bound proteins. The bound protein was then eluted from the column with buffer containing 50mM Hepes, 100mM NaCl, 500mM Imidazole, 10% glycerol pH 8.0. The eluted protein was treated with 2 mg/ml of TEV protease to cleave the 6His tag. The protein was then incubated with TEV for 2 hrs at 22 °C or overnight at 4 °C. An additional round of Ni-NTA chromatography was carried out to remove 6His-TEV.
The protein was further purified using a Superdex S200 size exclusion column (GE life sciences) equilibrated with 50 mM Hepes, 100 mM NaCl, 10% Glycerol pH 8.0. The corresponding protein fractions from size exclusion column were collected and verified using SDS-PAGE and UV-vis spectroscopic methods. The pure protein was then concentrated to the desired concentration to carry out further experiments.

All constructs of FKF1 in pGST vector were purified using glutathione resin (QIAGEN) at 22 °C. After binding protein was treated on column with 2 mg of TEV protease per milliliter of resin overnight at 4 °C. Proteins were then washed from the column with buffer. An additional round of Ni-NTA chromatography was conducted to remove 6His-TEV. The proteins was further purified using Superdex S200 size exclusion column as mentioned.

3.6.2 Size-exclusion chromatography:

To assay the formation of a light-state dimer, SEC was conducted using a Superdex 200 10/300 analytical column. The column was equilibrated with the buffer containing 50mM Hepes, 100Mm NaCl, 10% Glycerol pH 8.0. Light-state samples were generated by irradiation on ice for 5 min with a broad-spectrum flood light (150W). All samples were injected in 500 µl aliquots with a concentration range of 5–500 µM. Apparent MWs of all protein constructs were measured relative to gel filtration standards (Sigma Aldrich). Apparent MWs from SEC traces were used to confirm the identity of protein peaks relative to SEC on the Superdex 200 used in all other studies. Approximate dissociation constants were obtained via a linear fit of $1/[M_F]$ vs. the mole fraction term as shown

$$K_d = m^2[M_F]d(m+2d)$$ (7)
3.6.3 Determination of dissociation constants (Kₐ)

The dimer dissociation constants were determined for LOV domains of LKP2 and FKF1. The proteins were run on a Superdex 75 10/300 analytical column in reduced light state at various concentrations. The calculation of dissociation constants from SEC was described in

The derivation of Kₐ was described as following:

For a monomer-dimer equilibrium:

\[ [D] \rightleftharpoons 2[M] \quad (8) \]

Where and are dimer and monomer concentrations respectively. At equilibrium, the dissociation of dimer into monomers can be represented in terms of dissociation constant (k_d) as following:

\[ K_d = \frac{[M]^2}{[D]} \quad (9) \]

If the protein samples contain both monomers and dimers, the total moles of monomer-dimer can be represented by \( T_m \). Therefore,

\[ T_m = \text{moles of monomer} + \text{moles of dimer} \quad (10) \]

\[ \frac{\tau_m}{T_m} = \frac{\text{moles of monomer}}{T_m} + \frac{\text{moles of dimer}}{T_m} \quad (11) \]

\[ 1 = m + d \quad (12) \]
Where \( m \) and \( d \) are mole fractions of monomer and dimer respectively. The total number of moles of free monomer can be evaluated by the following equation:

\[
m_F = m_T m + 2d_T m\tag{13}
\]

\[
T_m = \frac{m_F}{m + 2d}\tag{14}
\]

Equation (13) in terms of mole fractions and volume of sample (V):

\[
K_d = \frac{(\frac{m_T m}{V})^2}{(\frac{d_T m}{V})}\tag{15}
\]

\[
K_d = \frac{m^2 T_m}{d\cdot V}\tag{16}
\]

Using equation 13 and 14 we have the following expression:

\[
K_d = \frac{m^2}{d} \cdot \frac{m_F}{(m + 2d)\cdot V}\tag{17}
\]

Since the equation 17 can be rearranged with an additional concentration of the free monomer term:

\[
K_d = \frac{m^2 [M_F]}{d(m + 2d)}\tag{18}
\]
REFERENCES


CHAPTER 4:

STRUCTURAL CHARACTERIZATION OF LOV PHOTORECEPTOR ENV

This chapter includes figures and data previously published in the journal *Structure* and is referenced as: (17)


4.1 Introduction to the fungal circadian clock

Plants, bacteria, and fungi utilize light-oxygen voltage (LOV) domain photoreceptors to integrate the information of blue-light intensity into their circadian mechanism. The circadian clock of a filamentous fungi *Neurospora crassa* is well characterized and serves as a model to understand the fungal circadian system. The *Neurospora* clock is primarily maintained by two LOV containing photoreceptors White Collar-1 (WC-1) and VVD (1). WC-1 a modular protein comprising an LOV domain, PAS domain, DNA-binding region, nuclear localization sequences and a transcriptional activator domain forms a dimer with White Collar 2 (WC-2) via PAS: PAS interface on to initiate gene transcription of frequency gene (FRQ) and various clock-controlled genes.
After FRQ is produced it interacts with FRQ-interacting RNA helicase (FRH) to inhibit the white-collar complex (WCC) (2). This forms a negative element to the positive transcription-translation feedback loop of WCC.

Additionally, VVD an isolated LOV protein regulates photo-adaptation by desensitizing the WCC to light input pathway (1,3). VVD tunes blue-light responses and modulates gating of the circadian clock. The expression of vvd gene itself is controlled by WCC. The first light pulse from the environment initiates the expression of vvd gene. After VVD protein is transcribed it physically interacts with WCC to repress it. In a vvd, null strain amplified light responses are seen. A vvd null strain appears bright yellow due to the continuous synthesis of carotenoids as compared to a pale-yellow appearance of a WT strain. Therefore, VVD regulates the activity of WCC to function as transcriptional activators (3). Previous studies have shown that upon light activation the LOV domain of WC-1 of WCC homodimerizes with LOV domain of WC-1 on the other WCC to form a WCC dimer to initiate gene transcription. In vitro structural and solution studies have shown that VVD forms homo-dimer upon light activation (44). This lead to a hypothesis that upon light activation the LOV domain of VVD heterodimerizes with LOV domain of WC-1 on WCC to inhibit the formation of WCC dimer and thus negatively regulates gene transcription. Also, studies have shown that by swapping the LOV domain of WC-1 with that of VVD maintains regular functions. Therefore, VVD maintains a balance in gene transcription in the presence of continuous light by negatively regulating the WCC complex. The negative regulation by VVD is transient thereby allowing WCC to function at certain light intensities and abolishing activity when required to maintain balance. Therefore, VVD provides additional negative regulation by competitive dimer formation and is essential for response to changing levels of light and for adaptation under constant light.
Envoy (ENV1) a PAS/LOV protein was first identified and cloned due to its lack of expression in cellulase negative strain of fungus *Trichoderma reseei* (5). *T. reseei* is an industrially relevant fungus and is used widely in industry for the production of cellulolytic enzymes (6). ENV is an orthologous protein to VVD, and they are members of same phylogenetic sister clade that includes WC-1 photoreceptor. Previous studies have shown that ENV plays a role in cellulase induction in the presence of cellulose. ENV lacking strain shows a delay in induction of cellulase enzyme when grown on a media containing cellulose in the presence of light. Also, ENV lacking strain show growth defects in the presence of constant light but not under constant darkness.

Further, genetic studies of cellulase expression, metabolic function, sexual development, growth and stress responses in this fungus indicate that these processes are regulated by blue-light and nutrient signaling through two LOV proteins, ENV1 and Blue-light receptor 1 (BLR1) (7,8,9). BLR1 is orthologous to *N. crassa* WC-1, and it regulates blue light mediated gene transcription in a similar fashion to Neurospora. BLR-1 forms a heterodimer with Blue-light receptor 2 (BLR2).
orthologous to *N. crassa* WC-2 to initiate gene transcription. The negative regulation of BLR complex is facilitated by ENV1 similar to VVD. Thus, *T. reesei* is presumed to follow an analogous mechanism to *Neurospora*. However, mutational studies indicate differences in signal transduction must exist (8,10,11). First, ENV1 is incapable of rescuing VVD null mutants, thereby indicating alternative signal transduction mechanisms (5). Second, whereas VVD is only functional under light-state conditions, ENV1 retains some functions even in the dark (13,14). Third, ENV1 and BLR1 are capable of integrating nutrient sensing, blue-light sensing, and oxidative stress into a coherent signal. To evaluate the differences in signaling mechanisms, we conducted solution and biophysical studies of ENV1.

Currently, little is known about ENV1 signaling at the molecular or chemical level. However, some insight can be garnered from *N. crassa*. Both, light and dark state crystal structure of VVD and other variants are available that provide detailed information about signaling events occurring upon light activation. VVD serves as a model for LOV domain signaling. Light activation of VVD leads to cysteine-flavin adduct formation and protonation of the N5th position of the flavin that further leads to translation of the N-terminus from the LOV core and dimer formation (4,15).

**Figure 17: Dimer Formation:** Blue Light displaces the N-cap (pink) from the LOV core (blue), which docks to the hinge region of another subunit to form a dimer. The two αα helices from each subunit also create a dimer interface.
The hydrogen bonding rearrangement starting from flavin active site leads to the translation of the N-terminal. The N-terminal cap contains a helical element (αα), a short strand (ββ) and an N-terminal latch that wraps around the LOV core Flavin N5 protonation following blue light activation causes a conserved Gln182 residue that is hydrogen bonded to N5 with its amino side chain in the dark to flip and form a hydrogen bond with carboxylic side chain of Ala72 residue on the N-terminal hinge. The LOV domain of VVD is connected to the N-terminal by this hinge region. Formation of this hydrogen bond causes the neighboring Cys71 residue to move from a buried to an exposed conformation. The movement of this Cys71 residue is critical to the release of the N-terminal from the LOV-core and dimer formation. Mutation of this Cys71 to Ser decoupled structural changes following blue light activation (16). The crystal structure of mutant VVD C71S shows that hydroxyl group of serine forms a firm hydrogen bond with the Asp68 carbonyl than Cys71 which does not change hydrogen bonding partners upon light activation that stabilizes the loop structure and blocks the N-terminal release and subsequent dimerization (16).

Further size-exclusion chromatography of the mutant VVD C71S showed that it forms a monomer even in the light state. In contrast, mutation of Cys71 to a Valine leads to partial dimerization even in the dark (16). The valine side chain can mimic both the conformations of Cys71 in light and dark, therefore, leads to latch release and partial dimer formation in the dark. Cys71 switches hydrogen bond from Asp68 carbonyl to Asp68 amide in the light state. These changes cause a 2Å shift in ββ towards the PAS core by movement of Pro66. The shifts in ββ causes shifts in αα by allostery which further leads to the release of N-terminal latch. The N-terminal latch now docks on to the hinge region of the adjacent subunit forming dimer contacts.

Size exclusion chromatography studies of VVD WT showed that it elutes at a volume which is intermediate between dimer and monomer on an S200 analytical column (4). This is due
to low dimer affinity and a rapid exchange equilibrium between dimer and monomer (4,16). Thus, structural and solution studies show that light activation leads to release of N-terminal from the LOV core to form a dimer. Further *in-vivo* studies show that N-terminal release is essential for signaling. The mutant strain carrying C71S mutation is not able to down-regulate carotenoid production and appears bright yellow like mutant strain containing a deletion of VVD gene (16). Therefore, the N-terminal release is vital for VVD to regulate its function *in-vivo*. This further supports the hypothesis that N-terminal release followed by dimer formation with LOV domain of WC-1 downregulates the transcriptional activation of WCC.
4.2 Crystal structure of ENV1

To understand the activation and signaling mechanisms of ENV, various constructs of different length such as N-terminal truncated protein, ENV-55 and ENV-64 (56-207, 65-207) and the full length (FL 1-207) were cloned and expressed. All constructs revealed LOV-type
photochemistry as described earlier. Upon light activation, a cysteine-flavin adduct was formed, which decayed with a time constant of 25 minutes. In contrast to VVD, the photocycle of ENV1 is 15 times faster (17). Due to intrinsic disorder in the N-terminus like VVD, attempts to crystallize ENV-55 and ENV-FL were unsuccessful (16). Crystals were obtained for ENV-64 in its dark state, which diffracted up to 2.2 Å. The crystal structure of ENV-64 contained four molecules per asymmetric unit. Each monomer unit was comprised of the core PAS fold similar to VVD consisting of a 5-stranded anti-parallel β-scaffold and two α helices, which create an interface for binding FMN cofactor. The N-terminal cap (N cap, residues 65-99) contains a helical element (aa), a short strand (bβ) and an N-terminal latch (65-74). The N-cap is connected to the PAS core through an N-terminal hinge like VVD. The FMN cofactor is bound to the LOV domain by amino acid residues that are conserved in photo-tropin like LOV domains as discussed before (18). Hydrogen bonding interactions with pyrimidine ring of FMN are made by amino acid residues on Gβ, Hβ, Iβ and Eα and Fα helices. Ile 187, Ile 161 and Ala 110 make hydrophobic contacts with the dimethylbenzene moiety of FMN.
Figure 19: ENV1 Photocycle and crystal structure: A) Light (red) and dark (black) state spectra of ENV1 FL exhibit an LOV type chemistry. The dark state shows a free flavin peak at 450nm and two shoulder peaks at 425nm and 478nm. Light absorption and cysteine-flavin adduct formation lead to bleaching of peaks at 450nm, 425 and 478nm and formation of a single peak at 380nm. B) Dark state recovery rate of ENV1 FL after light activation is calculated by measuring the absorbance at 450nm (black) and 478nm (red) using a spectrometer at different time intervals. The exponential fitting of the trends of absorbance observed at 450nm, and 478nm yielded a time constant for dark state recovery to be approximately 25 mins. C) Dark state crystal structure of ENV-64 contains four molecules per asymmetric unit. The E-F loop is mainly disordered and absent in three of four molecules.
Figure 20: Crystal structure of ENV1: A) ENV1 has a typical LOV fold consisting five antiparallel β strands and four α helices that are connected to the N-terminal shown in pink by a hinge region colored in red. The N-terminal latch is shown in yellow B) The FMN cofactor is bound to the LOV domain by various hydrogen bonding interactions with amino acid residues that are conserved in phototropin like LOV domains. The N5 of flavin is hydrogen bonded to Gln204 on Iβ. Asn173 and Asn183 on Gβ and Hβ make hydrogen bonds with O2, N3, and O4. Gln139 on Eα makes contacts with N1. The positive guanidium groups Arg136 and Arg158 on Eα and Fα helices make salt bridges with phosphate moiety of FMN.
4.2.1 Crystallographic dimer interface

The ENV-64 structure was composed of two symmetric dimers that buried an average of 1477 Å² that accounts for a -30.9 kcal/mol free energy of formation (17). Various modes of dimer association have been observed in PAS domains. PAS: PAS homo-dimer and heterodimer involves utilizing the hydrophobic β scaffold or the N/C-terminal in mediating dimer contacts. The crystallographic dimer observed in ENV1 utilizes the N-terminal to mediate dimerization like VVD. The dimer interface is formed by hydrophobic interactions between α helices of two subunits and interactions between N-terminal latch and PAS core of another subunit. The Tyr67 residue of N-terminal latch forms extensive hydrogen bonds to the PAS core of another subunit through a hydrophobic cleft created by translation of β strand of the N-terminus. Furthermore, the Ser68 and Ser70 of the latch region formed hydrogen bonds to Asp95 residue present at the hinge region of the neighboring subunit.

Therefore, the hydrogen bonds and hydrophobic interactions between amino acid residues of two monomer subunits contribute to dimer formation in dark state ENV-64. The dimer interactions of ENV 64 are similar to dimer interactions formed by VVD in light state (15). Therefore, ENV undergoes a very modest conformational change upon light activation, unlike VVD which undergoes a significant structural change that promotes constitutive activity that facilitates ENV dark state functions.
Figure 21: N-terminal forms important dimer contacts: A) The N-terminal α helices from two subunits form hydrophobic contacts via Leu78 residue. The N-terminal latch from one subunit forms contacts with N-terminal hinge and β scaffold of the other subunit. The N-terminal latch is shown in blue and N-terminal hinge in pink. Tyr67 a latch residue forms hydrogen bonds with β scaffold residues Thr186 and Val184. Tyr67 also forms a hydrogen bond with a hinge residue Asp95. Asp95 on the other-hand forms extensive hydrogen bonds with latch residues Ser68 and Ser70. B) Upon closer examination, a Pro93 residue on the hinge of one subunit forms hydrogen bonds with latch residue Ser68 of another subunit.
4.2.2 ENV-64 dark state dimer resembles VVD light state dimer

Although the crystals of ENV-64 were obtained in dark state, it resembled the light state conformation of VVD. Comparison of the ENV-64 dimer with dark- and light-state VVD revealed the orientation of N-cap and dimer contacts were nearly identical to the VVD light-state dimer (VVD-LSD) (Figure 20B) (15) but differed substantially from dark-state VVD (Figure 20C). The lifetime of ENV is too short (25 mins) to obtain a light state crystal and also the active site cysteine positions the thiol away from the C4a position in all four molecules unlike in light state where it positions the thiol towards the C4a position. Therefore, ENV may form a constitutive dimer.

Figure 22: Comparative Studies of VVD ENV dimer: A) Crystal structure of ENV-64 in dark state reveals a dimer interface formed by N-terminal helices (yellow) and also by movement of N-terminal Latch (red) into the hydrophobic cleft. B) Crystal structure of VVD light state shows identical dimer contacts as an ENV-64 dark state. C) Crystallographic dimer of VVD dark state is distinctly different from the ENV-64 dark state the N-terminal latch is packed against the LOV core unlike in VVD light state and ENV-64 dark state.

On further sequence comparison of ENV and VVD revealed variations in amino acid residues in ENV that are conserved and important for causing structural changes in VVD (Figure 20). These variations in amino acid residue may be responsible for promoting light state like conformation in ENV even in its dark state. In VVD light promotes flavin N5 protonation that causes the Gln182 amide that is previously hydrogen bonded to N5 to flip and form a hydrogen bond with Ala72 carbonyl on the N-terminal hinge that connects the N-cap and LOV core. Rotation
of Gln 182 and formation of a new hydrogen bond with a hinge residue Ala72 on light activation causes the neighboring Cys71 to change hydrogen bonding partners and induce a translation in N-cap leading to dimerization (4,15,16). These changes in hydrogen bonding upon light activation in VVD leads to movement of N-terminal hinge and bβ towards the PAS core and release of the N-terminal latch to form dimeric contacts. Upon sequence comparison and crystal structure analysis, it was found that ENV1 has a Serine 99 residue instead Ala72 on the hinge.

Crystal structure analysis showed that ENV1 consists of a hydrogen bond network similar to VVD light state; it contains a hydrogen bond between Gln204 and Ser99 in its dark state (Figure 21), which corresponds to the hydrogen bond between Gln182 and Ala72 in light state VVD. Presence of a serine residue instead of alanine in ENV allows the Gln204 carbonyl to hydrogen bond with the hydroxyl group of Ser99 in its dark state, which is similar to hydrogen bonding network of VVD light state. Further, ENV1 has a valine (Val98) residue at the hinge region instead of a cysteine (Cys71) in VVD (Figure 20), which as seen earlier is key to initiate the structural changes. Val98 mimics both positions of the Cys71 (VVD) thiol in light and dark state. Previous studies have shown that mutation of Cys71 to valine in VVD predisposed it to dimerization and formed some dimer even in the dark. Therefore, the presence of valine and a serine residue in ENV1 leads to rotation of hinge residues 94-98 towards the PAS core and break an Asp95-Ser97 H-bond present in both light and dark-state conformations of VVD. Disruption of these interactions promotes translation of bβ to mimic VVD-LSD in dark-state ENV. An additional 2.4 Å shift because of disruption of the Asp95-Ser97 interaction stabilizes contacts between the N-latch and N-hinge at the dimer interface. Additionally, a Cys residue (Cys96) is shifted two residues as compared to Cys71 in VVD that moves it to a more surface-exposed position on the hinge. Although, Cys96 is incapable of being involved in H-bond switching like Cys71, the
exposed site on the N-hinge positions the residue ideally to respond to additional environmental stimuli that is discussed later in this section.

**Figure 23:** Sequence alignment of ENV1 and VVD: ENV1 has a Ser99 instead of Ala72 in VVD and also Val98 instead of Cys71. Cys96 in ENV1 corresponding to Cys71 in VVD is two residues shifted on the hinge.

**Figure 24:** Comparison of the ENV-64 crystal structure with VVD: A) Comparisons of the hinge regions of ENV1 (yellow) with dark-state VVD (pink) showing the translation of bβ and N-terminal hinge in ENV1 relative to VVD dark state. B) Comparisons of the hinge regions of ENV1 (yellow) with light-state VVD (gray) shows very little difference as compared to VVD dark state (pink) suggesting the formation of a VVD light state like dimer in the ENV1 dark state. N5 protonation and rotation of Gln204 initiate signal propagation. ENV1 contains an H-bond between Gln204 and Ser99 in the dark state that induces a shift in bβ. Val98 mimics the light-state conformation of Cys71, leading to movement of bβ and creation of a light-exposed cleft. Ser97 is rotated away from bβ to facilitate interactions across the dimer interface.
4.3 Size Exclusion Chromatography of ENV1

Based on the crystallographic data, it was predicted that ENV-64 should exhibit light-state VVD-like behavior under dark-state conditions. Therefore, by analogy to VVD, ENV-64, and FL-ENV1 should constitutively form rapidly exchanging dimers even in the dark. By contrast, size exclusion chromatography (SEC) revealed that dark-state FL-ENV1 is mostly monomeric. It purifies as a mixture of a slow- and fast-eluting species that multiangle light scattering confirmed as a monomer-dimer mixture. The percentage of dimer in the dark state is relatively low. Exposure to blue light led to a dose-dependent increase in the dimeric fraction. In the presence of reducing agents like dithiothreitol (DTT) ENV1 eluted as a monomer in the light as well as the dark state. Therefore, ENV1 requires light as well as oxidative conditions to form a dimer. Further, the N-terminal is involved in mediating important dimer contacts as shown by the crystallographic data. Size exclusion chromatography of N-terminal truncated constructs showed that dimer affinity and percent dimer formation was reduced in ENV1-64 construct as compared to ENV1-55 construct. Thus, a short peptide (residues 55-64) may form stabilizing dimer interactions. Such data indicate that the N-terminus of ENV1 is essential for mediating conformational changes. Similar data was observed in VVD, where dimer affinity was strongly dependent on the length of the N-terminus. Truncation of N-terminus affects dimer formation (4).
Figure 25: Size exclusion chromatography of ENV1 FL and N-terminal truncated constructs: A) MALS of ENV1 FL(1-207) in light state elutes a mixture of dimer and monomer. B) Dark-state FL ENV1 (black) elutes as a monomer and dimer mixture. Light increases the population (red) of the dimeric component. Continued exposure to blue light for an additional 10 min (magenta) leads to a dose-dependent increase in dimerization. The reduced light state (green) and dark state (blue) are monomeric in the presence of 5 mM DTT. C) ENV-FL in light state (red) has a dimer percentage of 56.2% and dimer percentage of 6.1% in dark state D) ENV-55 in light state (red) has a dimer percentage of 54.2% and dimer percentage of 7.6% in dark state E) ENV-64 in light state (red) has a dimer percentage of 36.1% and dimer percentage of 4% in dark state
However, in ENV1 dimer formation requires oxidative stress to allow the formation of an analogous competitive model of blue-light signaling in *T. reesei*. This suggests that ENV1 adopts an alternative signaling mechanism than VVD that is tailored to the native environment to incorporate oxygen sensing into blue light signaling. The formation of a stable light- and oxygen-dependent crosslinked dimer suggests the involvement of a Cys residue in ENV1 activation. Upon examination of crystal structure, a Cysteine 96 residue that occupies a surface-exposed position in the dimer interface positioning the sulfhydryl-moieties within 6.5 Å (17). Such close contacts would allow for disulfide-mediated cross-linked dimer formation with only modest rearrangement of the N-cap. To investigate if Cys96 plays a role in disulfide linked dimer formation, it was mutated to a threonine residue that is native to VVD and size-exclusion chromatography was performed. Indeed, the introduction of a C96T variant abolished light-induced dimer formation.

![Figure 26](image)

**Figure 26: Cys96 mediates disulfide-linked dimer formation:** A) Elution profile of mutant ENV 1FL C96T shows that it is not capable of dimerization in both light (magenta) and dark state (black) therefore the cysteine is indeed involved in formation of a disulfide-linked dimer B) Crystal structure reveals two cysteine residues (yellow) in close proximity that upon light activation can form a disulfide-linked dimer.
If light-induced disulfide formation plays a role in ENV1 signaling, three requirements would have to be met. First, the Cys residue should be confined to a protein region undergoing a conformational change following adduct formation. Second, the reduction potential of disulfide bond formation should be within physiologically relevant ranges. Third, disruption of disulfide bond formation should have an effect on T. reesei physiology.

The crystal structure analysis proves the first criteria; the Cys96 is located on the N-terminal that undergoes conformational changes upon light activation and Cys96 is also within the dimer interface. The reduction potential of dimer formation was calculated by performing redox titrations. The redox titrations were performed by incubating the protein in the light as well as the dark state with varying ratios of oxidized glutathione GSSG and reduced glutathione GSH. The GSSG/GSH ratios were varied in such a way to obtain a reduction potential in the range of -390 to 230 mV. The protein was incubated in such conditions for 2 hrs. After two hr incubation, SDS PAGE was performed on protein samples to analyze the fraction dimer present under each condition. A gel analyzer software was used to predict the amount of dimer and monomer from the intensity of the bands. The fraction dimer under each condition was calculated and plotted against the reduction potential. The midpoint potential was obtained by fitting points to a sigmoidal curve. The midpoint potential for the Cys96 disulfide was on the order of 266 ± seven mV at pH 7.0. To the best of our knowledge, the cellular redox potential of Trichoderma has not been measured, but the cytosolic redox potential of the related fungus M. oryzae has been measured as 290 mV (at pH 7.0) and shifts dramatically depending on metabolic, reproductive, and pathogenic states (20). Therefore, the reduction potential of the Cys96 disulfide is well poised to sense the metabolic state of the cell and mediate responses to oxidative conditions. The experiments were
done to show if disruption of disulfide bond affected organism physiology was done at our collaborator Monika Schmoll's lab and will be discussed in the following section.

Figure 27: Calculation of Midpoint Reduction potential of disulfide-linked dimer formation: A) An SDS PAGE of ENV1 FL protein samples incubated in various reduction potential of GSSG/GSH pair varying from -390mV to -230mV. The intensity of the dimer peak increases as the reduction potential decreases. The amount of dimer and monomer in each lane was estimated by a gel analyzer software, and fraction dimer was calculated. B) The dimeric fraction of FL ENV1 in light state as a function of the reduction potential of a GSH/GSSG couple at pH 8.0. The experiment was completed in triplicate, and the midpoint potential was determined to be -325 +/- 7 mV at pH 8.0 in the light state via fitting with a titration curve. Using a pH dependence of the GSSG/GSH couple of -59 mV/pH unit (53) that correlates to a reduction potential of -266 mV at pH 7.0. C) Dimeric fraction of FL ENV1 in dark state as a function of the reduction potential of a GSH/GSSG couple at pH 8.0. This experiment was done in triplicates, and the midpoint potential was determined to be -338mV +/-7 mV at pH 8.0. The midpoint potential at pH 7.0 was calculated to be -279mV.
4.4 In-vivo significance of disulfide bond formation

To evaluate if disruption of disulfide bond formation affected organism physiology an env gene harboring Cys96Thr mutation was introduced into a strain lacking ENV1 in the background of the wild-type strain called as QM6a (21). Then the growth (hyphal extension on plates) of the resulting strains (QEC96T) was evaluated in darkness, low light (1,500 lux), high light (8,000 lux), and varying carbon sources. These carbon sources mimic low-nutrient (minimal medium with glucose), nutrient-rich (malt extract), and natural conditions with a recalcitrant cellulosic substrate requiring specific enzymes for degradation (minimal medium with carboxymethylcellulose). Oxidative stress was evaluated by addition of 25 mM menadione (final concentration). The resulting in vivo data was consistent with a distinct role of Cys96 in integrating oxidative stress, metabolism, and light intensity. QEC96T rescues growth defects of Δenv1 on malt extract and minimal medium, indicating successful complementation of T. reesei with Cys96Thr.

Furthermore, the lack of growth defects in QEC96T verifies that Cys96 does not have a general effect on organism growth. In addition, under constant darkness, QEC96T strains did not show significantly different growth patterns compared with the wild-type in the presence of menadione. Upon exposure to low light, however, we found that Cys96 is essential to the function of ENV1 in response to oxidative stress with glucose (p = 0.002) or cellulose (p < 0.001) as carbon sources. Specifically, complementation of Δenv1 with the Cys96Thr allele did not alleviate the growth defect of Δenv1 in the presence of menadione. Wild-type growth of QEC96T was observed on the complex medium malt extract under low light. Under high light conditions, no statistically significant effect of Cys96 was observed, consistent with a role of ENV1 and Cys96 in differentiating carbon sources and light intensity. In general, our analyses clearly revealed differences in the relevance of Cys96 in oxidative stress responses depending on the carbon source.
and light intensity. This finding supports the hypothesis of ENV1 connecting light and stress response with carbon sensing and utilization.

Figure 28: C96T mutants are able to complement growth defects in high light and darkness: The growth defect of Δenv1 is rescued by complementation with an allele carrying the C96T mutation. Upon growth on glucose (A), cellulose (B) or malt extract (C) in light, QEC96T shows equal growth as the wild-type and significantly enhanced growth compared to its parent strain Δenv1. In darkness, the deletion of env1 has no effect on growth. In addition, Cys96 is not relevant in high light or darkness. Upon growth on glucose (D, G), cellulose (E, H) or malt extract (F, I) in the presence of 25 mM menadione in high light (D, E, F) or darkness (G, H, I), QEC96T shows equal growth compared to the wild-type and significantly enhanced growth compared to its parent strain Δenv1. Error bars reflect standard deviations
A Competitive dimer model of LOV function in filamentous fungi is modified to sense oxidative stress:

Since, the disulfide bond formation by Cys96 residue meets all the three criteria mentioned above it plays a significant role in allowing organism to adapt and integrate light and oxidative stress signaling in one protein. Dual sensing roles of a blue-light and oxidative stress has also been observed in *Rhodobacter sphaeroides*, where the BLUF protein AppA acts in a ternary complex with PpsR and DNA to affect photosynthesis. Further, upon sequence comparison of VVD and ENV it was found that Cys 96 is two-residue shifted than Cys71 in VVD to form a CxV motif. This two-residue shift in ENV1 positions Cys96 to be more solvent exposed on N-terminal hinge

**Figure 29: C96T mutant is incapable of rescuing growth defects in low light:** A–C) Relevance of Cys96 to growth in the presence of menadione in low light (1,500 lux) on different carbon sources. An ENV1-null strain (Δenv1) has a growth defect compared with WT (QM6a) grown with glucose (A), cellulose (B), or malt extract (C) as the sole carbon source. Complementation with an ENV1 Cys96Thr variant is incapable of rescuing the growth defect with glucose or cellulose but demonstrates enhanced growth with malt extract. The data indicate that ENV1 is involved in coupling light, oxygen, and metabolism, which is dependent on the presence of Cys96. The statistical significance (p ≤ 0.005) of growth patterns is reflected by different small letters for wild-type and mutant strains. Growth patterns were evaluated for 3, 4, and 5 days after inoculation for statistical analysis. Error bars represent SD.
to respond to oxidative stress and form a disulfide linked dimer, whereas Cys71 in VVD is inaccessible to form a disulfide linked dimer.

Previous studies have shown that ENV1 follows an analogous mechanism to VVD and plays a role in negative regulation of BLR1:BLR2 complex. Our results lead to a hypothesis that in presence of oxidative stress ENV1 tends to form a disulfide linked homodimer instead of heterodimer with BLR1. During oxidative stress conditions ENV1 is sequestered as inactive homodimer thereby allowing BLR1:BLR2 complex to regulate the transcription of genes that can relieve the oxidative stress. Therefore, oxidative stress allows for sequestration of ENV1 in an inactive state without negatively affecting BLR1 signal transduction. Also, the Cys71 native to VVD is not two residue shifted in BLR1 so ENV1 cannot form a disulfide linked heterodimer with BLR1. Although ENV1 is homologous to VVD evolutionary pressure has caused it to respond to additional stimuli. The two-residue shift of Cys96 residue in ENV1 has aided the fungi to diverge from signaling mechanisms observed in model organism Neurospora and respond to oxidative stress.

According to phylogenetic classification, fungus T.reesei belongs to the phylum of ascomycota, class of Sordariomycetes and order of Hypocreales. Upon sequence alignment of ENV1 homologues in fungi belonging to the order of Hypocreales, the two-residue shift in cysteine residue corresponding to a CxV motif was observed in certain species of plant pathogens such as Fusarium and Verticillium. Therefore, this two-residue shift may be relevant in plant pathogens to adapt to oxidative stress when they first attack the plant. Although, the importance of this two-residue shift should be verified by understanding their signaling mechanisms. Thus, ENV1 has divergent signaling mechanisms from its homologue VVD where it integrates light sensing, metabolic signaling and oxidative stress responses.
Figure 30: Model of divergent signaling in Trichoderma reesei: A) Sequence alignment for selected sequences of the N-hinge of ENV1 (TrENV) and homologs from *F. oxysporum* (FoVVD), *F. fujikuroi* (FfVVD), *V. alfalfa* (VaVVD), *N. crassa* VVD (NcVVD), and *N. crassa* and *T. reesei* BLR1 (NcWC1 and TrWC1). Although bβ (blue) is conserved, Hypocreales (Tr, Fo, Ff, and Va) contain a two-residue shift of the key Cys residue (+) and likely retain ENV1-like signaling mechanisms. All WC1 species contain a Cys residue in a position analogous to VVD. Neighboring regions (orange) are often involved in tuning LOV photochemistry lifetime. (B) ENV1 is capable of integrating signals in the light and dark, including survival of increasing levels of blue light. This is achieved through interactions with the white collar complex (BLR1:BLR1:BLR2). In analogy to *N. crassa*, light promotes dimerization of BLR1 to induce gene expression. ENV1 is presumed to repress gene expression through competitive dimerization with BLR1. Here we propose an additional level of control to protect against oxidative stress. This is achieved by sequestration of ENV1 in a homodimer under oxidative stress conditions to block formation of ENV1:BLR1 (X) and subsequent repression of BLR1:BLR2.
4.6 Cys96 Is conserved in closely related sordariomycetes

Because our experiments showed a biological relevance of Cys96, we were interested in whether the function of Cys96 is relevant for evolution. Therefore, we performed a phylogenetic analysis of the ENV1 homologs of 36 Sordariomycetes as available at JGI Mycocosm. Alignment of their amino acid sequences revealed partial conservation of Cys96 (Figure 26). The phylogenetic tree showed that species with Cys at the position corresponding to 96 in T. reesei cluster separately from those with other amino acids at this position, including N. crassa (Figure 27). Specifically, our data indicate that Cys96 is conserved in Hypocreales (except for I. spp.), Glomerellales, and Diaporthales, whereas other amino acids are present at this site in Sordariales, Coniochaetales, Xylariales, Magnaporthales, and Ophistomales. The species present in both groups include beneficial and model fungi (N. crassa, Trichoderma spp., P. anserina, A. montagnei, or D. eschscholzi) as well as pathogens (Fusarium spp., C. parasitica, C. ligniaria, or M. grisea). Because knowledge of the physiology of most of these fungi is very limited and because their strategies for response to oxidative stress are unknown, we cannot draw any conclusion as to the physiological relevance of Cys96 that leads to this evolutionary phenomenon. We propose that the function of Cys96 in oxidative stress and metabolic processes played an important role in the evolution of species within Sordariomycetes.
Figure 31: Sequence alignment of selected Sordariomycetes: Trichoderma reesei (ENV1), Fusarium oxysporum (plant pathogen), Verticillium alfafae (plant pathogen), Neurospora crassa VVD (reference), Colletotrichum gloeosporioides (plant pathogen), Villosiclava virens (plant pathogen), Claviceps purpurea (plant pathogen), Metarhizium anisopliae (biodiesel, lipases), Beauveria bassiana (insecticide; insect pathogen), Cordyceps militaris (insect pathogen, cellulase), Zymoseptoria tritici (plant pathogen) ENV1 homologs. The position corresponding to Cys96 in Trichoderma reesei is denoted with a black arrow. All organisms contain the consensus CxV motif (red), with a two-residue shift in the Cys residue compared to N. crassa. Thus, these plant and insect pathogens as well as lipase/cellulase producing organisms may retain oxidative stress sensing mechanisms analogous to ENV1. Conserved residues (yellow) primarily involve structural or flavin binding residues. Tyr67 ($), which forms key contacts in the VVD/ENV1 dimer interface is 100% conserved. The consensus LOV NCRFLQ motif (blue) is conserved in all species with exception of the insect pathogens Beauveria bassiana and Cordyceps militaris (green). The lack of the Cys residue required for adduct formation (&) may indicate these species no longer function as light-sensors, but rather only function through oxygen sensing. Interestingly, the Gln residue (Q204 ENV1 number) that H-bonds to Ser99 (%) is not well conserved in these species. Given the importance for this residue in signaling in other LOV proteins the lack of conservation is consistent with altered signaling mechanisms within the VVD/ENV1 superfamily.
Figure 32: Evolutionary Relationships of ENV1 Homologs in Sordariomycetes: Species names are given along with the protein ID of the respective JGI genome database (http://genome.jgi-psf.org/programs/fungi/index.jsf). In the bar, the amino acid present at the site corresponding to Cys96 in T. reesei ENV1 is shown. The evolutionary history was inferred using the minimum evolution (ME) method (19). The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units and those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the close neighbor interchange algorithm at a search level of 1. The neighbor-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the data set (complete deletion option). A total of 117 positions were in the final data set.
4.7 Materials and methods

4.7.1 Cloning and purification

Constructs were designed based on sequence homology to *Neurospora crassa* VVD and consist of full-length (1-207) and N-terminally truncated constructs ENV-64 (65-207) and ENV-55 (56-207). All constructs were cloned from cDNA obtained from *T. reesei* grown on cellulose using standard PCR approaches. Envoy constructs were subsequently cloned into the pGST vector using NcoI and XhoI cut sites and were verified by DNA sequencing (Genewiz). The C96T point mutant was introduced in FL ENV1 using the Quickchange protocol (Stratagene).

ENV1 constructs were expressed in *E. coli* JM109 cells. Cells were grown at 37 °C until reaching an OD 600 of 0.6. At this point, the temperature was decreased to 18 °C for 40 min. At 18 °C 0.3 mM isopropyl thiogalactoside (RPI) was added to initiate protein expression. Pellets were harvested after 22 hrs and stored in 100 mM NaCl, 50 mM Hepes (pH 8), and 10% glycerol (ENV-buffer). All constructs were purified with glutathione affinity resin (Qiagen). To avoid copurification with *E. coli* contaminants, ENV1 constructs were eluted on column via treatment with 2 mg of TEV protease per milliliter of resin. The columns were then incubated for 2 hrs at 22 °C. Cleaved proteins were eluted in ENV-buffer containing 5 mM DTT. An additional round of Ni-NTA chromatography was conducted to remove His6-TEV prior to final purification with a Superdex S200 size exclusion column equilibrated with ENV-buffer. Purified constructs were concentrated to ~ 100 μM and used within 8-hours to minimize oxidation.
4.7.2 Structural analysis

ENV-64 crystals were obtained with reservoir solutions containing 100 mM Hepes pH 7.5 buffer, 6% (v/v) Peg400 and 1.4 M ammonium sulfate using 1.5 µL well solution with 1.5 µL of ENV-64 at 6 mg/ml. Protein for crystallographic studies were purified in ENV-buffer with 5 mM TCEP (RPI). Crystal trays were then set in a dark room illuminated with a red safe light.

Diffraction data was collected at the F1 beamline at the Cornell High-Energy Synchrotron Source (CHESS). Data was collected at 100 K with 20 % ethylene glycol as a cryoprotectant. Resultant data was scaled and reduced in HKL2000 (22). Phase information was obtained via molecular replacement with PHASER (23) and PHENIX (24) and a search model of VVD-36 (pdb 3D72). Iterative rebuild cycles were completed in COOT (25) and refinement with REFMAC5 and PHENIX (24). Final structures (PDB ID: 4WUJ) were solved to 2.23 Å with R and Rfree of 20.6 and 26.0% respectively.

4.7.3 Size-exclusion chromatography

To assay the formation of a light-state dimer, SEC was conducted using a Superdex 200 10/300 analytical column. The column was equilibrated with ENV-buffer. Light-state samples were generated by irradiation on ice for 5 min with a broad-spectrum flood light (150W). 100 µM light and dark-samples were run in the presence and absence of 5 mM DTT. To assay reversibility, previously irradiated samples were stored in the dark for 2-hrs prior to injection.

Apparent MWs of all protein constructs were measured relative to gel filtration standards (Sigma Aldrich). Absolute MWs for ENV constructs were measured using SEC in tandem with Multi-Angle Light scattering (MALS). Protein samples at 100 µM were applied to a Superdex 75 10/300 analytical column equilibrated with 50 mM Hepes pH 8, 100 mM NaCl, and 5 % glycerol, followed by light scattering and refractive index (RI) measurements on a Wyatt Minidawn light
scattering instrument. Samples were run at a flow rate of 0.35 ml/min and MWs determined using ASTRA software from Wyatt technologies. Apparent MWs from SEC traces were used to confirm the identity of protein peaks relative to SEC on the Superdex 200 used in all other studies.

4.7.4 Spectroscopy and kinetics

Photophysical properties of ENV1 and its variants were obtained using protocols detailed in Chapter 2 and previous literature (26,27). UV-Vis spectra were obtained on an Agilent 8453 spectrophotometer. Light-state samples were prepared via irradiation on ice with a broad-spectrum flood-light (150 W) prior to spectra acquisition. Kinetics of adduct decay were obtained by measuring the absorbance at 450 and 478 nm at 296 K. Frequency of data collection was altered to allow ~10-15 data points per half-life to avoid repopulation of the light-state adduct. Resultant data was fit with a mono-exponential decay and kinetic parameters reported in Supplementary Table 1.

4.7.5 Disulfide reduction potential

Redox–titrations were used to extract the midpoint reduction potential of the C96 disulfide. A constant total glutathione concentration of 5 mM was maintained under aerobic conditions. Total reaction volumes of 500 µl were prepared for all 100 µM protein samples. The GSSG/GSH ratio was manipulated to achieve reduction potentials in the range of -390- -230 mV. Samples were incubated in the light- or dark for 2-hours prior to running on a non-reducing SDS PAGE gel. Monomer and dimer fractions were calculated using GelAnalyzer software and plotted as a function of the GSSG/GSH reduction potential. The resulting plots were fit with a titration curve and the midpoint potential was extracted. The procedure was repeated in triplicate and the average reported +/- 1 standard deviation.
4.7.6 Strains and cultivation conditions

*Trichoderma reesei* QM6a (ATCC13631) was used as parental strain and complementation with mutated alleles was performed in QM6a Δenvl, which lack the whole reading frame encoding ENV1 and shows an equal phenotype to similar mutant strains in the background of QM9414 (5,13).

For assessment of stress response, strains were cultivated on plates with malt extract medium (3% w/v) or Mandels-Andreotti minimal medium (28) with glucose (1%w/v) or carboxymethylcellulose (1% w/v) as carbon source containing 25 mM menadione (oxidative stress). All combinations were assayed in constant darkness, low light (1500 lux) and high light (8000 lux) at 28 °C. Colony diameters were measured after two days in constant darkness and after 3, 4 and 5 days in low light and high light. QM6a and QM6a Δenvl were used as controls for every set. At least five replicates were analyzed for each set and strain. For complementation strains, three different recombinant strains were used at least in duplicates. Statistical evaluation of growth assays was performed using PSPP version 0.7.9 (one-way ANOVA, post-hoc tests; http://www.gnu.org/software/pspp/).

For investigation of cellulase gene expression, strains were grown on Mandels Andreotti minimal medium with 1% (w/v) microcrystalline cellulose in constant light and constant darkness as described previously (5).

In all cases, plates for inoculation were pre-grown in constant darkness for 14 days to avoid interference by circadian rhythmicity and harvesting or measurements were taken at similar times every day.
4.7.7 Construction of ENV1-C96T complementation strains in T. reesei

For construction of the complementation vector pENV1C96TamdS the adapted yeast-based recombination method was applied as described earlier (32) using primers Env1C96T3F (5’ ATTGACCTTGCCCTCTCGACACCTC 3’), Env1C96T3R (5’ GATGACATTCTACTCAAGACGACGAACCGCTACACAAAC 3’), Env1C96T5F (5’ GTAACGCCAGGGTTTTCCCATCTCGGACAATTCGACAGACG 3’) and Env1C96T5R (5’ AGACGGAGGTGTGCAGAG 3’) which contain overlap sequences for combination with the *amdS* marker cassette (22). The vector was sequenced to confirm the presence of the mutation and an amplified fragment was used for protoplast transformation of QM6a Δ*env1*. Copy numbers of integration were evaluated for all strains to be used (30,31) and strains containing more than one copy were excluded from further analyses. Again, ENV1 was sequenced to verify the presence of the desired mutation without further errors.

4.7.8 Isolation of total RNA and qRT-PCR

For isolation of RNA, mycelia were harvested and frozen in liquid nitrogen after 72 hrs of growth either in light or under red safety light (darkroom lamp, Philips PF712E, red, E27, 15 W) in case of cultures grown in constant darkness. Isolation and quality control of total RNA as well as qRT-PCR were performed as described previously with the ribosomal gene *l6e* as light-independent control (30). Statistical evaluation of results was performed using qBase+ (ANOVA, Tukey-Kramer post-test, p<0.05).
REFERENCES


CHAPTER 5:

CHARACTERIZATION OF ENV1 AND VVD MUTANTS

Both ENV and VVD are short LOV proteins from fungus *Trichoderma Reseii* and *Neurospora crassa*. They are homologous and regulate gene transcription by blue light regulation of dimer formation (1,2). Despite their homology, differences in signaling mechanisms exist. VVD plays a role in phase resetting of the circadian clock by regulating the transcriptional activity of WCC (3). VVD competes with WC-1 to form a heterodimer with WC-1 on WCC to suppress gene activation (3). Although, VVD allows for gene activation under certain light intensities (4). Therefore, VVD resets the core circadian clock's sensitivity to light (4). The existence of circadian rhythms in all filamentous fungi has not been proven yet. However, the core blue light photoreceptors WC-1 and VVD are retained.

Recent research on an economically important filamentous fungi *Trichoderma reseei* show that it retains the two-core blue light photoreceptors but is adapted to integrate blue light signaling and oxidative stress responses (2). In, *Trichoderma*, blue right receptor-1 (WC-1 homologue) and blue light receptor-2 (WC-2 homologue) initiate gene transcription by dimer formation and ENV1 (VVD homologue) negatively regulates gene transcription by competing with BLR2 (5,6). In addition to blue light responses, we have seen that ENV1 also helps the organism in coping with oxidative stress (2). ENV1 has evolved from its homologue protein VVD to respond to additional stimuli. Sequence alignment of ENV1 and VVD show that specific amino acids that are key in signal translation in VVD are different in ENV1. These amino acid substitutions in ENV1 has enabled it to respond to additional stimuli.
5.1 Solution characterization of ENV1 S99A mutant

As discussed in the previous chapter due to particular amino acid variations the dark state crystal structure of ENV-64 resembles the light state crystal structure of VVD. ENV1 has a Ser99 residue on the N-terminal hinge that connects PAS core to N-cap that corresponds to Ala72 in VVD. On examination of the crystal structure of ENV-64 it was observed that the hydroxyl group of Ser99 forms a hydrogen bond with the carboxylic side chain of Gln204. In VVD, the carbonyl side chain of Ala72 on the hinge abuts the carbonyl side chain of Gln182 that creates an unfavorable interaction (7). Upon light activation and flavin N5 protonation, the amide side chain of Gln182 rotates away from N5 and forms a hydrogen bond with carbonyl side chain of Ala72 that relieves the unfavorable interaction (7,8).

In ENV1 the hydroxyl group of Ser99 already forms a favorable interaction with carbonyl of Gln204 (2). Therefore, a favorable hydrogen bond between a Gln204 on the Iβ sheet and Ser99 on the N-terminal hinge is present in the dark state of ENV1. Also, neighboring to Ser99 ENV1 has a Val98 residue that corresponds to Cys71 in VVD. In VVD Cys71 sulfhydryl group breaks a hydrogen bond with carbonyl of Asp68 and swivels to form a hydrogen bond with the amide side chain of Asp68 in light state (8). The two methyl groups of Val98 can mimic both the positions of sulfhydryl group of Cys71 in light and dark state. The amide side chain of Val98 in ENV1 also forms a hydrogen bond with carbonyl side chain of Asp95 (Asp68 in VVD). These substitutions in the ENV1 lead to rotation of hinge residues 93-98 towards the PAS core that is seen in VVD light state structure (2). The position of Pro98 is similar to Pro66 in VVD light state. In VVD, the changes in hydrogen bond arrangement on light activation cause a 3.5 Å movement in Pro66 that leads to movement of bβ towards the β sheet and destabilizes the N-terminal latch contacts to LOV core which then swivels away from LOV core to form dimeric contacts (7). Therefore, due to
Ser99 and Val98, the hydrogen bond arrangement in the ENV1 dark state is similar to VVD light state and causes the N-terminal to adopt a similar confirmation like VVD light state. Further, in VVD light and dark state, there is a hydrogen bond between Asp68 and Ser70 whereas in ENV1 Ser 97 corresponding to Ser70 is rotated away from bβ and does not form hydrogen bonds with Asp95 (2,8). Disruption of these interactions promotes translation of bβ to mimic VVD-LSD in dark-state ENV1. Further disruption of the Asp95-Ser97 interaction stabilizes contacts between the N-latch and N-hinge at the dimer interface in ENV1.

In order to show that differences in signaling mechanisms between ENV1 and VVD stem from amino acid variations we performed mutations on ENV1 and VVD and investigated their behavior. We mutated a Ser99 in ENV1 to an Alanine which is native to VVD on a full-length protein construct ENV (1-207). The mutation to an alanine would disrupt the dark state hydrogen bond between Gln204 on Aβ sheet of LOV core and Ser99 on N-terminal hinge. The disruption of hydrogen bond between LOV core and N-terminal hinge in dark state could restore VVD like signaling mechanism. We expressed and purified ENV (1-FL) S99A and performed size exclusion chromatography on a Superdex 200 10/300 analytical column to analyze its solution behaviour. Previous experiments of size exclusion chromatography have shown that ENV1 WT requires light as well as oxidative conditions to form a dimer. In the presence of reducing agents like DTT dimer formation was wholly abolished. The mutant ENV(1-FL) S99A exhibited abherant elution profiles. Dark-state ENV (1-FL) S99A elutes as a monomer-dimer mixture, however the dimeric state is more compact than that observed in WT proteins. Light-activation promotes efficient formation of an expanded dimeric-state that resembles the WT. Treatment with DTT, substantially reduces dimer formation, but results in an expanded monomeric state.
Figure 33: ENV1 dark state resembles VVD light state like conformation: A) ENV1 dark state structure (yellow) is overlapped on VVD dark state (grey) structure. The hinge region of ENV1 is more shifted towards the PAS core as compared to the VVD hinge region. The N-terminal latch of ENV1 is away from the PAS core whereas in VVD it wraps around the PAS core. B) ENV1 dark state (yellow) structure is overlapped on VVD light state structure (pink). The hinge and latch regions of ENV1 adopt similar conformation to VVD light state. The Ser99 hydrogen bonds to Gln204 similar to Gln182 and Ala72 bond in VVD. The methyl group of Val98 of ENV1 adopts similar conformation to the thiol group of Cys71. Pro93 of ENV1 also has a similar conformation to Pro66 of VVD.
The ENV1 (1-FL) S99A exhibited a shift in its elution volume even in presence of DTT at higher concentrations of protein showing concentration-dependent dimer formation like VVD. Mutation from serine to alanine makes ENV1 undergo similar conformational responses like VVD upon light activation. Therefore, a shift in elution volumes is observed even in the presence of DTT due to conformational changes occurring at the N-terminal hinge. Upon light activation due to protonation of flavin N5 position Gln204 flips to form a hydrogen bond with Ala99 on the N-terminal hinge (7). Such conformational changes do not occur in ENV1WT upon light activation due to presence Ser99 it already has a conformation similar to light state VVD in its dark state. Therefore, in ENV1 S99 stabilizes the dark-state, allowing dark-state functions but disrupting VVD-like signaling. Combined, these results suggest that S99 H-bonds promote an expanded light state like conformation of VVD. In the absence of the stabilizing H-bond, ENV1 adopts a more compact structure that can be released via a VVD-like mechanism forming an H-bond between Gln204 and Aβ.
5.2 Characterization of ENV1 V98C mutant

We also mutated Valine98 to a cysteine residue on an N-terminal truncated construct (65-207) to observe its solution characteristics. ENV-65 V98C eluted as a mixture of dimer and monomer in its light state. Although, ENV-65 V98C formed a very weak dimer in the light state the dimer percentage was only about 30%. Thus, mutation of valine to cysteine in ENV1 abrogates dimerization and shows that valine is a key residue involved in signal translation. Val98 predisposes ENV1 to dimerization, and thus it requires very few conformational changes to form a dimer upon light activation. As seen earlier due to S99 residue there is already a hydrogen bond...
between the β sheet of LOV core and N-terminal hinge in the dark state of ENV1. So, a new hydrogen bond to the N-terminal hinge is not formed on light activation such that it does not perturb the conformation of the neighboring cysteine residue in the V98C mutant. In VVD, the formation of hydrogen bond between Gln182 and Ala72 on light activation causes the neighboring Cys71 to change from a buried to an exposed conformation to initiate dimerization. In ENV1 absence of formation of a new hydrogen bond to the hinge, residues do not disturb the cysteine residue in V98C mutant and predispose it to dimerization. Therefore, V98C mutant forms a very weak dimer as the cysteine residue may not change its conformation upon light activation.

5.3 Importance of HI loop in signaling

The crystal structure of ENV1 revealed multiple conformations of the loop connecting H, and I beta sheets. The conformation of the HI loop affects the pitch of the αα helix in the N-cap. These lead to differences across the dimer interface in comparison to VVD light state dimer. Upon superimposing the light state structure of VVD on ENV1, a subtle alteration in the dimer interface of ENV1 was observed due to rotation of αα helix. The differences are centered around Pro189

Figure 35: Size exclusion chromatography of ENV-65 V98C: A) ENV-65 V98C forms a very weak dimer on light activation (red) it is mostly monomeric. In dark state, it elutes as a pure monomeric species.
(Pro167 in VVD) and W191, which distorts Hβ leading to rearrangement of the H-I loop and associated contacts with the N-cap. The proline 189 in the HI loop flips its carbonyl group in 3 out of 4 molecules that disrupt the contacts to β sheet resulting in multiple conformations of the loop. In the fourth molecule, the carbonyl moiety of Pro189 is positioned to retain H-bonding patterns indicative of β-sheet formation (Fig. 36B). The HI-loop residues make contact with the αα helix of the N-cap. Therefore, altered conformations of HI-loop affects its contacts to the αα helix and thereby changing the helix pitch. Helix movement is coupled to weak interactions between Arg80 of αα helix and Asp193 of HI loop that allow the formation of a π-cation interaction between Arg80 and Trp191. Such multiple conformations in HI loop are not observed in VVD.

**Figure 36: HI loop conformations:** A) The multiple conformations of HI loop are shown in two different molecules of ENV1WT. The different conformations of HI loop affect the helical pitch of the αα helix. B) Pro189 flips its peptide in two different molecules which leads to varying conformations of Trp191. E) The alternative conformations lead to movement of the αα helix away from the LOV core. Helix movement is coupled to weak interactions between R80 and D193 that allows the formation of a π-cation interaction between R80 and W191. Altered contacts are depicted with dashed lines (black-primary conformation, red alternative conformation)
We also mutated Trp191 to an Arg native to VVD, on a ENV1 (1-FL) S99A construct to observe its behavior. The S99A:W191R mutant undergoes a rapid monomer:dimer interconversion like VVD under light-state conditions. In the presence of light and reducing conditions, it exhibits a concentration-dependent elution profile. At high protein concentrations like 300µM the elution volume represented the molecular weight of an intermediate between dimer and monomer. The dimer dissociation constant of S99A:W191R is 20 µM similar to VVD WT with a dimer affinity of 13µM. Thus, mutant S99A:W191R is able to restore VVD like signaling in ENV1. Thus, both the N-hinge (S99A) and H-I loops stabilize the ENV1 fold in a light state like VVD conformation and inhibit VVD-like signal transduction. Disruption of both sites promotes VVD-like behavior. Combined, the data implies that contacts between the H-I loop and N-cap may mediate signal transduction in VVD-like species. Further, signal propagation involves coupling between the β-scaffold, H-bonding to the active site Gln and structural rearrangement of the N-terminal hinge motif (1,8,9). In this manner, ENV1 indicates conformational dynamics of VVD-like LOV domains may be more substantial than indicated in light-state crystal structures but, can be modified to incorporate additional environmental stimuli (oxidative stress).
Figure 37: Size exclusion chromatography of ENV1 1-FL S99A:W191R
A) The light state (red) of ENV (1-FL) S99A:W191R elutes as a mixture monomer and dimer. The dark state elutes as a monomer (black). The reduced light state (green) in the presence of 5Mm DTT (green) elutes with a slight shift in elution volume representing maybe a mixture of dimer and monomer compared to reduced dark state (blue). B) Elution profile of ENV (1-FL) S99A:W191R in reduced light state conditions at injected concentrations of 300µM (black), 100µM (red), 50µM (blue), 20µM (magenta), 10µM (green), 5µM (navy-blue) show concentration dependence like VVD. The change in elution volumes at different concentrations may be due to changes occurring at the N-terminal due to light activation unlike in ENV1WT.
5.4 A72S mutation blocks VVD dimerization

In attempts to make VVD have ENV1 like signaling, an alanine to serine mutation was made. As seen earlier, Ala72 is a residue on the N-terminal hinge which is crucial for signaling and establishing contacts between LOV core and N-terminal. The VVD A72S mutant was expressed and purified. The mutant protein was expected to have contacts between LOV core and N-terminal hinge in its dark state as seen in ENV1. The Ser99 in ENV1 forms a hydrogen bond with Gln204 in its dark state which makes ENV1 have a light state like conformation of VVD. Although, VVD A72S was expected to have contacts between the N-terminal hinge and LOV core in its dark state. Surprisingly, VVD A72S mutant eluted as a monomer in the light as well as dark state.

![Size exclusion chromatography of VVD A72S and VVD WT](image)

**Figure 38: Size exclusion chromatography of VVD A72S and VVD WT:** A) The light (green) and dark (blue) state of VVDA72S elutes as a monomer similar to the dark state (black) of VVD WT. The light state elution profile of VVD WT in comparison to VVD A72S is shown in red.
We also attempted to crystallize VVD A72S mutant. The crystals of VVD A72S in its dark state were obtained in space group P1211 same as VVD WT. The crystals were obtained by slightly modifying the buffer conditions in which crystals for VVD WT were obtained (1). The crystallographic dimer interface of VVD A72S was like the one observed in crystals of dark state VVD WT.

**Figure 39: VVD A72S dark state dimer resembles VVD WT dark state dimer:** A) Crystal structure of VVD A72S in its dark state. The dimer interface of VVD A72S is similar to VVD WT dark state. B) Crystal structure of VVD WT in its dark state where the N-terminal is packed against the LOV core. C) Crystal structure of light activated VVD WT dimer the N-terminal is displaced away from the LOV core to form dimer interface.

On observing the crystal structure of VVD A72S, it was found that it has similar hydrogen bond between the hydroxyl group of Ser 72 and the carboxylic side chain of Gln182 in its dark state as observed in ENV-64 between Ser99 and Gln204. The presence of hydrogen bond between LOV core and N-terminal caused ENV-64 in its dark state to have a light state like conformation of VVD due to mimicking of light state like hydrogen bond network of VVD. Although the mutant VVD A72S has a hydrogen bond network similar to VVD light state its conformation resembles that of a dark state VVD-WT. Upon superimposing structures of VVD A72S and VVD WT dark, it was found that they are almost identical. The conformation of Cys71 in VVD A72S is like conformation of Cys71 in VVD WT-dark state. It was previously reported that Cys71 changes its conformation from buried to a more exposed position upon light activation. The conformation of
Cys71 was found to be very important for initiating structural changes in VVD upon light activation. Although the presence of hydrogen bond network like light state VVD the conformation of Cys71 in VVD A72S is identical to its conformation in VVD-WT dark state structure.

Therefore, there may be two possibilities due to which VVD A72S is not capable of dimerization 1) In the dark state as seen in crystal, structure Ser72 forms a hydrogen bond with the carboxylic side chain of Gln182 similar to ENV1. Upon light activation as Gln182 flips the amide side chain forms a hydrogen bond with the carboxylic side chain of Ser 72. As the hydrogen bond is always present in both light and dark state between the LOV core and N-terminal hinge, the neighboring Cys71 is not disturbed, and it remains in its buried conformation. Therefore, VVD A72S is not capable of dimerization as Cys71 does not change conformation. Also, maybe a double mutant A72S:C71V will be capable of dimerization as ENV1 has a valine residue instead of cysteine. The valine residue can mimic both the conformations of Cys71 and as shown in previous studies, a C71V mutation in VVD lead to partial dimerization even in its dark state. 2) Also, it could be that A72S mutation makes VVD more ENV1 like in a way that it is capable of dimerization only in the presence of both light and oxidative stress.
Figure 40: Superposition of VVD WT dark and VVD A72S: A) The crystal structure of VVD WT dark state (gray) is superposed on VVD A72S dark state (yellow). The two structures are identical with no differences in conformation of the N-terminal hinge and N-terminal latch. In VVD A72S the Ser 72 forms a hydrogen bond with Gln182. The conformation of Cys 71 in VVD A72S is similar to Cys 71 from VVD WT dark state structure.
Figure 41: Superposition of VVD WT light state and VVD A72S: A) The crystal structure of VVD light state (blue) is superimposed on VVD A72S dark state (yellow). The N-terminal hinge in light state VVD is more shifted towards the LOV core than VVD A72S. The N-terminal latch of VVD light state is rotated away from the LOV core. There is a significant difference in rotation of Pro66 residue in both structures. Although there is a similar hydrogen bond network between Ala72 and Gln182 in VVD light state and Ser72 and Gln182 in VVDA72S the conformation of Cys71 varies in both structures.
Figure 42: Superposition of VVD A72S and ENV-64: A) The crystal structure of ENV-64 dark state (blue) is superimposed on VVD A72S dark state (yellow). The N-terminal latch of ENV-64 rotated away from the LOV core. There is also a difference in position of Pro66 and Pro98 in two structures. Ser 99 from ENV-64 and Ser72 from VVD A72S form a hydrogen bond with Gln 204 and Gln 182 respectively in dark state. One methyl group, Val98 of ENV-64, mimics the Cys71 position of VVD A72S the other methyl group resembles the position of Cys71 in VVD light state structure.
Therefore, mutation of key amino acids in ENV1 that are involved in signaling to residues native to VVD exhibited similar solution behavior like VVD (8). The double mutant ENV1 S99A: W191R exhibited similar rapid exchange behavior like VVD and had similar dimer affinity. Although, these mutants were not tested to see how it affected the biological function \textit{in-vivo}.

Attempts to make VVD have ENV like signaling behavior were not successful. Mutation of Ala72 to a serine residue native to ENV decoupled structural changes following light activation in VVD. Thus, A72S mutation disrupted structural changes and dimer formation \textit{in-vitro}. Although, A72S mutant has similar hydrogen bond network to ENV and VVD light state between the LOV active site and N-terminal in its dark state, due to the absence of allosteric effects on light activation the Cys71 does not change conformation and initiate structural changes.
5.5 Materials and methods

5.5.1 Cloning and purification

An N-terminally truncated VVD construct was obtained (VVD-36). A point mutation Ala72Ser was introduced in the VVD-36 using Quick-Change protocol (Strata-gene). The DNA sequence was verified by sequencing (Genewiz). All ENV constructs consisting of full-length (1-207) and N-terminally truncated constructs ENV-64 (65-207) were cloned from cDNA obtained from T. reseei. Envoy constructs were cloned into pGST parallel vector using Nco1 and Xho1 sites.

Point mutations such as Ala72Ser and Ser99Ala were introduced into VVD-36 and ENV respectively using Quick-Change protocol (Stratagene). All constructs were verified by DNA sequencing (Genewiz). VVD constructs were expressed in E.coli BL21 DE3 cells under constant light conditions. ENV constructs were expressed in E.coli JM109 cells. Cells were grown at 37 °C until the O.D reached to 0.6. The temperature was then decreased to 18 °C for 40 min and cells were induced with 0.2 Isopropyl thiogalactoside (RPI). After 22 hrs of protein expression, ENV pellets were harvested and stored in buffer containing 50 mM Hepes, 100 mM NaCl, 10 % Glycerol pH 8.0 and VVD pellets were stored in buffer containing 50 mM Hepes, 150 mM NaCl, 13% glycerol pH 8.0.

All ENV constructs were purified on a glutathione affinity resin (QIAGEN). The protein was eluted from the column by treatment with 2 mg TEV protease per milliliter of resin for 2 hrs at 22°C. Cleaved protein were eluted in buffer containing 50 mM Hepes, 100 mM NaCl, 5 mM DTT 10 % Glycerol pH 8.0. An additional round of nickel-nitriloacetic acid chromatography was conducted to remove His6 TEV prior to final purification on Superdex S200 size exclusion column. Purified constructs were concentrated to ~ 100µM and used within 8 hrs to avoid oxidation. All VVD constructs were purified by nickel-nitriloacetic acid chromatography. The
protein was eluted from the column with buffer containing 50 mM Hepes, 150 mM NaCl, 250 mM imidazole 13% glycerol pH 8.0. After elution protein was treated with thrombin overnight at 4 °C prior to final purification on Superdex S200 size exclusion.

5.5.2 **Size exclusion chromatography**

Size exclusion chromatography was used to observe conformational changes following light excitation in VVD and ENV variants such as ENV S99A and VVD A72S. All samples were injected in 500 µl aliquots with a concentration range of 5–500 µM. Light-state samples were generated by irradiation on ice for 5 min with a blue laser. 100 µM light and dark samples were run in the presence and absence of 5 mM DTT for all ENV constructs. Spectra were obtained using Agilent 8453 spectrophotometer to verify the light and dark state of protein. the Apparent molecular weights (MWs) of all protein constructs were measured relative to gel filtration standards (Sigma-Aldrich). Protein samples at 100 µM were applied to a Superdex 75 10/300 analytical column equilibrated with suitable buffer. Apparent MWs from SEC traces were used to confirm the identity of protein peaks relative to SEC on the Superdex 200 used in all other studies. Approximate dissociation constants were obtained via a linear fit of $1/[M_F]$ vs. the mole fraction term as shown

$$K_d = m^2[M_F]d(m+2d)$$

5.5.3 **Determination of dissociation constants ($K_d$)**

The dimer dissociation constants were determined for ENV-WT and mutant proteins ENV-S99A and ENV-S99A: W191R. The proteins were run on a Superdex 75 10/300 analytical column in reduced light state at various concentrations. The calculation of dissociation constants from SEC was described in Chapter 3 methods.
5.5.4 Spectroscopy and kinetics

The photophysical properties of ENV, VVD and its variants were obtained using protocols mentioned before. UV-visible spectra were obtained on an Agilent 8453 spectrophotometer. Light-state samples were prepared via irradiation on ice with a blue laser prior to spectrum acquisition. Kinetics of adduct decay were obtained by measuring the absorbance at 450 and 478 nm at 296 K. The frequency of data collection was altered to allow 10–15 data points per half-life to avoid repopulation of the light-state adduct. The resultant data were fit with a mono-exponential decay and kinetic parameters were reported.

5.5.5 Structural analysis

VVD-A72S crystals were obtained with reservoir solutions containing 100 mM Sodium citrate dihydrate pH 5.6, 18% (w/v) Peg 4000 and 100 mM ammonium acetate using 1.5 µL well solution with 1.5 µL of VVD A72S at 5 mg/ml. Protein for crystallographic studies were purified in 50 mM Hepes pH 8.0, 100 mM NaCl and 13 % Glycerol. Crystal trays were then set in a dark room illuminated with a red safe light. Crystals were obtained in about a week at room temperature.

Diffraction data was collected at the F1 beamline at the Cornell High-Energy Synchrotron Source (CHESS). Data was collected at 100 K with 20% ethylene glycol as a cryoprotectant. Resultant data was scaled and reduced in HKL2000 (10). Phase information was obtained via molecular replacement with PHASER (11) and PHENIX (12) and a search model of VVD-36 (pdb 3D72). Iterative rebuild cycles were completed in COOT (13) and refinement with REFMAC5 (14) and PHENIX (12). Final structures were solved to 2.4 Å with R and R_free of 19.6 and 30.0.0% respectively.
REFERENCES


CHAPTER 6:
ENV1 PHOTOCHEMISTRY AND OSMOTIC STRESS SIGNALING

This chapter includes figures and data previously published in the *Journal of Biological Chemistry* and is referenced as: (1)

6.1 ENV Photochemistry

LOV domain chemistry is characterized by blue-light induced formation of a covalent adduct between a bound flavin cofactor (FMN, FAD or riboflavin) and a conserved Cys residue in a GXNCRFLQ motif. Concomitant with adduct formation is protonation of the N5 position of the isoalloxazine ring, which has been argued to impart signal transduction in LOV proteins (2,3,4). The covalent adduct is defined by a broad UV-vis absorption band centered at approximately 390 nm (LOV$_{390}$). Upon return to the dark, the adduct-state spontaneously decays to an oxidized flavin (LOV$_{450}$) on the timescale of seconds-to-days (5,6). Currently the biological role of the wide range in photocycle lifetimes is unknown, however several studies have suggested that they facilitate adaptation to changing levels of light intensity (5,7,8). For these reasons, chemical tuning of the LOV photocycle lifetime through understanding of the adduct decay mechanism has been attempted in several systems (5,6,9,10,11,12,13).
Several lines of reasoning have led to a general mechanism of adduct decay. First, solvent isotope effect (SIE) experiments indicate that a single proton abstraction event is rate limiting (5,6). Second, adduct decay can be catalyzed by the presence of small molecule bases such as imidazole (14). Third, residue substitutions at regions that regulate solvent access to the flavin active site have a substantial effect on LOV photocycle lifetimes (6,15). Combined these, experiments implicate N5 deprotonation as the rate-determining step in adduct decay. Consistent with such a model, mutation of residues that regulate accessibility of small molecules to the N5 position or that tune hydrogen bonding characteristics affect kinetics of LOV proteins (5,6,10,11,13,15,16). Importantly, the natural base responsible for N5 deprotonation remains to be determined, however several possibilities have been suggested that involve elements in the immediate vicinity of the active site flavin (6,15).

In LOV proteins most residues near the N5 position are hydrophobic, with the exception of a conserved Gln residue important for signal transduction (17). That has led to two proposed models regulating N5 deprotonation. 1) Gln facilitated transfer of a proton from the N5 position to the active site Cysteine (6). Consistent with the conserved Gln acting as a proton transfer agent, Gln→Leu and Gln→Asn mutations have a large effect on adduct decay kinetics (18,19). 2) The involvement of ordered water molecules within the flavin binding pocket (15). Crystal structures partially support the latter mechanism, where LOV proteins conserve an ordered solvent channel leading to the active site Cys residue (2,6). However, the water molecules do not penetrate deep enough to directly affect the N5 proton, leading to ambiguity as to their role in adduct decay. Recent FTIR studies indicate these ordered water molecules play a key role in regulating LOV
lifetime (15); however, whether the effect is direct or indirect remains to be determined. Further, analysis of a bacterial sLOV protein (McLOVn) indicates that introduction of a Thr residue in the vicinity of the N5 position can abrogate base catalysis in some systems (20).

The crystal structure of ENV1 indicate that it can function as a model system for the exploration of the effect of ordered water on LOV kinetics (21). Specifically, similar to McLOVn, a hydrophilic residue (Thr101) in the vicinity of the N5 position provides a hydrogen bond donor/acceptor that may enable solvent recruitment near N5 (20). Initial photochemical characterization confirmed LOV type chemistry, with a time constant for adduct scission of ~1500 seconds (21). Such kinetics are 15-fold faster than its homolog, VVD, despite high sequence conservation (21,6). To examine the origin of the altered rate of adduct decay, we conducted a comprehensive study of the ENV1 photocycle and corresponding kinetics.

Full-length (FL; 1-207) and an N-terminally truncated construct ENV-64 were used for measuring photo-cycle kinetics. Using both constructs allows confirmation that photocycle properties are conserved following truncation and allows direct computational studies of the photocycle using the ENV-64 crystal structure (PDB ID 4WUJ). As reported previously (21), both constructs purify with a bound FMN cofactor and demonstrate spectra consistent with blue-light induced formation of a covalent cysteiny1-flavin C4a adduct (Figure 43A). Interestingly, depending on purification conditions, the FMN cofactor either purifies as a mixture of oxidized FMN and reduced neutral semiquinone (in the presence of reduced glutathione) or oxidized FMN (no glutathione) (Figure 43A, B). Such behavior suggests that the oxidation potential of the FMN cofactor may reside within physiologically relevant ranges as has been observed previously in LOV protein variants (6,22), and has recently been shown of being competent for signal transduction (23). Thus, ENV1 may employ the semiquinone in dark-state signaling. Notably,
after cycling through one photocycle, we have so far been unable to chemically re-reduce the FMN cofactor, suggesting that oxidation of an intrinsic element may abolish subsequent chemical reduction.

Comparisons of the rate of adduct decay in the presence (reduced) and absence (oxidized) of glutathione demonstrates similar rates of adduct decay under all conditions and constructs tested. However, the rate of oxidation of the semiquinone displays distinctly different kinetic properties. Kinetic traces obtained at 450 and 478 nm from light-activated oxidized ENV-64 and FL ENV1 reveal time constants of 1500 and 1600 seconds respectively (Figure 43D). In contrast kinetic traces of reduced samples are biexponential. The minor component (30%) demonstrates a time constant comparable to oxidized samples (1100 seconds), whereas the major component (70%) has a time constant of 180 seconds (Figure 43D). Analysis of the kinetic trace at 622 nm, which isolates the spectral signature of the neutral semiquinone, reveals mono-exponential decay with a time constant of 180 seconds (Figure 43D). We thus assign the major component as resulting from FMN oxidation and the minor component resulting from adduct decay. We conclude that the rate of oxidation proceeds ~10-fold faster than adduct decay pathways in ENV1. Similar mixtures of oxidized and reduced semiquinones were observed in slow cycling variants of *N. crassa* VVD, where the kinetics of adduct decay and oxidation seemed to occur on similar time scales (6,22).
Figure 43: Kinetics and spectra of ENV1 proteins: A) Light (red) and dark (black) spectra of ENV 1-207 in the absence of glutathione. The light state spectra contain a single peak at ~380 indicative of C4a adduct formation. B) In the presence of glutathione the dark-state is still consistent with oxidized flavin (black curve), however in the light (red) C4a adduct formation competes with semiquinone formation (broad absorption band above 500 nm). Kinetic traces at 450 (black), 478 (red) and 622 (blue) nm confirms competition between flavin semiquinone formation and C4a adduct formation. In the absence of glutathione (C), ENV1 kinetics are mono-exponential. In contrast, kinetics of adduct decay (450 and 478 nm) in the presence of glutathione (D) are biexponential with one component matching the rate of semiquinone oxidation (from 622 nm kinetics).
Sensitivity of WT ENV1 to reducing conditions and a 15-fold faster rate of adduct decay demonstrate key photocycle differences compared to its homolog VVD. To better understand factors affecting the ENV1 photocycle, we probed the reaction landscape using Eyring and Arrhenius analyses. Kinetic studies revealed energies of activation comparable to other circadian clock photoreceptors, but divergent compared to typical LOV proteins. Specifically, both FL and ENV-64 demonstrated lower than expected energies of activation of 67 and 70 kJ/mole respectively (Figure 42). Arrhenius parameters for most LOV proteins, of similar photocycle lifetimes, have reported activation barriers on the order of 80-100 kJ/mole (5,11). Notably, studies of other circadian clock photoreceptors indicate atypically low energy barriers that Eyring analysis reveals are compensated by large unfavorable entropies of activation (5). Similarly, Eyring analysis of ENV1 indicates that both FL and ENV-64 exhibited low enthalpies (ΔH‡) of activation (68 kJ/mole and 72 kJ/mole respectively) that are compensated by large unfavorable entropies (ΔS‡) (-75 J mol⁻¹ K⁻¹ and -63 J mol⁻¹ K⁻¹ respectively) (Figure 44B). Studies of other LOV proteins with entropy-compensated photocycles revealed alterations in H-bonding to the active site flavin (5,11). These alterations in H-bonding often affect or are affected by solvent accessibility to the active site (5,11,13). Examination of SIE’s in FL ENV1 is consistent with proton abstraction being rate limiting (SIE=4.1), thus solvent likely plays a key role in regulating ENV1 kinetics. To better examine the factors altering the ENV1 reaction landscape we examined solvent accessibility in ENV1.

Base catalysis studies were performed by addition of varying concentrations of the small base imidazole. Imidazole has been reported as an efficient enhancer of dark state recovery in LOV proteins (14). While no direct conclusions can be drawn concerning recovery mechanism, the sensitivity of LOV kinetics to changes in imidazole concentration can indicate alteration of solvent
site accessibility (25,5). Base catalysis of FL ENV1 and ENV-64 revealed a modest effect of imidazole on ENV1 kinetics (Figure 43). Importantly, ENV1 is approximately 15-fold more sensitive to imidazole than its homolog VVD. Thus, the difference in photocycle kinetics in ENV1 and its homolog VVD may be explained by the difference in solvent accessibility of these proteins.

6.2 Photocycle kinetics of ENVT101I

To examine the origin of the variation in solvent accessibility, we examined fungal sequences for residue substitutions in vicinity of flavin cofactor. Alignment of ENV1 homologues from different genera and phylogeny revealed that the Threonine at position 101 is confined the fungi *Trichoderma, Cordyceps, and Beauveria* as well as an isolated member of the bacteria *Methylcystis* (Figure 44). Residue substitutions at this site are known to have large effects on LOV kinetics in VVD and other proteins (6,26). Specifically, an Ile74Val variant of VVD accelerates the photocycle 24-fold, however it has no effect on solvent accessibility (6). To the best of our knowledge, Ile74Thr or related substitutions have not been studied in VVD, however similar variants have been shown to substantially affect photocycle kinetics in a bacterial sLOV protein McLOVn (20).

Spectroscopic investigation of Thr101Ile indicated several differences compared to WT. First, the alternative pathway involving the reduced neutral semiquinone was abolished and kinetics were characterized as mono-exponential. Consistent with our predictions that Thr101 dictates accelerated photocycle kinetics via solvent recruitment, introduction of a Thr101Ile variant resulted in a 250-fold decrease in the rate of adduct decay. Base catalysis studies of the Thr101Ile variant confirms that the large effect on the kinetics of adduct scission correlates with a large effect on solvent accessibility. Specifically, the Thr101Ile variant is 100-fold less sensitive to imidazole (a=14). Thus, the presence of a Thr residue in the LOV active site confers additional
sensitivity to imidazole. Notably, the 250-fold change in photocycle kinetics and only a 100-fold effect on solvent accessibility indicates that other factors may be involved in tuning LOV photochemistry in these systems. Moreover, in contrast to McLOVn, where a Thr abrogates imidazole catalysis, in ENV1 a Thr enhances a base catalyzed mechanism (20). Combined the data indicates that the site adjacent to N5 is involved in the LOV mechanism but, acts in concert with additional sites.
Figure 44: Eyring plots and T101I kinetics: A, B) Eyring plots of ENV1-FL (A) and ENV-64 (B). Both constructs demonstrate moderate $\Delta H^\ddagger$ (FL: 68 kJ/mole; ENV-64: 72 kJ/mole) and large unfavorable $\Delta S^\ddagger$ (FL: -75 J mol$^{-1}$ K$^{-1}$; ENV-64: -63 J mol$^{-1}$ K$^{-1}$). C) Absorbance at 450 nm (black) and 478 nm (red) as a function of time for ENV-FL T101I in the presence of 50 mM imidazole. Kinetics at 0 mM are too long lived to reach full recovery on a reasonable time scale. The lifetime of adduct decay is 250-fold longer than WT ENV1.

Figure 45: Imidazole catalysis of ENV1: Dependence of the rate constant for adduct decay on imidazole concentration. Both FL and ENV-64 (A and B respectively) demonstrate similar effects of imidazole on the rate constant for adduct decay. The results are consistent with a solvent accessibility factor of 1500. In contrast, a T101I leads to increased solvent protection with an accessibility factor of 15 (C). D) Active site of ENV1 (PDB 4WUJ). An ordered solvent channel (red spheres) leads to the flavin active site. T101 contacts C135 and Q204 (black lines), leaving a small pocket adjacent to the N5 position. S99 anchors Q204 to a specific conformation allowing an H-bond to N5.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Time Constant (s)</th>
<th>Rate (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL ENV1</td>
<td>1445</td>
<td>$6.9 \times 10^{-4}$</td>
</tr>
<tr>
<td>ENV-64</td>
<td>1500</td>
<td>$6.6 \times 10^{-4}$</td>
</tr>
<tr>
<td>FL T101I</td>
<td>~400000</td>
<td>$2.5 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

Table 4: Kinetics of thermal reversion for ENV1 constructs and variants at 296 K.

Figure 46: Evolutionary relationships of fungal ENV1 homologs. Sequences were retrieved from the JGI Mycocosm database and species names are given along with the protein ID as provided in the respective JGI genome database (http://genome.jgi.doe.gov/programs/fungi/) The bar on the right side represents the amino acid present at the position corresponding to T101 in *T. reesei* ENV1. Phylogenetic analysis was performed as described previously (21) using the minimum evolution algorithm of MEGA4 with 500 bootstrap cycles.
Examination of the ENV1 crystal structure reveals several possible conclusions that may have implications for LOV kinetics and adduct scission mechanism throughout the LOV family. Previous studies indicate a possible role of ordered solvent in adduct decay mechanisms (25,6,15), however a precise locale of action has remained elusive. Given that Ile74Val variants in VVD have no effect on imidazole access, but introduction of a corresponding Thr to the ENV1 active site imparts a large effect on LOV kinetics, we propose that hydrophilic residues at this position may facilitate recruitment of solvent to the active site (Figure 45D, 47A-C). Consistent with such a mechanism, a Thr at position 101 allows sufficient space to order a water molecule in H-bonding contact to Thr101, the flavin N5 position and Gln204, which have all been implicated in the rate limiting steps of adduct decay as well as signal transduction (6,10,11,14). To probe such a mechanism further, computational studies were conducted at our collaborator’s lab (Dr. Roberto Bogomoni) to probe the reaction landscape and corresponding ability to recruit solvent to the active site to increase the rate of adduct decay.

6.3 Computational analysis for water occupancy at FMN active site

Based on our kinetic analysis of the ENV1 photocycle, we hypothesized that the addition of a large, hydrophobic side chain in the flavin-binding pocket disrupts recruitment of ordered water. In turn, the reduction in area available for water to reside creates a higher energy barrier for water to cross, reducing the velocity of dark state recovery. Although it has been previously noted that water diffuses into the active site (27), no energetics associated with this process have been measured. To test our hypothesis, we used molecular dynamics simulations to calculate the free energy change for water binding in the flavin HN5 region (protonated N5). In addition, we correlate recovery kinetics with the stability of water within the flavin-binding pocket.
Comparison of WT ENV-64 and the Thr101Ile variant indicate distinct differences in the energetics of active site water that have implications for both water occupancy and possible signal transduction mechanisms in LOV proteins. In both proteins, HN5 dictates the affinity of water for the flavin active site, where water occupancy is most dense between 200-300 pm from the HN5 position (Figure 47D-G). Despite similarities in the location and duration of active site water occupancy, WT and Thr101Ile show different energies of water binding (Figure 47D, E). Specifically, WT-ENV1 demonstrates a lower binding energy compared to Thr101Ile (11.5 kJ/mol and 15.5 kJ/mol respectively). The energy difference of 4.0 kJ/mol corresponds to the lack of a hydrogen bond to water from Ile101.
Figure 47: Wild type ENVOY active site: A) The dark-state crystal structure (PDB 4WUJ) of ENV1 contains a buried conformation of Gln204, consistent with all other LOV domain structures. In the buried conformation Gln204, H-bonds to the unprotonated N5. B) MD simulations of ENV1 demonstrate recruitment of water to the flavin active site following adduct formation. A water molecule is ordered near HN5 by H-bond interactions with Thr101 and Gln204. Water recruitment involves rotation of Gln204, from the buried conformation (A) to an exposed conformation that H-bonds with O4. C) Alternative low occupancy position of water in the flavin active site for the Thr101Ile variant. The lack of H-bonding to Thr101 in the Thr101Ile variant destabilizes water occupancy near HN5, allowing access to a low occupancy site in the vicinity of O4. The decreased occupancy near HN5 limits the ability of water to function as a direct base catalyst in deprotonating HN5, thereby decreasing the rate of adduct decay in the Thr101Ile variant. D) Radial distribution of water about HN5. WT proteins show higher probabilities for water occupancy near HN5. E) Potential of mean force (PMF), in kJ/mol for solvent near HN5. Water is stabilized by H-bonds to Thr101 in WT proteins. The Thr101Ile variant destabilizes water in the vicinity of the HN5 position. F, G) Water-HN5 distances plotted versus HN5-Q204 distance. Color indicates number of frames within a given hexagonal bin. F) wild type envoy. G) Thr101Ile mutant.
Further examination of molecular dynamics simulations provides insight into the coupling of ordered water in LOV active sites and conformational changes in the adjacent Thr101 and Gln204 sites. Currently, no structures of LOV proteins exist with an ordered solvent molecule adjacent to N5. The lack of water likely results from a dynamic competition between a buried conformation of Gln204 and solvent, where the relative populations vary based on the residue identity at position 101. Instead, ordered water in crystal structures is confined to a solvent channel adjacent to the flavin ribityl moiety (Figure 47D). MD simulations indicate that water accesses the flavin active site through two primary modes. The predominant access point is via a channel above the active Cys, opposite the ribityl group, between the Dα and Eα helix loops. In addition, water occasionally enters from a channel near the ENV1 C-terminus. In both cases, solvent access to the active site is coupled to rotation of Thr101 and movement of Gln204 that have implications for signal transduction and activation of adduct scission (47A-C).

Water recruitment by Thr101 results in rotation of the hydroxyl group to H-bond to the bound water. H-bond formation in turn results in local ordering of Thr101. In contrast, a Thr101Ile variant is ordered regardless of water occupancy (Table S1). Further upon binding, the HN5/Thr101-coordinated water recruits additional waters to the active site. Thus, Thr101 stabilizes ordered water to the active site and lowers the energy barrier for additional dynamic water near the flavin active site. Consistent with their dynamic nature, these additional waters are not fixed in one place, but rather readily diffuse into and out of the flavin-binding pocket. These dynamic waters in turn are coupled to the residue identity at position 101 and conformational changes in Gln204.

Rotamer analysis of Gln204 indicates that adduct formation and water recruitment to HN5 is coupled to rotation of Gln204, from a buried position (H-bonding to N5), to an exposed position.
(H-bonding to O4) (Figure 47). Such conformational changes have been observed in other MD simulations and have been implicated as a possible signal transduction mechanism (27). However, crystal structures and MD simulations indicate that signal transduction may rather involve only rotation of Gln204 to H-bond to the newly protonated HN5 (2,22,28). Interestingly, an examination of HN5-water and HN5-Gln distances reveals a single site of water near HN5 that is stabilized by local ordering of Thr101 (5F, G). In contrast, the lack of an H-bond in Thr101Ile allows release of water from HN5 to a position near O4 and Gln204. In this manner, the HN5, Gln204, Thr101 locus is implicated in both local structural ordering (near position 101) conformational changes to the C-terminus (Gln204) and activation of HN5 deprotonation through water recruitment to the flavin active site and stabilization adjacent to HN5.

6.4 Effect of T101I mutation in-vivo

Kinetic and computational studies report a role of Thr101 in recruiting solvent to the active site, which may in turn alter LOV allostery. To examine if Thr101 is important for LOV signaling in T. reesei, we constructed a QM6aΔenv1 strain complemented with the mutated Thr101Ile allele. The dramatic decrease of photocycle lifetime in ENV1 compared to VVD should alter gene expression of target genes due to stabilized ENV1 in the night. Previously, ENV1 was shown to influence a considerable number of metabolic genes and shows a severe growth phenotype on diverse carbon sources (29). Hence, we were interested whether the strain bearing the mutated allele (Thr101Ile) would show an altered reaction to changing light conditions or light-dark cycles. Therefore, we tested growth during 2-hour cycles of low light (200 lux) – high light (5000 lux) with a ramping period of 2 hours, similar 3 hour cycles without ramping, 12 hour cycles of darkness – light, constant low light and constant high light conditions. As nutritional conditions
we chose malt extract medium or minimal medium with carboxymethylcellulose (CMC), glucose or glycerol as carbon source.

Predicting effects of the Thr101Ile variant necessitates considering both the effect of kinetics (long lifetime) and the effect of solvent induced protein dynamics (allostery). Any effects of the long lifetime should be confined to changes in the light-dark cycle, particularly light intensity. We would expect the Thr101Ile to stabilize the light state and lead to higher light-state populations at low light intensity. Consistent with light-dependent activity of ENV1, we found that a strain lacking env1 shows differences in hyphal extension rates compared to the wild-type under the different light conditions, especially upon growth on CMC and glycerol. Surprisingly, despite this indication that ENV1 might influence tolerance of different light intensities and reaction to changes in light intensities, the T101I mutant behaved like wild type under all conditions. Thus any biological effects of T101I are likely light-independent and confined to effects on protein allostery.

6.5 T101 is involved in tolerance of osmotic stress in T. reesei

As previous analyses indicated an involvement of ENV1 in reaction to stress conditions, we tested the relevance of Thr101Ile in this process. Based on the role of Thr101 in regulating solvent access and dictating solvent dependent conformational changes, we predicted that the Thr101Ile variant may affect osmotic stress responses. Given the close proximity of Thr101 to Cys96, which has been implicated in oxidative stress (21), we cannot rule out effects on general stress responses. In examining the effect of Thr101Ile on oxidative stress, we found no significant differences in hyphal extension in darkness, low light (1500 lux) or high light (8000 lux) in the strains bearing the Thr101Ile allele under oxidative stress upon growth on malt extract medium or minimal medium with either glucose or glycerol as carbon source.
In contrast, examination of osmotic stress under analogous conditions revealed a direct role of Thr101 in osmotic stress responses. Specifically, we observed an increased hyphal extension compared to the wild-type strain in high light upon growth on carboxymethylcellulose (Figure 6) with all three conditions inducing osmotic stress (1 M NaCl, 1 M sorbitol or 1 M KCl). No significant changes were observed in low light or darkness on any carbon source. These results are consistent with MD simulations and kinetic studies indicating that the Thr101Ile variant disrupts solvent recruitment to the active site. Under osmotic stress conditions solvent occupancy would be diminished in Thr101, leading to alteration of the Gln204 conformation. The occlusion of solvent by Thr101Ile, would mimic the decreased solvent occupancy predisposing Gln204 to an “osmotic-stress-like” conformation and a constitutive amplified response. These should be exacerbated under high-light conditions. Indeed, we see hyperactivity in response to osmotic stress in these variants that are specific to high-light intensities. Thus, we conclude that Thr101Ile is involved in regulation of tolerance to osmotic stress in high light in a carbon source dependent manner. These results demonstrate that Thr101 acts in a signaling region to alter integration of environmental stress.
Figure 48: Hyphal extension under different condition: A) hyphal extension of Δenv1 compared with wild type on different carbon sources. Strains were grown on plates containing malt extract agar (MEX, 3%, w/v), Mandels Andreotti minimal medium with carboxymethylcellulose (MA CMC, 1%, w/v), glucose (MA GLU, 1%, w/v), or glycerol (MA GLY, 1%, w/v) as carbon source. Light conditions were 2-h cycles of low light (200 lux) to high light (5,000 lux) with a ramping period of 2 h (2sRAMP), similarly 3-h cycles without ramping (3hRAMP), 12 h cycles of darkness to light with 3 h ramping from dark to light (12DLsR), constant low light (LL20), and constant high light (LL100). At least three biological replicates were considered. B–D, hyphal extension under different conditions of osmotic stress relative to wild type. The strains (wild-type QM6a, Δenv1, and the mutant bearing the T101I allele QET101I) were grown under high light conditions (8,000 lux) in Mandels Andreotti minimal medium with carboxymethylcellulose (1% w/v) as a carbon source and NaCl (B), sorbitol (C), or KCl (D) (1 M each) to impose osmotic stress. Hyphal extension rate was monitored over 14 days. At least three biological replicates were considered, and three different transformants of QET101I were used. The statistical significance of growth patterns is shown by asterisks for mutant strains. *, $p \leq 0.05$; **, $p \leq 0.005$; ***, $p \leq 0.0005$. Error bars represent S.D.
6.6 Materials and methods

6.6.1 Cloning and protein purification

ENV1 constructs were designed based on sequence homology to *Neurospora crassa* VVD. Initial constructs focused on full-length (1-207) and an N-terminally truncated construct homologous to VVD-36, ENV-64 (65-207). All constructs were cloned from cDNA obtained from *T. reesei* grown on cellulose using standard PCR approaches. Envoy constructs were cloned into pGST vector using NcoI and XhoI cut sites and were verified by DNA sequencing (Genewiz). Point mutation T101I was introduced into ENV1 (1-207) using the Quick-change protocol (Stratagene). All mutants were also verified by DNA sequencing (Genewiz).

All constructs were expressed in *E. coli* JM109 cells. Cells were grown at 37° C until reaching an optical density at 600 nm of 0.6. The temperature was then decreased to 18° C for 40 min to assure thermal equilibrium. At 18° C 0.3 mM isopropylthiogalactoside (IPTG) (obtained from RPI) was added to initiate protein expression. After 22 hrs the bacterial pellets were harvested and stored in 100 mM NaCl, 50 mM Hepes (pH 8), and 10% glycerol. All constructs were purified with glutathione affinity resin (Qiagen). After binding, the columns were treated with 2 mg of TEV protease per milliliter of resin at 22° C for 2 hrs to cut the GST tag from the proteins. Cleaved proteins were eluted in buffer containing 100 mM NaCl, 50 mM Hepes (pH 8), 10% glycerol. An additional round of Ni-NTA chromatography was conducted to remove His6-TEV prior to final purification with a Superdex S200 size exclusion column equilibrated with 100 mM NaCl, 50 mM Hepes (pH 8), and 10% glycerol.

6.6.2 Spectroscopy and kinetics

UV−visible absorbance spectroscopy for all constructs and mutants of ENV were conducted on an Agilent 8453 spectrophotometer. Photophysical properties of all constructs were
verified to show LOV type chemistry exhibiting a single characteristic peak around 380-400 nm for C4a adduct and two broad absorption peaks for the ground state. The ground state spectra consisted of a peak at 450 nm and two vibrational bands at 425 nm and 475 nm. Light-Dark recovery rates were studied with imidazole to expedite the conversion rates. Imidazole acts as a base to catalyze adduct scission by abstraction of N5. Stock solutions containing 1.0 M imidazole, 100 mM NaCl, 50 mM Hepes (pH 8.0), and 10% glycerol were prepared. Light-Dark recovery rates were measured at various concentrations of imidazole ranging from 0-500 mM imidazole. Each data set at a particular concentration of imidazole was repeated in triplicate. All samples were exposed to a broad-spectrum white floodlight source (150 W), while being incubated on ice to populate the light state. For Eyring analysis light-dark recovery rates were studied at temperatures between 288 and 306 K. Kinetics of thermal and base catalyzed reversion were obtained from the absorbance at 450 and 478 nm as a function of time. All values for 450 and 478 nm were corrected for deviations in the baseline by subtracting the absorbance at 600 nm. Full spectra were collected at varying times, such that a minimum of 10 data points per half-life was obtained. Data were fit using mono- and biexponential equations as required to extract kinetic parameters. All time constants were reported as \( 1/k_{\text{adduct scission}} \) that are averaged between the values obtained at 450 and 478 nm. We note that biexponential kinetics were only observed in data sets where reduction competed with adduct formation. Experimental methods for dealing with the biexponential nature of those data sets are discussed in the main text.

6.6.3 Molecular dynamics simulations

Molecular simulations were run on the University of California, Santa Cruz, computing cluster, using the charmm27 force field (30) as implemented in the GROMACS 4.6.5 program suite (31,32,33,34,35), with the addition of flavin-cysteinyl adduct parameters (27). The starting
coordinates were taken from the crystal structure of ENV-64 wild type, and a mutation at Thr101 generated using Chimera (36). For examining solvent dynamics, ten starting positions within the first and second coordination spheres about the flavin were randomly selected as shown in Figure S1.

Starting configurations with one water molecule inside the flavin pocket were manually constructed. An equilibration procedure of energy minimization, then 100 pico second (ps) simulation in the fixed N, V and T (number; N, volume; V and temperature; T) ensemble with the protein atoms positions restrained, then 100 ps with fixed N, P and T (pressure; P) and protein position restraints, followed by 750 ps of NPT ensemble with no position restraints was used, the last step was a 1 ns production run. 100 individual production runs were stitched together using the gromacs utility trjcat (34). Using the GROMACS radial distribution function tool (34) a plot of the density of water molecules around flavin N5 was created. From the radial distribution function, \( g(r) \), we calculated the energy, \( w(r) \), to pull a water from infinite distance to a given \( r \) from N5, using the equation \(-RT \ln \, g(r) = w(r)\) (37). The difference between the lowest and highest part of these curves were taken as the energy barrier to bound water exchange with the bulk solvent. The correlation between water-HN5 distance and HN5-Gln204 distance was examined by measuring the two distances in all frames of the trajectory, using the gromacs utility g_dist, and a simple python (ww.python.org) script to extract the distances and couple the distances from the same frame together. These bivariate data sets were then binned into hexagonal bins and heatmaps constructed, with number of frames having values within the bin defining the color of said bin, using the R (R Core team, 2015) package hexbin (24). Using a python (ww.python.org) script, frames from the trajectories were binned into one of 2 bins according to water-HN5 distance, distances of less than 360 picometers were considered to be hydrogen bonds. All frames in the bin
were then stitched together into a new trajectory, then the gromacs utility g.chi was called, and the order parameter for residue 101 was extracted.

6.6.4 Strains

The parental strain used for the osmotic stress assay was *T. reesei* QM6a (ATCC13631). The mutant strain QM6aΔenv1, lacking the whole open reading frame of ENV1 (38) was complemented with the mutated allele.

6.6.5 Stress response

Analysis of response to oxidative stress was performed as described previously (21). The osmotic strain assay for the parental and mutant strains was performed on plates with malt extract medium (3 % w/v) or with Mandels-Andreotti minimal medium with carboxymethylcellulose sodium salt (Sigma, St. Louis, MO, USA) (1 % w/v) or glucose (Roth, Karlsruhe, Germany) (1 % w/v) as a carbon source supplemented either with NaCl (Roth, Karlsruhe, Germany) (1M), sorbitol (Sigma, St. Louis, MO, USA) (1M) or KCl (Sigma, St. Louis, MO, USA) (1M) to impose osmotic stress. Strains were grown in constant darkness, constant low light (1500lux; Philips Master TL5 HO) and in constant high light (8000 lux) at 28 °C for 5 days. As this prescreening indicated a relevance of T101 for tolerance of osmotic stress in high light, we repeated the experiment for this condition with 14 days of incubation. The control strains used for all combinations of the assay were QM6a and QM6aΔenv1, and 3 biological replicates were used in every experiment. Three different recombinant strains harboring the T101I mutations were considered in each experiment and used at least in duplicates. Statistical analysis of hyphal extension after 14 days was performed using PSPP version 0.9.0 (oneway ANOVA, post hoc tests Tukey, Bonferroni and Scheffe).
6.6.6 Construction of ENV1-T101I complementation strains in T. reesei

The complementation vector pENV1T101IamdS was constructed by yeast recombination essentially as described previously (39) using the primers Env1T101I3F (5’ CTCCCTCATTTGTGCAGATTC 3’), Env1T101I3R (5’ GATGACATTACACTCAAGACGTGATGTTCAACCCGAGTCC 3’), Env1T101I5R (5’ CTGATATCGCACAAGATGAG 3’), and Env1T101I5F (5’ GTAACGCCAGGTTTTCAGTAAGAAGACTGCTTTATTATTGC 3’). The primers Env1T101I5F and Env1T101I3R contain overlapping sequences (for combination with the amdS marker cassette (40). The mutation was confirmed by sequencing the vector, and the amplified cassette was used for protoplast transformation of QM6aΔenv1 (41).
REFERENCES


Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig,
M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J.,
Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B.,


for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory
Comput* **4**, 435-447


34. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith,


Hypocrea jecorina (Anamorph Trichoderma reesei), modulates cellulase gene transcription


A.1 Peroxide production in LOV proteins

All LOV proteins bind to a flavin cofactor. Previous studies have shown that photoreduction of flavoproteins following exposure to blue or UV light results in production of peroxide. In order to investigate if LOV proteins produce peroxide upon photo-reduction we used Amplex red assay kit. The amplex red assay kit consists of Amplex red reagent (10-acetyl-3,7-dihydroxyphenoxazine) to detect hydrogen peroxide. In presence of peroxidase enzyme, Amplex red reagent reacts with hydrogen peroxide in 1:1 ratio to form a red fluorescent product called resorufin. Since, resorufin has a high extinction coefficient of $58,000 \pm 5,000 \text{ cm}^{-1}\text{M}^{-1}$, the assay can be performed fluorometrically or spectrophotometrically. The excitation and emission maxima of resorufin is approximately 571 nm and 585 nm, respectively. This assay is very sensitive and can detect as little as 10 pm of H$_2$O$_2$ in a 100 µL volume.

We performed this assay spectrophotometrically in dark as the amplex red reagent is sensitive to light in a total volume of 100 µL. The red fluorescent product formed due to presence of peroxide can be detected with spectrophotometry at an emission spectral peak at 572 nm. We took 50 µl of concentrated protein in its dark state and incubated with 50 µl of Amplex reagent and another 50 µl of protein was taken in its light state (light state was obtained by shining a blue laser for a minute) and incubated with 50 µl of Amplex reagent for thirty minutes. To check if peroxide is produced by protein and not by buffer in which it is contained we took 50 µl of protein buffer and irradiated with blue laser for a minute and 50 µl of Amplex reagent was added and incubated for thirty minutes. A negative control was included which contained 50 µl of reaction
buffer and 50 µl of amplex reagent solution was added to it and incubated for thirty minutes. A positive control was also included which contained 50µl of reaction buffer with 5 µM peroxide and 50 µl of amplex reagent was added and incubated for thirty minutes. After thirty-minute incubation of samples in dark spectra was measured to detect peroxide production. This assay was performed for ZTL and FKF1 28-174. Both of these LOV proteins showed peroxide production in its light state and not the dark state. The protein buffer did not show peroxide production confirming peroxide production from flavin containing LOV proteins.

To quantify the amount of peroxide produced by the protein various standard solutions containing peroxide from 5 µM to 12.5 µM were prepared. These solutions were incubated for thirty minutes with amplex reagent and then spectra was taken. A graph was made by plotting concentration of peroxide in sample versus the absorbance at 572 nm. A linear graph was obtained using the line equation y=mx+c peroxide produced in the protein sample was quantified. The peroxide concentration of FKF1 in light state was found to be 4.82 µM and in ZTL it was 11 µM.
Figure 49: Peroxide production in LOV proteins: A) Absorbance spectra of various standards samples containing peroxide at a concentration range from 0-12.5 µM. B) Linear plot of concentration of peroxide in the standard sample v/s Absorbance at 572nm to estimate the quantity of peroxide produced in the protein samples. C) Light state spectra (red) of FKF1 has a peak at 572 nm confirming peroxide production whereas dark state (black) has a small peak comparable to the sample containing the Hepes buffer (blue) and a negative control sample (pink) has no peak. D) Light state spectra of ZTL (red) has a peak at 572 nm corresponding to peroxide production the dark state (black) has small peak comparable to the Hepes buffer (pink) and a negative control sample (blue) has no peak.
A.2 Photochemical Characterization of ENV and VVD mutants

Earlier in Chapter 5 we discussed about ENV and VVD mutants that affect signaling behavior. We also studied their photo-cycle kinetics to investigate if these mutations affect the lifetime of adduct state. All LOV proteins as discussed before form a cysteine-flavin adduct upon light activation. The lifetime of the adduct state depends upon various factors such as nature of residues surrounding the active site and solvent accessibility. Since these mutations were made in vicinity of the flavin active site we wanted to investigate if the mutations affected the lifetime of adduct state. Photocycle kinetic studies of mutant VVD A72S exhibited a time constant for adduct decay to be ~5 hrs which is same as the wildtype VVD protein. Therefore, mutation of Ala72 to a serine does not affect photocycle kinetics but blocks light activated dimer formation. Similarly, we investigated the photocycle of mutant ENV S99A. The time constant for adduct decay of the mutant was ~50 mins which twice as much longer than the wildtype ~26mins. Therefore, mutation from a polar serine residue to a non-polar alanine residue affected the photo-cycle. Similarly, the double mutant S99A:W191R has longer adduct state lifetime of ~ 1.2 hrs which is 2.7 times longer than the wildtype ~26 mins.
Figure 50: Photocycle kinetics of ENV and VVD mutants: A) Light-dark state recovery rates show first-order kinetics. The absorbance trace of ENV 1-FL S99A at 450nm (black) and 478nm (red). B) Dark state recovery of ENV 1-FL S99A W191R at 450nm (black) and 478nm (red). C) Dark state recovery of VVD A7S at 450nm (black) at 478nm (red). D) Rate constants and recovery times of ENV and VVD mutants. The VVD A72S has similar recovery rates compared to VVD WT. The recovery rates of ENV 1-FL S99A and ENV 1-FL S99A W191R are slower compared to ENV 1-FL.