University of California
Santa Cruz

Examination of the hydrophobicity of the LOV (Light-Oxygen-Voltage) protein active site, using computational chemistry and Raman spectroscopy

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

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Rafael Ibrahim Silverman y de la Vega

December 2017

The dissertation of Rafael I. Silverman y de la Vega
is approved:

_____
Professor David Kliger, Chair

_____
Professor Carrie Partch

_____
Professor Roberto Bogomolni

_____
Tyrus Miller
Vice Provost and Dean of Graduate studies

Table of contents

List of figures

Abstract

Examination of the hydrophobicity of the LOV (Light-Oxygen-Voltage) protein active site, using computational chemistry and Raman spectroscopy

by

Rafael Ibrahim Silverman y de la Vega

LOV (Light-Oxygen-Voltage) domains are common blue light sensing proteins, found in plants, fungi and both kingdoms of prokarya. LOV proteins are also commonly used in engineered proteins, usually to impart blue light sensitivity. Despite their ubiquity and utility, the mechanism by which they turn light into a biochemical signal is contested. The initial photophysics, and the allosteric protein effects are well documented; but the key chemical step in between these two processes is unknown. In this step, a triplet state on the flavin chromophore creates a chemical bond to a nearby cysteine. This thesis describes the use of computational chemistry and Raman spectroscopy to explore the role of water in the LOV photocycle. Both molecular dynamics, and quantum chemistry was applied to this effort. GROMACS was the MD engine used, and the CHARMM27 force field was employed. The quantum chemistry program Orca was chosen for the density functional theory described in this document. Both of these computational methods show that water is stable in the LOV active site. Raman vibrational spectroscopy of free cysteine as a model compound show that the catalytic Cys inside the LOV active site need not be protonated for LOV photochemistry to occur.

Acknowledgements

This dissertation contains an adaptation of "A Native Threonine Coordinates Ordered Water to Tune Light-Oxygen-Voltage (LOV) Domain Photocycle Kinetics and Osmotic Stress Signaling in Trichoderma reesei ENVOY."[43,] and is reproduced with the permission of the other authors. An email from the corresponding author, Brian Zoltowski, of the paper above describing the permission to adapt this work has been submitted as a supplemental file along with this document. The author of this dissertation contributed all of the molecular dynamics simulations, and wrote the corresponding sections of the paper listed above. The author of this dissertation did not do any of the biochemistry or spectroscopy described in the paper above. The text of the journal article above was altered significantly when it was incorporated into this dissertation.

Section 1.1

Introduction to light sensing, and light sensing proteins

The great nuclear reaction in the sky produces large amounts of electromagnetic radiation, which hits the earth in cycles. This light is a vital biological cue, as changes in light often signal a need to shift gears. For some organisms more light means danger is near; for other creatures light indicates a time to wake and begin feeding. Light itself can pose a threat, in the case of UV light. Light is even the food source for the photosynthetic, which in turn provide food for nearly every other living thing on earth. Light sensing is thus an important function. But, how does an organism turn photons into a biological signal? Light is ephemeral, and capturing photons requires a much more specialized apparatus relative to the absorption of nutrients by an organism utilizing a generic set of proteins.

Francis Crick's tongue in cheek "Central Dogma of Biology" states that DNA codes for RNA, which codes for proteins, and proteins are the molecules that do everything else. So, according to this theory, a protein will sense light, and generate a signal that turns on a response, either by changing the function of another protein, or by changing which DNA regions are transcribed to RNA, thus changing the protein makeup of the system. In fact, proteins are the biological molecules that sense light.

There are a large variety of protein photoreceptors, and they are found throughout the tree of life. Highly differentiated Eukaryotes, such as vertebrates, have specialized organs that house photoreceptors. Other, less differentiated creatures such

as filamentous fungi, have the photoreceptors spread out in most or all of their cells. Photoreceptors may be floating in the cytoplasm, anchored to a membrane, or membrane integral. A wide variety of functionality exists across the diversity of photoreceptors.

Rhodopsins are a very common group of photoreceptors, which pump ions across a membrane, creating electric and chemical potential gradients. The most common photoreceptor group consists of the photosystems used by photosynthetic organisms. This type of protein complex captures visible light and harnesses this energy to generate proton gradients, which drive carbon fixation. Another method employed by photoreceptors acts by allosteric inhibition of their downstream partner. A more complex manner in which photoreceptors affect downstream signaling partners is by altering their chemistry: light activation causes the photoreceptor to add or remove prosthetic groups. The change in prosthetic group in turn changes the behavior of the downstream partner.

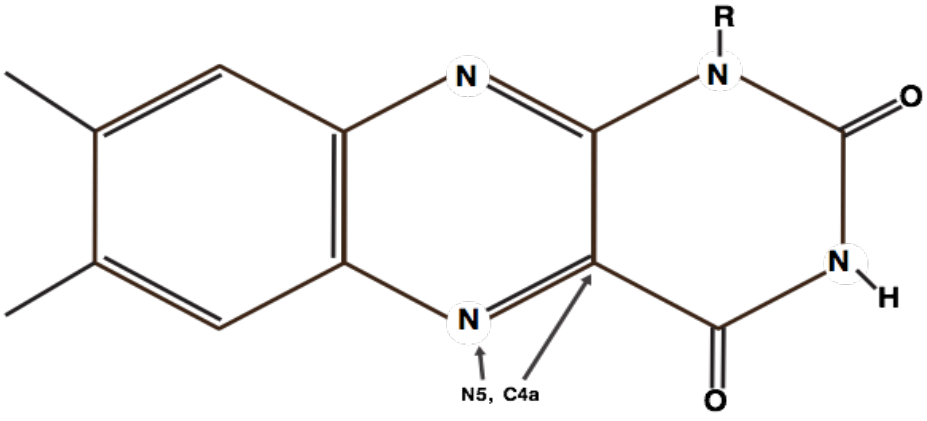


Figure 1.1: Flavin, with N5 and C4a labeled

But, light sensing depends on absorption of light, yet proteins are composed of amino acids which cannot absorb visible light. So, in order to accomplish light sensing, proteins must utilize cofactors. One common light absorbing cofactor is flavin (figure 1.1). Flavins have a triple fused isoalloxazine ring, which is aromatic and also N heterocyclic. There are two carbonyl groups on one of the fused rings, and two methyl groups on the opposite ring. Biological flavins have additional moieties hanging off of N1. In the case of the archaically named Flavin MonoNucleotide (FMN), N1 is bonded to a phospho-ribityl (a phosphorylated derivative of ribose) moiety. Flavin Adenine Dinucleotide (FAD) has, additionally, adenine bonded via a diphosphate linkage to the ribityl group. The list of the different flavins goes on, but further enumeration of them is superfluous.

Flavins readily undergo redox chemistry; their large, heterocyclic and aromatic ring system easily stabilizes extra electrons and protons (figure 1.2)[29]. Their ability to undergo one and two electron reduction enables flavins to function as biological electron carriers, and they are often used to reduce or oxidize other biomolecules[29]. The different oxidation states of flavin are termed quinone, semiquinone and hydroquinone, in order from fully oxidized to fully reduced. Flavin's conjugated ring system means some of their electronic energy states have relatively low energy, allowing them to absorb visible light. Promotion to the first singlet excited state requires a blue photon. This absorption enables flavin's key biological function as it relates to this thesis: light sensing.

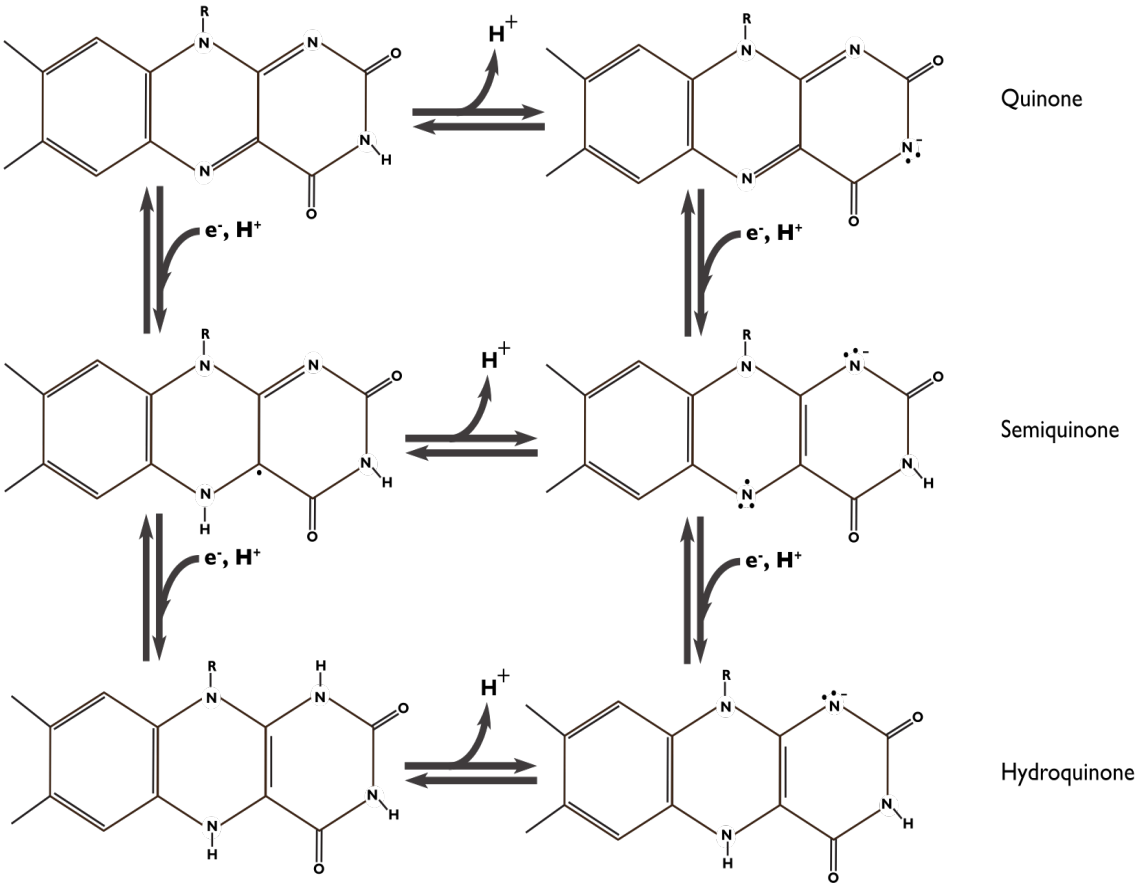


Figure 1.2, Oxidation states of flavin. Quinone is the most oxidized, addition of an electron puts flavin into the semi-quinone state, and addition of a second electron puts flavin into the hydroquinone state.

There are 3 major classes of light sensing flavoproteins: Cryptochrome/photolyase proteins; BLUF proteins (Blue-Light-Sensing-Flavoproteins)(figure 1.3); and LOV proteins (Figure 1.4). Cryptochromes have 2 chromophores, either a folate and a FAD or two flavins. The mechanism of light sensing by cryptochromes involves the stabilization of a radical electron on the flavin. Both BLUF and LOV domains contain the same type of chromophore, flavin, but

Figure 1.3, a BLUF protein, with FMN (magenta) bound between β sheet and αhelices, this shows a typical PAS α/βfold.

they differ in their mechanism of activation. BLUF proteins sense blue light by forming a neutral semiquinone radical, probably by reduction by a nearby Tyr residue. This rearranges the protein hydrogen bond network[49]. BLUF and LOV domains have a general structure that is very similar, and in fact they are both grouped into the PAS (Per-Arnt-Sim[Period-Aryl hydrocarbon Receptor Nuclear Translocator-SIngle Minded]) protein superfamily. [51, 76,41,47]

PAS domains are an extremely large group, with diverse functions and mechanisms[51]. They often act as dimerization interfaces and sensors[51]. In fact, PAS sensing often involves dimerization as the output from the sensing input. All PAS

domains have a conserved α/β fold (figures 1.3 and 1.4) that often functions as the binding site for a prosthetic group. In the case of LOV domains[47], a flavin, usually FMN, is held here (figures 1.4 and 1.9) As mentioned earlier, flavins are capable of absorbing blue light and undergoing redox chemistry. What transpires in the LOV active site after blue light illumination is the focus of this document.

Section 1.2

LOV domains and their photocycle

LOV domains are the primary photorecepters in a variety of blue light sensing pathways[4]. The kingdoms Eubacteria, Archaebacteria, Plantae, and Fungi all have LOV proteins, but for some unknown reason, Animalia lacks them. Eukaryotic LOV domains fall into 4 evolutionarily distinct clades [41]: 1) the plant phototropins, which regulate the various plant responses to blue light, such as phototropism and stomatal opening, this category contains LOV2; 2) the zeitlupe and flavin-binding Kelch repeat F-box protein group, which are involved in plant circadian rhythm; 3) aureochromes, which regulate photomorphogenesis in the heterokont (formerly called stramenopile) algae; 4) the white collar group of fungal circadian rhythm proteins, which contains ENVOY and VVD. Eukaryotes likely acquired LOV domains early in evolution. Most likely they were a product of horizontal gene transfer from prokaryotes, perhaps when endosymbiosis occurred with plastids or mitochondria.

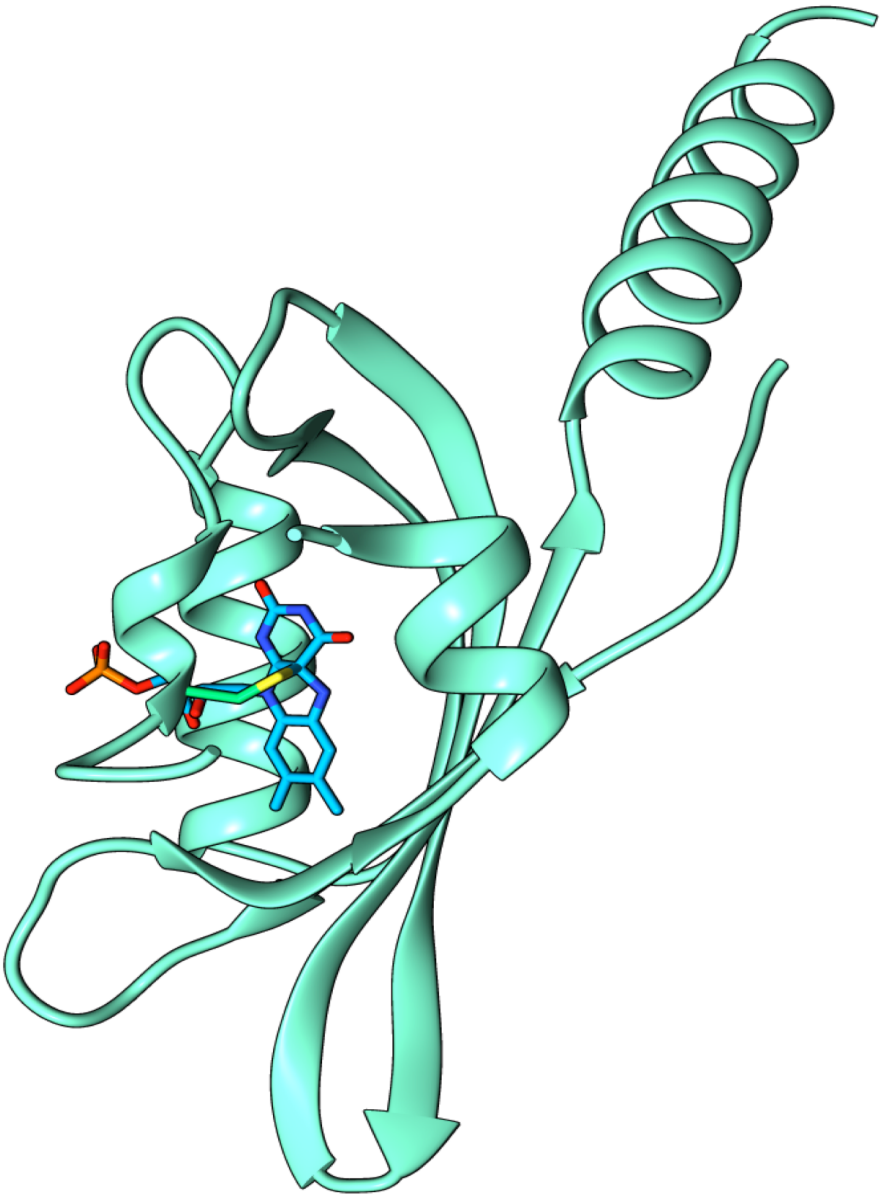


Figure 1.4, A LOV protein, YTVA from *Bacillus subtilus*. The signal transducing j-helix is protruding to the left.

LOV proteins are have been a popular target for bioengineers. iLOV[16] is an engineered mutant used as a fluorescent tag, which has superior characteristics compared to green fluorescent proteins; it is smaller, less toxic, and doesn't require maturation time. A LOV protein, miniSOG was derived from phototropin and used as a singlet oxygen producer[71]. MiniSOG's original purpose was to activate cell fixing chemicals, but miniSOG singlet oxygen production has been used for far more[67, 62, 22, 59]. Perhaps the most interesting use of miniSOG was to selectively kill cells in *C. elegans*[84], using a laser.

By far the most common engineered use of LOV domains is in optogenetics. At the time of the writing of this thesis (summer 2017) there were about 2,940 results from a google scholar search for "LOV optogenetics". Optogenetic tools are used to bring genes under the light activation. Controlling gene activation with light is useful because it allows precise temporal and/or spatial control over gene expression. There are native LOV proteins that already have DNA binding activity, such as El 222[53], or anti-anti-sigma factor effector domains[21,13], so it is a natural progression to engineer LOV proteins to interact with gene targets.

So how do LOV domains turn a photon into a biological signal? The absorption of a blue photon by the flavin of a LOV domain imparts enough energy to the flavin to form a metastable bond between the flavin and a nearby cysteine[77],[2] The adduct is N5 protonated on the isoalloxazine, unlike the dark state. This change in hydrogen bond availability on the flavin creates a chain reaction of rotamer shifts[27] to

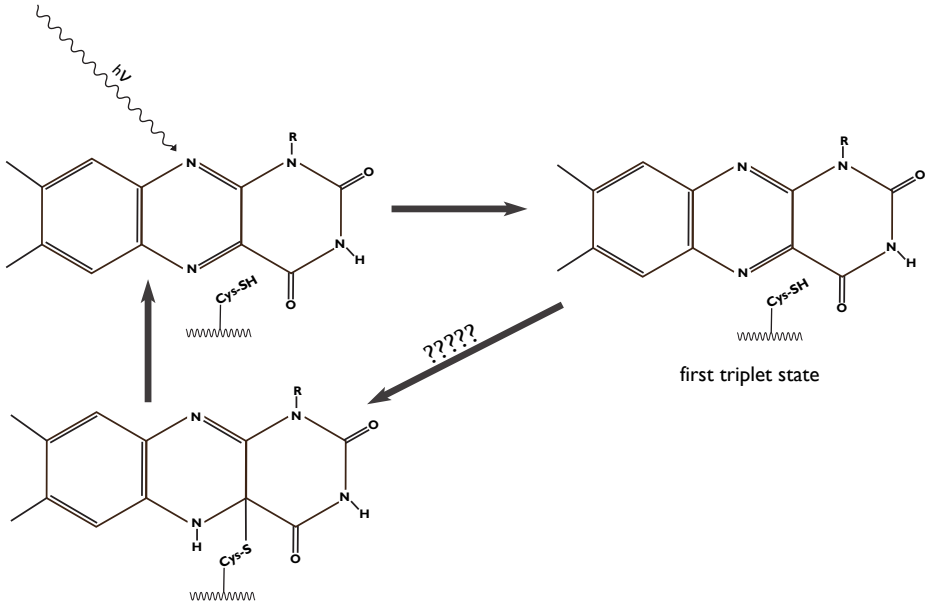


Figure 1.5: Highly simplified LOV photocycle. Blue light absorption followed by intersystem crossing puts flavin into the first triplet excited state. The triplet decays into the adduct by an unknown mechanism, and the adduct decomposes thermally to the ground (dark) state.

unlock an α helix (figures 1.4) that docks on the opposite side of the β-sheet. This helix can be N or C terminal to the main LOV PAS fold, but all LOV proteins have one[28].[21] The allosteric release of this helix allows downstream partners to receive the signal. Eventually, after a half life ranging from seconds to hours[28],[21], the photo-adduct degrades, and the protein returns to the dark state.

The LOV photocycle (Figure 1.5) is not completely understood at this date. Although the initial photophysics[72,32] and the downstream signal transduction are relatively well-characterized[41,30,28], the key catalytic step is a point of contention between the Bogomolni research group and several other research groups. When

activating light is absorbed by a LOV domain's flavin, a triplet state is formed. This triplet then decays into the adduct state by an unknown mechanism. The debate on this unknown mechanism hinges on the behavior of perhaps 5 atoms: the sulfur of a conserved cysteine residue, the proton on that Cys, N5 and C4a of flavin, and a proton that is on N5 in the adduct, but absent in the dark state[1,38,28].

The ground state of LOV proteins exists as a singlet. The photocycle (Fig 1.5) of LOV proteins begins with the absorption of a photon in the blue or UV wavelength. This puts the flavin into either the first or second singlet excited state. The second singlet rapidly decays into the first singlet excited state. The first singlet either decays to the ground state, or undergoes intersystem crossing to form the first triplet state. The triplet state decays into the photoadduct: a weak bonding state between the carbon at 4a position of flavin, and the sulfur of the conserved cysteine. The photoadduct S-C bond is metastable, and decays back into the ground state. The lifetime of this adduct varies considerably among different LOV domains. The general range is from seconds to hours, depending on the protein and experimental conditions.

Section 1.3

LOV photo-adduct formation and breakage hypotheses

Although LOV proteins were first described in the literature nearly twenty years ago, hypotheses of both the formation and the scission mechanisms of the S-C4a

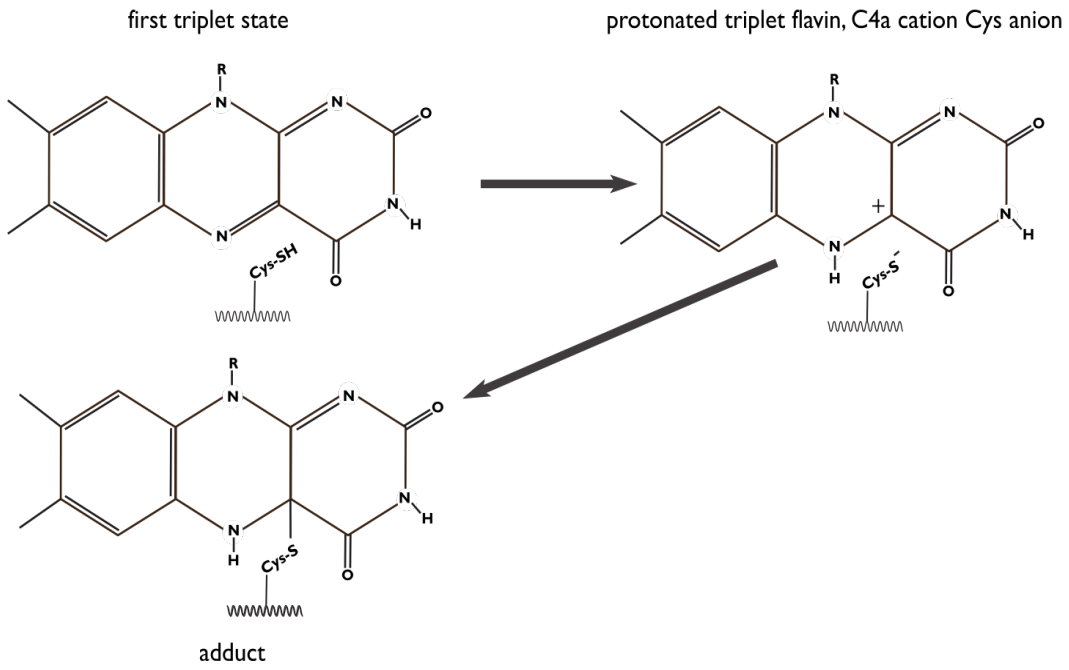


Figure 1.6: The ionic mechanism of LOV photo-adduct formation. The triplet state becomes N5 protonated, then a nucleophilic Cys thiolate anion attacks C4a.

adduct remain contested. There are two popular formation mechanism hypotheses[38]. One proceeds through a ionic step, and is called the "ionic mechanism" (figure 1.6); the other proceeds through a radical step, and is called the "radical mechanism." (figure 1.7)

Maybe because of the lack of competing hypotheses regarding its mechanism, the reverse reaction mechanism is far less debated than the forward. In adduct scission it has been hypothesized that a base deprotonates N5. The identity of this base could be any number of nearby amino acids, or a water [88]. Although there isn't much debate over the hypothesis that HN5 deprotonation triggers adduct breakage, the base that

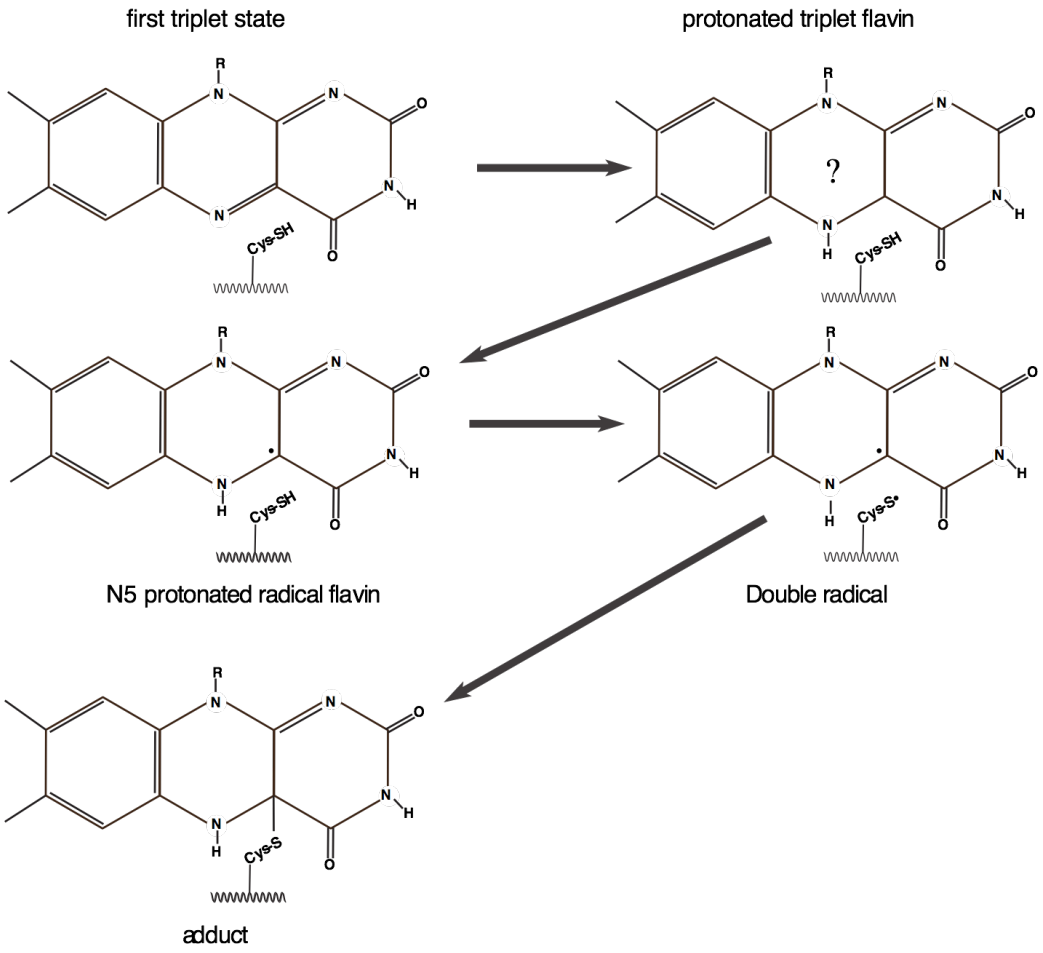


Figure 1.7: The radical mechanism of LOV photo-adduct formation. Triplet flavin (probably) becomes N5 protonated, then abstracts an electron or hydrogen atom from Cys, and somehow maintains unpaired electron spin, the radical pair undergoes recombination forming the adduct.

removes this proton is unknown. Once N5 is deprotonated, the electrons freed then reform a pi bond between N5 and C4a. The two formerly bonded electrons between C4a and Cys localize on this cysteine's sulfur, leaving no net bond between the Cys and flavin. Finally, the Cys thiolate regains a proton, returning to thiol state… or it doesn't. Data from the Bogomolni lab shows that the Cys doesn't need to exist as thiol in order to cycle, and the protonation state of the Cys thiol is explored in this

thesis. Water is needed for the reverse reaction to proceed [12]. The location of this water is addressed in chapter two of this thesis [43], and we propose that water is indeed the catalytic base in LOV deactivation.

Adduct formation has been investigated by many labs, and a variety of techniques have been utilized in examinations of the photocycle. There have been three techniques that have provided the most useful information: UV-visible absorption, infrared, and quantum chemistry. UV-vis absorption has been one of the most commonly used techniques[28], and the identity of most of the intermediates in the photocycle were determined with it. Optical absorption has the advantage of being fast and easy to implement, and produces fairly simple spectra. However, it is limited by the simplicity of the spectra as it's a low information density technique. Basically, determination of the energy levels of a few loosely held valence electrons of the molecule does not provide much detail. In addition to normal limitations of UV-vis absorption, the absorption bands of the various hypothesized flavin states are often quite similar to each other. For example, the radical and the triplet flavin both absorb at nearly the same region in the red. This hinders attempts to positively identify absorption bands.

The high information density method most commonly used to study LOV proteins is infrared spectroscopy. Both plain FTIR (Fourier Transform InfraRed) and step scan FTIR have been employed successfully in the elucidation of kinetics between different states [85], and have even been used to speculate on the identity of

states[68,33,80, 39]. Unlike optical absorption, an absolute spectrum such as just the dark state is fairly useless because of its complexity. The sheer number of peaks in certain regions often completely obscures the bands of interest. The broad absorption of the amide I and amide II vibrations in particular overlap with flavin bands.

Plain FTIR is hindered by the speed at which the path length in the interferometer is varied. Thus, regular FTIR spectra describe well states that live longer than the time that the motors take to go through an entire cycle. Step scan infrared circumvents this issue by taking snapshots step by step throughout a wavelength range. Since at every time point a spectrum must be taken, every point needs a laser flash. Thus, sample has to be replaced for every single time point. If the system under study takes a long time to return to the dark state large amounts of sample, and clever means to cycle the sample is required. Thus, step scan experiments take much longer than plain FTIR, and often require specialized devices to cycle the sample, especially if it is a dried protein film. Ironically, despite the longer duration of step scan IR experiments, much faster reaction time scales can be observed with this technique. Step scan IR is limited by just a few nanoseconds on the fastest machines. The fast time scale, and stepwise data available from step scan IR allow great insight into reaction mechanisms.

One challenge of spectroscopic data is that the peaks are not always easily identifiable. For example, the flavin radical as well as the flavin triplet state absorb in the same red region[29]. A popular method for identifying potential intermediates is

quantum calculations. Predictably, quantum chemistry has been applied to the LOV photocycle. These calculations are useful because they are definite; there is no questioning whether an intermediate is a radical or ion. However, many things, including poor experiment design, limit them. Additionally, all quantum methods utilize approximations, beginning with the Born-Oppenheimer approximation and likely including many more. On top of the inherent limitations of calculating quantum effects, one can easily create unphysical systems, and have no gauge on how reasonable they are. When energies of different reaction pathways are similar, drawing a conclusion that one is the true path is quite a stretch. Although their accuracy is often questionable, quantum calculations are extremely precise. On their own there is no way to confirm the results of quantum calculations, but when combined with other techniques they become very valuable tools. These calculations can be used to explain the results from other methods, and other techniques can be used to bolster results from quantum calculations. An interesting combination of infrared and quantum chemistry was described by Alexandre et all in 2009[2].

The first spectroscopic studies of LOV photochemistry used time resolved UV-vis absorption, and identified many important features. The adduct absorbs(figure 1.8) at a maxima of 390 nm, and is a singlet state termed S_{390}. The ground state is a singlet, and is dominated by flavin's absorption at 360 and 450 nm, with vibronic (Figure 1.8) structure related to the 450 maxima at 425 and 475 nm [77] and is termed S_{450}. Excitation at either of these two bands (450 or 360) produces the

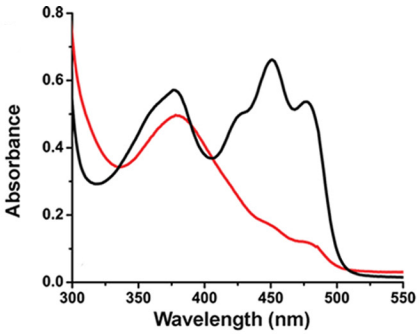


Figure 1.8 LOV UV-visible absorption spectra. Black is the dark State, red is the spectrum of the adduct

excited singlet states, which decompose into the first triplet excited state, which absorbs at 715 nm. T_{715} is immediately followed by S_{390}. A (slightly) later study[38] confirmed the identity of these states. Perhaps tellingly, no intermediates have ever been definitively identified between the T_{715} and S_{390}, despite two decades of attempts.

Infrared studies on the mechanism of adduct formation have focused on the protonation state of flavin immediately preceding S_{390}. A recent study found what appears to be a protonated triplet flavin [80]. The Bogomolni lab has unpublished FTIR data that do not always show the prominent S-H vibrational band at 2660, a band that is always found in the spectra published by other groups[2,5,33,34,Kikuchi 2009},[68,80,85.] This data is discussed in chapter 4 and the conclusion of this document. Although other features are prominent in LOV IR spectra, no other ones are as relevant to determining the reaction mechanism.

Quantum mechanical studies [25, 55] examined the relative energies associated with a radical versus an ionic transition state. They found the radical pathway to have slightly lower energy. However, these studies didn't include a stabilizing group to interact with the bare negative charge of the thiolate anion. Since bare negative charges sometimes produce unphysical simulations[36], these studies are inconclusive.

LOV photochemistry shows an extreme resilience to mutation. As long as there is a cysteine close enough to flavin C4, the adduct forms[87]. It need not be in the canonical position above flavin, nor even on that side of the isoalloxazine ring; the photochemistry proceeds with surprisingly similar kinetics regardless[87]. Even BLUF domains undergo LOV like photoadduct formation when a Cys is introduced near enough to the flavin [76]. No other residue is required, as long as the protein maintains structural integrity, i.e. doesn't flop apart and lose the flavin. But, even the similar residue serine, which has a alcohol group instead of cysteine's thiol, doesn't undergo LOV photochemistry when replacing Cys in the active site[73].

An overview on the techniques used to examine the LOV photocycle would be lacking without a mention of electron spin resonance. Since radicals are states with unpaired electrons, if the exist for long enough they can be detected by a similar method to nuclear magnetic resonance. Bittl and coworkers [70] used electron paramagnetic resonance (EPR) to look for a radical intermediate, and found no evidence of radicals on the timescale available. Although radicals have been photoinduced in mutant LOV domains lacking the reactive Cys needed to form the adduct[73,70] to date, no radical electron has ever been proven to be relevant to the LOV photocycle.

Both the radical and the ionic mechanisms (figures 1.6 and 1.7) for adduct formation are reasonable hypotheses, and neither can be easily dismissed, despite claims to the contrary[46]. There are several data that are usually mentioned as

indicating the invalidity of the ionic mechanism. The most commonly mentioned is the protonation state of N5. The flavin appears to exist with a protonated N5 in the moment immediately preceding adduct formation. The conclusion that this disfavors an ionic intermediate hinges on the assumption that there is no place for a proton to sit in the stages in between Cysteine deprotonation (which is the first step in the ionic mechanism) and adduct formation, except on N5. That is to say, the assumption is that Cys must donate its proton to N5, and not to another nearby proton acceptor such as a polar side chain or a water.

Based on this assumption several groups have tried to identify a radical intermediate. The only evidence that they found that remotely looks like a radical is a very noisy, TROA spectrum[6], and even this could be triplet flavin or noise. Furthermore, the authors of that study have had five years from the release of this study to publish better quality spectra showing this intermediate, but have not. Quantum studies for the most part favor a radical intermediate[2, 15,25,55], But, as mentioned several times earlier in this document, no wet lab technique has ever been able to confirm unpaired electrons[46].

Another interesting thing that can be seen in LOV FTIR is the protonation state of Cysteine. Data from the Bogomolni group, some of which is shown in the conclusion of this text, show deprotonated cysteine in some constructs of the second LOV domain in plant phototropin (LOV2) but not in others. This feature has never been seen in other LOV domains by the Bogomolni group, and has never been

reported by other research groups. If the Cys can exist in an already deprotonated state, yet still undergo the same photochemistry, then a radical mechanism seems less likely. In particular, the often proposed mechanism involving the transfer of a hydrogen atom from Cys seems much less likely, if in fact Cys can be deprotonated before adduct formation.

The ionic mechanism (figure 1.6) is in no way hindered by a deprotonated Cys. The proton formerly on the Cys thiol could sit on a nearby proton sink. And, in fact, strong evidence[80] exists that N5 is protonated immediately before adduct formation. So, where would N5 get a proton? Certainly Cys is a strong candidate. Cys could either donate a proton to triplet flavin directly, or it could donate a proton to a nearby water or amino acid, which then delivers the proton to triplet flavin. A nearby water or water cluster could also provide a proton to the triplet flavin; this hypothesis is addressed in this thesis.

These points about Cys and N5 protonation hinge on the possibility of a nearby site where a proton can be easily accepted by and donated to, whenever needed. This sink would need to be stable for an indefinite time with, or without, said proton. Either water or a nearby amino acid residue could act as this proton sink. It is unlikely to be a position on flavin other than N5, since the odd flavin species produced would probably be detected by FTIR. Proponents of the radical mechanism claim[34, 68, 86,85, 33] that the flavin binding pocket is far too hydrophobic for water to linger. They argue that it is thus also far too hydrophobic for a Cys thiolate to exist.

And, lacking a suitable proton source or sink, a radical is the only way for the adduct to form.

Although the back reaction is not nearly as contentious a topic as adduct formation, the reverse reaction is also not completely understood. Solvent isotope effects on kinetics have implicated deprotonation of N5 as the rate limiting step in adduct scission[38, 19, 77, 90]. Although direct dependence on solvent pH is not seen in LOV proteins, addition of a small base can increase the back reaction rate by several orders of magnitude [1,90].

Water has been implicated in the LOV back reaction[12]. The Bogomolni group using FTIR showed a LOV protein subject to decreasing partial pressure of water underwent correlated slowing of adduct scission. The stepwise decrease in rate indicated the dependence of the back reaction on a water cluster[12]. A frame from an unpublished molecular dynamics simulation of LOV protein with multiple waters within the flavin binding pocket is shown in figure 1.9.

Perhaps the most insightful study was performed by Brian Zoltowski on the fungal LOV protein VVD[90]. Zoltowski used mutagenesis to tune the back reaction kinetics. This identified several key regions where the residue type had large effects on adduct stability. Interestingly, no mutation, other than mutating the active Cys out entirely, has ever entirely stopped the back reaction without also denaturing the protein. One could argue that mutating the Cys out entirely doesn't stop the back reaction, it prevents the adduct from forming, but this is just semantics.

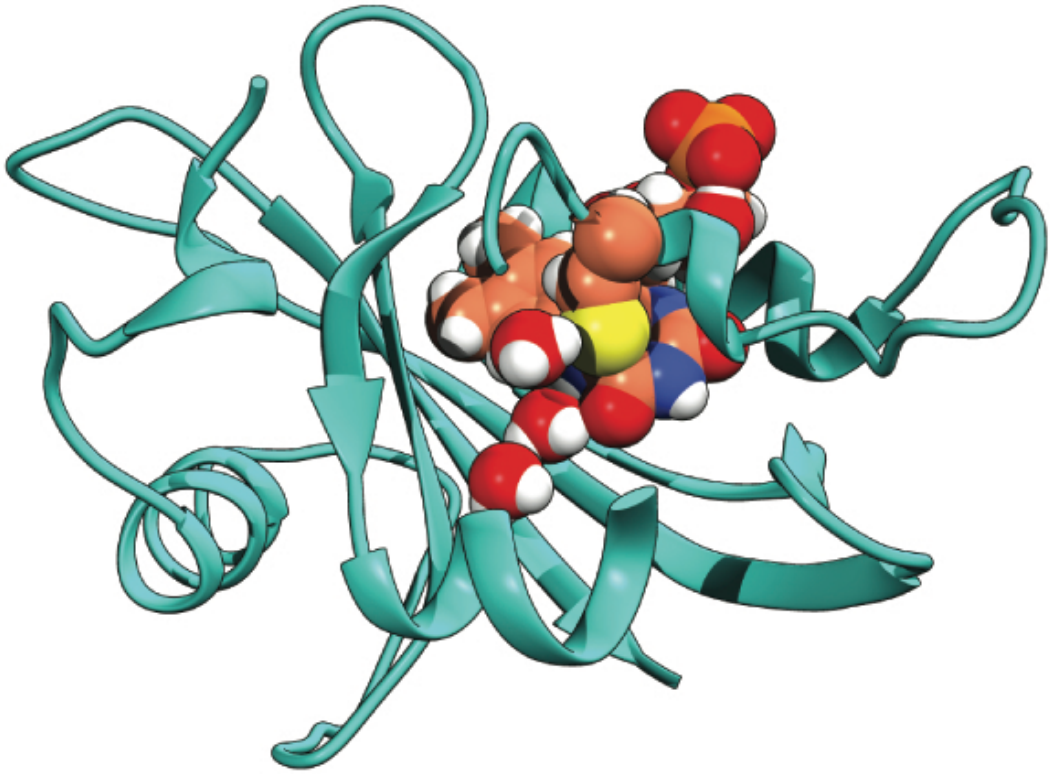


Figure 1.9: LOV domain with 3 waters clustered around flavin's N5. This datum was a single frame from an unpublished molecular dynamics simulation. Part of a loop region is cut away, because it is in front of the water cluster.

The resilience of the back reaction to mutation indicates that no residue plays an irreplaceable role, even if the replacement has very different side chain characteristics. This likely means that it is not an amino acid side chain that deprotonates N5, but rather it is a water, or water cluster that plays this role. In this hypothesis, the nearby amino acids dictate how close the water "likes" to get, thus determining kinetics.

Structured water is a central theme in this thesis. A water molecule, or a water cluster, (figure 1.9) proximal to N5 plays a central role in the hypothetical back reaction, as well as in the ionic mechanism of adduct formation.

Experimental evidence already has been published that support the conclusion that this cluster does in fact exist and play a central role in adduct scission[12]. It is reasonable to believe that this water cluster exists during the dark state of LOV proteins as well, but there is no evidence that this is in fact the case. One could argue that the water cluster enters the protein after the adduct forms. But, even the radical mechanism likely passes through a HN5 flavin intermediate, which also needs a proton source. This thesis examines these arguments about the environment within the flavin binding pocket. A general hypothesis explored in the following chapters is that the flavin binding pocket does contain water, and is in fact hydrophilic enough to accommodate a thiolate anion.

Section 1.4

Overview of chapters 2 through 5

Chapter two of this thesis identifies a key residue that controls the stability of water in the LOV active site; mutation of this residue also alters the adduct lifetime. Using molecular dynamics simulations, the change in adduct lifetime is correlated closely with the energetics of water in the active site. We infer that it is indeed this water that catalyzes the reverse reaction. Chapter three examines the protonation

state of the active site Cys, by using Raman spectroscopy of a monomeric Cys amino acid model system. Key vibrational bands were identified in the Cys Raman spectrum. Chapter four focuses on quantum chemistry calculations to explain the Raman spectrum of Cys and LOV domains. These calculations were used to identify normal modes of Cys, and to calculate theoretical Raman and IR spectra of a stripped down LOV active site model. Chapter five, the conclusion, takes the various data from earlier chapters, and ties them all together into a coherent argument that the LOV active site is able to accommodate water, and this water plays an important role in the LOV photocycle. The final section of chapter five describes some future directions for the study of the LOV adduct formation and scission mechanisms.

Chapter 2

This chapter is an adaptation of the journal article "A Native Threonine Coordinates Ordered Water to Tune Light-Oxygen-Voltage (LOV) Domain Photocycle Kinetics and Osmotic Stress Signaling in *Trichoderma reesei* ENVOY."[43], and is reproduced with the permission of the other authors.

Section 2.1

A detailed look at hypotheses and published data related to LOV adduct breakage, and the relevance of the LOV protein ENVOY

Several lines of reasoning have led to a general mechanism of LOV adduct decay. First, solvent isotope effect (SIE) experiments indicate that a single proton abstraction event is rate limiting [64, 90]. Second, adduct decay can be catalyzed by the presence of small molecule bases such as imidazole [1]. Third, residue substitutions at regions that regulate solvent access to the flavin active site have a substantial effect on LOV photocycle lifetimes [12, 90]. 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 [12, 17, 19, 64, 66, 88, 90]. Importantly, the natural base responsible for N5 deprotonation remains to be determined, however several possibilities have been suggest that involve elements in the immediate vicinity of the active site flavin.

In LOV proteins most residues near the N5 position are hydrophobic, with the exception of a conserved Gln residue important for signal transduction. 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 [90]. Consistent with the conserved Gln acting as a proton transfer agent, Gln to Leu and Gln to Asn mutations have a large effect on adduct decay kinetics [52, 56]. 2) The involvement of ordered water molecules within the flavin binding pocket [12]. Crystal structures partially support the latter mechanism, where LOV proteins conserve an ordered solvent channel leading to the active site Cys residue [89, 90]. 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 [12]; however, whether the effect is direct or indirect remains to be determined. Further, analysis of a bacterial sLOV protein (McLOVn) indicates that the introduction of a Thr residue in the vicinity of the N5 position can abrogate base catalysis in some systems [35]. Unfortunately, structures of McLOVn do not exist and thus cannot be used as a model for the role of solvent in base catalysis.

In this chapter, we present a photochemical study of the sLOV protein ENVOY (ENV1) from the filamentous fungi *Trichoderma reesei*. Recent crystal structures indicate that ENV1 may function as a model system for the exploration of the effect of structural water on LOV kinetics [44]. Specifically, similar to McLOVn, a

hydrophilic residue (Thr101) in the vicinity of the N5 position provides a hydrogen bond acceptor that may enable solvent recruitment near N5 [35]. Kinetic studies reveal that Thr101Ile variants demonstrate a 250-fold slower photocycle. Further, Thr101 directly regulates base catalysis and solvent access to the N5 position, but in a manner opposite to that observed in McLOVn. In ENV1, kinetic studies and computational analysis of structural water indicate that Thr101 affects both solvent access to the active site and ordering of water molecules adjacent to the N5 position. In these regards, ENV1 validates a role of structural water in functioning as an intrinsic base to tune LOV chemistry.

Recent studies of a *Neurospora crassa* Vivid (VVD) homolog from *T. reesei* (ENV1) revealed divergent signaling mechanisms within closely related fungal LOV domains [44]. Initial photochemical characterization confirmed LOV type chemistry, with a time constant for adduct scission of ~1500 seconds [44]. Such kinetics are 15-fold faster than its homolog, VVD, despite high sequence conservation [44, 90]. To examine the origin of the altered rate of adduct decay, we conducted a comprehensive study of the ENV1 photocycle and corresponding kinetics. The author of this document provided the computational data and analysis in this chapter, while Professor Zoltowski and his students did all the wet lab work described in this section.

Section 2.2

Materials and methods used to examine the role of water in ENVOY photo-adduct breakage

Protein preparation:

Full-length (FL; 1-207) and an N-terminally truncated construct ENV-64 (65-207) were cloned, expressed and purified to homogeneity. ENV Constructs were designed based on sequence homology to *Neurospora crassa* VVD. Initial constructs focused on the 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 (provided by Monika Schmoll) 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 ENV (1-207) using the Quick-change protocol (Stratagene). All mutants were also verified by DNA sequencing (Genewiz).
All constructs were expresses *in Escherichia coli* JM109 cells. Cells were grown at 37C until reaching an OD at 600nm of approximately 0.6. The temperature was then decreased to 18C for 40 min to assure thermal equilibrium. At 18C 0.3 Mm isopropylthiogalactoside (RPI) was added to initiate protein expression. After 22 hrs the bacterial pellets were harvested and stored in100 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 22C for 2hrs to cut the GST tag from the proteins. Cleaved proteins were then 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.

Spectroscopy and Kinetics:

UV–visible absorbance spectroscopy for all constructs and mutants of ENV were conducted on Agilent 8453 spectrophotometer. Photophysical properties of all constructs were verified to show LOV type chemistry exhibiting a single characteristic peak around 380-400nm for C4a adduct and two broad absorption peaks for the ground state. The ground state spectra consisted of a free flavin peak at 450nm and two vibrational bands at 425nm and 475 nm. 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 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 0mM-500mM Imidazole. Each data set at a particular concentration of Imidazole was repeated thrice. 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 $1/k_{adduct\_scission}$ that are averaged between the values obtained at 450 and 478nm. 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 [44], both constructs purify with a bound FMN cofactor and demonstrate spectra consistent with blue-light induced formation of a covalent cysteinyl-flavin C4a adduct (figure 2.1). 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 2.1). 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 [81, 90]. Notably, after cycling through one photocycle, we have so far been unable to chemically re-reduce the FMN cofactor.

Comparisons of the rate of adduct decay in the presence (reduced) and absence (oxidized) (figure 2.1 C and D) of glutathione demonstrates similar rates of

adduct decay under all conditions and constructs tested. However, the rate of oxidation of the neutral semiquinone displays distinctly different kinetic properties. Kinetic traces obtained at 450 and 478 nm from light-activated oxidized ENV-64

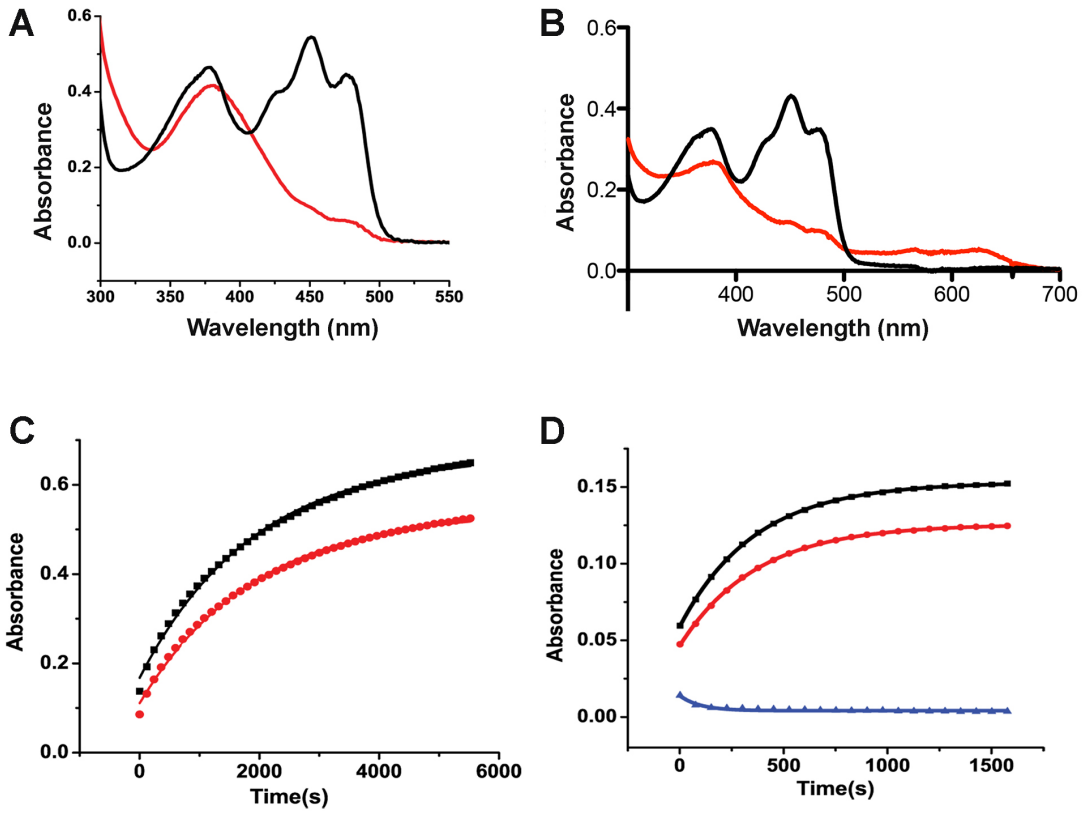


Figure 2.1: Kinetics and Spectra of ENV1 Proteins. A) Light (black) and dark (red) spectra of ENV 1-207 in the absence of glutathione. The light state spectra contains a single peak at ~380 indicative of C4a adduct formation. B) In the presence of glutathione the dark-state is still consistent with oxidized 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).

and FL ENV1 reveal time constants of 1500 and 1600 seconds respectively (Figure 2.1 C and D). 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 2.1x). Analysis of the kinetic trace at 622 nm, which isolates the spectral signature of the neutral semiquinone, reveals monoexponential decay with a time constant of 180 seconds (figure 2.1 C). We thus assign the major component as resulting from oxidation of the reduced semiquinone and the minor component resulting from adduct decay. We conclude that the rate of oxidation proceeds ~10-fold faster than adduct decay 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 [81, 90].

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 (figure 2.2). Kinetic studies revealed energies of activation comparable to other circadian clock photoreceptors. Specifically, both FL and ENV-64 demonstrated lower than expected energies of activation of 67 and 70 kJ/mole respectively. Arrhenius parameters for most LOV proteins of similar photocycle lifetimes have reported activation barriers on the order of 80-100 kJ/mole [64, 66]. Notably, studies of other circadian clock photoreceptors

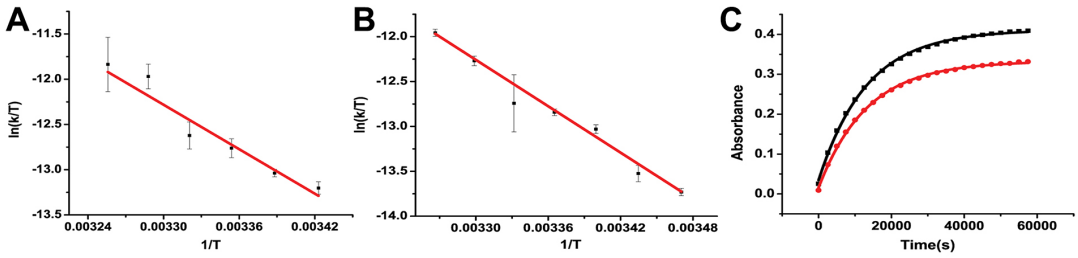


Figure 2.2: Eyring Plots and T101I Kinetics. A,B) Eyring plots of ENV1-FL (A) and ENV-64 (B). Both constructs demonstrate equivalent ΔH‡ (FL=68 kJ/mole; ENV-64=72 kJ/mole) and ΔS‡ (FL=-75 J mol^{-1} K^{-1} ; ENV-64= -63 J mol^{-1} K^{-1}). C) Absorbance at 450 nm (black) and 478 nm (red) as a function 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.

indicate atypically low energy barriers that Eyring analysis reveals are compensated by large unfavorable entropies of activation [64]. 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).

Such behavior renders ENV1 less sensitive to modest fluctuations in temperature and may facilitate temperature compensation of circadian clocks [64]. Studies of other LOV proteins with entropy-compensated photocycles revealed alterations in H-bonding to the active site flavin [64, 64, 66]. These alterations in H-bonding often affect or are affected by solvent accessibility to the active site [64, 66, 88]. To better examine the factors altering the ENV1 reaction landscape we examined solvent accessibility in ENV1 with computational tools.

Molecular dynamics simulations were run on the University of California, Santa Cruz computing cluster, using the charmm27 force field [10] as implemented in the GROMACS 4.6.5 program suite [7, 42, 82, 31, 61], with the addition of flavin-cysteinyl adduct parameters[27]. The starting coordinates were taken from the crystal structure of ENV1-64 wild type, and a mutation at Thr101 generated using Chimera [60]. 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&T (Number of particles, Volume, & Temperature) ensemble with the protein atoms positions restrained, then 100 ps with fixed N,P&T (Number of particles, Pressure, & Temperature) 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 into one trajectory using the gromacs utility trjcat [61].

Using the GROMACS radial distribution function tool [61] a plot of the density of water molecules around flavin N5 was created(figure 2.4c). From the radial distribution function, g(r), we calculated the free energy, w(r), (figure 2.4d) to pull a water from infinite distance to a given r from N5, using the equation -RT ln g(r) = w(r) [14]. The difference between the lowest and highest part of these curves was 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 (www.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 (figure 2.5), with number of frames having values within the bin defining the color of said bin, using the R (Team, 2015 #8} package hexbin [11]. Using a python (www.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 stiched together into a new trajectory, then the gromacs utility g_chi was called, and the order parameter for residue 101 was extracted.

Section 2.3

Discussion of ordered water in ENVOY active site, and its implications

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 [1]. 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 [64, 65]. Base catalysis of FL ENV1 and ENV-64 revealed a modest effect of imidazole on ENV1 kinetics (figure 2.3). Specifically, both FL and ENV-64 exhibit accessibility factors of ~1500, which is comparable to fast cycling bacterial LOV proteins [64], but moderate compared to the highly sensitive truncated LOVK [65]. Importantly, ENV1 is

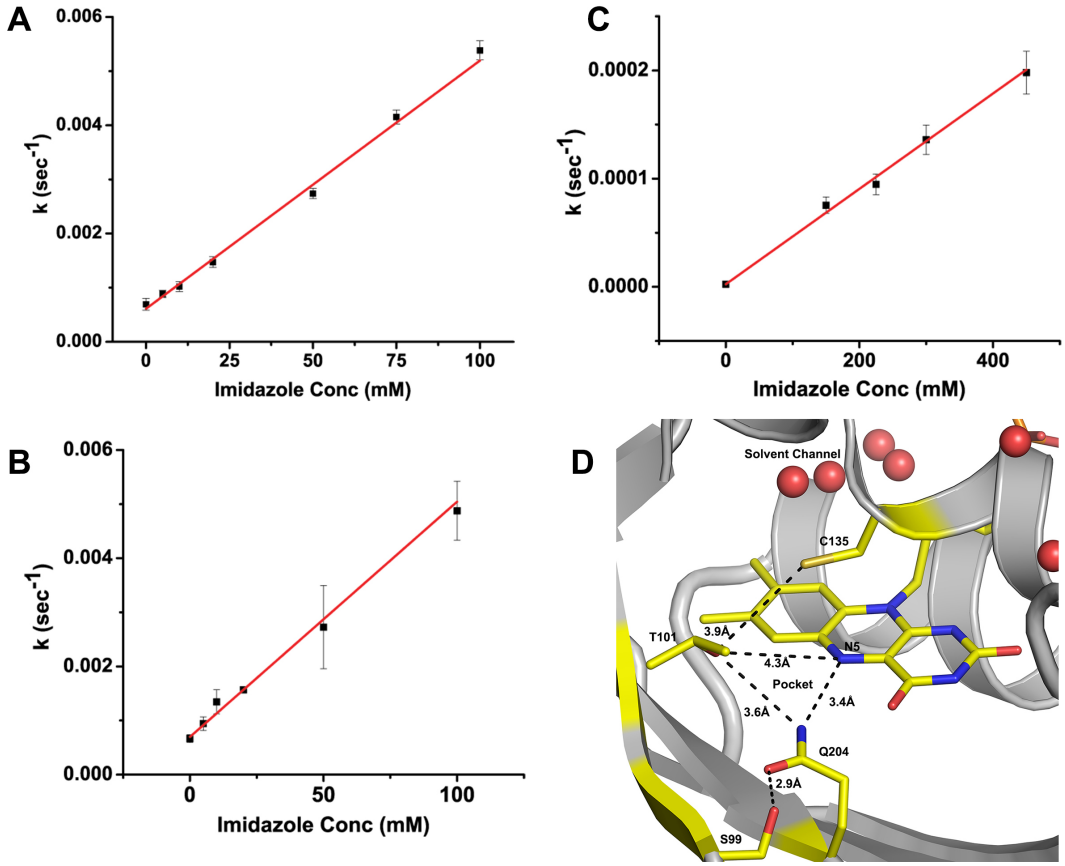


Figure 2.3: Imidazole Catalysis of ENV1. Dependence of the rate constant for adduct decay on imidazole concentration. Both FL and ENV-64 (A and B respectively) demonstrate similar effects of imidazole on the rate constant for adduct decay. The results are consistent with a solvent accessibility factor of 1500. In contrast, a T101I leads to increased solvent protection with an accessibility factor of 15. D) Active site of ENV1. 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.

approximately 15-fold more sensitive to imidazole than its homolog VVD (figure 2.3). Thus, the 15-fold difference in photocycle kinetics in ENV1 and its homolog VVD may be explained by the difference in solvent accessibility of these proteins.

To examine the origin of the variation in solvent accessibility, we examined VVD and ENV1 sequences for residue substitutions in the vicinity of the flavin cofactor. Importantly, one residue caught our attention. ENV1 contains a Thr residue at position 101 (Ile74 in VVD). Residue substitutions at this site are known to have large effects on LOV kinetics in VVD and other proteins [90]. Specifically, an Ile74Val variant of VVD accelerates the photocycle 24-fold, however it has no effect on solvent accessibility [90]. 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. In McLOVn the presence of a Thr residue abrogated imidazole catalysis [35]. Thus, to examine whether the Thr at position 101 may affect solvent accessibility and lead to a faster photocycle in ENV1, we studied a Thr101Ile variant in FL ENV1 to extract relevant photocycle parameters.

Indeed, 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. 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 yet 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 [35]. Combined the data indicate that the site adjacent to N5 is intricately involved in the LOV mechanism, but acts in concert with additional sites.

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 indicated a possible role of ordered solvent in dark-state decay, however a precise locale of action has remained elusive. Given that Ile74Val variants 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. 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 [1, 17, 66, 90]. To probe such a mechanism further, we conducted computational studies to probe the reaction landscape and corresponding ability to recruit solvent to the active site to increase the rate of adduct decay.

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 structural 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-ENV164 and the Thr101Ile variant indicate distinct differences in the energetics of active site water that have implications for both water occupancy (figure 2.4) and possible signal transduction mechanisms (figure 2.5) in LOV protein. In both proteins, HN5 dictates the affinity of water for the flavin active site, where water occupancy is most dense ~200 picometers (2 Å) from the HN5 position . Despite similarities in the location and duration of active site water occupancy, WT and Thr101Ile show different energies of water binding. 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. 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. Rather, ordered water in crystal structures are confined to a

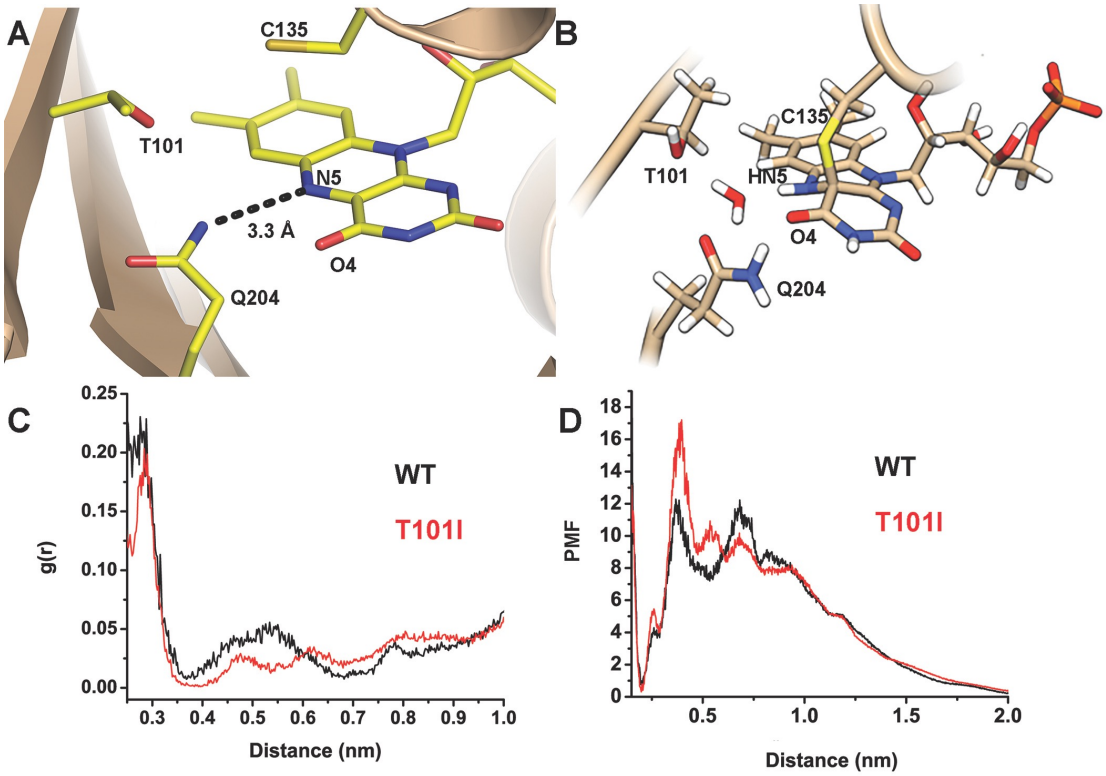


Figure 2.4: Energetics and location of water in the active site A) The dark-state crystal structure of ENV1 contains a buried conformation of Gln204, consistent with all other LOV domain structures. In the buried conformation Gln204, H-bonds to the unprotonated N5. B) MD simulations of ENV1 demonstrate 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) Radial distribution of water about HN5. WT proteins show higher probabilities for water occupancy near HN5. D) 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.

solvent channel adjacent to the flavin ribityl moiety. 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.

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 (figure 2.5). 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 (figure 2.5) 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). 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 recent MD simulations indicate that signal transduction may rather involve only rotation of Gln204 to H-bond to the newly protonated HN5 [9, 81, 89]. 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. In contrast, the lack of an

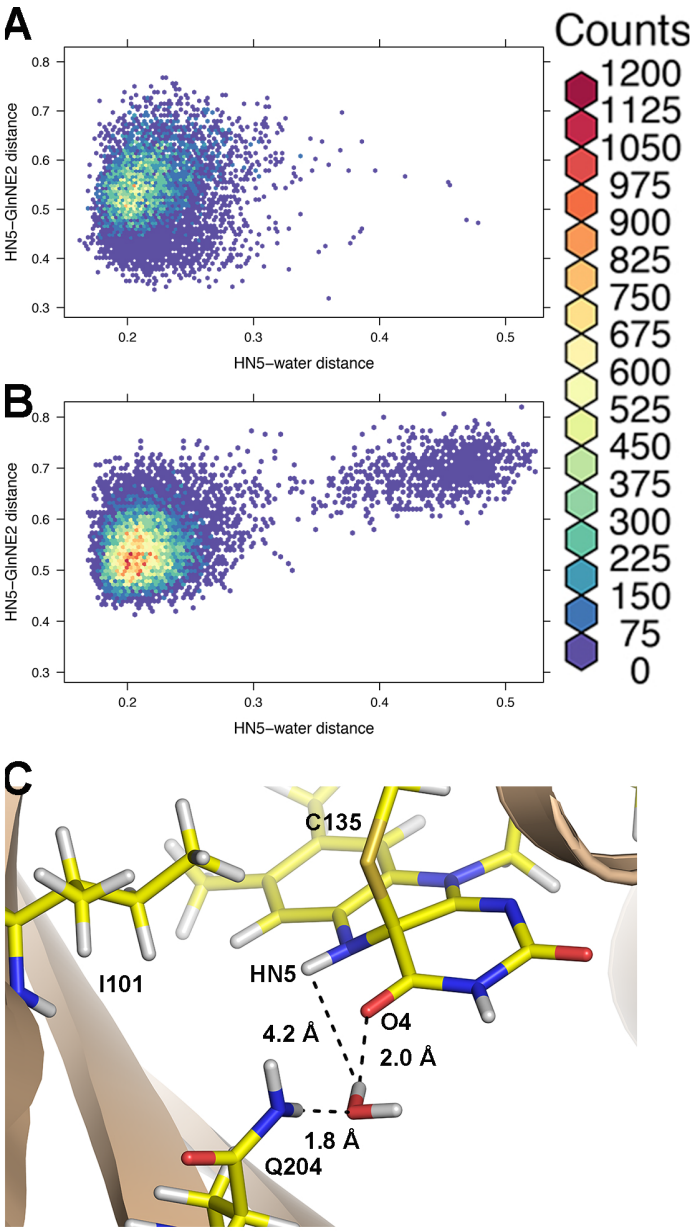


Figure 2.5: Water-HN5 and Q204-HN5 distances. A & B, Water-HN5 distances plotted versus HN5-Q204 distance. Color indicates number of frames within a given hexagonal bin. A) Wild type envoy. B) Thr101Ile mutant. C) Alternative low occupancy position of water in the flavin active site for the Thr101Ile variant, this structure is from a bin in the middle of the cluster in the upper right of figure B. 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 Thr101Ile ENV1.

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.

Due to the utility of LOV proteins in optogenetic tools, several studies have aimed to improve the fidelity of LOV optogenetic devices by either tuning photocycle parameters or signal transduction pathways; these studies are discussed by Pudasaini in 2015[63]. These efforts have been hampered by several fundamental elements. First, often it is difficult to decouple effects of solvent, steric constraints and electronic effects on photocycle parameters. These problems are exacerbated by an incomplete understanding of the role of water in mediating adduct scission in LOV proteins. Second, the effect of these photocycle altering variants has only been studied in a structural context in a few systems and the structural effect on signal transduction has been poorly explored [87, 88, 90].

In the present study we have employed ENV1 as a model protein for examining the role of a residue at the juncture between the LOV core and an N-terminal cap (position 101) on adduct decay kinetics. Position 101 has been implicated in tuning LOV domain kinetics in numerous systems, but its role in activating water has not been explored [35, 37, 64, 90]. Similarly, its effect on signaling mechanisms is not well known. Kinetic studies and molecular dynamic simulations of

ENV1 indicate that position 101 is ideally positioned to tune adduct decay pathways by altering solvent-HN5 interactions. Further, solvent access to HN5 is coupled to local order within the N-terminal region, Ncap, through Thr101 as well as conformational changes within a conserved Gln residue (Gln204) implicated in signal transduction in LOV proteins. Combined, these studies have implications for adduct decay pathways and signal transduction in LOV proteins.

Thr101 leads to stabilization and recruitment of water adjacent to HN5. Solvent recruitment directly correlates with a decrease in stability of the Cys-flavin C4a adduct, leading to acceleration in adduct decay kinetics. Thus, water likely functions as the intrinsic base deprotonating HN5 in the rate limiting step of adduct decay in LOV proteins.

In addition, Thr101 may mediate signal transduction in some LOV proteins. Activation of LOV proteins is typically coupled to signal transduction through two primary mechanisms. 1) Alteration of protein stability through order-disorder transitions following adduct formation. 2) Conformational changes in Ncap or the homologous C terminal region Ccap, elements, which hinge on rotation of an active site Gln (Gln204) following adduct formation and N5 protonation [88]. Molecular dynamic studies of ENV1 indicate that the residue identity at position 101 directs light induced changes in local order near the Ncap. Specifically the presence of the additional methyl functionality in Thr101Ile variants leads to ordering of position 101 (and the Ncap) in both the light and dark. In contrast, Thr101 is more disordered in

the dark, but following adduct formation and water recruitment is more ordered. Thus, side chain identity at position 101 may dictate order-disorder transitions in some LOV proteins. Similarly, water recruitment to HN5 is coupled to rotation of Gln204, whereby adduct formation leads to Gln204 rotations from the buried conformation present in all LOV structures, to a more exposed position interacting at the flavin O4 position. These movements are reminiscent of results of other MD simulations, which indicated that the active site Gln rotates out of the flavin binding pocket following adduct formation [27,21]. Notably, such conformations have not been observed in any crystal structures, so their role in signaling pathways remains unknown. We note that the role of position 101 in possibly both signal transduction and photocycle tuning indicates great care needs to be employed in studying rate altering variants, as many sites may have unexpected consequences in signal transduction.

Section 2.4

Chapter 2 summary

In this chapter, a slow cycling mutant of the LOV protein ENVOY was described. A threonine is proposed to regulate the photocycle speed. It does this by holding an ordered water in a position such that it is able to abstract the proton from N5 of flavin. Mutation of this Thr to Ile drastically changes the stability of water, and this directly affects the adduct lifetime.

Chapter 3

Section 3.1

Vibrational spectroscopy of LOV domains, and why Raman spectroscopy of cysteine was pursued in this thesis

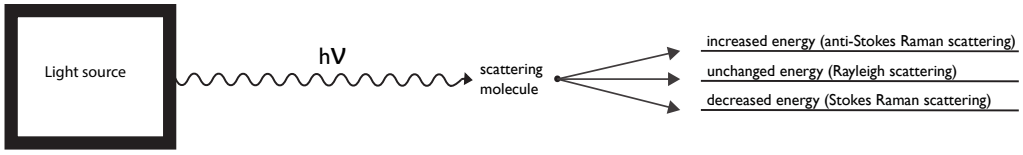
There are two main and opposed hypotheses regarding the formation of the adduct from the triplet state. The ionic mechanism (fig 1.6) proposes that the Cys is deprotonated, with a negative charge residing on the sulfur. The flavin becomes protonated at N5, then carbon C4a undergoes nucleophilic attack from a lone pair on the negatively charged Cys thiolate, forming the adduct. The radical mechanism (fig 1.7) proposes that a protonated Cys has either a radical electron or a hydrogen abstracted by the triplet flavin. The system then undergoes radical recombination, forming the adduct. Recent vibrational studies have made a strong case for the N5 of the flavin being deprotonated before adduct formation[80], but this doesn't rule out extremely fast N5 protonation, or protonation after adduct formation. Regardless, the group reporting this drew the conclusion[80] that the ionic mechanism was untrue.

Vibrational spectroscopy uses photons to infer something about the vibrational states present in a molecule. There are two common types, Infrared and Raman spectroscopy. Infrared spectroscopy is a type of absorbance spectroscopy, in which infrared photons are passed through a sample, and the wavelengths that are

absorbed by the sample correspond to the energy of the different vibrational modes of the sample. Raman spectroscopy is fundamentally different than IR, although similar information about energy states is inferred. Since the Raman effect is a scattering phenomenon, it is not the absorbed light that carries the vibrational information, rather, it is the scattered light that does so. When individual molecules scatter light they can do so with no change in frequency, elastic scattering, or with a change in frequency, inelastic scattering. These are termed Rayleigh and Raman scattering, respectively (figure 3.1). Given the topic of this thesis, LOV domains, which detect blue light, it is interesting to note that the color blue inspired the discoveries of both Rayleigh and Raman scattering. Lord Rayleigh wanted to explain the color of the sky, and C.V. Raman was fascinated by the bright blue Mediterranean sea[45].

Raman spectroscopy relies on inelastic scattering of light; the change in frequency is gained from or lost to vibrational energy from the scattering molecule. The terms Stokes and anti-Stokes refer to photons with less or more energy that the incident light; the term Stokes originates in a description of fluorescence. The Raman effect can be easily explained using classical mechanics. Light is a self propagating magnetic and electric dipole. The electric dipole can interact with the electric dipoles of a molecule and distort the electron and proton structure. The electrons are

Figure 3.1 Raman and Rayleigh scattering



polarized by photon absorption, and when they squish back into their unpolarized shape they emit a photon; since their movement back into unpolarized state is a changing electric dipole, it makes a changing magnetic dipole. But, the molecule's proton arrangement can also change, and thus change the vibrational state of the molecule. If the vibrational state of the molecule changes, the emitted photon has different energy than the absorbed photon, this is inelastic scattering.(figure for Stokes and anti-Stokes?) If the molecule doesn't change vibrational state the scattering is elastic. If the incident light approaches the energy of an electron transition to a higher state the scattering efficiency increases drastically; this is termed resonance Raman. There are other types of Raman scattering, such as multiple photon and stimulated Raman, but these are not important to this thesis, and are not discussed.

Raman spectra are usually presented with intensity of the scattering as an arbitrary Y axis unit, and inverse centimeters, cm^{-1} (wavenumber) on the X axis. Wavenumber equals frequency, scaled by the speed of light, has units of energy, and is directly proportional to energy. Wavenumber is superior to using the wavelength to describe vibrational spectra because wavelength is not linear with respect to energy.

But, why bother with Raman spectroscopy at all? Raman spectra of proteins are crowded. In particular, the amide I, II and III, which are overtones of the amide vibrations, completely obscure many useful bands. A useful trick can be used to reduce or even eliminate annoying vibrational absorption bands: difference Raman.

Difference Raman works by subtracting the dark state protein spectrum from the light. In this way positive and negative bands appear where changes occurred and bands that are unchanged upon excitation are removed. This only works if the states are homogenous. If the two spectra are composed of many different protein states the subtraction result will be intractable, there will be no way to assign any band to a particular state. For this reason light activated proteins are well suited for these studies, as it is easy to keep the proteins under study in their active or dark state. In this way one knows that all negative bands in a difference spectrum are from the dark state, and all positive bands are contributed solely by the light state.

Another common problem in vibrational spectroscopy is identifying bands. Common methods to to assign bands to particular amino acids within a protein, such as mutagenesis and stable isotope labeling, are both expensive and time consuming. Introduction of non-native amino acids as vibrational probes is one way to identify vibrational bands, but this method suffers from the same drawbacks as regular mutagenesis. Artificial amino acid probes have been used to determine the electrostatic potential within a protein, by using the Stark effect[75,48]. But, non-natural amino acids perturb the interior of the protein. We circumvent the issues of these mutagenesis methods by using the vibrations of the endogenous catalytic cysteine of LOV proteins. We use this Cys to determine not only the protonation state of the same Cys, but also to examine the hydrophobicity of the interior of the protein.

Cysteine is a canonical amino acid, with thiomethane as the sidechain. The thiol moiety has a pKa of roughly 8. Thus, at biological pH one would expect the sulfur to be deprotonated 90% of the time. However, in the radical mechanism for LOV adduct formation, the Cys reacts from a protonated state[68]. It has been proposed [33,34,39,85] that the electrostatic potential of the flavin binding pocket is very inhospitable to water. This argument continues to state that the protein environment makes deprotonation of the cysteine essentially impossible[33,34,39,85]. It was also argued that there is no nearby base to accept the thiol's proton [68]. However, an earlier Molecular Dynamics (MD) study[27] noted that a water is occasionally present in the flavin binding pocket. Additionally, as described in chapter two of this thesis, we demonstrated[43], the highly probable existence of a low energy binding site for water near the active Cys thiol, and coordinated to N5. **This evidence suggests that the hypothesis stating the binding pocket is too hydrophobic for water needs to be reexamined.**

The limitations of MD simulations prevented its use in predicting the actual protonation state of the reactive cysteine, since that involves quantum effects. In particular, since MD simulations approximate chemical bonds as unbreakable, they are not useful to model the formation and breakage of said bonds. More specifically, MD does not model proton transfers to nitrogen lone pairs adequately. That doesn't rule out other forms of computational chemistry, and another chapter in this thesis describes the use of quantum chemistry to identify bands in vibrational spectra.

The Bogomolni group, using FTIR spectroscopy reported an ordered water cluster required for photocycle completion[12]. Given the computational evidence[43] that water spends a lot of time coordinated to N5, we propose that this water cluster is 1) the base that deprotonates N5 during adduct breakage, 2) the base that deprotonates the thiol of the reactive cysteine, and 3) the proton donor for N5.

Difference (light minus dark) infrared (IR) spectra of LOV domains produced by other groups invariably[33, 34, 68, 85] show a negative band at ~2560 cm^{-1} that was attributed to the thiol SH stretch. Our lab has sometimes, but not always, seen this band. The presence of this band has been reported [34] as evidence that the Cys must be protonated, and thus that the ionic mechanism of adduct formation must be false. Since we have not always seen this band, we haven't been convinced that the ionic mechanism must be false. Also, this band has not been studied in detail, and its identity should be confirmed in order to use it as a strong argument for the validity of the ionic mechanism. Although both infrared and Raman spectra show this band in light minus dark difference spectra, water also absorbs in this region of the IR. However, water doesn't interfere too much in Raman spectra, thus, Raman is a good tool to examine this vibration. Whether the 2560 cm^{-1} vibration is a useful target to examine the protonation state of the Cys in the LOV active site is explored in this chapter.

Another vibrational band that that our group has been interested in exists at ~670 cm^{-1}. This band has been attributed to a SC stretching mode[58]. It was suspected

by members of the Bogomolni gorup that it would have been possible to deduce whether the sulfur was protonated by calculating the spring constant for this vibration, and then calculating how much the mass of the oscillator changed by using a simple harmonic oscillator approximation.

The data presented in this chapter are relevant to the LOV photocycle in that they represent a model system to show both the identity and the utility, or lack thereof, of the two Cys vibrational bands mentioned above. To accomplish this we utilized monomeric Cys amino acid as a model system. The Raman spectrum of Cys was taken at differing pH, in order to quantify the behavior of the 670 and 2560 cm^{-1} bands at different protonation states.

Section 3.2

Materials and methods used to collect Raman spectra of free, monomeric cysteine

Because the scattering intensity increases with the number of molecules in solution, all Raman experiments comparing peak size had to be carried out with the maximum possible precision in Cys concentration. In order to determine the ideal concentration of cysteine, its free amino acid monohydrate was put into aqueous solution. Roughly 3 molar (moles per liter, M) Cys gave the best Raman spectra, however, at pH above 4, the highly concentrated 3M Cys formed an unidentified precipitate, sometimes fairly quickly. This precipitate often had a sulphurous odor and interfered with the Raman scattering. 1M Cys was found to be the ideal

concentration, high enough to give decent scattering, but not so concentrated as to precipitate at higher pH during the experiment. All Cys Raman data presented in this chapter were collected from 1 M solutions. In order to produce precise concentrations of Cys, all preparations were made using the same 5 mL volumetric flask, and the same analytical balance.

pH adjustments were made by dissolving the Cys along with NaOH pellets, rather than dissolving the Cys in pre-titrated basic solution. This was because for some unknown reason the Cys didn't dissolve well in pre-prepared NaOH solutions, and often recrystallized nearly immediately. 1M Cys with 4, 3 & 2 M NaOH was used in Raman experiments. The pelletized nature of the NaOH made measuring exactly the mass needed to make precisely integer values of Molarity impossible. The NaOH was also old, and thus of unknown water percent. For these reasons, the pH was measured after dissolving the Cys along with the hydroxide. The measured pH for the approximately 4, 3, 2, and 1 M NaOH were 10, 9, 8, 4, respectively. The Cys used was its free acid monohydrate, so the carboxylate, the thiol, and the amide were all protonated, and thus 1 mol of Cys was capable of neutralizing 3 mol of hydroxide. It is also likely that the pH of the de-ionized water used affected the pH of the Cys solutions. The pH of the de-ionized water used was not measured every time, but was often around pH 5.

1 M Cys that was dissolved in pure water, (without any base) had a very low pH, close to 1. Cys was also dissolved in pure D_2O, its pD was assumed to be close to

1 as well, and was used at 1M Cys. Once the Cys sample was dissolved it was placed into a small capillary tube using a microliter syringe, the end of the capillary was sealed with tape; since the samples were only used for one collection they did not need to be hermetically sealed.

All spectra were collected using the macro stage on a Thermo-Nicolet Raman spectrometer, using the 780 nm laser. The aperture was set as a 50 micron slit. The Raman laser was focused using a two-step process. First the spotting laser was used to align the sample in the collection chamber, while open. Next, the X, Y and Z focus of the Raman laser was done to maximize signal with the chamber door closed. Each presented spectrum was the average of 4000 10 second scans, with "moderate" suppression of cosmic rays.

A series of spectra were produced with Cys at pH 10, 9, 4 and 1, these are shown in figure 3.4. Only the region of the 2560 cm^{-1} band is shown in figure 3.4, for simplicity. A comparison of different states of the 670 cm^{-1} band is in figures 3.2 and 3.3. Figure 3.2 compares the 670 cm^{-1} vibration in Cys at pH 10 to Cys at pH 1, and figure 3.3 compares Cys 670 cm^{-1} vibration in D_2O to H_2O.

Section 3.3

Discussion of the Raman spectrum of cysteine, and how this is useful for examining the protonation state of the LOV reactive cysteine

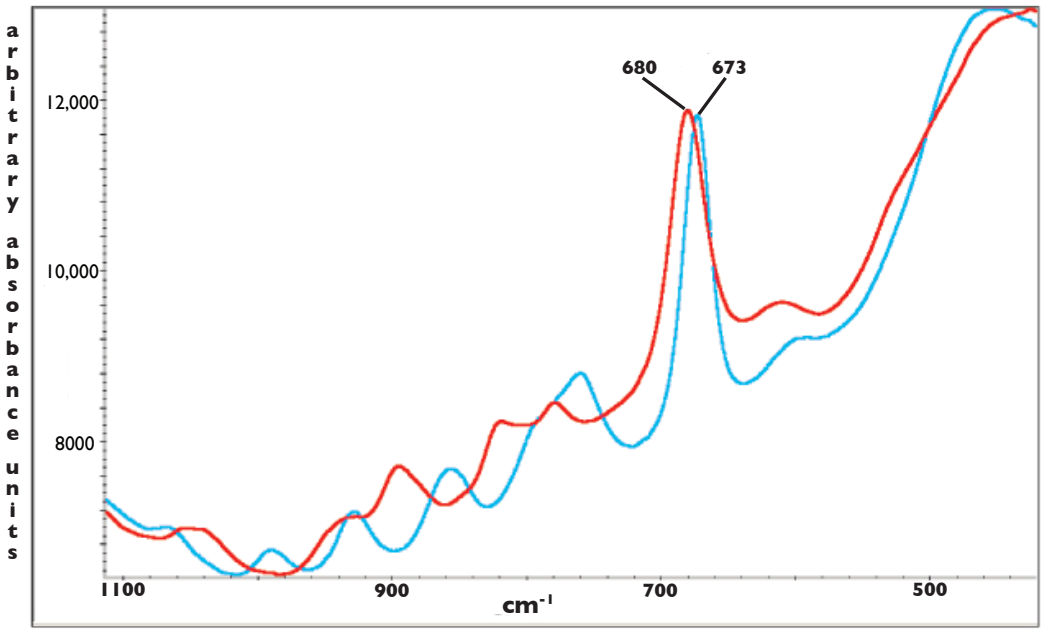


Figure 3.2, Cysteine Raman spectra, the 670 cm^{-1} band, pH 11 in red, pH 1 in blue

The 670 cm^{-1} band (figures 3.2 & 3.3) does not appear to behave as if the SH is a simple mass on a spring attached to the alpha carbon. This contrasts with older literature[58] that described it as SC stretching mode. In fact, upon deprotonation (figure 3.2) the band is shifted to higher wavenumber. From the computational analysis presented in chapter 4, we determined that this mode is a SC stretch, in that the S - C bond does, in fact, stretch. But, the vibrational mode is spread across the entire molecule, and upon deprotonation the entire mode changes (figure 4.5). The same unexpected shift in the Raman occurs with D_2O compared to H_2O (figure 3.3),

and the SH vibrates with more energy than SD. We conclude, based on the computational analysis in chapter 4, and from the unexpected shifts presented here, that the band at 670 cm^{-1} is useless for probing the protonation state of the reactive Cys in LOV domains.

Figure 3.4 shows the 2560 cm^{-1} band at differing pHs. The peak is located in the same place, but its intensity is dependent on the fraction of the sidechain that is protonated. At pH 10, close to 99% of the amino acid exists as S^- thiolate. Even at this pH, the band is not entirely gone. The fact that the peak is still visible even when only 1% of the sample is able to vibrate in that mode shows its strong oscillator

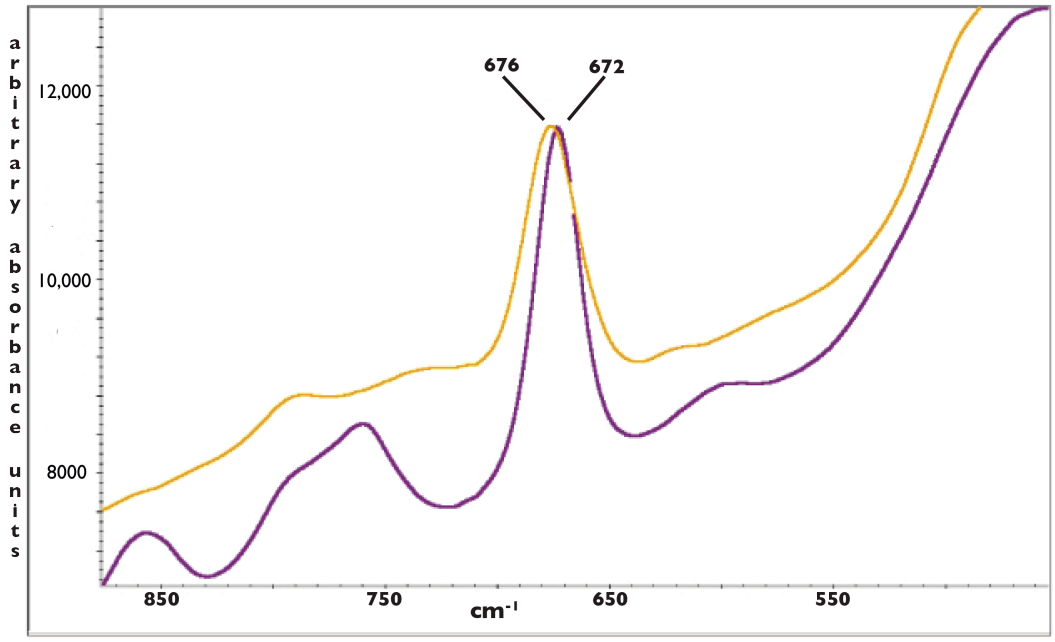


Figure 3.3 Cysteine Raman spectra, the 670 cm^{-1} band. Cys is dissolved in D_2O in orange, in H_2O in purple

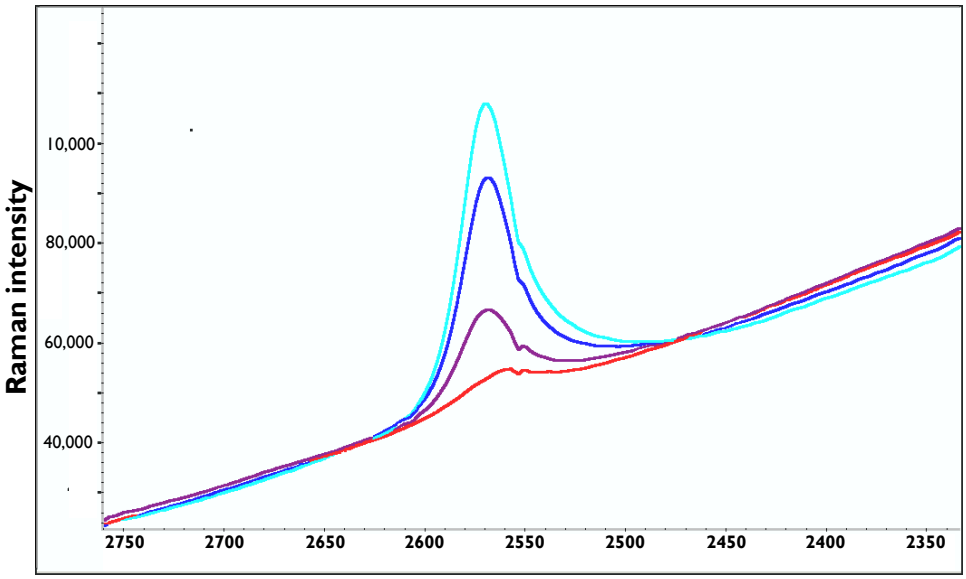


Figure 3.4 Cysteine Raman spectra. This is the 2560 cm^{-1} band, at pH 10, 9, 4 and 1, in red, purple, blue, aqua, respectively

strength. When the proton concentration of the solution is increased there is a corresponding increase in 2560 cm^{-1} peak size.

The use of the 2560 cm^{-1} band in earlier arguments[33, 34, 68, 85] about the adduct formation mechanism is supported by this Raman data. With the 2560 cm^{-1} band positively identified and its behavior understood, we can use it as an endogenous probe of the protonation state of the reactive Cys within the LOV domain. The implications of the behavior of this band in actual LOV vibrational difference spectra is given in the conclusion of this thesis.

Section 3.4

Summary of chapter three

In this chapter we showed that the 2560 cm^{-1} of cysteine is very useful in arguments about the protonation state of Cysteine in the LOV active site. However, the band at 670 cm^{-1} is useless for this same purpose.

Chapter 4

Section 4.1

Quantum chemistry's application to understanding the LOV photocycle

As described in the previous chapter, there is a debate on the capacity of the LOV active site to accommodate water, or an ionized Cys thiol. The ability of the LOV active site to harbor water or a thiolate has implications for both the forward and reverse reactions in the LOV photocycle. Vibrational spectroscopy is often used in the debate over these things. In order to make LOV vibrational spectra more tractable, the Raman of Cys at differing pH was examined in the previous chapter. Certain vibrational bands did not behave as expected, and an explanation for this was needed. This chapter describes the use of quantum chemistry calculations to explain the Raman spectrum of Cys. Theoretical vibrational spectra of a model LOV active site are also presented, with the goal of using them to explain previous vibrational data from whole LOV domains.

Computational studies have been undertaken of the reaction coordinates between flavin and Cys[39, 78,2, 24] but used nothing to stabilize the bare negative charge on the thiolate, which is unphysical. Even beyond mere unphysicality, bare negative charges are known to be, at best, poorly modeled with quantum chemistry software. Also, previous computational studies [39, 78, 2, 24, 15] of the LOV photo adduct neglected modeling water in the binding pocket, and thus were not very useful in assessing the validity of an ionic mechanism, if water is part of said mechanism.

Members of the Bogomolni group have proposed the SC stretching mode of Cys at 670 cm^{-1} in the Raman spectrum could be used to demonstrate the protonation state of the LOV reactive cysteine. By using the free amino acid Cysteine as a model compound, the force constant of the S-C vibration at 670 cm^{-1} would be calculated from the first order harmonic oscillator approximation. This calculation would be carried out twice, once with the mass of the attached H added to the mass of the S to model protonated Cys; and once with only the mass of the sulfur, to model deprotonated Cys. Using this force constant, the protonation state of Cys in the difference Raman spectra of LOV domains would be determined. However, Raman spectra of Cys titrated at different pH levels showed an odd effect; the 670 cm^{-1} vibration was shifted to the red in protonated form, when compared to high pH Cys (figure 3.3).

Another feature of LOV vibrational spectra is a band at 2560 cm^{-1}. This band has been described[33, 34, 68, 85] as a SH stretch. The presence of the 2560 cm^{-1} band has been used as an argument[33, 34, 68, 85] that the active Cys is always protonated in the LOV dark state, and thus the ionic mechanism of LOV photoadduct formation must be false[33, 34, 68, 85]. This argument is described in much more detail in the previous chapter of this thesis.

So, here is a summary of the questions addressed in this chapter. What are the vibrations appearing at 670 and 2560 cm^{-1}? Can predicted difference spectra be created, for comparison to wet lab data, to evaluate the prediction from MD that

water is stable in the flavin binding pocket? Can these predicted spectra also be used to predict the stability of a thiolate ion within the active site? We use quantum chemistry to answer these questions, but what is quantum chemistry?

Quantum chemistry is a category of computational chemistry. The two main categories of computational chemistry are Molecular Dynamics (MD), and quantum chemistry. MD is covered in more detail in another chapter of this thesis. This thesis doesn't present new computational tools, but describes the use of existing programs for computational quantum chemistry applied to biochemical questions about the photocycle of LOV proteins.

Modern quantum calculations are far too complex to be done by hand, so computers must be used to carry out the calculations. In fact, computational chemistry is nearly as old as digital computers themselves. It has grown from extremely simple electron energy and structure calculations for tiny molecules, into much more. This broad field spans from what is more properly called physics, all the way into biochemistry. For example, protein structure predictions and statistical mechanics of fluids can both be considered at times to be computational chemistry, despite their dissimilarities to each other. What they have in common is a need for chemical information about structure and reactivity, and that the number crunching must be done on a computer.

Digital computers were first built during World War 2 and the first digital-programmable computer was built in England in 1943. The first computers took

paper tape as their input. For another couple decades computer input/output consisted mainly of punch cards and punch tape. It was during this period that computational chemistry was developed, but it was considered merely a way to calculate, rather than a field unto itself. It wasn't until 1988 when the Nobel prize in chemistry was awarded to Pople and Kohn that computational chemistry was officially recognized as its own discipline. Since then, the massive increase in computational power has made the use of computational chemistry available to anyone with the time to click a few times in the graphic user interface available for some programs.

Quantum mechanics was developed around the turn of the 20th century, to explain phenomena that classical physics failed to describe correctly, such as blackbody radiation, and the hydrogen emission spectrum. Perhaps the most revolutionary aspect of this new physics was the concept of restriction of particles to discrete energy levels. This so-called quantization gave the new physics its name. Attempts by DeBroglie to incorporate wave mechanics into quantum mechanics laid the foundation for Erwin Schrodinger to create the now eponymous equation.

Despite the apparent simplicity of the Schrodinger equation, solving it for more than 2 particles is intractable[50] due to the correlation of the position and momentum of the particles. The nearly infinite possibilities for even a tiny system of 3 particles, renders solving for the position and momentum computationally prohibitive.

In order to circumvent this problem, approximations must be introduced[50]. The most common, and in fact, almost universally one used is the Born-Oppenheimer approximation[50]. This takes that the nuclear and electrons wavefunctions are separable. This is based on the (not too severe) assumption that the electrons are moving much, much faster than the nuclei[50]. The electrons can be approximated as seeing fixed nuclear coordinates, and the nuclei approximated as seeing a smeared out electric field from the electrons.

There are two main methods to solve for electron structure, after using the Born-Oppenheimer approximation[20]. The first, called Hartree-Fock (HF) theory, takes the electrons to be uncorrelated with each other[20]. Although the limitation of neglecting electron-electron correlation is severe, [20] this can be compensated by additional terms added to HF calculations to make up for the deficiency[20]. HF calculations are often useful in that they are cheaper[20] in terms of computational power compared to the other common quantum chemistry method, Density Functional Theory (DFT).

DFT takes the electrons to be correlated with each other[20], but treats them not individually, but as a cloud of electron density. From this cloud of electrons, a functional that describes the density is created. This means that DFT doesn't solve for a wavefunction, it optimizes the electron density[20]. Both HF and DFT method iteratively solve equations, using the resultant as the input for the next round, until the energy converges to a set tolerance. In order for this type of iterative approach to

work, the variational principle must hold for that system[20], otherwise one has no good means of gauging the quality of the results. These approaches are termed self-consistent fields, because they (hopefully) lead to the global energy minimum for that system, so are self consistent, if not actually physically true[20].

DFT was the method of choice for this thesis. The reason for this is purely functional (no pun intended). Previous use of DFT functionals has produced good results[74] for biochemical systems in which hydrogen bonding is important. A DFT calculation begins with the construction of a starting electron density from a basis set[20]. A basis set is a group of linearly independent mathematical functions. A linear combination of the members of the basis set is used to represent the electron density. Old basis sets were the hydrogen wavefunctions[20]. Construction of modern basis sets utilizes the mathematical principle that states any function can be approximated using a linear combination of other functions. Since integrating Gaussian functions is much cheaper computationally than integrating the hydrogen wavefunctions, Gaussian basis sets are used almost[20] exclusively now.

As outlined in the first part of this chapter, as well as in chapter 3 of this thesis, knowing the vibrational mode associated with bands in vibrational spectra of LOV proteins is very useful. We can take these and argue that certain residues are protonated or not at key moments in the photocycle. This, in turn, allows us to evaluate the viability of hypotheses regarding the mechanism of formation and scission of the flavin-cysteine photoadduct. In addition, knowledge of the stability of

water reinforces arguments made using molecular dynamics regarding the hydrophilicity of the flavin binding pocket. The methods of quantum chemistry outlined above provide the theoretical framework in which the following data was collected.

Section 4.2

Materials and methods of quantum chemistry used

Orca [54] is a fast, free quantum chemistry package, and is the only quantum chemistry software used to create the data in this chapter. Orca was picked over Gaussian, despite the UCSC campus license for Gaussian. The reason for this is Gaussian is proprietary, and a philosophical decision to use open source software was made. All molecular geometries were created using Gabedit [4]. Calculations were carried out on a Dell T7400 with dual quad core Xeon processors, running Ubuntu Linux version 14.04.

Orca's implementation of COSMO [40] implicit solvation was used, with calculations at 2.2, and 80 (unitless) dialectric constant to investigate whether the vibrational bands of Cys of interest were different in water versus the interior of the protein. Two forms of Cys were analyzed, protonated and deprotonated SH. In both sulfur protonation states the nitrogen was deprotonated to more closely mimic the vibration of polymerized Cys within the protein. The carboxyl was protonated in the calculation of the SH vibration. Despite the slight unphysicality of the protonation of

the carboxylic acid when the amine is deprotonated, a bare negative charge is well known to produce bad results in these types of quantum calculations. Normal mode analysis was carried out after "tight" self consistent field convergence of geometry and energy, using the Def2-TZVP basis set [83] with the Alrich TZP auxiliary set [69] and an integration grid of size 4.

COSMO was used with a dialectric of 2.2 in calculations of coordinated water in a stripped down version of the active site. Almost all residues needed to be excluded in the calculations. Initially, only hydrophobic residues were excluded, since they don't contribute much more than a structural role in the catalytic cycle, and too many atoms makes calculations intractably long. Convergence issues necessitated the removal of all but the catalytic Cys residues as well, because even position restraints didn't suffice to create a stable system. The key water was included. Lumiflavin, a flavin with a methyl group bonded to N1, undergoes the same adduct formation as FMN, so to save computing resources it was used instead of FMN. Input geometries and converged geometries are given in figures 4.4 and 4.5. The adduct geometry is not shown since it is quite similar to the dark state.

The crude output files were plain text, these were converted into animations using Orca's normal mode analysis program. These animations were visualized using a version of Pymol [23] compiled from source code. The normal modes were inspected visually. The normal mode figures (figures 4.3 and 4.4) were made using a picture of Cys rendered in Chimera [60], with further images added using Adobe Illustrator.

Orca also has a program that can output vibrational spectra. This was used with a peak widening option, to replicate actual spectra visually. This program outputs a text file with two columns, which were converted to CSV format using vim. The CSV files of the spectra were further processed using R [79]. R was used to create difference spectra from the individual calculations by subtracting the relevant columns from the CSV files. R was also used to plot the spectra. R's peak labeling package was broken for the Linux distribution used in the Bogomolni lab, so a work around utilizing the R package Shiny was undertaken. The plots were then exported to Adobe Illustrator, and combined into the figures in this chapter.

The converged geometries are compared to the starting geometries in figures 4.2 and 4.3. The band at 2560 cm^{-1} in Cys vibrational spectra (figure 3.2) was demonstrated to be a SH stretching mode, as had been previously reported (figure 4.3A). However, the 670 cm^{-1} band (figure 4.3B and 4.4), which was thought to be a SC stretching mode, turns out to be much more complicated when the calculated vibrational mode is inspected. The 670 cm^{-1} band does involve CS stretching (figure 4.5), but the mode encompasses the entire molecule, and the deprotonated Cys vibrates in a very different way. This calculation could explain the unexpected shift of this band in the Raman spectrum of Cys described in the previous chapter of this thesis. Clearly, the SC mode encompasses the entire molecule, while the 2560 cm^{-1} mode only involves the thiol moiety. Also to be noted, the band at 2560 cm^{-1} is present only in the protonated Cys.

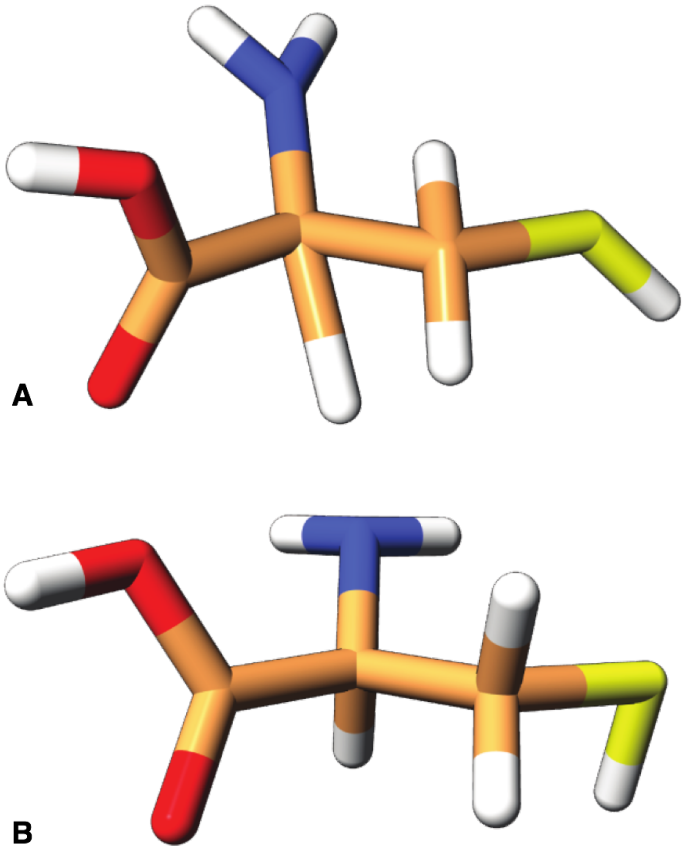


Figure 4.1, Cys geometry, A is the input, B is optimized

We proposed in the previous chapter to use Cys as a model compound to argue that Cys can exist as a thiolate in LOV proteins. The electrostatic potential of the active site of proteins is the source of the catalysis, and, as stated earlier, both the stability of water and the ability of Cys to exist as a thiolate are determined by local electrostatics. Since we carried out the Raman spectroscopy of the model compound Cys in H_2O and D_2O, we couldn't merely argue that the bands must be the same within the flavin pocket as they are in water, since the interior of proteins has ε close

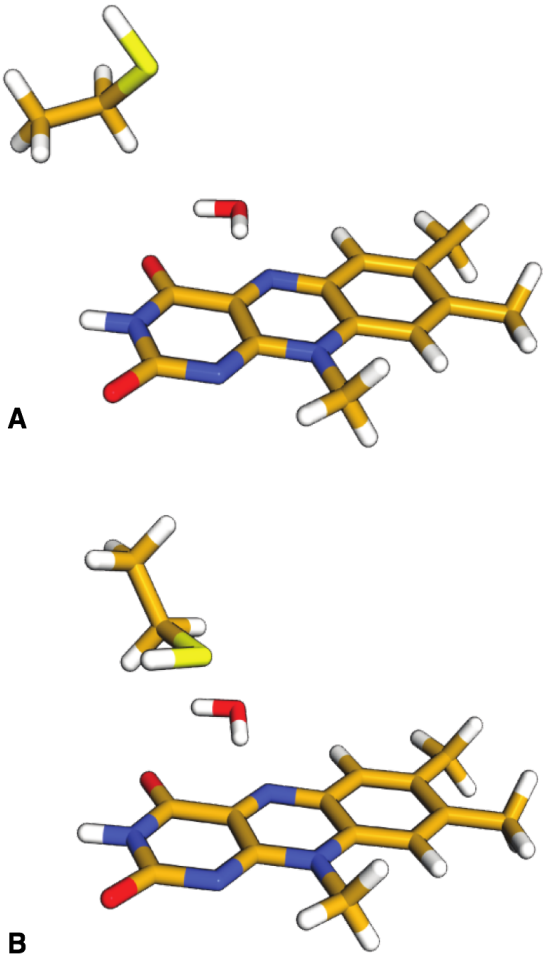


Figure 4.2, dark state geometry. A is the crude input, B is the optimized geometry

to 2.2, whereas water it is ~80 (these numbers are unit-less, since they describe a relative permittivity). The comparison of these two conditions was carried out computationally, and the data is in table 1.

An absolute vibrational spectrum of a protein is very complex. Moreover, it is a static picture. Difference spectra are far more informative of dynamic processes. By taking the light spectrum and subtracting that from the dark we can see the bands that are changed upon light activation. By calculating the vibrational modes of flavin

and key catalytic Cys, as well as the hypothesized ordered water, and subtracting it from the vibrations calculated from a model of the adduct, theoretical difference spectra were produced (figures 4.6 and 4.7).

Section 4.3

Discussion of quantum chemistry analysis of cysteine and of the LOV active site

A main argument against an ionic mechanism of adduct formation is that in the dark state, the key catalytic Cys amino acid is always in its protonated form. Many researchers [2, 33, 34, 68, 85] have claimed that the negative band at 2560 cm^{-1} in vibrational spectra is always present in LOV vibrational difference spectra, and its identity as a SH stretching mode proves that the key Cys is always protonated. The data in this chapter show that indeed, this is a SH stretch, as the spectra presented here all agree with physical spectra. However, our conclusions are quite different than those from other groups.

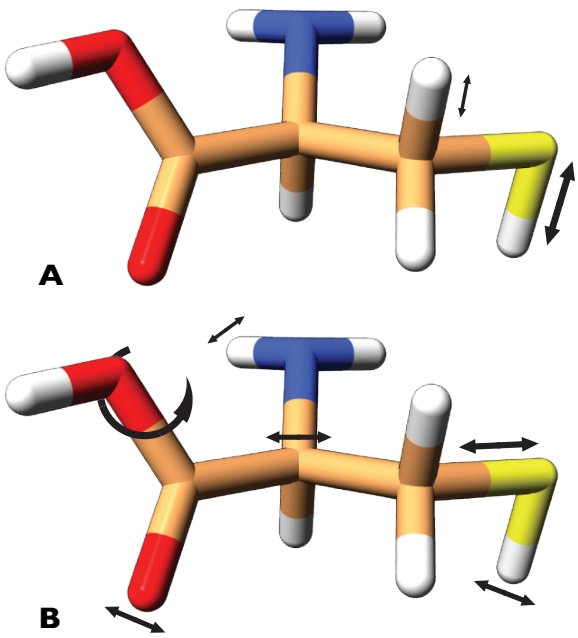


Figure 4.3, visualization of normal modes. A is the 2560 cm^{-1} vibration, B is the 670 cm^{-1} mode

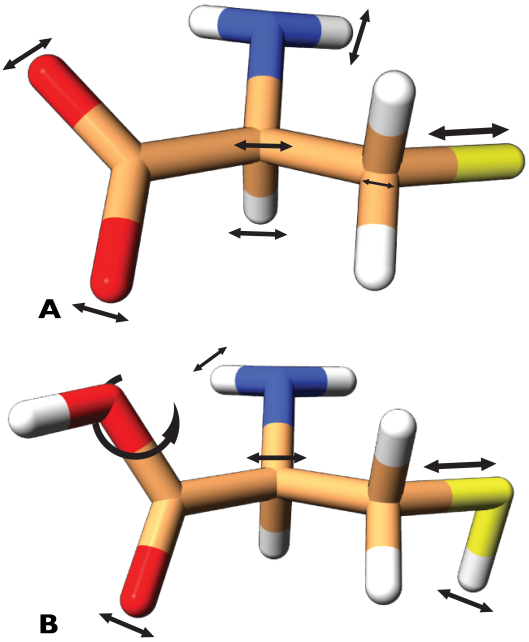


Figure 4.4, vibrational modes of Cys near 670 cm^{-1}. The size of the arrow indicates a larger displacement, A, deprotonated thiol, B, protonated thiol

All the theoretical spectra in this thesis were produced to replicate experimental conditions, both inside the protein, and with Cys as a monomeric amino acid model compound. These spectra show the same 2560 cm^{-1} band, which is, unsurprisingly, from the same vibrational mode. One might argue that although these theoretical spectra agree with measured ones, the theory could be misleading about the identity of the peak at 2560 cm^{-1}. It is also entirely possible that the peak at 2560 cm^{-1} isn't the Cys SH stretch at all, and the disappearance of the 2560 cm^{-1} band was from the protonation of a different amino acid that changes when the adduct form.

These arguments can be easily answered. By using the free amino acid cysteine as a model compound, we examined vibrational peaks without the confounding effect of the holoprotein. When the vibrational bands of Cys in water were compared to Cys in a protein interior-like environment (table 4.1), most differences were only a few cm^{-1}, and all the bands were present in both the modeled water dialectric and at

dialectric of 2.2. The results of the quantum calculations affirm the identity of the peak at 2560 cm^{-1} in LOV domains as a SH stretch (figure 4.3A).

We can also use another peak in the Cys Raman spectrum to test how well the theory agrees with reality. Earlier literature on Cys Raman spectroscopy [58] identified a peak at 670 wavenumber as a SC stretching mode. It was expected that upon deprotonation of the SH the oscillation would change, and be red shifted, with the assumption that the SH would vibrate as a single mass, and the thiolate must vibrate with less energy. But, when titrating in base, the 670 cm^{-1} band was shifted the opposite way (figure 3.2). Further confusion was added with the D_2O/H_2O comparison; D_2O also shifted the oscillation to the red (figure 3.3). Examination of the mode at 670 cm^{-1} (figure 4.4) shows why deprotonation affects the frequency. Change in sulfur protonation changes the mode itself; the simple model of the thiolate S- acting as merely a lighter version of the SH is too limited. Because the vibration encompasses much of the molecule, the mode is not simply a SC stretch. That the computational model both predicts the unexpected shift, as well as explains it helps justify the use of the model to interpret other results.

Based on the consistency of the computational results when compared to physical spectra, it can be concluded that Cys in water is a reasonable model to examine the vibrational bands of the LOV protein interior. Further directions include recalculations with more amino acids in the flavin binding pocket, and recalculations of Cys in 2.2 dialectric, since the presence of several imaginary (non-real) vibrational

modes may indicate the calculation converged on a local minimum, rather than a global one. The best direction to take quantum calculations of LOV proteins is to model the reaction coordinates between the adduct and the dark state, as no good studies have done this.

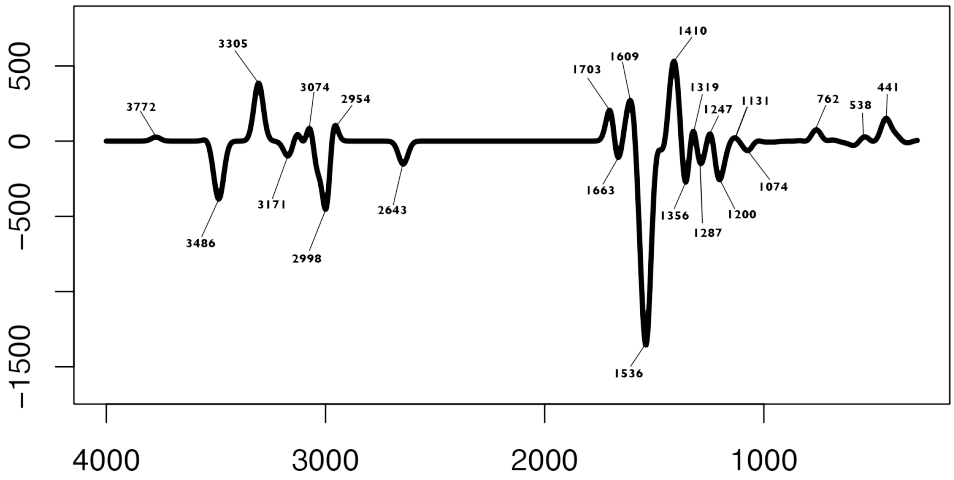


Figure 4.6, theoretical Raman light minus dark difference spectrum

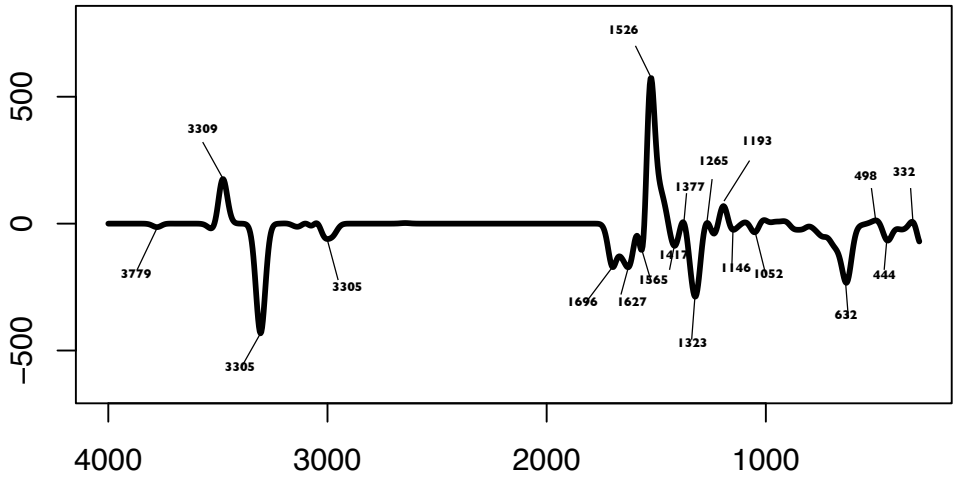


Figure 4.6 theoretical light minus dark infrared difference spectrum

Section 4.4

Summary of chapter 4

Chapter four explains the results of chapter three, by showing how some of the normal modes of cysteine change with protonation state. Chapter four also presents theoretical Raman and Infrared spectra of a stripped down LOV active site containing flavin, Cys, and a water coordinated to N5

Table 4.1, table of calculated Cys vibrations calculated at different COSMO dialectric (top row) numbered in order of energy, the numbering doesn't add up correctly because of imaginary modes that appeared in the 2.2 dialectric calculation

2.2		80	
mode	freq	mode	freq
0	0	0	0
1	0	1	0
2	0	2	0
3	0	3	0
4	0	4	0
5	0	5	0
6	-3593.5	6	29.47
7	-3488.33	7	98.04
8	-193.53	8	161.73
9	-88.84	9	204.9
10	64.36	10	245.58
11	94.62	11	257.53
12	186.73	12	307.5
13	264.05	13	395.6
14	282.14	14	517.27
15	307.14	15	572.18
16	428.13	16	604.37
17	482.69	17	671.94
18	516.78	18	743.26
19	584.38	19	774.36
20	627.63	20	832.99
21	679.65	21	885.64
22	729.92	22	910.41
23	781.1	23	1017.42
24	789.5	24	1083.08
25	832.1	25	1123.08
26	940.14	26	1174.6
27	997.84	27	1247.55
28	1111.7	28	1256.38
29	1145.62	29	1326.06
30	1233.33	30	1327.72
31	1254.98	31	1361.07
32	1343.89	32	1436.25
33	1604.49	33	1606.53
34	1686.69	34	1692.5
35	2559.59	35	2644.97
36	2985.69	36	2948.14
37	3323.92	37	3040.39
38	3459.07	38	3104.78
39	3551.74	39	3401.96

Chapter 5

Section 5.1

a review of hypotheses of adduct formation and breakage, and how the data in this thesis fits with these hypotheses

As described in the introduction to this thesis, LOV domains are important blue-light photoreceptors. Blue light fuels photochemistry that ultimately produces a metastable bond from the flavin prosthetic group to a nearby cysteine. But the hypothesis describing the manner in which LOV proteins turn photons into this adduct is contested. It is certain that a triplet flavin precedes adduct formation[38,77], but how does the forward reaction proceed from the triplet flavin state to the S-C4a adduct? (figure 1.5) Flavin most likely passes through an unidentified intermediate state before forming the adduct. This thesis attempts to reveal what this intermediate is. The ionic mechanism (figure 1.6) states that the intermediate has a positively charged C4a, whereas the radical mechanism (figure 1.7) claims that flavin becomes a radical. These two mechanisms have been debated for nearly 2 decades[46, 47].

The ionic mechanism (figure 1.6) proposes that the reaction proceeds by a nucleophilic Cys sulfur attacking an electrophilic C4a. N5 protonation could occur contemporaneously, or be prefaced by N5 protonation. The radical mechanism (figure 1.7) depends on the transfer of either a radical electron or radical hydrogen atom from the Cys thiol to flavin. Radical recombination and proton transfer follow

this, if the proton has not already been transferred along with the single electron (as a hydrogen atom transfer).

So what are the published data that support or detract from these hypotheses? The three main methods used to examine the short-lived intermediate(s) are infrared spectroscopy (IR), time resolved UV-visible absorption spectroscopy, and quantum chemistry. IR and computational chemistry are more information dense, but IR is a time consuming method, and quantum chemistry does not necessarily provide an accurate description of physical systems. Let us examine the quantum chemistry and the UV-vis absorption data before discussing the more informative IR.

The flavin radical unfortunately has a similar UV-vis absorption spectrum[29] as the triplet state, and this makes differentiating them quite difficult. This problem is evident in a recent study [6] that claimed to have evidence for the flavin radical in an absorption spectrum. The title of the paper calls this spectrum an indication, which it may be, but it certainly isn't proof. This paper provided the most evidence of a radical ever seen in a UV-vis spectrum. Their data suggests that the radical is either not important for the LOV photocycle, or is so transient that it is undetectable.

Quantum chemistry calculations[2, 55, 26, 25] have been used to argue that the ionic mechanism is false for many years. These studies have come to a bewildering variety of conclusions. For example, Domratcheva[25] claimed that the formation of the adduct in the radical mechanism was dependent on the proper orientation of the molecular orbitals of the Cys-S and the flavin. However, mutation of the Cys to other

places, even on the other side of the flavin still results in adduct formation[55, 24]. Furthermore, Domratcheva[25] utilized calculation methods that inadequately model bond breakage, according to another LOV quantum chemistry researcher[25]. Another common problem was the theoretical treatment of the thiolate. Bare negative charges are well known to be poorly treated by DFT, especially when modeling in the gas phase, as was done in by Alexandre in 2006[2].

A very recent paper [15], published July 10, 2017, in Angwandte Chemie used high level theoretical methods to claim adduct formation proceeds via radical proton transfer to flavin, followed immediately by SC bond formation without an energy barrier. Chang et al [15] also proposed an explanation for why the hypothetical intersystem crossing required during radical recombination occurs; their analysis found the energy of the singlet adduct was very close to the radical, which they argue allows for the intersystem crossing. Chang [15] argues that the protonated flavin exists for an incredibly short period of time, which is a direct contradiction of a recent study [80] that shows a protonated triplet flavin, which was also supposedly non-radical.

The recent study by Thoing[80] that contradicts Chang[15] 2017, used IR to make a strong case that the adduct is formed from a N5 protonated triplet state of flavin. The protonation of N5 does raise some interesting ideas. First, let us examine the relevance of N5 protonation before adduct to disprove the ionic hypothesis. One can imagine N5 protonated before the thiolate's nucleophilic attack of C4a. This point was mentioned by Thoing in 2013,[80] but for unclear reasons the same paper

concluded that the ionic mechanism was false. The protonation of N5 also raises a big problem with the argument that the flavin binding pocket (figure 5.1) is too hydrophobic to hold an ionized Cys. If a cation can exist, why not an anion? Both of these points were used to attempt to disprove the ionic mechanism. But, as mentioned earlier in this paragraph, N5 protonation only rules out one ordering of the steps in the ionic mechanism. The other point, the claim that the Cys must be protonated, is contested by data presented in this thesis.

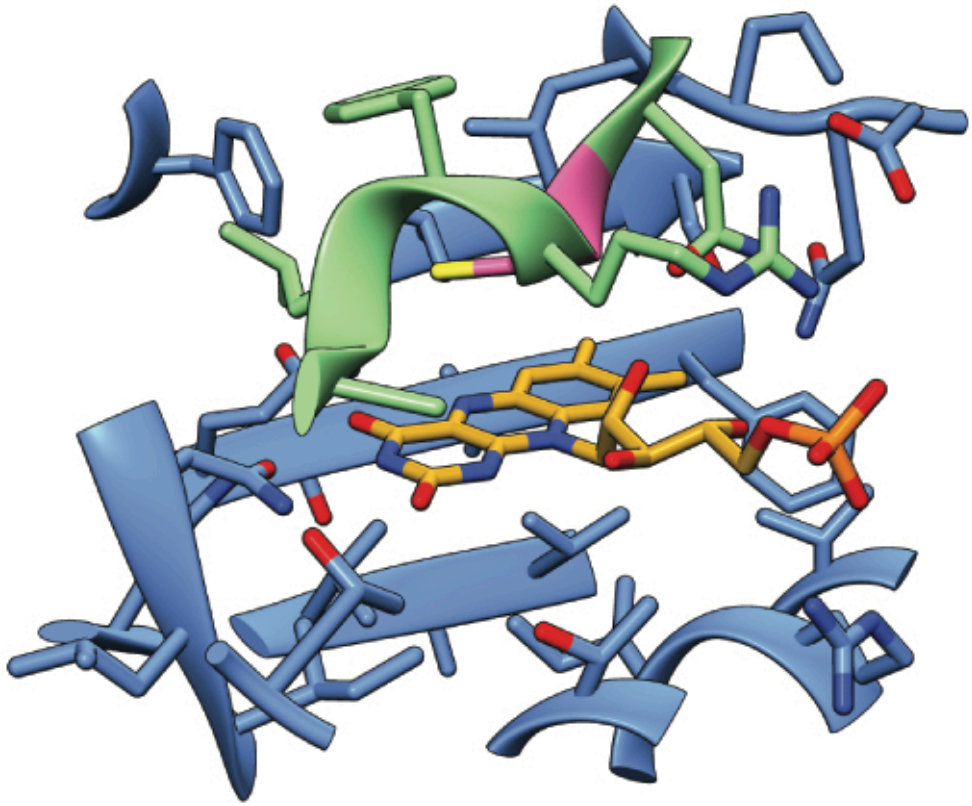


Figure 5.1, LOV flavin binding pocket of El222. FMN is held against the PAS â sheet, the LOV consensus sequence. NCRFL is shown in green, except for the active Cys, which is pink

Another question is if the N5 protonated triplet state is stable enough to be detected, what does the rest of the flavin look like? It can't be a radical, since the lifetime of the protonated triplet is long enough for a radical to be detected by EPR[8, 70]. And again, the relatively long lifetime of N5 protonated flavin apparently contradicts Chang's 2017[15] claim that N5 is protonated extraordinarily rapidly by a radical hydrogen atom and adduct formation proceeds from N5 protonation with no energy barrier. Thoing, in their 2013 paper[80], also ruled out the N5H flavin as being an anion, based on vibrational modes. The non-radical triplet N5H state flavin would be either hydroquinone, or cationic (figure 1.2). If protonated triplet flavin is a hydroquinone (figure 1.2), this would involve odd proton transfers from and to N1 at other parts of the photocycle. A much simpler explanation is that the protonation of N5 produces cationic C4a, as implicated in the ionic mechanism(figure 1.6). Cationic states of other atoms seem very improbable.

Several infrared studies[39, 33, 68, 34] have claimed that the Cys must be protonated immediately before adduct formation. We now address the protonation state of the active Cys directly. The role of water is undeniable in the reverse reaction[13]. However, other groups have argued[34] inexplicably that the flavin binding pocket is far too inhospitable to water for it to exist often enough to be a base capable of deprotonating the catalytic Cys thiol. This argument continues to state that since there are no appropriate amino acids nearby, the SH must always exist as a thiol. Visual inspection of the flavin binding pocket (figure 5.1) shows that although the most dominant

amino acid population consists of hydrophobic residues, there are polar amino acids, as well as polar functional groups on flavin itself. Even in the active site shown in figure 5.1, which is one of the least polar of all known[27] LOV active sites, polar amino contacts still exist. Furthermore, a previous molecular dynamics study[27] even reported the occasional presence of water over a simulated microsecond. This study was focused on signal transduction after adduct formation, not on adduct formation and breakage hypotheses, so it didn't explore further the energetics or dynamics of the observed water.

In chapter two of this thesis, the role of water in the lifetime of the adduct was examined. It was shown that the presence of a key stabilizing residue in the short LOV protein ENVOY holds water next to N5 (figure 2.3). Molecular dynamics over hundreds of nanoseconds of simulation time show that water readily diffuses from the bulk solvent into the flavin binding pocket. This data suggests that H_2O is able to exist indefinitely within the flavin binding pocket. By using the radial distribution function of water about N5 (figure 2.3 B), we calculated the Helmoholtz (figure 2.3 C) free energy of binding for the slow cycling mutant and wild type. The mutant had a longer adduct lifetime, which was explained by its lessened ability to stabilize water in the key position where it could deprotonate N5 and trigger SC bond breakage. This is seen in the higher energy barrier to entry for water in the Ile mutant.

The data described in the previous paragraph makes a convincing case that water could diffuse into the LOV active site, and that it plays a key role in the

photocycle. Once inside, there is no reason it couldn't deprotonate Cys, potentially immediately after protonating N5. The question then is, can LOV proteins undergo photochemistry with a deprotonated catalytic Cys?

A previous graduate student in the Bogomolni lab generated a Raman difference spectrum[18] of LOV2 that shows photocycling, but with the SH stretch at 2560 cm^{-1} absent. These data deflate the argument that the Cys must exist as thiol, and never as thiolate. If a thiol was present, the 2560 cm^{-1} would be visible. It was present in many other spectra produced on the same equipment, so the absence of the 2560 cm^{-1} band cannot be attributed to instrument error.

The peak identification of 2560 cm^{-1} was previously based partly on non-protein studies[68,34], so we wanted to confirm its identity. We also hoped to use another Cys vibrational band, previously reported to be a SC stretch[39], at 670 cm^{-1} (figures 3.2 and 3.3). Raman of acid/base titration of monomeric free Cys amino acid as a model for protein interior presented in thesis show that the SH stretch at 2560 cm^{-1} wavenumber is very useful in the examination of the protonation state of the catalytic Cys in LOV domains. We additionally show that the 670 cm^{-1} band is useless for the same purpose. The 670 cm^{-1} mode changes when protonated (figure 3.3) in that the band's center is shifted, but not in the expected direction. The thiolate mode has more energy than the protonated CSH. Whereas, the degree of protonation changes the intensity of the 2560 cm^{-1} band, but not its center (figure 3.2).

We used density functional theory to model a stripped down version of the active site. Although convergence issues necessitated the removal of all the amino acids except the catalytic Cys, water was very stable when hydrogen bonded to both protonated and deprotonated N5. Converged adduct and dark state electron structures and geometries were obtained, and from these, vibrational modes and spectra were calculated.

Similar calculations were carried out on free Cys. By examining the visualized modes of cysteine (figures 4.3 and 4.4), the odd titration behavior of the 670 cm^{-1} and 2560 cm^{-1} bands (figures 3.2, 3.3) when titrated with hydroxide can be explained. The 2560 cm^{-1} band is simple, and cannot exist without the proton, whereas the 670 cm^{-1} band can accommodate a lack of H, but the mode itself changes when the proton is absent.

While a few different mechanisms could be produced to explain how the light induced changes were indeed due to adduct formation, by using the difference spectra from the theoretical calculations in this thesis we can make a good case that the sample was indeed undergoing the typical LOV photochemistry. An overlay of the theoretical data with the actual data is presented in figure 5.2. The lack of the SH stretch 2560 cm^{-1} in the actual Raman light minus dark spectrum, and its presence in the theoretical one is the key feature.

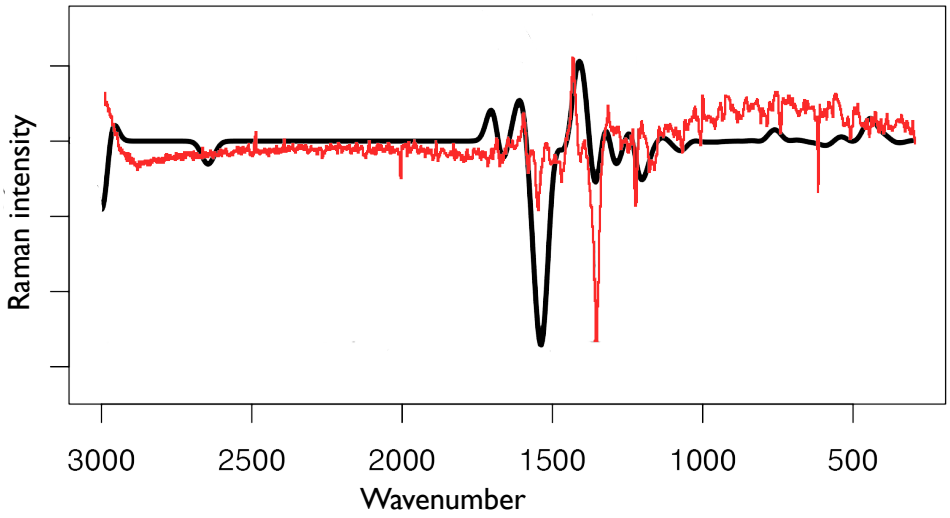


Figure 5.2, Theoretical LOV Raman of stripped down active site, in black, actual Raman light minus dark difference spectrum in red. Scale on the Y axis is not shown, since the scale of the two spectra are different

This thesis presents several key points relating to the active site of LOV domains. The first point is that water is quite stable within the active site. Molecular dynamics simulations and quantum chemistry show water latching onto N5 in both protonated and deprotonated N5 states. This shows the flavin binding pocket is not, in fact, hydrophobic. The second point is that the catalytic Cys residue does not need to exist entirely as a protonated thiol. Experiments done in the Bogomolni lab using Raman spectroscopy of Cys combined with quantum chemistry make a good case demonstrating the lack of a Cys thiol vibrational band.

These points can be combined into a coherent hypothesis: Water is a key part of the LOV photocycle, acting both to stabilize the Cys thiolate anion in the forward reaction, and to remove the N5H proton to trigger the reverse reaction.

The data in this thesis certainly disprove the argument that the LOV flavin binding site is too hydrophobic for even water, let alone a thiolate. But, this doesn't prove conclusively that the ionic mechanism is true. In fact, there is no way to prove that any hypothesis is true; all science can do is disprove explanations. So, let us explore ways in which the two adduct formation hypotheses can be disproven.

Section 5.2

Future directions to examine the LOV photo-adduct formation and breakage reaction mechanisms

A relevant concept to this thesis is that protonation state refers to only that of a single protein. The Raman data presented in this thesis is from a macroscopic number of proteins. For any given conjugate acid/base pair, at 3 pH points higher than pKa, there would be 1 protonated per 1,000 deprotonated. Thus, even when the 2560 cm^{-1} band is well below the noise, there are potentially more than a trillion active Cys residues that are in the thiol from. This brings the question as to whether the photocycle is affected by pH of the solvent. The so-called small base effect has been known for a decade; a small base such as imidizole added to solution increases the rate of adduct formation[88,46,3,2]. However, the increase isn't drastic enough to

prove a reaction mechanism. In fact, the pH of the bulk solvent probably doesn't have a huge effect on the local electrostatics, which surely govern the pK_b of any base inside the protein. The small base effect has been hypothesized[90] to be due to the base lodging along the ribityl group of the flavin. This either changes local electrostatics, or serves to aid in the transfer of a proton away from the active site.

The question as to whether enough of the protein was in a state to form the adduct is basically moot regarding the Raman spectra discussed in this document due to the experimental setup. Since a laser must be applied to the sample for more than a few seconds, essentially all of the protein is put into the light state before the light spectra were taken. Even if only 0.1% of the protein was in a productive state at any given nanosecond, by the time 1 second has passed an appreciable amount of adduct will have formed. This argument cuts both ways: the presence or absence of the SH vibration in a difference vibrational spectrum does not mean that enough Cys wasn't in the correct protonation state to form adduct, regardless of what that protonation state is. Basically, vibrational spectroscopy involving macroscopic quantities of sample isn't quite precise enough to conclusively disprove hypotheses about protonation states.

So, what kinds of evidence could be hypothetically collected that disproves the radical mechanism? Lets start with less convincing, but easier to accomplish experiments. The most obvious continuation of the work presented in this thesis would be more quantum chemistry. The end and starting points of the reaction

coordinates are already complete, and calculating the energy barriers along the ionic and radical mechanisms would be interesting. However, even if one pathway is across a lower energy path, it doesn't mean that the reaction proceeds that way. These calculations often assume that a radical exists already, and the formation of the radical probably involves a higher energy state energy than the barrier to forming the adduct from the radical pair. Also, computations rely on a number of approximations; not least among them is the use of less than the entire protein.

A major role of protein structure is the maintenance of the correct orientation of reactive species for catalysis. Therefore, better than using a stripped down active site as this thesis describes, mixed QM/MM with most of the protein modeled classically, but with the flavin, ordered water, and catalytic Cys modeled quantum mechanically, would provide better data. Of course, these types of calculations are expensive computationally. Since the photocycle continues even when the Cys is mutated into different positions [87], the protein structure is probably more important to prevent the flavin from diffusing out than to maintain perfect alignment of the catalytic groups. Considering the strong evidence that N5 is protonated before the adduct forms, any future calculations should include N5 protonated flavin along with the deprotonated version used in previously published studies. Examination of the electron density around C4a and N5 in protonated triplet flavin could answer the question regarding the degree of positive charge on C4a in the N5 protonated species, although this has been done to some extent already.[2]

A major problem with most wet-lab experiments is that they involve macroscopic numbers of molecules. There are several reasons it is hard to reduce the number of molecules in an experiment. The Raman effect in particular is an inefficient process; the number of scattered photons is much less than the incident radiance; and Rayleigh scattered photons are far more numerous than inelastically scattered ones. This also puts a limitation on the shortness of a Raman experiment. Since a certain number of photons have to hit the detector for a reasonable signal to noise ratio to be achieved, long time averages must be taken. This results in a data that is an average across a population as opposed to an examination of what is occurring at the level of an individual molecule.

Identifying short-lived intermediates is impossible if they don't build up in sufficient quantity for the measurement apparatus to detect. Single molecule experiments can address this, but no suitable single molecule method exists to disprove either adduct formation hypothesis. But, there are tricks to increase the number of Raman scattered photons. Normal resonance enhancements are not useful, since fluorescence of the flavin begins to dominate the signal. Fluorescence also prevents the use of UV lasers for Raman, as UV light can excite flavin into electronic states above the first singlet, which subsequently undergo fluorescence. Although at laser wavelengths in what used to be called pre-resonance frequencies fluorescence isn't a big problem, the chromophore Raman bands are not of particular interest, and those are the ones that would be enhanced by resonance of the laser to the electronic

excitation bands. Surface Enhanced Raman Scattering (SERS) relies on surface plasmons[57] to enhance the number of Raman scattered photons. By binding a small number of LOV domains onto an appropriate nano-particle it might be possible to determine whether the reactive Cys residue is protonated based on vibrational bands. Unfortunately we don't expect SERS to allow for significantly shorter experiments, which is why it wasn't pursued further in this thesis.

The radical mechanism, of course, involves radical electrons. These are typically extremely reactive, which is the reason given by proponents of said mechanism that they haven't been detected. There are 2 common ways to test for radicals: scavengers, and Electron Parmagnetic Resonance (EPR). To our knowledge, radical scavengers haven't been utilized in an attempt to quench the reaction. The main problem with a scavenger experiment is that a null result doesn't disprove the radical mechanism, therefore it is too easy to explain the lack of effect as being due to the scavenger not being able to penetrate to the interior of the protein. The scavenger may also perturb the system in uncontrolled ways, and may reversibly inhibit the protein, so even a positive result might not be due to actual radical scavenging. Reversible inhibition that wasn't due to actual scavenging would be difficult, if not entirely impossible to differentiate from scavenging activity.

The definitive experiment that would disprove the false mechanism is EPR. EPR is based on the same phenomenon that allows for Nuclear Magnetic Resonance spectroscopy. Basically, a particles spin can be affected by a magnetic field. By

inducing a split in the degenerate spin states of an electron, a microwave absorbance spectrum can be created. There would be no splitting if the electron spin was paired, so this technique would detect radical electrons. These experiments have been attempted already, and have not detected any radicals. But, these experiments rely on time averages, and the hypothesized LOV radical could exist for far too short a time to be detected with this method. This limitation is purely an instrumental one, and the time resolution of EPR apparatus has recently improved. Therefore, in the future it is very likely that EPR will be used to finally disprove either the radical or the ionic mechanism.

1. Alexandre, M. T.; Arents, J. C.; van Grondelle, R.; Hellingwerf, K. J.; Kennis, J. T., A base-catalyzed mechanism for dark state recovery in the Avena sativa phototropin-1 LOV2 domain. *Biochemistry* **2007,** *46* (11), 3129-3137.

2. Alexandre, M. T.; Domratcheva, T.; Bonetti, C.; van Wilderen, L. J.; van Grondelle, R.; Groot, M.-L.; Hellingwerf, K. J.; Kennis, J. T., Primary reactions of the LOV2 domain of phototropin studied with ultrafast mid-infrared spectroscopy and quantum chemistry. *Biophysical journal* **2009,** *97* (1), 227-237.

3. Alexandre, M. T. A.; Arents, J. C.; van Grondelle, R.; Hellingwerf, K. J.; Kennis, J. T. M., A Base-Catalyzed Mechanism for Dark State Recovery in the Avena sativa Phototropin-1 LOV2 Domain. *Biochemistry* **2007,** *46* (11), 3129-3137.

4. Allouche, A. R., Gabedit—a graphical user interface for computational chemistry softwares. *Journal of computational chemistry* **2011,** *32* (1), 174-182.

5. Ataka, K.; Hegemann, P.; Heberle, J., Vibrational spectroscopy of an algal Phot-LOV1 domain probes the molecular changes associated with blue-light reception. *Biophysical journal* **2003,** *84* (1), 466-474.

6. Bauer, C.; Rabl, C. R.; Heberle, J.; Kottke, T., Indication for a radical intermediate preceding the signaling state in the LOV domain photocycle. *Photochemistry and photobiology* **2011,** *87* (3), 548-553.

7. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995,** *91* (1–3), 43-56.

8. Bittl, R.; Kay, C. W.; Weber, S.; Hegemann, P., Characterization of a flavin radical product in a C57M mutant of a LOV1 domain by electron paramagnetic resonance. *Biochemistry* **2003,** *42* (28), 8506-8512.

9. Bocola, M.; Schwaneberg, U.; Jaeger, K.-E.; Krauss, U., Light-induced structural changes in a short light, oxygen, voltage (LOV) protein revealed by molecular dynamics simulations—implications for the understanding of LOV photoactivation. *Frontiers in molecular biosciences* **2015,** *2*.

10. Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., CHARMM: the biomolecular simulation program. *Journal of computational chemistry* **2009,** *30* (10), 1545-1614.

11. Carr, D., Binning and plotting functions for hexagonal bins. Now uses and relies on grid graphics and formal (S4) classes and methods. 2015.

12. Chan, R. H.; Bogomolni, R. A., Structural water cluster as a possible proton acceptor in the adduct decay reaction of oat phototropin 1 LOV2 domain. *The Journal of Physical Chemistry B* **2012,** *116* (35), 10609-10616.

13. Chan, R. H.; Lewis, J. W.; Bogomolni, R. A., Photocycle of the LOV‐STAS Protein from the Pathogen Listeria monocytogenes. *Photochemistry and photobiology* **2013,** *89* (2), 361-369.

14. Chandler, D., *Introduction to Modern Statistical Mechanics*. Oxford University Press, Inc: 189 Madison Ave, New York, New York 10016-4314, 1987.

15. Chang, X.-P.; Gao, Y.-J.; Fang, W.-H.; Cui, G.; Thiel, W., Quantum Mechanics/Molecular Mechanics Study on the Photoreactions of Dark- and Light-Adapted States of a Blue-Light YtvA LOV Photoreceptor. *Angewandte Chemie* **2017**, n/a-n/a.

16. Chapman, S.; Faulkner, C.; Kaiserli, E.; Garcia-Mata, C.; Savenkov, E. I.; Roberts, A. G.; Oparka, K. J.; Christie, J. M., The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences* **2008,** *105* (50), 20038-20043.

17. Christie, J. M.; Corchnoy, S. B.; Swartz, T. E.; Hokenson, M.; Han, I.-S.; Briggs, W. R.; Bogomolni, R. A., Steric interactions stabilize the signaling state of the LOV2 domain of phototropin 1. *Biochemistry* **2007,** *46* (32), 9310-9319.

18. Ciorciari, J. M. Vibrational spectroscopy of the LOV2 protein domain from the plant blue-light receptor phototropin 1 : light-induced structural changes. University of California, Santa Cruz, 2004.

19. Corchnoy, S. B.; Swartz, T. E.; Lewis, J. W.; Szundi, I.; Briggs, W. R.; Bogomolni, R. A., Intramolecular proton transfers and structural changes during the photocycle of the LOV2 domain of phototropin 1. *Journal of Biological Chemistry* **2003,** *278* (2), 724-731.

20. Cramer, C. J., *Essentials of computational chemistry: theories and models*. John Wiley & Sons: 2013.

21. Crosson, S.; Rajagopal, S.; Moffat, K., The LOV Domain Family: Photoresponsive Signaling Modules Coupled to Diverse Output Domains. *Biochemistry* **2003,** *42* (1), 2-10.

22. Deerinck, T.; Bushong, E.; Lev-Ram, V.; Shu, X.; Tsien, R.; Ellisman, M., Enhancing serial block-face scanning electron microscopy to enable high resolution

3-D nanohistology of cells and tissues. *Microsc. Microanal* **2010,** *16* (S2), 1138-1139.

23. DeLano, W., The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific; 2002. Accessed 6/25/2007. Available at http://www. pymol, org: 2010.

24. Dittrich, M.; Freddolino, P. L.; Schulten, K., When light falls in LOV: A QM/MM study of photoexcitation in Phot-LOV1 of C. Reinhardtii. *The journal of physical chemistry. B* **2005,** *109* (26), 13006.

25. Domratcheva, T.; Fedorov, R.; Schlichting, I., Analysis of the primary photocycle reactions occurring in the light, oxygen, and voltage blue-light receptor by multiconfigurational quantum-chemical methods. *Journal of chemical theory and computation* **2006,** *2* (6), 1565-1574.

26. Fedorov, R.; Schlichting, I.; Hartmann, E.; Domratcheva, T.; Fuhrmann, M.; Hegemann, P., Crystal Structures and Molecular Mechanism of a Light-Induced Signaling Switch: The Phot-LOV1 Domain from Chlamydomonas reinhardtii. *Biophysical Journal* **2003**
, *84* (4), 2474-2482.

27. Freddolino, P. L.; Gardner, K. H.; Schulten, K., Signaling mechanisms of LOV domains: new insights from molecular dynamics studies. *Photochemical & Photobiological Sciences* **2013,** *12* (7), 1158-1170.

28. Glantz, S. T.; Carpenter, E. J.; Melkonian, M.; Gardner, K. H.; Boyden, E. S.; Wong, G. K.-S.; Chow, B. Y., Functional and topological diversity of LOV domain photoreceptors. *Proceedings of the National Academy of Sciences* **2016**, 201509428.

29. Heelis, P., The photophysical and photochemical properties of flavins (isoalloxazines). *Chemical Society Reviews* **1982,** *11* (1), 15-39.
30. Herrou, J.; Crosson, S., Function, structure and mechanism of bacterial photosensory LOV proteins. *Nature reviews microbiology* **2011,** *9* (10), 713-723.

31. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (3), 435-447.

32. Holzer, W.; Penzkofer, A.; Hegemann, P., Photopysical and photochemical excitation and relaxation dynamics of LOV domains of phot from Chlamydomonas reinhardtii. *Journal of luminescence* **2005,** *112* (1), 444-448.

33. Iwata, T.; Nozaki, D.; Tokutomi, S.; Kagawa, T.; Wada, M.; Kandori, H., Light-induced structural changes in the LOV2 domain of Adiantum phytochrome3 studied by low-temperature FTIR and UV− visible spectroscopy. *Biochemistry* **2003,** *42* (27), 8183-8191.

34. Iwata, T.; Tokutomi, S.; Kandori, H., Photoreaction of the cysteine S− H group in the LOV2 domain of Adiantum phytochrome3. *Journal of the American Chemical Society* **2002,** *124* (40), 11840-11841.

35. Kaley, K.; Pudasaini, A.; Zoltowski, B. D., Short LOV proteins in methylocystis reveal insight into LOV domain photocycle mechanisms. *PloS one* **2015,** *10* (5), e0124874.

36. Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A., Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *Journal of Computational Chemistry* **2002,** *23* (16), 1515-1531.

37. Kawano, F.; Aono, Y.; Suzuki, H.; Sato, M., Fluorescence imaging-based high-throughput screening of fast-and slow-cycling LOV proteins. *PLoS One* **2013,** *8* (12), e82693.

38. Kennis, J. T. M.; Crosson, S.; Gauden, M.; van Stokkum, I. H. M.; Moffat, K.; van Grondelle, R., Primary Reactions of the LOV2 Domain of Phototropin, a Plant Blue-Light Photoreceptor. *Biochemistry* **2003,** *42* (12), 3385-3392.

39. Kikuchi, S.; Unno, M.; Zikihara, K.; Tokutomi, S.; Yamauchi, S., Vibrational Assignment of the Flavin− Cysteinyl Adduct in a Signaling State of the LOV Domain in FKF1. *The Journal of Physical Chemistry B* **2009,** *113* (9), 2913-2921.

40. Klamt, A.; Schüürmann, G., COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2* **1993,** (5), 799-805.

41. Krauss, U.; Minh, B. Q.; Losi, A.; Gärtner, W.; Eggert, T.; von Haeseler, A.; Jaeger, K.-E., Distribution and Phylogeny of Light-Oxygen-Voltage-Blue-Light-Signaling Proteins in the Three Kingdoms of Life. *Journal of Bacteriology* **2009,** *191* (23), 7234-7242.

42. Lindahl, E.; Hess, B.; van der Spoel, D., GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* **2001,** *7* (8), 306-317.

43. Lokhandwala, J.; Hopkins, H. C.; Britton, C. W.; Rodriguez-Iglesias, A.; Bogomolni, R.; Schmoll, M.; Zoltowski, B. D., A Native Threonine Coordinates Ordered Water to Tune Light-Oxygen-Voltage (LOV) Domain Photocycle Kinetics

and Osmotic Stress Signaling in Trichoderma reesei ENVOY. *Journal of Biological Chemistry* **2016,** *291* (28), 14839-14850.

44. Lokhandwala, J.; Hopkins, H. C.; Rodriguez-Iglesias, A.; Dattenböck, C.; Schmoll, M.; Zoltowski, B. D., Structural biochemistry of a fungal LOV domain photoreceptor reveals an evolutionarily conserved pathway integrating light and oxidative stress. *Structure* **2015,** *23* (1), 116-125.

45. Long, D. A., Quantum Mechanical Theory of Rayleigh and Raman Scattering. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules* **2002**, 49-84.

46. Losi, A.; Gärtner, W., Solving Blue Light Riddles: New Lessons from Flavin-binding LOV Photoreceptors. *Photochemistry and Photobiology* **2017,** *93* (1), 141-158.

47. Losi, A.; Mandalari, C.; Gärtner, W., The Evolution and Functional Role of Flavin-based Prokaryotic Photoreceptors. *Photochemistry and Photobiology* **2015,** *91* (5), 1021-1031.

48. Ma, J.; Pazos, I. M.; Zhang, W.; Culik, R. M.; Gai, F., Site-Specific Infrared Probes of Proteins. *Annual review of physical chemistry* **2015,** *66*, 357-377.

49. Mathes, T.; Götze, J. P., A proposal for a dipole-generated BLUF domain mechanism. *Frontiers in Molecular Biosciences* **2015,** *2*, 62.

50. McQuarrie, D. A., *Quantum chemistry*. University Science Books: 2008.

51. Mei, Q.; Dvornyk, V., Evolution of PAS domains and PAS-containing genes in eukaryotes. *Chromosoma* **2014,** *123* (4), 385-405.

52. Nash, A. I.; Ko, W.-H.; Harper, S. M.; Gardner, K. H., A conserved glutamineplays a central role in LOV domain signal transmission and its duration. *Biochemistry* **2008,** *47* (52), 13842-13849.

53. Nash, A. I.; McNulty, R.; Shillito, M. E.; Swartz, T. E.; Bogomolni, R. A.; Luecke, H.; Gardner, K. H., Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proceedings of the National Academy of Sciences* **2011,** *108* (23), 9449-9454.

54. Neese, F., The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012,** *2* (1), 73-78.

55. Neiß, C.; Saalfrank, P., Ab Initio Quantum Chemical Investigation of the First Steps of the Photocycle of Phototropin: A Model Study¶. *Photochemistry and Photobiology* **2003,** *77* (1), 101-109.

56. Nozaki, D.; Iwata, T.; Ishikawa, T.; Todo, T.; Tokutomi, S.; Kandori, H., Role of Gln1029 in the photoactivation processes of the LOV2 domain in Adiantum phytochrome3. *Biochemistry* **2004,** *43* (26), 8373-8379.

57. Otto, A.; Mrozek, I.; Grabhorn, H.; Akemann, W., Surface-enhanced Raman scattering. *Journal of Physics: Condensed Matter* **1992,** *4* (5), 1143.

58. Pawlukojć, A.; Leciejewicz, J.; Ramirez-Cuesta, A.; Nowicka-Scheibe, J., L-cysteine: neutron spectroscopy, Raman, IR and ab initio study. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2005,** *61* (11), 2474-2481.

59. Perkins, G. A.; Jackson, D. R.; Spirou, G. A., Resolving presynaptic structure by electron tomography. *Synapse* **2015,** *69* (5), 268-282.

60. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.

61. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013,** *29* (7), 845-854.

62. Proshkina, G.; Shilova, O.; Ryabova, A.; Stremovskiy, O.; Deyev, S., A new anticancer toxin based on HER2/neu-specific DARPin and photoactive flavoprotein miniSOG. *Biochimie* **2015,** *118*, 116-122.

63. Pudasaini, A.; Kaley, K.; Zoltowski, B. D., LOV-based optogenetic devices: light-driven modules to impart photoregulated control of cellular signaling. *Frontiers in molecular biosciences* **2015,** *2*.

64. Pudasaini, A.; Zoltowski, B. D., Zeitlupe senses blue-light fluence to mediate circadian timing in Arabidopsis thaliana. *Biochemistry* **2013,** *52* (40), 7150-7158.

65. Purcell, E. B.; McDonald, C. A.; Palfey, B. A.; Crosson, S., An analysis of the solution structure and signaling mechanism of LovK, a sensor histidine kinase integrating light and redox signals. *Biochemistry* **2010,** *49* (31), 6761-6770.

66. Raffelberg, S.; Mansurova, M.; Gärtner, W.; Losi, A., Modulation of the photocycle of a LOV domain photoreceptor by the hydrogen-bonding network. *Journal of the American Chemical Society* **2011,** *133* (14), 5346-5356.

67. Ryumina, A. P.; Serebrovskaya, E. O.; Shirmanova, M. V.; Snopova, L. B.; Kuznetsova, M. M.; Turchin, I. V.; Ignatova, N. I.; Klementieva, N. V.; Fradkov, A. F.; Shakhov, B. E., Flavoprotein miniSOG as a genetically encoded photosensitizer for cancer cells. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2013,** *1830* (11), 5059-5067.

68. Sato, Y.; Iwata, T.; Tokutomi, S.; Kandori, H., Reactive cysteine is protonated in the triplet excited state of the LOV2 domain in Adiantum phytochrome3. *Journal of the American Chemical Society* **2005,** *127* (4), 1088-1089.

69. Schäfer, A.; Horn, H.; Ahlrichs, R., Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *The Journal of Chemical Physics* **1992,** *97* (4), 2571-2577.

70. Schleicher, E.; Kowalczyk, R. M.; Kay, C. W. M.; Hegemann, P.; Bacher, A.; Fischer, M.; Bittl, R.; Richter, G.; Weber, S., On the Reaction Mechanism of Adduct Formation in LOV Domains of the Plant Blue-Light Receptor Phototropin. *Journal of the American Chemical Society* **2004,** *126* (35), 11067-11076.

71. Shu, X.; Lev-Ram, V.; Deerinck, T. J.; Qi, Y.; Ramko, E. B.; Davidson, M. W.; Jin, Y.; Ellisman, M. H.; Tsien, R. Y., A Genetically Encoded Tag for Correlated Light and Electron Microscopy of Intact Cells, Tissues, and Organisms. *PLOS Biology* **2011,** *9* (4), e1001041.

72. Silva‐Junior, M. R.; Mansurova, M.; Gärtner, W.; Thiel, W., Photophysics of Structurally Modified Flavin Derivatives in the Blue‐Light Photoreceptor YtvA: A Combined Experimental and Theoretical Study. *ChemBioChem* **2013,** *14* (13), 1648-1661.

73. Song, S. H.; Dick, B.; Penzkofer, A.; Hegemann, P., Photo-reduction of flavin mononucleotide to semiquinone form in LOV domain mutants of blue-light receptor phot from Chlamydomonas reinhardtii. *Journal of Photochemistry and Photobiology B: Biology* **2007,** *87* (1), 37-48.

74. Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P., Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes. *The Journal of Chemical Physics* **2003,** *119* (23), 12129-12137.

75. Suydam, I. T.; Boxer, S. G., Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins. *Biochemistry* **2003,** *42* (41), 12050-12055.

76. Suzuki, H.; Okajima, K.; Ikeuchi, M.; Noguchi, T., LOV-like flavin-Cys adduct formation by introducing a Cys residue in the BLUF domain of TePixD. *Journal of the American Chemical Society* **2008,** *130* (39), 12884-12885.

77. Swartz, T. E.; Corchnoy, S. B.; Christie, J. M.; Lewis, J. W.; Szundi, I.; Briggs, W. R.; Bogomolni, R. A., The Photocycle of a Flavin-binding Domain of the Blue Light Photoreceptor Phototropin. *Journal of Biological Chemistry* **2001,** *276* (39), 36493-36500.

78. Swartz, T. E.; Wenzel, P. J.; Corchnoy, S. B.; Briggs, W. R.; Bogomolni, R. A., Vibration spectroscopy reveals light-induced chromophore and protein structural changes in the LOV2 domain of the plant blue-light receptor phototropin 1. *Biochemistry* **2002,** *41* (23), 7183-7189.

79. Team, R. C., R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., 2015.

80. Thoing, C.; Pfeifer, A.; Kakorin, S.; Kottke, T., Protonated triplet-excited flavin resolved by step-scan FTIR spectroscopy: implications for photosensory LOV domains. *Physical Chemistry Chemical Physics* **2013,** *15* (16), 5916-5926.

81. Vaidya, A. T.; Chen, C.-H.; Dunlap, J. C.; Loros, J. J.; Crane, B. R., Structure of a light-activated LOV protein dimer that regulates transcription in Neurospora crassa. *Science signaling* **2011,** *4* (184), ra50.

82. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005,** *26* (16), 1701-1718.

83. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005,** *7* (18), 3297-3305.

84. Xu, S.; Chisholm, A. D., Highly efficient optogenetic cell ablation in C. elegans using membrane-targeted miniSOG. *Scientific Reports* **2016,** *6*, 21271.

85. Yamada, D.; Kandori, H., FTIR Spectroscopy of Flavin-Binding Photoreceptors #. In *T Flavins and Flavoproteins*, 2014; Vol. 1146, pp 361-376.

86. Yamamoto, A.; Iwata, T.; Sato, Y.; Matsuoka, D.; Tokutomi, S.; Kandori, H., Light signal transduction pathway from flavin chromophore to the Jα helix of Arabidopsis phototropin1. *Biophysical journal* **2009,** *96* (7), 2771-2778.

87. Zayner, J. P.; Sosnick, T. R., Factors that control the chemistry of the LOV domain photocycle. *PLoS One* **2014,** *9* (1), e87074.

88. Zoltowski, B. D.; Nash, A. I.; Gardner, K. H., Variations in Protein–Flavin Hydrogen Bonding in a Light, Oxygen, Voltage Domain Produce Non-Arrhenius Kinetics of Adduct Decay. *Biochemistry* **2011,** *50* (41), 8771-8779.

89. Zoltowski, B. D.; Schwerdtfeger, C.; Widom, J.; Loros, J. J.; Bilwes, A. M.; Dunlap, J. C.; Crane, B. R., Conformational switching in the fungal light sensor Vivid. *Science* **2007,** *316* (5827), 1054-1057.

90. Zoltowski, B. D.; Vaccaro, B.; Crane, B. R., Mechanism-based tuning of a LOV domain photoreceptor. *Nat Chem Biol* **2009,** *5* (11), 827-834.