Spring 5-4-2018

Modeling and analyzing an optogenetic system for photoactivatable protein dissociation

Anvin Thomas
anvin.thomas@uconn.edu

James Schaff
schaff@uchc.edu

Honors Scholar Thesis

"Modeling and analyzing an optogenetic system for photoactivatable protein dissociation"

Anvin Thomas

Department of Molecular and Cell Biology

College of Liberal Arts and Sciences

University of Connecticut

Honors Thesis Advisor

Professor James Schaff

Department of Cell Biology

Center for Cell Analysis and Modeling

UConn Health

Honors Advisor

Dr. Johann Peter Gogarten

Department of Molecular and Cell Biology

College of Liberal Arts and Sciences

University of Connecticut

Spring 2018

Abstract

Computational modeling of cell-cell interactions can grant clues and can answer questions about an experiment, especially for observations about binding interactions and kinetics. This approach was used to investigate an interaction between a light-oxygen-voltage (LOV) domain and an engineered protein called Zdark (Zdk). The LOV domain is membrane-bound while Zdk is cytosolic. The LOV domain and Zdk bind strongly in dark (K_d 26.2 nM), and weakly upon exposure to blue light (K_d > 4 μM). Total internal reflection fluorescence (TIRF) images are acquired of Zdk, the fluorescent species bound to a mCherry tag, and the loss of fluorescence is observed upon illumination. However, secondary binding sites of transfer-messenger RNA SsrA and stringent starvation protein B SspB can impact the off rate. SsrA is an RNA binding protein that is bound to the LOV molecule, and SspB, which enhances recognition for the SsrA binding site, is bound to the Zdk peptide. The affinity between the two binding sites alters the kinetics of the dissociation of the LOV domain and Zdk. To note, the SsrA and SspB binding sites can still be connected when the LOV domain and Zdk dissociate.

Modeling of this experiment was primarily done through Virtual Cell. VCell is a free computational tool for modeling and simulation of various cell biology systems that can be used to perform and analyze simulations. This was used to mathematically model the microscopy experiment, with goals of gaining new insight to the underlying biology and evaluating quantitative analysis of the experimental data. Several important tasks include gathering theoretical values for the kinetics of the systems in terms of the on rate and off rate and looking for observations about the interaction of the secondary binding sites. This will all be used in an attempt to construct an experimental protocol utilizing this optogenetic system, especially for lower affinity binding.

Acknowledgments

I would like to first acknowledge the strong support I received from my family and friends in writing my thesis. I would not be in the position I am now without their help. I would also like to thank the University of Connecticut and the Honors Program for providing me the opportunity to conduct and write about my research for a thesis. Without the Honors Program, I would not have been able to connect with so many peers and professors.

I would also like to thank the UConn Health Research Program, for allowing me the chance to work and conduct research at UConn Health for a year. It was an honor to be able to interact and learn with some of brightest researchers I've ever met. At UConn Health, I need to thank James Schaff for being my mentor over the year and for being my thesis advisor. Without his assistance, I wouldn't have learned nearly half as much as what I did. Thanks also go to the people I've met at UConn Health for giving me advice and other resources to reference in my research.

Thanks also goes to Dr. Johann Peter Gogarten, for not only being my assigned Honors advisor but for reading this thesis and providing suggestions and comments. His guidance has been very helpful in navigating my undergraduate career.

Table of Contents

Introduction

The importance of computational modeling in biology cannot be understated. It allows researchers to delve into unexplored reactions between molecules, to see what information tweaking their experimental procedure can provide, as well as to create thought experiments to better understand their subject of research. The case here is no different; computational modeling permits us to explore a unique binding interaction between two proteins in an optogenetic system, as well as the limitations and usefulness of this system.

LOV2 Trap and Release of Protein (LOVTRAP) is an optogenetic approach to measuring a reversible, light-induced protein dissociation[1]. Optogenetic refers to the biological technique which utilizes light to control cells to behave in a specific way. Here, the aspect being controlled are the binding kinetics of a small protein, named Zdark (Zdk), and a LOV2 (light-oxygen-voltage) domain. Zdk is a protein created by mRNA screening of a library derived from the Z component of protein A. The LOV2 domain is a photo-sensing molecule from *Avena* sativa phototropin 1.

LOV2 and Zdk bind strongly in dark with a dissociation constant (K_d) of 26.2 nM. However, upon exposure to blue light (irradiation at 450-490nm), LOV2 and Zdk bind weakly and tend to dissociate (K_d > 4 μM) [1]. The affinity is over 150 times stronger in dark than that in light. LOVTRAP was developed in order to isolate a protein of interest (POI) using the aforementioned “dark state” affinity. The POI is bound to one of these, depending on which component is anchored to the cell membrane – this is up to the experimenter to decide, as plasmids of both situations can be easily created[2]. The constructs in the experiment are made with use of the TOM2 fragment, which is a mitochondrial import receptor subunit which localizes the components to the mitochondrial membrane[2]. The system used the n-termini of

LOV2 and Zdk1 to interchangeably label with either mCherry or to bind to a fixed structure such as the mitochondria or the plasma membrane (Fig. 1). A major aspect of this approach is how broadly applicable it can be. All that is needed is the POI being fused to either LOV or Zdk. This leads to precise diffusion limited activation kinetics that can be tightly controlled, with deactivation and recovery rates that can vary from seconds to minutes with the presence of mutations[1].

The method used and analyzed was this aforementioned LOVTRAP system, with LOV2 anchored to the plasma membrane while Zdk is present in the cytosol. Total internal reflection fluorescence (TIRF) microscopy was used to measure and detect for this LOVTRAP system, utilizing a trap and release mechanism for the fluorescent protein mCherry[2]. The protein of interest is bound to Zdk, where a binding site for the POI exists bound to LOV2. The purpose of this constructed system is to measure translocation kinetics between the membrane and cytosol with and without the secondary binding of the POI to characterize the affinities. The binding site of the protein of interest has two experimental affinities. There is a high affinity state (K_d = 100 nM) and a low affinity state (K_d = 800 nM). It is expected that the low affinity state will have a relatively small impact on the dark-state binding of LOV2 and Zdk but will have an affect on the release kinetics when the light is turned on. However, the light activation of the LOV domain in this reaction is independent; it does not depend on binding for this activation to occur.

The main goals of the research into this system is to estimate the in vivo binding affinities of two proteins. Protein “A” is Ssra, an RNA binding protein, while Protein “B” is Sspb, which enhances recognition of Ssra-tagged proteins[1]. This estimation is done by observing the optogenetic translocation of LOV/Zdk. The influence of this binding on the optogenetic protocol will also be observed in order to compare the bindings of Zdk alone and Zdk and protein B. This

will achieve our main goal of seeing if this LOVTRAP system is viable to measure lower affinity interactions as well as high affinity interactions. Only high affinity interactions have been researched and tested in literature thus far.

This analysis was conducted through Virtual Cell (VCell), a platform made by the Richard D. Berlin Center of Cell Analysis and Modeling for analyzing and modeling cell biological systems[3]. Several biomodels of the LOVTRAP system were constructed using this software, which were utilized for numerous simulations and calculations to allow us to further investigate the underlying kinetics. This is crucial, especially in using stochastic and deterministic approaches and applications to help solve for the parameters of the system. A self-created profile likelihood program and functionality was also added to VCell in order to help solve for these parameters.

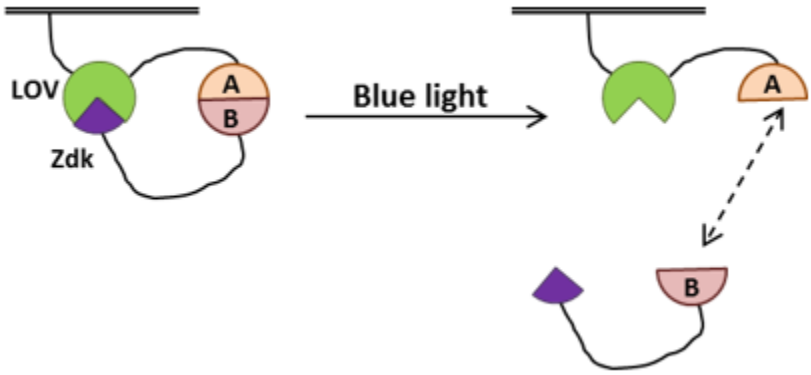


Figure 1. A diagram demonstrating the basic binding interaction of LOV and Zdk as well as the protein A and B components in response to light activation. Upon exposure to blue light, LOV and Zdk components separate due to a high dissociation constant. Secondary sites A and B can be bound or unbound after exposure of blue light, depending on whether it is strong or weak affinity.

Understanding this interaction can help better design binding mechanisms that can utilized in the field of medicine. Examples including modulating G-Protein Coupled Receptor (GPCR) signaling[4] and antitumor immunity[5]. To note, a graduate student, Abhijit Deb Roy, as part of the Center of Cell Analysis and Modeling collected this data by use of Total Internal

Reflection Fluorescence (TIRF) microscopy in the mechanism shown above. TIRF microscopy provided imaging at high resolution directly at the membrane by only exciting fluorophores in a close proximity to the coverslip. This is done by reflecting light at an incident angle greater than the critical angle onto the membrane. Data was collected from these trials, including average fluorescent intensities across the membrane, threshold areas, and time stamps. This data will be implemented into the biomodel for parameter estimation. The translocation kinetics between the membrane and cytosol in this system was better understood in respect to a mathematical model of phophoinositide turnover – the approach used there in terms of constructing a model was also used here[6].

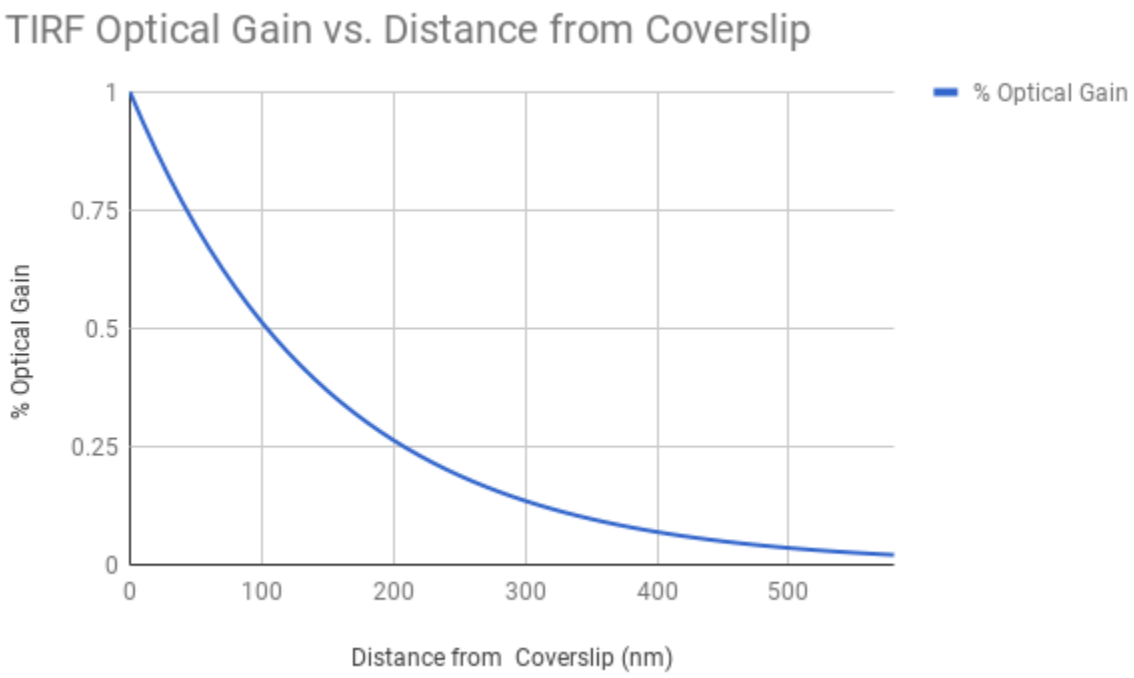


Figure 2. This represents the characteristic exponential decay of the evanescent field in Total Internal Reflection Fluorescence (TIRF) microscopy for this system, based on our calculations. This was done by using a variation of the formula of exponential decay with the fraction of distance from the coverslip divided by the distance of the coverslip (150 nm in this case). The fluorescence observed within the experiment will actually be based on factors of photon collection efficiency and the observed forces of light.

Methods

Several types of models were used to fully investigate this optogenetic system. The main two used included a rule-based model and a simple non-rule-based model of the binding interaction. There are several advantages and disadvantages of each type of model. A non-rule-based model, as the name suggests, does not use a set of rules to determine its corresponding reaction network. The main benefit of using this method is that the user must understand the system they are working with in order to translate the data and experimental procedure into Virtual Cell accurately. It also allows the user to be freer and to have more control in determining which reactions to include. The other main method of modeling is rule-based modeling. This method utilizes a set of rules to specify all the possible outcomes of a biological model. The rules then create the patterns of the reaction network. The benefit here is a strict reaction network based on the set of rules, but the rules can in turn lead to outcomes and secondary reactions that one may have not considered. An example of rule-based modeling can be seen in Figure 3, as well as a reaction diagram in Figure 4.

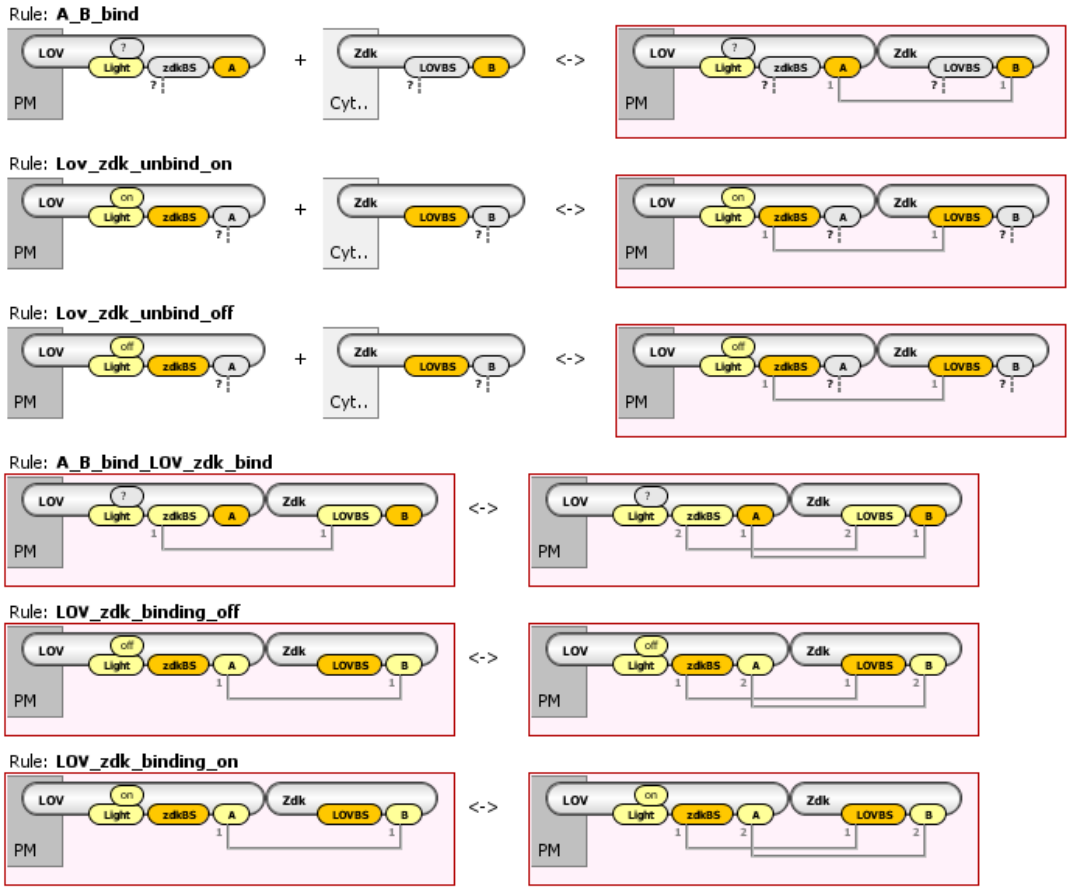


Figure 3. An example of rule-based modeling. The reactions for this molecule are governed by the set of rules seen above, which the differences highlighted in orange. The "?" symbol within the rules represent a possible bound state, where the model has situations where a receptor is bound and one where it is not. Lines show interactions and bindings between two different sites. Each species also shows where it is present in the cell – in this case, either the plasma membrane (PM) or the cytosol (cyt).

Deterministic applications within each type of modeling were then used. Deterministic applications solve and create outcomes of the model through known relationships among reaction rates and concentrations, where uncertainty in the parameters can lead to different solutions. This type of application is ideal for this system, as the basic binding interactions have been studied and are known, but certain factors are unknown. Results of a simulation within a deterministic application can be seen in Figure 5.

Within the models, parameters and functions are created to help govern the reactions as well as to explore various aspects of the model. A goal of this investigation is to determine which

parameters within the system are identifiable and important; which parameters, when solved for, create the biggest impact on the system when changed?

There are three different variations of the main biomodel utilized. There is a control model, where there are no secondary A and B binding sites; a high affinity model, where the A and B binding sites have a K_d of 100nM; and a low affinity model, where A and B binding sites have a K_d of 800nM. These represent variations in the optogenetic system. Several aspects of the control model can be seen represented in Table 1 and Figure 4. The goal of LOVTRAP being used for low affinity binding will be tested by the respective low affinity model. If we are able to measure the decay of fluorescence and demonstrate that it is comparable to the decays of high affinity and control models, then we can consider the low affinity model as successful in terms of measurement and usage.

However, in creating rule-based and non-rule-based models, there are several assumptions that need to be made in order to facilitate that creation. These assumptions are based on prior research and the best judgment of the researchers making the model. The first main assumption made is that LOV and Zdk do not form polymers; every LOV domain has just one Zdk binding site in our example. It is stated that there is a limited amount of LOV molecules and even less so of Zdk^1. Hence, to create the greatest fluorescence they must interact in a one-to-one ratio, since the fluorescent tag is bound to Zdk and this will create more fluorescence at the membrane. Another main assumption made is that the molecules are in steady state and in equilibrium before reactions begin and before the exposure to blue light. Doing so facilitates mathematical analysis tremendously and can simply show the changes in each species. Related to this is the assumption that molecules are well mixed before the reactions begin as well – the

actual experiment begins a certain time period after all the molecules are associated with one another. This leads to the prevalence of the inactive bound complex before the exposure to light.

For kinetics, the basic forward and reverse rates of the overall reaction network are based on relations to other reactions. To elaborate, the forward activation rates are the same, the binding kinetics are the same numerically as they are not affected by activation (the blue light stimulus). The binding interaction rates are greater than that of the activation rates comparatively. The process of light-induced dissociation, is a multistep process compared to one step for initial binding. With these kinetics in place, when we are looking at a model with the additional secondary binding sites, the reactions act as a pseudo first order. This will make our analysis simpler.

The last main, important assumption to note is that in regards to the binding of LOV and Zdk, the tether between this complex and the secondary B protein can freely stretch and move in a limited hemisphere space. This space is defined in the following equation, which is half the volume of a sphere of r is the radius of this hemispherical space based on a maximal tether length.

$$Space = \frac{2}{3}\pi r^3$$

Within the model, an example of parameters and functions used for analysis are listed below in Table 1. From the list of parameters, the most important and identifiable parameters need to be found. Some parameters will be controlled by the experimentalist in usage of this protocol, as they will affect the optogenetic system significantly. An example would be the forward rates of activation, as well as the intensity of the blue light. Some parameters are made due to the intrinsic properties of the system. What is being measured is the detection of the fluorescence by the low affinity system. There is a degree of freedom within our protocol in

terms of the recovery rate (from activated state to deactivated state), as there are various isoforms that can be used within this experimental procedure[1]. For an experimentalist's purpose, they should use whatever isoform is most convenient to them in terms of timescale – either 1.7, 18.5, or 496 seconds[1]. For our purpose, we used the slowest time scale as it was used to approximate our pseudo steady state due irreversible activation. A faster time scale can be used to reach equilibrium faster, which can be a different time scale all on its own. A pseudo steady state would not be needed, as one is able to approximate a true steady state.

Table 1. An example of several parameters and functions used within the model.

Original Parameter	**Function**	**Current Value/Function**
Kd1	Kd in dark	0.026 μM
K_bind	Kf of forming inactive LOV-Zdk	$4\ \frac{1}{s*\mu M}$
K_unbind1	Rxn rate of unbinding	Kd1 * K_bind
K_unbind2	Rxn rate of unbinding	Kd2 * K_bind
Kf_activation	Kf of light activation	$10\ \frac{1}{s}$
Kd2	Kd in light	4.0 μM
Act_light	Parameter meant to trigger event when light is turned on	0
BoundLov	Concentration of bound LOV-Zdk	(InactiveBound + ActiveBound)
LOV_fluorescence	Fluorescence	(Gain * (BoundLov + (cyt_zdk_near * UF)))

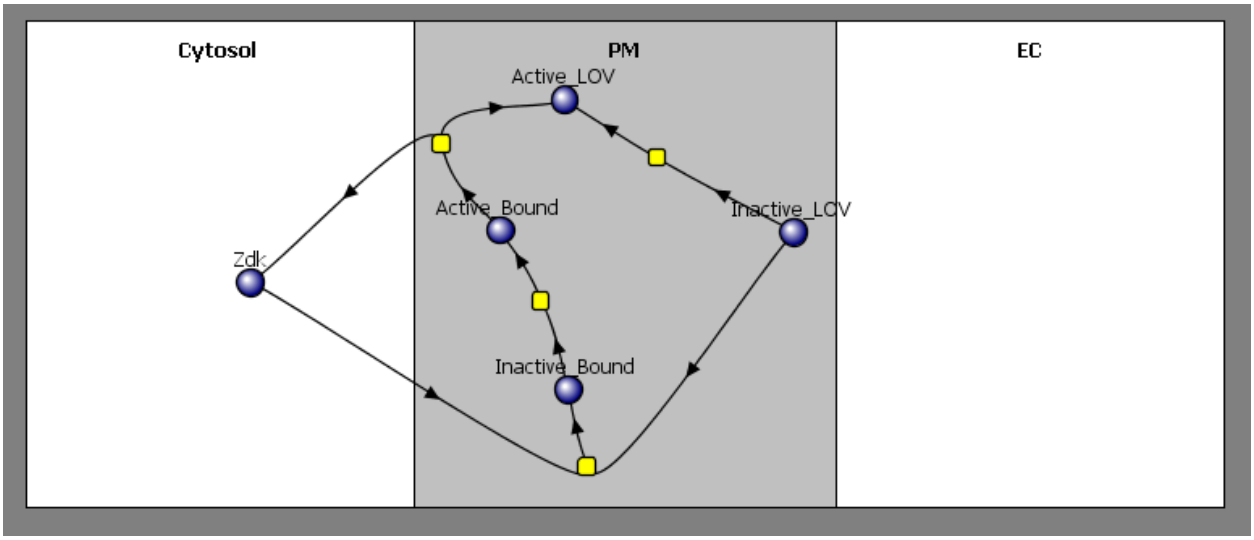


Figure 4. An example reaction diagram of the reaction defined in Figure 1 between LOV and Zdk within VCell. This does not include the secondary binding sites of A and B. Here, Zdk and Inactive_LOV bind to create the Inactive_Bound complex, which upon exposure to blue light, becomes activated and dissociates into Zdk and the Active_LOV forms. The yellow boxes correspond to nodes that represent the reaction, and it is where reaction kinetics can be modified. The purple circles correspond to the species created and changed during all the reactions.

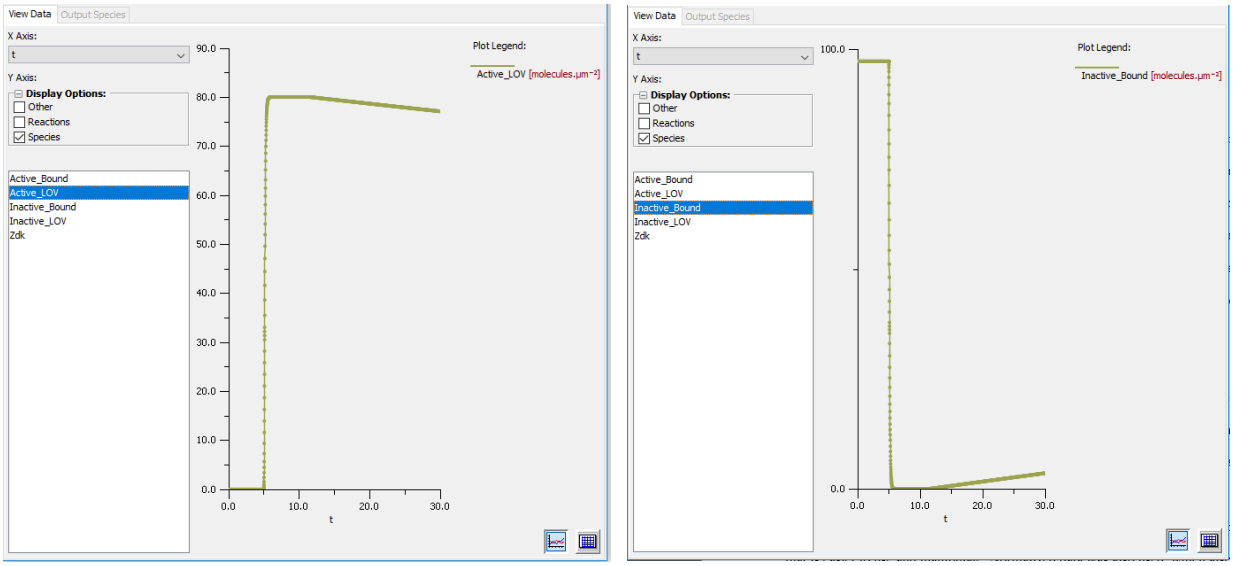


Figure 5. Simulation results that were constructed by the biomodel. Each diagram corresponds to a different species (Active_LOV and Inactive_Bound respectively). Each species responds to the blue light drastically, which is expected. The concentration of Active_LOV is expected to increase after blue light exposure, while Inactive_Bound is expected to decrease as it becomes activated and forms the active species.

In exploring the model, one important aspect of using this model is the ability to apply pulses of light. The pulse represents the signal where light is reintroduced – in 2 second intervals. To clarify, at one second, the light is turned on, while at 1.1 the light is turned off, then at 3.0

seconds the light is turned on again. Within the main model, there is an application that explores the effect of pulsing to see how pulsed duration and amplitude affect the concentrations of species. It will also be used to judge how quickly the system slowly returns to its pseudo-equilibrium (disregarding the long timescale of return to inactivation of LOV), as well as which parameters are most sensitive to the addition of light. This was later used to evaluate which species may be approximated as being constant over time.

The collected fluorescence data of mCherry-Zdk is measured in terms of average fluorescent intensity over time. However, the fluorescence contains reflection from the blue light. The increase in fluorescence intensity due to reflection is relatively constant for while the light is turned on, but it creates discontinuities in the data, which can negatively impact the analysis (see Fig 9). This is adjusted for by removing those jumps in intensity to create representative data that is easier to use and manipulate. Normalized data was also used, which helps match the model and experiment.

Another method used to help solve for the most identifiable parameters of the system was a profile likelihood program that works through VCell. A profile likelihood refers to a function within a statistical model that is able to reduce the number of unnecessary parameters by determining which parameters are useful and important to the model to reproduce a particular data set. The profile likelihood program is connected to the parameter estimation functionality of VCell, utilizing the output of the estimation. The output created the starting point for the profile likelihood – the program will alter an aspect of the parameter estimation output and rerun the parameter estimation until a range of acceptable values is found for each desired parameter. This is done by increasing or decreasing the value of a parameter by a preset multiplier until it reaches a significant error amount with respect to data. The preset multiplier will define how specific a

range is examined in terms of each parameter. The parameters included kinetics, initial concentrations, and parameters used to convert concentrations to fluorescence. This also helped us determine the sensitivity of fit to the parameters (see Fig. 6 and Fig. 7).

```
last increasing run (KfActive2 = 4.061421944962974)
     Run:  objectiveValue = 5505271.6412
params  [Kd1, KfActive2, KfActive3, Kd2, Offset, Gain, zdk_init, Li_init]
best**  [0.025, 0.49835434729, 0.930305073101, 3.9, 18717.411946, 3918.89902692, 1.39952601506, 5753.22349365]
fitted  [0.0250119261076, 3.85835084771, 0.630939031631, 4.02402253942, 18630.1995944, 3925.0100035, 1.31416221691, 4472.2836885]
low     [0.025, 3.8583508477148247, 0.1, 3.9, 17000.0, 3000.0, 0.5, 10.0]
initial [0.025, 4.061421944962974, 0.930305073101, 3.9, 18717.411946, 3918.89902692, 1.39952601506, 5753.22349365]
high    [0.027, 4.2644930422111225, 5.0, 4.1, 20000.0, 6000.0, 5.0, 7000.0]

last decreasing run (KfActive2 = 0.14015754289749138)
     Run:  objectiveValue = 6064551.55289
params  [Kd1, KfActive2, KfActive3, Kd2, Offset, Gain, zdk_init, Li_init]
best**  [0.025, 0.49835434729, 0.930305073101, 3.9, 18717.411946, 3918.89902692, 1.39952601506, 5753.22349365]
fitted  [0.027, 0.147165420042, 1.85617272728, 4.1, 18723.6473341, 3888.29251346, 1.40222695264, 6034.13924838]
low     [0.025, 0.1331496657526168, 0.1, 3.9, 17000.0, 3000.0, 0.5, 10.0]
initial [0.025, 0.14015754289749138, 0.930305073101, 3.9, 18717.411946, 3918.89902692, 1.39952601506, 5753.22349365]
high    [0.027, 0.14716542004236596, 5.0, 4.1, 20000.0, 6000.0, 5.0, 7000.0]

0.14015754289749138,6064551.55289
0.14716542004236594,5479568.75681
0.15452369104448427,5311984.28375
0.16224987559670848,4768225.32767
0.17036236937654392,4450132.45325
0.17888048784537114,4230233.63251
0.18782451223763968,4064225.78544
0.19721573784952168,3953308.88418
0.20707652474199778,3882272.05043
0.2174303509790977,3865184.44797
0.2283018685280526,3867549.74115
0.23971696195445522,3864442.60504
0.251702810052178,3856086.08358
0.26428795055478693,3843283.694
0.2775023480825263,3826441.23621
0.2913774654866526,3806541.48617
0.3059463387609852,3784740.34844
0.3212436556990345,3762294.20418
0.33730583848398626,3739791.87254
0.3541711304081856,3718870.76263
0.3718796869285949,3699410.75929
0.3904736712750247,3682184.54416
0.409997354838776,3669548.99035
0.4304972225807147,3661936.77434
0.45202208370975056,3655605.21225
0.47462318789523805,3656883.41576
0.49835434729,3661136.08654
0.5232720646545,3657177.4079
0.549435667887225,3662532.00114
0.5769074512815863,3670647.68014
0.6057528238456656,3683850.21542
0.63604046503794 9,3702747.9362
```

Figure 6. An example of the profile likelihood functionality, which is used to determine realistic values of parameters within a certain error limit. The top half of the figure includes the output from the parameter estimation run in VCell, as well as the last increasing and decreasing runs of the profile likelihood. Several different parameter sets were created based on the profile likelihood. This includes a set compiling the highest and lowest possible values for each parameter before the error becomes significant, as well as a set containing the best values for each parameter based on an optimal objective value. The bottom half demonstrates the actual run, where a parameter value and error are displayed. Over this run, as the parameter value increases, the error decreases, which indicates the range of parameter values that leads to a good fit given a set threshold.

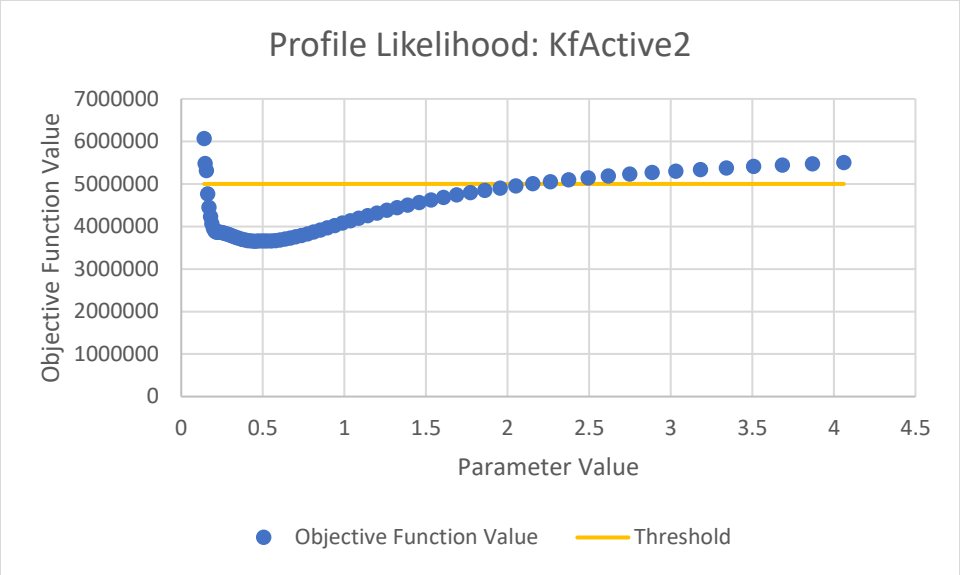


Figure 7. A visualization of the profile likelihood result, showing the realistic range in the kinetic rate constant with respect to the biomodel. This represents the expected error that the model has for a particular confidence interval. Any parameter value below the threshold results in a good fit, while any parameter value above the threshold leads to a bad fit. Here, the best range of this forward reaction rate is between 0.1 – 2 s^{-1}.

Results/Analysis

The first important result from observing this optogenetic system is establishing a steady state where future analysis can occur. This involves determining estimations of initial concentrations of LOV, Zdk, and the ratio of the size of the plasma membrane to the size of the cytosol. We derived these expressions based on the differential equations, set them equal to 0 (to represent the steady state), then algebraically reduce and combine these equations into a second order polynomial, the solution of which is in Figure 8.

$$A = Initial\ Concentration\ of\ LOV\ Domain$$

$$B = Initial\ Concentration\ of\ Zdk$$

$$C = \frac{Size\ of\ Plasma\ Membrane * Unit\ Factor}{Size\ of\ Cytosol}$$

$$K_d = 0.026\ \mu M$$

$$Steady\ State\ of\ LOV - Zdk\ Bound\ Complex = \frac{(B + (A * C) + K_d) - \sqrt{(A * C) + B + K_d)^2 - (4ABC)}}{2C}$$

$$Steady\ State\ of\ Zdk = B - (C * Steady\ State\ of\ LOV - Zdk\ Bound\ Complex)$$

$$Steady\ State\ of\ LOV\ Domain = A - Steady\ State\ of\ LOV - Zdk\ Bound\ Complex$$

Figure 8. The equations used to model the steady state of this reaction, based on solving for the initial concentrations known by the user.

This combination of factors derives a steady state before the light stimulus is expressed. This will also help us better evaluate the overall change in fluorescence. The last main effect of creating a steady state is that it allows us to model the experiment with the system initially at rest before the light is applied (second panel of Figure 9).

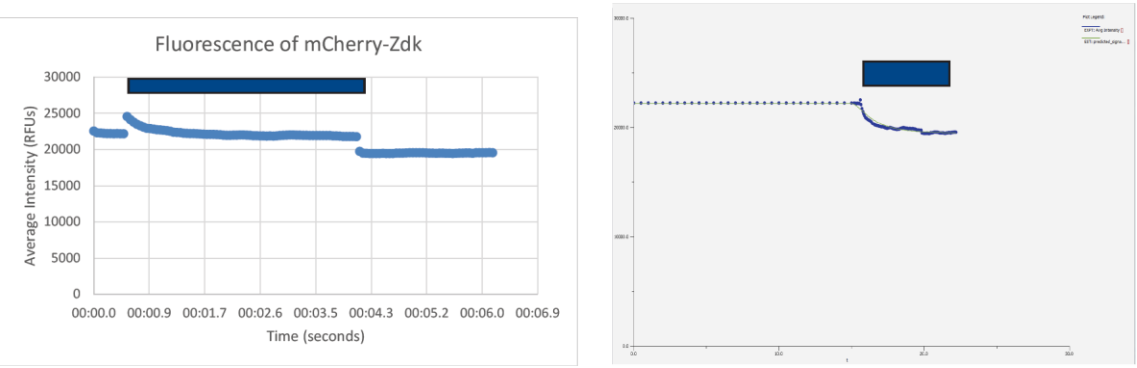


Figure 9. A conversion of the experimental fluorescence to the model fluorescence. The left image demonstrates the average intensity of a cell in a single trial, where the light is turned on at 0.6 seconds and stays on until 4 seconds. The right image shows the same trial in VCell, where the system is in steady state until the light is turned on. The blue bar in each image shows the time period where the blue light is shown. The offset in the experimental data is due to the background reflection from the blue light.

From our analysis, we learned more about several important aspects that can be useful for better understanding how LOVTRAP works. To start, we have better illustrated the representative binding curves of LOV and Zdk in the dark state and in light with respect to increasing concentration of Zdk (Fig. 10). This was constructed with the respective dissociation constants of the system in mind, with the dark state having a K_d = 0.026 µM and the light state having a K_d > 4 µM. The greatest difference in these binding curves will help us evaluate the concentration of Zdk where the greatest change in fluorescence over time, ΔF/F, is observed. This is the most sensitive state of the modeled fluorescence with respect to the dissociation constants.

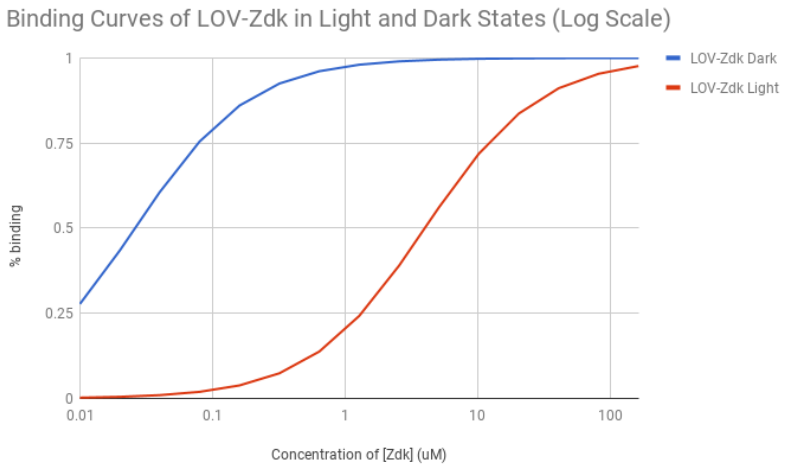


Figure 10. The change in fluorescence will be evaluated between the relationship of the fluorescence with a strong binding affinity (in dark) to the fluorescence with a weaker binding affinity (in light), shown by the respective binding curves.

$$[\mathrm{Zdk}_{\mathrm{free}}] = \frac{K^*[\text{LOV-Zdk}]}{[\text{LOV}]} + \frac{K_a[\text{LOV}_a\text{-Zdk}]}{[\text{LOV}_a]}$$

$$F_0 = ([\text{LOV-Zdk}]_0 + [\text{LOV}_a\text{-Zdk}]_0 + \frac{1}{2} {}^*]{}^*\mathrm{Zdk}_0) * \text{Gain}$$

$$F_1 = ([\text{LOV-Zdk}]_1 + [\text{LOV}_a\text{-Zdk}]_1 + \frac{1}{2} {}^*]{}^*\mathrm{Zdk}_1) * \text{Gain}$$

$$\frac{\Delta F}{F} = \frac{(F_1\text{-}F_0)}{\max(F_1, F_0)}$$

Figure 11. This figure demonstrates how we evaluated the change in fluorescence with respect to various important components in dark (1) and light (0) conditions. $\mathrm{Zdk}_{\mathrm{free}}$ represents the amount of free Zdk floating near the membrane but not bound. F_1 represents the steady state fluorescence of the high affinity binding state when the system is in the dark with a K_d of 0.026 μM. F_0 represents the steady state fluorescence of the low affinity binding rate when the system is exposed to light with a K_d of 4 μM. The steady state solutions follow the binding curves shown in Figure 10. *l* represents the characteristic TIRF length of 150 nm, which indicates the distance into the cell where TIRF excitation occurs. The fluorescence is proportional to the bound state of the complex and the active LOV state.

An important result that came about from our analysis was a justification for using a non-spatial model. Initially, both a spatial and a non-spatial deterministic model were used, with analysis shifting towards usage of the latter. This was done as it was discovered that at higher concentrations of Zdk (greater than 1 μM), the overall percent change in concentration over time was less than 1%. This indicates maximal saturation. Figures 13-14 demonstrate how little Zdk

changes based on the difference between the maximum and minimum values over time. As a further approximation to the non-spatial model, we observe that since the concentration of well-stirred cytosolic Zdk does not vary much over time (about 1% change over time), it can be treated as having constant concentration, with any changes being attributed to an acceptable error.

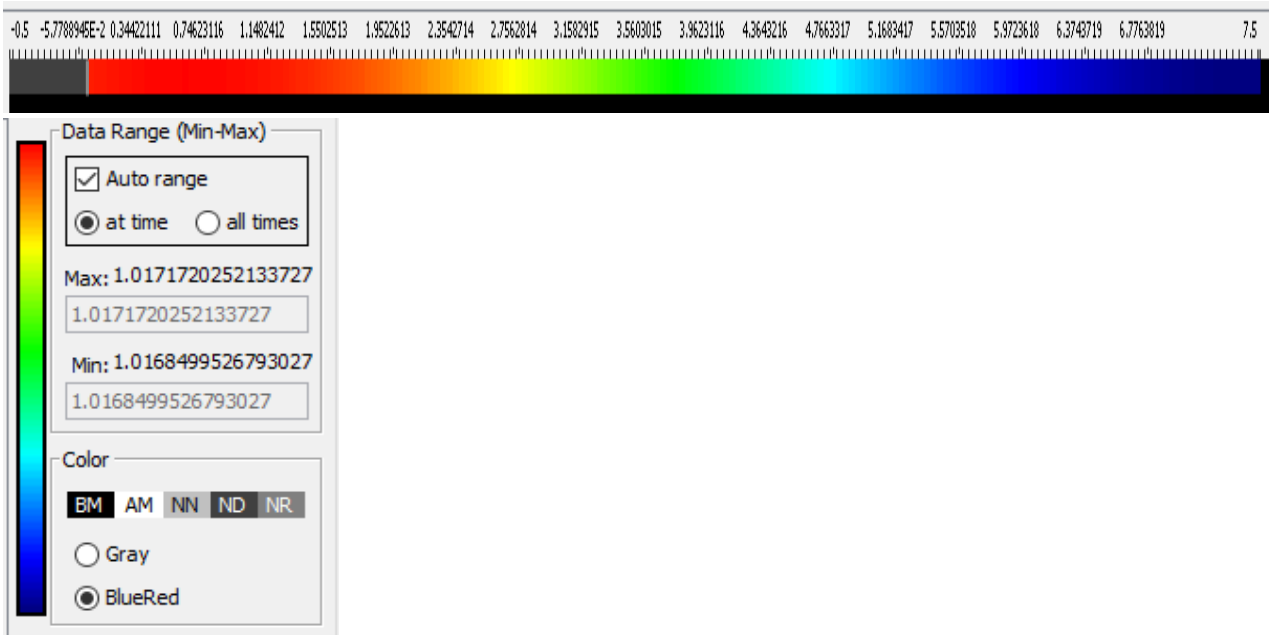
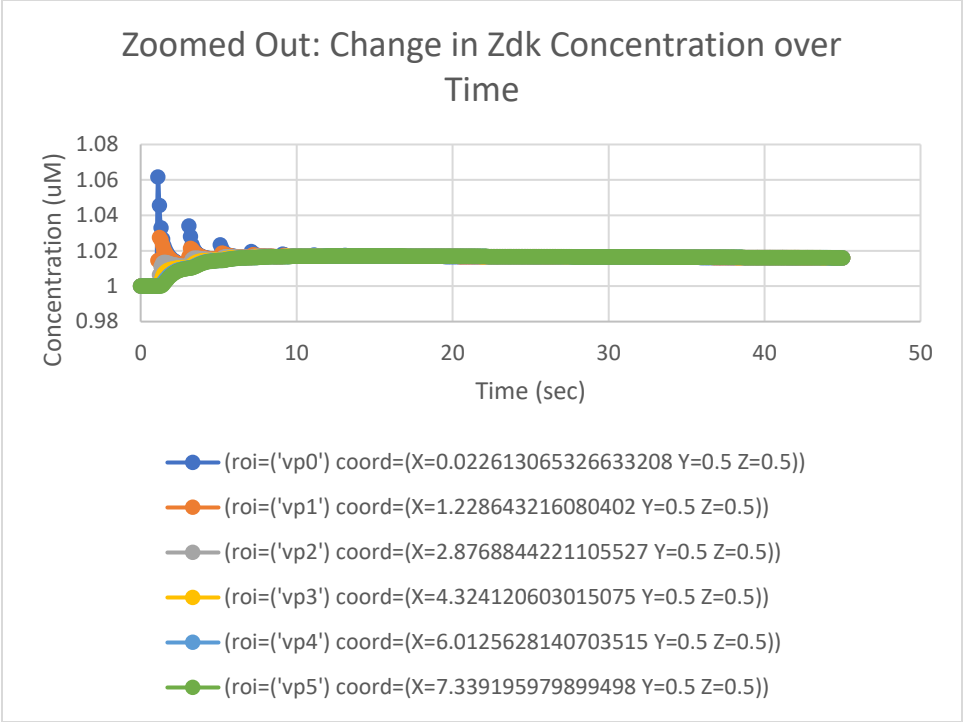


Figure 12. This is an example of a spatial application result conducted in VCell for the concentration of Zdk. Shown here is a one-dimensional look at the membrane/coverslip, as well as a distance of 7.5 μm into the cytosol. Based on the gradient and looking at the max and min of Zdk, one can see that the concentration of Zdk does not really change too much at this concentration upon exposure to light. At higher concentrations, one can assume that Zdk is constant through time.



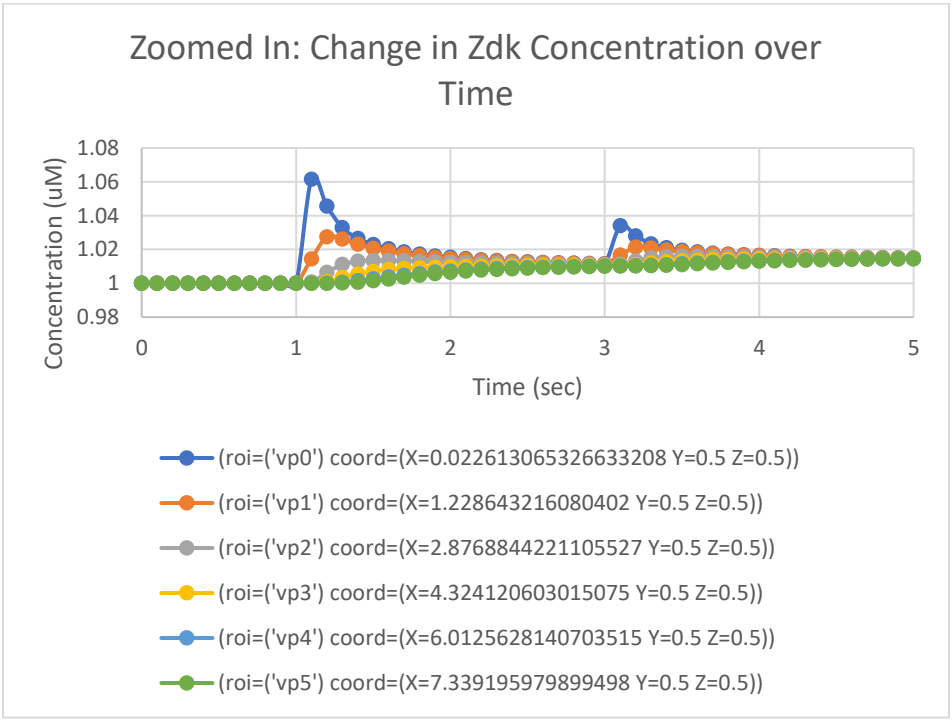


Figure 13-14. This figure shows the relative change in concentration over time for Zdk using the pulse protocol. At the first instance of the light pulse, the changes in concentration depending on the distance into the cytosol (the concentration gradient) is due to binding and unbinding activity at the membrane. Over time, the gradient diminishes as the membrane activity reaches an equilibrium. Overall, the change in concentration due to the diffusion gradient, present in the spatial model, is not much different (3% spatial difference averaged over pulse duration) than the effect of the species being well-mixed in the non-spatial model. Refer to Figure 12 for more context.

It is important to determine which parameters are necessary to the system, so that when conducting the experimental procedure of LOVTRAP an experimentalist can attempt to control for these factors. Two of these degrees of freedom include the initial concentrations of LOV and Zdk. It may seem obvious, but it is an important factor to consider how much of each molecule is incorporated in the cell from the start in order to favor a LOV-Zdk complex over a LOV-Zdk-LOV-Zdk… polymer. The initial distribution of each molecule is assumed to be in the inactive state of the overall system due to absence of blue light.

Another aspect about developing this experimental protocol is the issue with timescales. The pulse protocol exposes the differences between a fast and slow timescale – the duration and amplitude of the pulse of the light matters. The longer the stimulus of light with the same energy, the slower the timescale will be for the activation of LOV. A more intense stimulus of light with shorter duration will lead to a faster time scale for activation. Knowing these two factors can allow an experimenter to modulate the length of the experiment in order to better detect the fluorescence, either by making the experiment longer where there will be more data to collect, or one can have a fast protocol in which will expose the dynamics of the faster timescales in the model. The sensitivity of the sensor matters to monitor the reaction.

Another biomodel was made to demonstrate the bivalent binding capabilities of the LOV-Zdk complex, where the secondary binding sites A and B can bind, and show that there is a cooperativity of the respective kinetics. This cooperativity depends on several factors. The main factor that determines this cooperativity is the possibility of steric hindrance. If the secondary binding sites are not hindered by the orientation of the other binding site, they should bind and high cooperativity should be seen. This cooperativity was measured by a new parameter called alpha (α), which is introduced in the new biomodel.

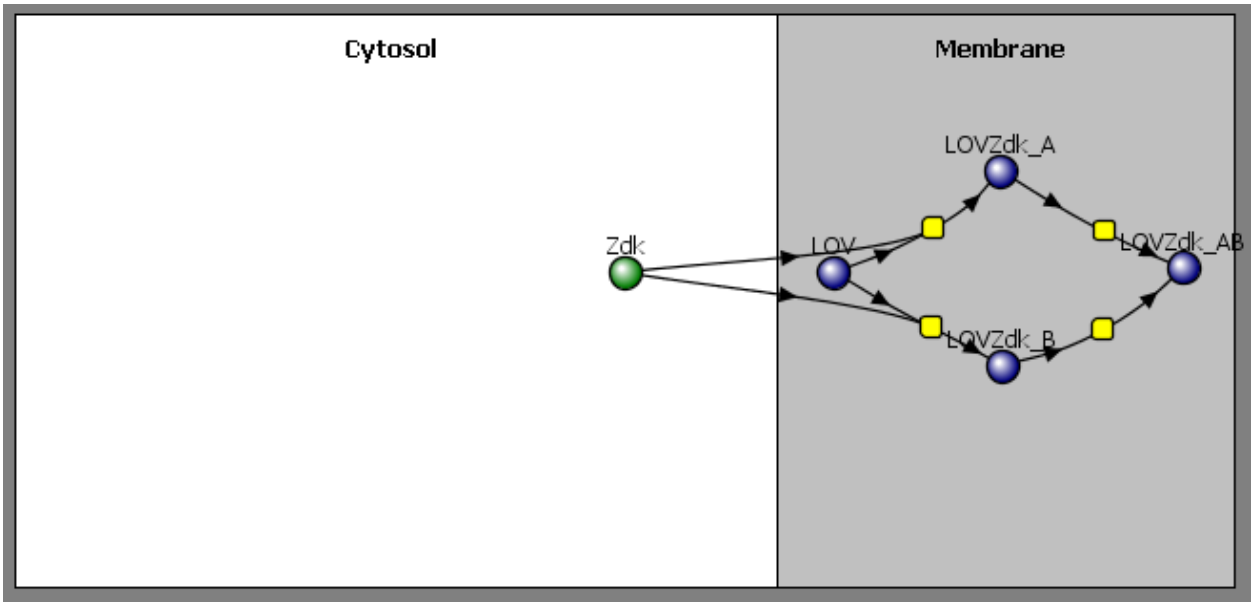


Figure 15. This biomodel is used to investigate the bivalent binding capabilities of the system, as well as measuring the possibility of cooperativity. This takes into consideration the conclusions of the simpler biomodel with respect to intramolecular binding of A and B secondary binding sites (to create LOVZdk_AB). For example, one can used the further approximation of constant cytosolic Zdk binding, which makes all reactions first order and therefore lead to easier analysis.

In this model, LOV and Zdk bind at the first node, and depending on the path, the secondary site A or B is bound. In the last two reaction pathways, the other secondary site is bound. The kinetics are similar to the previous biomodel at the first node, which is where the cytosolic Zdk and LOV are bound. They are similar in terms of being functions of the dissociation constants and will likely proceed at a much slower rate than the secondary binding reactions. There is an assumption in the last two reaction pathways that they have the same cooperativity effect, represented by the factor alpha. Alpha is calculated as the ratio between the effective concentration of a single Zdk tethered to its respective LOV molecule over the concentration of cytosolic Zdk. The cytosolic concentration refers to the Zdk that was measured to be constant over time, but not space. The tethered Zdk concentration is calculated to be the space that one molecule can travel within the tethered range. This equation can be seen below:

$$\frac{1}{\frac{2\pi r^3}{3}} \frac{molecules}{\mu m^3}$$

When this concentration is converted to the proper unit within VCell (µM), this concentration is calculated to be about 800,000 µM (0.8 M). This number may seem high at first, but the reasoning behind this number makes sense. Within the tether range, there is a higher probability of finding and binding Zdk. With respect to the cell and cytosol in general, where Zdk is located, the concentration within this tether range is going to be much, much higher. This is represented in Figure 16.

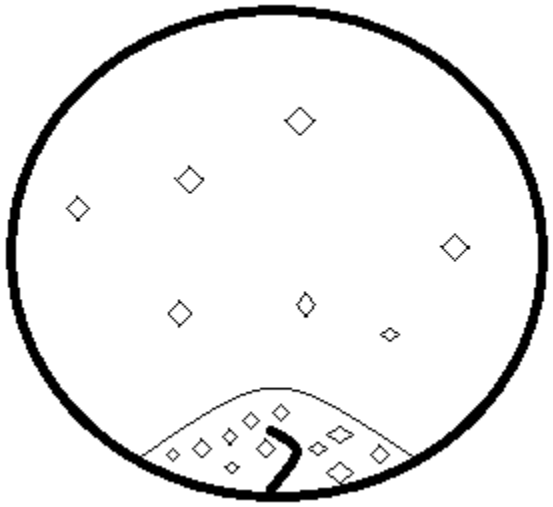


Figure 16. A visualization for why the concentration of Zdk (represented by the diamonds) near the tether is greater than the rest of the cytosol.

However, this alpha is representing a reasonable limit, if not an upper bound, of the possible cooperativity within the system. While the actual cooperativity cannot be determined currently in the system, knowing this upper limit can lend more information about controlling factors within the system. It can also be used as a new function, termed beta (β), which is the relationship between α and the dissociation constant, which can be used to determine the relationship between the activation of light and the change in cooperativity.

The last important matter to note is the issue of identifiable parameters. Mathematical analysis was conducted between five parameters that were established to be important based on profile likelihood: the total concentrations of LOV and Zdk, the two dissociation constants that causes optogenetic activation of the system, and δ, a variable that represents the fraction of LOV in an activated state compared to inactivated. δ, evaluated between 0-1 (0 being completely inactivated and 1 being completely activated) allows us to represent the two types of LOV in terms of the total LOV concentration, as such:

$$L_{a,tot} = \delta * L_{tot}$$

$$L_{i,tot} = (1 - \delta) * L_{tot}$$

Using these five parameters, a system of equations can be set up and solved for. The result and possible solution for this system of equations (a third order polynomial) can be seen below.

For $a * Z^3 + b * Z^2 + c * Z + d = 0$, where Z represents free cytosolic Zdk:

$$a = 1$$

$$b = K_i + L_{tot} - Z_{tot} + K_a$$

$$c = \delta * K_i * L_{tot} + (1 - \delta) * K_a * L_{tot} - (K_i + K_a) * Z_{tot} + K_i * K_a$$

$$d = -K_a * K_i * Z_{tot}$$

Here, a, b, c, and d represent coefficients to solve for the steady state and solve for the optogenetic system. *K* refers to the dissociation constant that represent the activated (K_a) and inactivated (K_i) states. *L* represents LOV, and Z represents Zdk.

To help solve for the system of equations, one can pulse (blast short bursts of light quickly within the system) to expose various levels of δ, each time activating a fraction of LOV. If δ is 0, meaning no activation, this considerably simplifies the system, as seen below:

$$a * Z^3 + b * Z^2 + c * Z + d = 0$$

$$a = 1$$

$$b = K_i + L_{tot} - Z_{tot} + K_a$$

$$c = -K_i * Z_{tot} + K_i * K_a + K_a * (L_{tot} - Z_{tot})$$

$$d = -K_a * K_i * Z_{tot}$$

If one knows the total concentrations of LOV and Zdk used during their experimental trials, the steady state can be easily solved for, as the dissociation constants are known variables. It's simply a matter of using the numbers correctly.

Discussion

Developing a variation of an experimental protocol is difficult, especially in terms of taking into account the possibility of measuring low affinity binding interactions. However, several conclusions can be made about the optogenetic system that can lend information to the experimenter. This will allow for a better understanding of the protocol as well degrees of freedom that can be changed and modified as the experimenter pleases.

When LOV is in excess and there is a moderate amount of Zdk, the fluorescence is more dependent on the bound states of LOV-Zdk rather than the unbound state, due to depletion of cytosolic Zdk. When Zdk is in excess and there is a moderate amount of LOV however, the fluorescence is going to be more dependent on the cytosolic Zdk. The fluorescence originates from the mCherry-Zdk, where cytosolic contribution is rather constant, but the fluorescence due to the bound complex changes between the light and dark states (which have different dissociation constants) at time points before and sometime after excitation. This can be seen by the equations of fluorescence (Figure 11).

In terms of the data, adapting the fluorescence data to VCell was possible, with the only issue being background reflection caused by the light source. Mathematically derived equations produced a model in an initial steady state to get a better fit of data for parameter estimation. This steady state shows the experimenter the relationship between each molecule, and it demonstrates again how changing one parameter affects another. Being easily able to adapt, import, and simulate data is one of VCell's strong points[3].

Observations can be made of the general trends of concentrations for the inactive LOV domain, bound LOV-Zdk complex, and active LOV domain. The concentration of inactive states decreases rapidly upon exposure to blue light, while the activated state concentration increases

quickly with the same exposure. After the light is turned off, slow recovery is seen back to the inactive states. This recovery can be controlled based on the mutant or variation used of LOV, which is modeled by changing the kinetics of the reverse of the activation reaction. In the model, the isoform that leads to the slowest recovery was chosen as this allowed for better timescale separation and understanding of the model in terms of kinetics.

The parameters that are the most sensitive and identifiable seemed to be the on rates of the inactive bound state and the active bound state, the initial concentrations of the Zdk protein and of the LOV domain, and the dissociation constants. Gain (an external parameter that represents the signal amplification at the detector) is a fitting variable, it can be removed from analysis as it does not have a meaningful effect on its own. These parameters which were found to be sensitive when fitting data were the same parameters which are present in the steady state equations (seen by Figure 8).

To see which of these parameters are unique, a profile likelihood analysis was done to look at the identifiability and sensitivity of these parameters. This was done by an implementation to VCell. The objective function value represents the error of the fit. As the parameter values are increased and decreased, the error will also increase until a range of best fit is determined. This program, if built into VCell in the future, can lead a user to better understand the bounds of the region of parameter space in the biomodel that influence the fit of the data.

The results of the time course and the experimental analysis were confirmed by another software called COPASI. COPASI is a software application for analysis and many types of simulations of biochemical networks[7]. Using this interface, a parameter scan was run using the same simulation file in .vcml format. With this, the intent is to independently validate the results

by using another software. However, this analysis was not completed due to lack of time, but more can be done in that respect.

Various factors of our analysis point towards LOVTRAP being a suitable protocol for low affinity binding. This was the original goal of the project, but this is not completely achieved. Some guidance however includes keeping Zdk as a concentration between the dissociation constants is one possibility that allows low affinity binding to be practical. Another aspect is having a LOV concentration high enough to prevent polymerization. Instead, the focus shifted more towards model analysis and carefully understanding the assumptions and dynamics behind the approximations.

Literature Cited

1) Wang H, Vilela M, Winkler A, et al. LOVTRAP, An Optogenetic System for Photo-induced Protein Dissociation. *Nature methods*. 2016;13(9):755-758. doi:10.1038/nmeth.3926.

2) Wang H, Hahn KM. LOVTRAP, a Versatile Method to Control Protein Function with Light. *Current protocols in cell biology*. 2016;73:21.10.1-21.10.14. doi:10.1002/cpcb.12.

3) Schaff JC, Vasilescu D, Moraru II, Loew LM, Blinov ML. Rule-based modeling with Virtual Cell. *Bioinformatics*. 2016;32(18):2880-2882. doi:10.1093/bioinformatics/btw353.

4) Spangler SM, Bruchas MR. Optogenetic approaches for dissecting neuromodulation and GPCR signaling in neural circuits. *Current opinion in pharmacology*. 2017;32:56-70. doi:10.1016/j.coph.2016.11.001.

5) Reimer M, Denby E, Zustiak SP, Schober JM. Ras GAP-related and C-terminal domain-dependent localization and tumorigenic activities of IQGAP1 in melanoma cells. Komarova Y, ed. *PLoS ONE*. 2017;12(12):e0189589. doi:10.1371/journal.pone.0189589.

6) Xu C, Watras J, Loew LM. Kinetic analysis of receptor-activated phosphoinositide turnover. *The Journal of Cell Biology*. 2003;161(4):779-791. doi:10.1083/jcb.200301070.

7) Mendes, P, Hoops, S, Sahle, S, Gauges, R, Dada, J, and Kummer, U. Computational modeling of biochemical networks using COPASI. *Methods Mol Biol*. 2009; 500, 17-59.