AFRL-RY-WP-TR-2019-0267

SELF-MODIFYING AND FAST ANALOG MOLECULAR COMPUTING VIA DESIGNED ENZYMES

Sagar D. Khare
Rutgers, The State University of New Jersey

Eduardo D. Sontag
Northeastern University

DECEMBER 2019
Final Report

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals.

Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RY-WP-TR-2019-0267 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//Signature//

ASHLEY M. DEMANGE
Program Manager
Sensor Subsystems Branch
Aerospace Components & Subsystems Division

//Signature//

TIMOTHY R. JOHNSON, Chief
Sensor Subsystems Branch
Aerospace Components & Subsystems Division

//Signature//

ADAM L. BROOKS, Maj, USAF
Deputy
Aerospace Components & Subsystems Division
Sensors Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show "//Signature//" stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
December 2019	Final	27 December 2017 – 31 May 2019

4. TITLE AND SUBTITLE
SELF-MODIFYING AND FAST ANALOG MOLECULAR COMPUTING VIA DESIGNED ENZYMES

5a. CONTRACT NUMBER
FA8650-18-1-7800

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
61101E

6. AUTHOR(S)
Sagar Khare (Rutgers, The State University of New Jersey)
Eduardo Sontag (Northwestern University)

5d. PROJECT NUMBER
1000

5e. TASK NUMBER
N/A

5f. WORK UNIT NUMBER
Y1QH

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rutgers, The State University of New Jersey
57 US Highway 1
New Brunswick, NJ 08901-8554

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command
United States Air Force

Defense Advanced Research Projects Agency (DARPA/MTO)
675 North Randolph Street
Arlington, VA 22203

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RYDR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RY-WP-TR-2019-0267

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. Report contains color. This material is based on research sponsored by the Air Force Research Labs (AFRL) and the Defense Advanced Research Project s Agency (DARPA) under agreement number FA8650-18-1-7800. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied of the Air Force Research Labs (AFRL), the Defense Advanced Projects Agency (DARPA), or the U.S. Government.

14. ABSTRACT
This work aimed to develop evolvable biological (enzymatic) circuits using designed/engineered proteins that can be made responsive to chosen input stimuli such as small molecule chemicals and/or light. These circuits would operate rapidly and perform complex integration of input signals to produce a desired easily-detectable output. This includes the design and optimization of elementary OR gates using chemical and light inputs, the development and validation of control-theoretic framework for parameter analysis of the designed gates, and the design and implementation of an AND gate. The proteins in each switch were successfully produced and demonstrated the desired input-output function in preliminary experiments. A mathematical framework was also successfully developed to analyze the functioning of each switch and their combinations into elementary logic gates.

15. SUBJECT TERMS
biological catalysts, protein engineering, enzymes, induced dimerization domain, proteases, evolvable networks, molecular circuit design, self-modifying and self-replicating programs

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ashley DeMange
Unclassified	Unclassified	Unclassified	SAR	24	19b. TELEPHONE NUMBER *(Include Area Code)* N/A

Table of Contents

Section **Page**

List of Figures

Figure **Page**

List of Tables

Table **Page**

1. SUMMARY

We proposed the development of evolvable biological (enzymatic) circuits using designed/engineered proteins that can be made responsive to chosen input stimuli such as small molecule chemicals and/or light. These circuits would operate rapidly and perform complex integration of input signals to produce a desired easily-detectable output. In other words, we took the first steps for developing the ability to perform computation with protein biomolecules, in particular, enzymes, using optochemical inputs. Control theoretic modeling was used to analyze and further refine our designs. A manuscript describing these results was submitted. In contrast with current state-of-the-art synthetic biological circuits, our proposed circuits are light-weight, requiring only a small number of components during operation *ex vivo*, will be able to perform faster computation and respond to chosen optical and chemical inputs: for example, concentrations of a set of chosen chemical molecules from a mixture can be integrated to add, multiply, or more generally, calculate a mathematical expression. Department of Defense (DoD) related applications of these circuits could involve rapid detection of chemicals and carrying out sense-and-neutralize functions.

In this project, we proposed the bottom up development of the elements that will be used to design these circuits. Our central players are multi-modular proteins (Figure 5) in which split enzymes that are "hooked up" to half of recognition domains.
The major objectives were:

1. Design and optimization of elementary **OR, XOR** gates using chemical and light inputs
2. Development and validation of control-theoretic framework for parameter analysis of the designed gates

Apart from the XOR gate, during the course of our work, we realized that an AND gate would have high dynamic range and other favorable characteristics, therefore we also designed and implemented an AND gate.

As seen in Figure 5, the **OR** gate is composed of two switches: a chemical input detecting switch and an optical input detecting switch. Both switches are in turn composed of two proteins that sense the input by forming a conditionally controlled complex, either of which can produce the output. After considerable optimization, **we have successfully produced the proteins in each switch and demonstrated that they are producing the desired input-output function in preliminary experiments**. We have also successfully developed a mathematical framework to analyze the functioning of each switch and their combinations into elementary **OR** and **XOR** gates. This framework was used to perform numerical simulations of proposed gates using experimental data we derived (and from the literature). Together, these studies have set the stage for building from the bottom up a variety of stimulus-responsive protein elements that can be programmed at molecular and supramolecular levels to build enzyme-based circuits.

2. EXPERIMENTAL APPROACH AND RESULTS

2.1 Overview of Experimental Approach

Approximately 100 proteins were expressed and purified. For each,

1. Synthesis of genes corresponding to the component proteins and recombinant expression in *E. coli*:

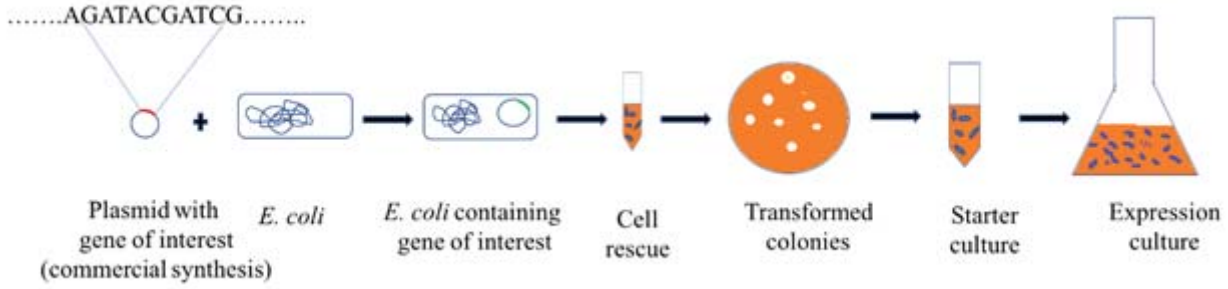


Figure 1: Overview Schematic of Gene Synthesis

2. Purification of protein produced in *E. coli*

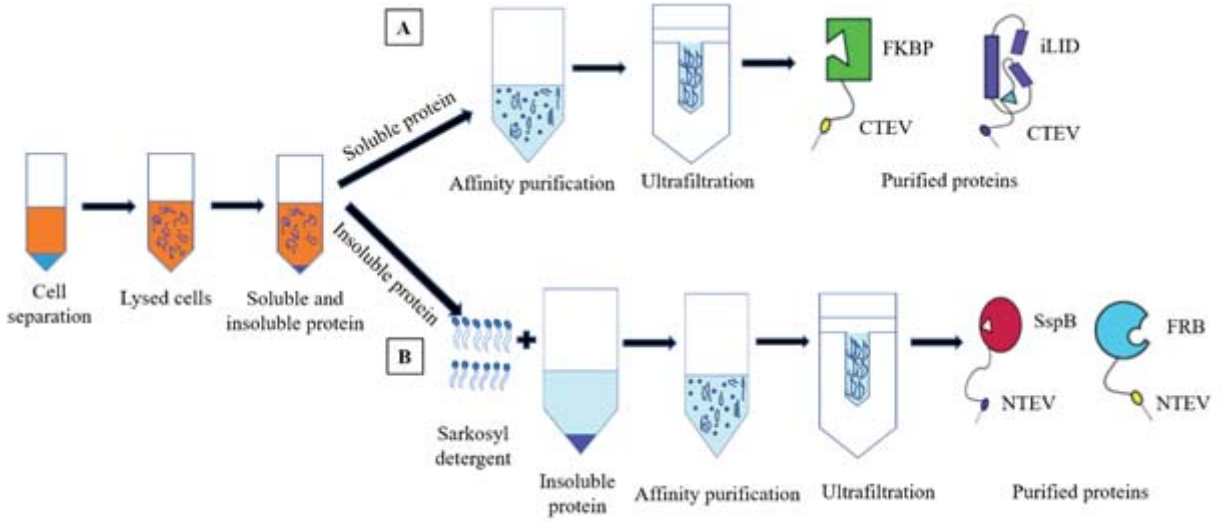


Figure 2: Overview Schematic of Protein Purification

The cells containing the protein of interest is separated for liquid medium by centrifugation. The separated cells are lysed to harvest the protein. The lysed cells containing soluble and insoluble protein were separated by centrifugation. (A) The soluble proteins in the supernatant is passed through affinity column to isolate the protein of interest from other proteins. The purified protein is concentrated by ultrafiltration. FKBP-CTEV and iLID-CTEV were purified from soluble fraction of lysed E. coli. (B) The insoluble protein was treated with sarkosyl to make the protein soluble. Once, the protein was solubilized, it was purified by regular purification method as earlier. SspB-NTEV and FRB-NTEV were purified from soluble fraction of lysed E. coli.

As an example, the purification of a protein component of an **OR** gate is shown in Figure 3. The existence of a single band in the gel shows that we successfully obtained this protein.

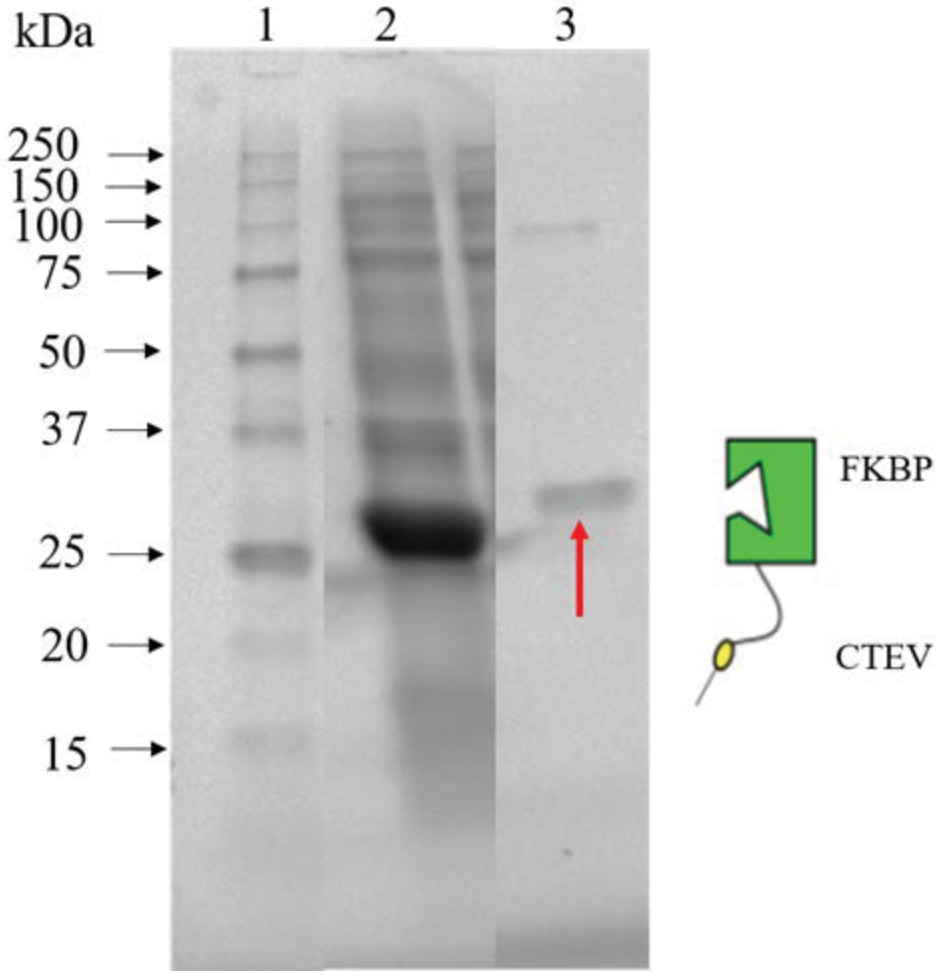


Figure 3: Purification from Soluble Protein Fraction

Purification from soluble protein fraction. Histidine tag purification of FKBP-CTEV protein from supernatant were carried out using nickel affinity column chromatography. The molecular weight of FKBP-CTEV is 24.7 kDa and the yield was 3 mg for 1 liter of expression. The graphics on the right shows the proteins being expressed. (1) Protein ladder, (2) Cell lysate, and (3) Purified protein.

In some cases, the protein produced in *E. coli* is entrapped in the water-insoluble cell debris. We used a detergent to solubilize the debris to obtain the protein. Two examples of proteins purified using this method is shown in Figure 4.

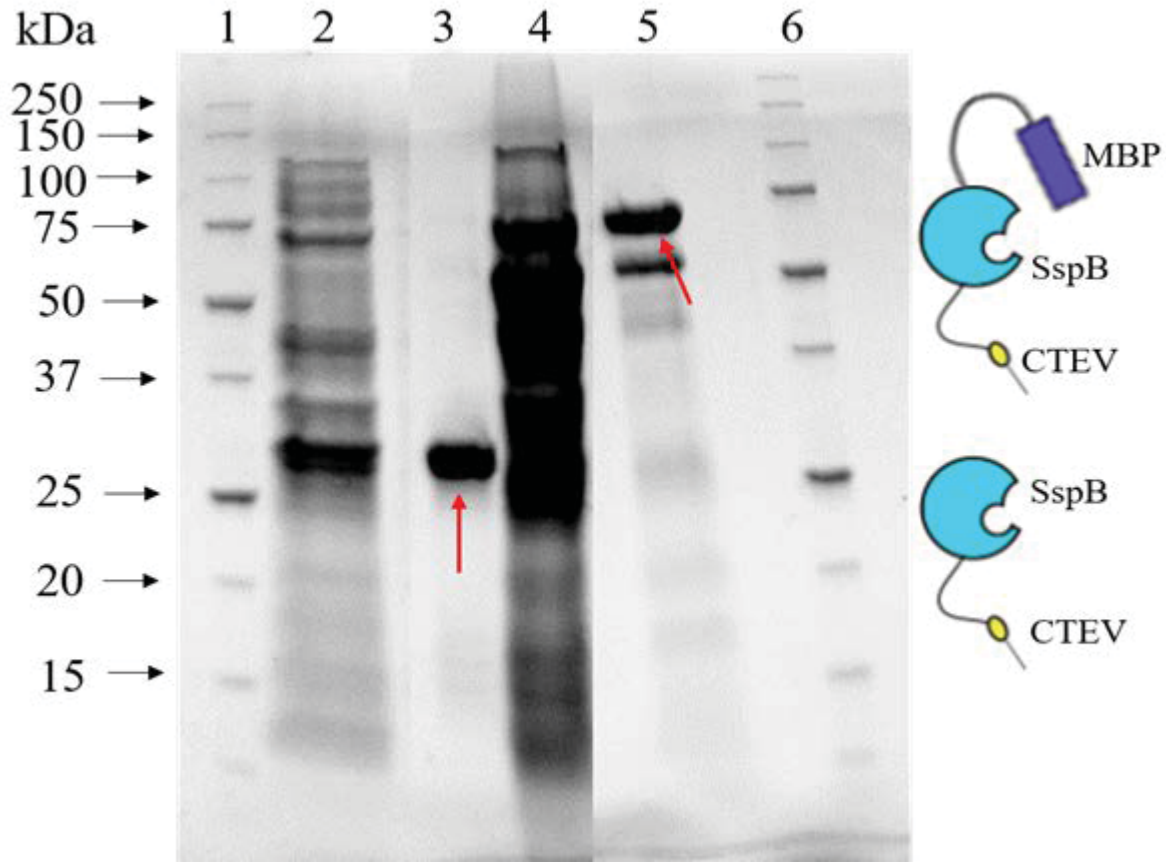
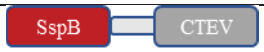


Figure 4: Purification from Insoluble Protein Fraction

Purification from insoluble protein fraction. The inclusion body protein in the cell pellet was treated with sarkosyl detergent to solubilize the protein. The solubilized protein was purified by affinity purification. The molecular weight of CTEV-SspB and MBP-CTEV-SspB is 28 kDa and 72 kDa. The graphics on the right shows the proteins being expressed. (1) Protein ladder, (2) CTEV-SspB flowthrough, (3) CTEV-SspB purified protein, (4) MBP-CTEV-SspB flowthrough, (5) MBP-CTEV-SspB purified protein, and (6) Protein ladder.

In a similar manner, **approximately 20 proteins** were tested for expression, and yield and purity in being produced by *E. coli* (see Table 1). In some of these proteins we added a MBP tag that helps solubilize them in aqueous solutions.

Table 1. Proteins Tested for Expression, and Yield and Purity

Constructs studied	Supernatant	Pellet	Supernatant purification	Sarkosyl purification
Chemical induced dimerization				
		✔		✔
	✔		✔	
		✔		
	✔			
		✔		
	✔			
	✔			
		✔		
Light induced dimerization				
		✔		
	✔		✔	
		✔		✔
	✔			
	✔			
		✔		✔
		✔		✔
		✔		✔

2.2 Testing the Function of OR Gate Switches

Having obtained the proteins required for constructing the switches, we used green fluorescent proteins (GFP) to detect the performance of the switch. It is expected that upon detecting signal, the protein components cluster together and the TEV protease is reconstituted. The reconstituted enzyme activity can be detected by measuring the amount of fluorescence produced upon cleavage of a GFP-REACh construct. In this construct the fluorescence of the GFP protein is quenched by the linked REACh domain, and cleavage liberates the GFP to turn the solution green.

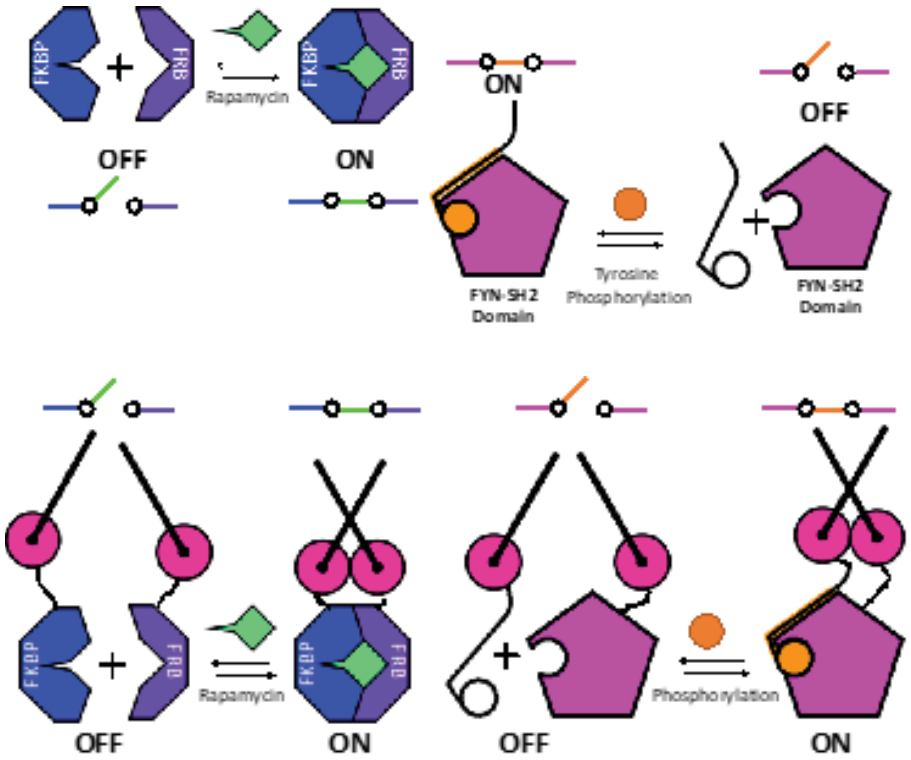


Figure 5: Overview Schematic of Recognition Elements

In OR protein logic gate, three components are necessary: two components are proteins and one component are different chemical inputs, which are the initiators needed for proteins to come together. (A) Is a chemical induced dimerization in which FRB and FBKP form a complex with rapamycin. This complex will bring together the NTEV and CTEV to form active TEV protease. (B) Is a phosphorylation induced dimerization in which an engineered SH2 domain and its phosphopeptide ligand form a complex in a phosphate-group dependent manner. (C) Both stimuli will bring together the NTEV and CTEV fragments to form active TEV protease.

First, we tested the **chemical switch (responds to rapamycin)**:

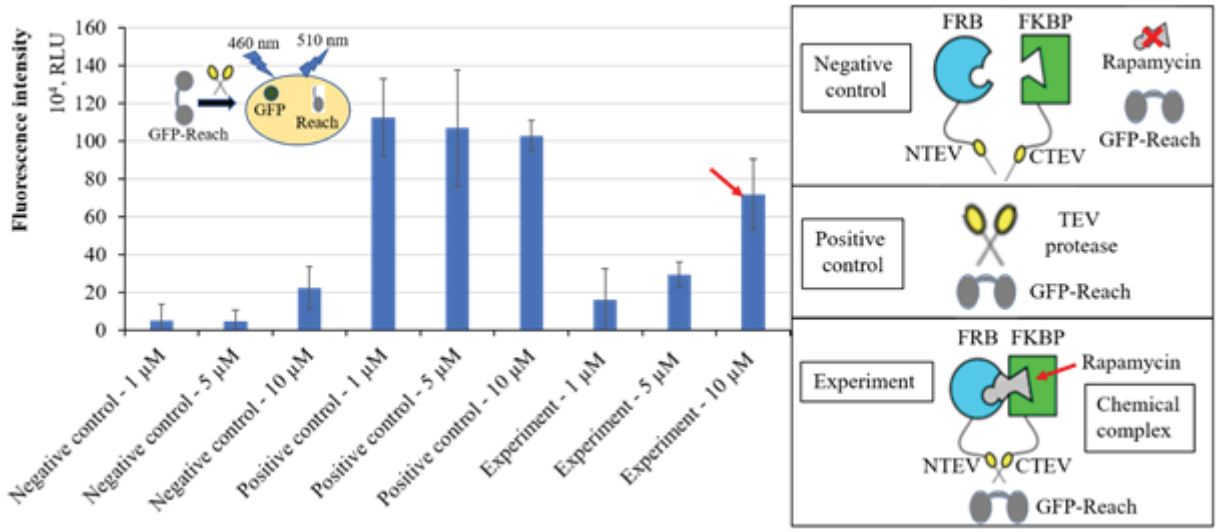


Figure 6: Chemical Induced Dimerization Activity Assay for OR Gate by Fluorescence Measurement

The samples were incubated overnight, and fluorescence intensity was measured using a plate reader. It was observed that for 10 μM of experiment sample there was 30% lower fluorescence intensity to corresponding positive control. The graphics on the right show the components of negative control, positive control, and experiment.

As shown above, we are able to see the conditional creation of TEV protease in response to the chemical.

2.3 Optimization of the Switch Performance using Dynamical Models

We developed ordinary differential equation (ODE) models to explore the dynamic response of this rapamycin switch. Based on this chemical reaction network, we ran a series of simulations varying the model parameters in order to optimize the concentrations of proteins and rapamycin. We were able to optimize the dynamic range of the switch by determining optimal protein and rapamycin concentrations. The modeling predicts that under certain conditions due to the binding properties of this system, the proteins will be saturated by the rapamycin, inhibiting protease reconstitution and fluorescence signal.

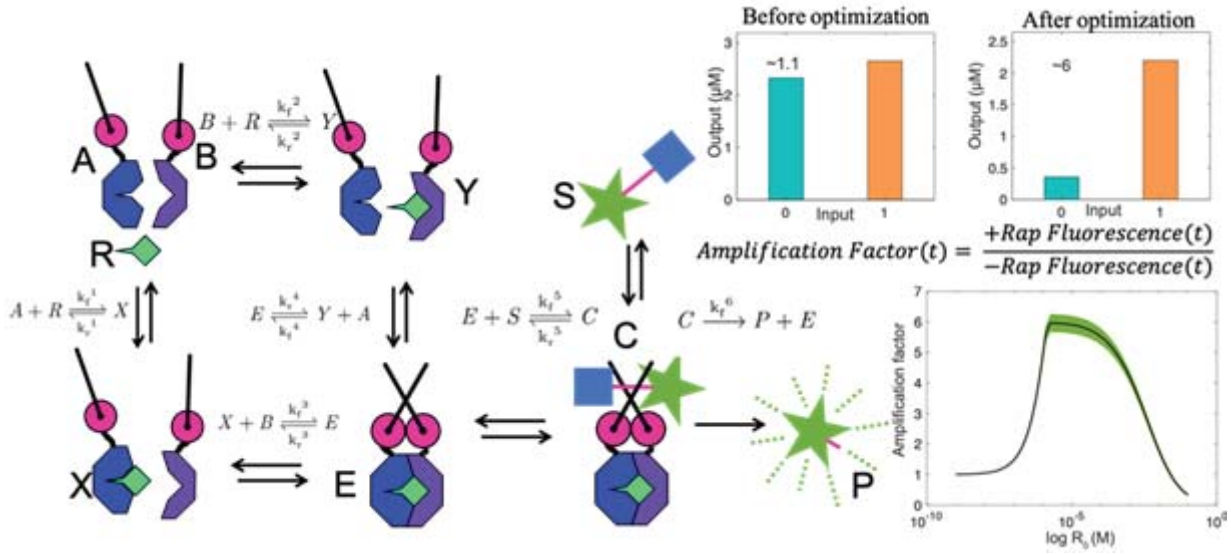


Figure 7: Optimization of the Rapamycin Switch Performance

We tested this prediction by varying rapamycin:protein ratio and found that it agreed with the modeling.

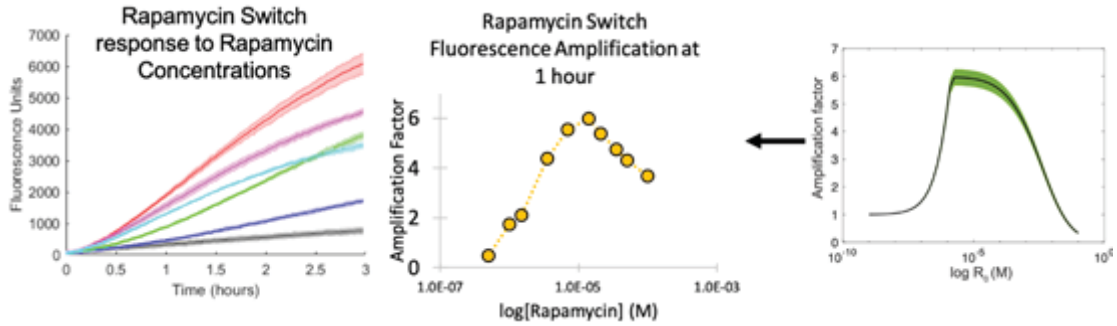


Figure 8: Rapamycin Switch Testing Results

Rapamycin concentrations were varied (grey = no rapamycin, blue= 3 μM, green = 5 μM, pink = 10 μM, red = 14 μM, cyan = 35 μM) at fixed concentrations (7μM) of the component proteins and protease activity was measured (left). Observed signal amplification factor as a function of rapamycin concentration (middle) and expected trend from modeling (right).

Similarly, we developed and optimized the phosphorylation-based protease switch:

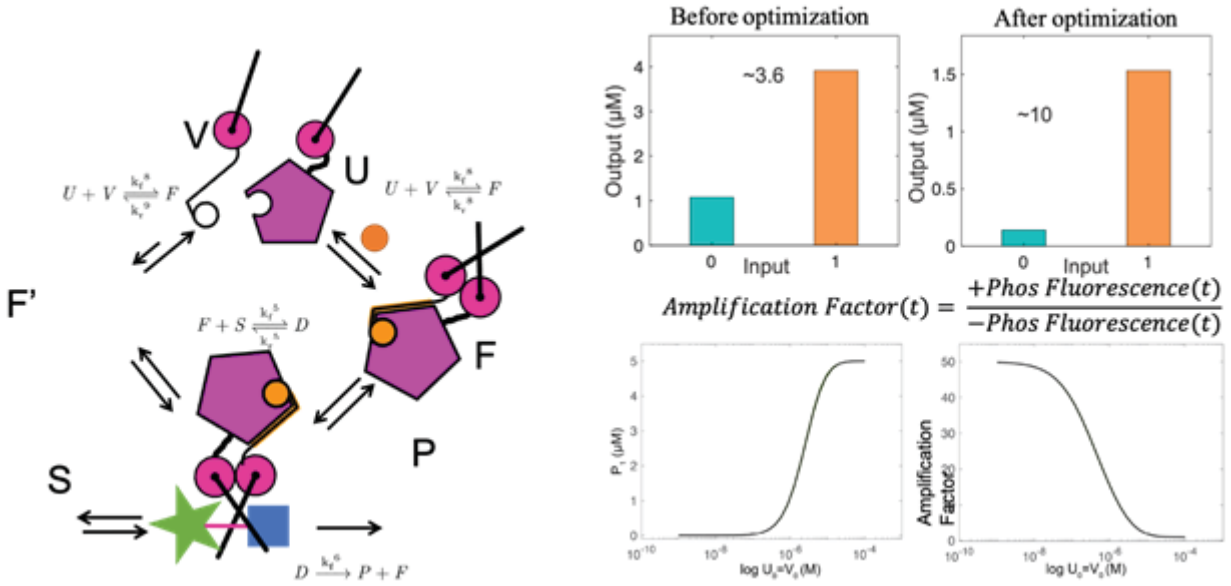


Figure 9: Optimization of the Phosphorylation Switch Performance

Experimental studies agreed with the predicted behavior and optimized response from the model:

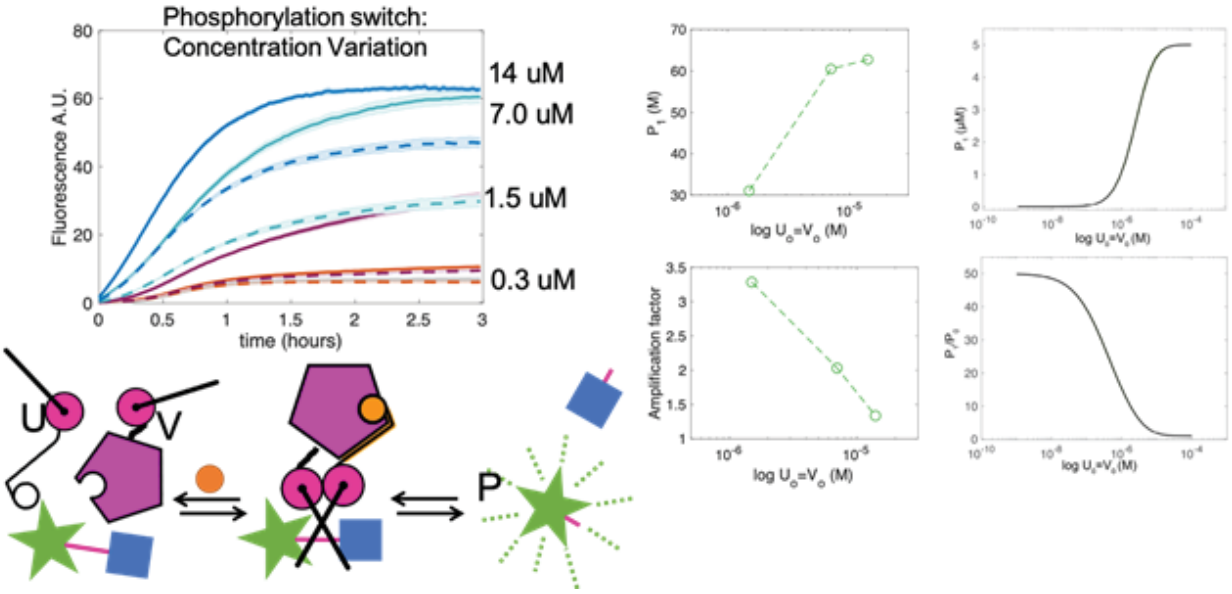


Figure 10: Phosphorylation Switch Testing Results

Performance of the switch at varying concentrations of components. Solid lines correspond to indicated concentrations and dashed lines in the identical color correspond to the observed signal with non-phosphorylated components at the same concentrations.

Thus mathematical modeling has informed the optimization of individual circuit elements.

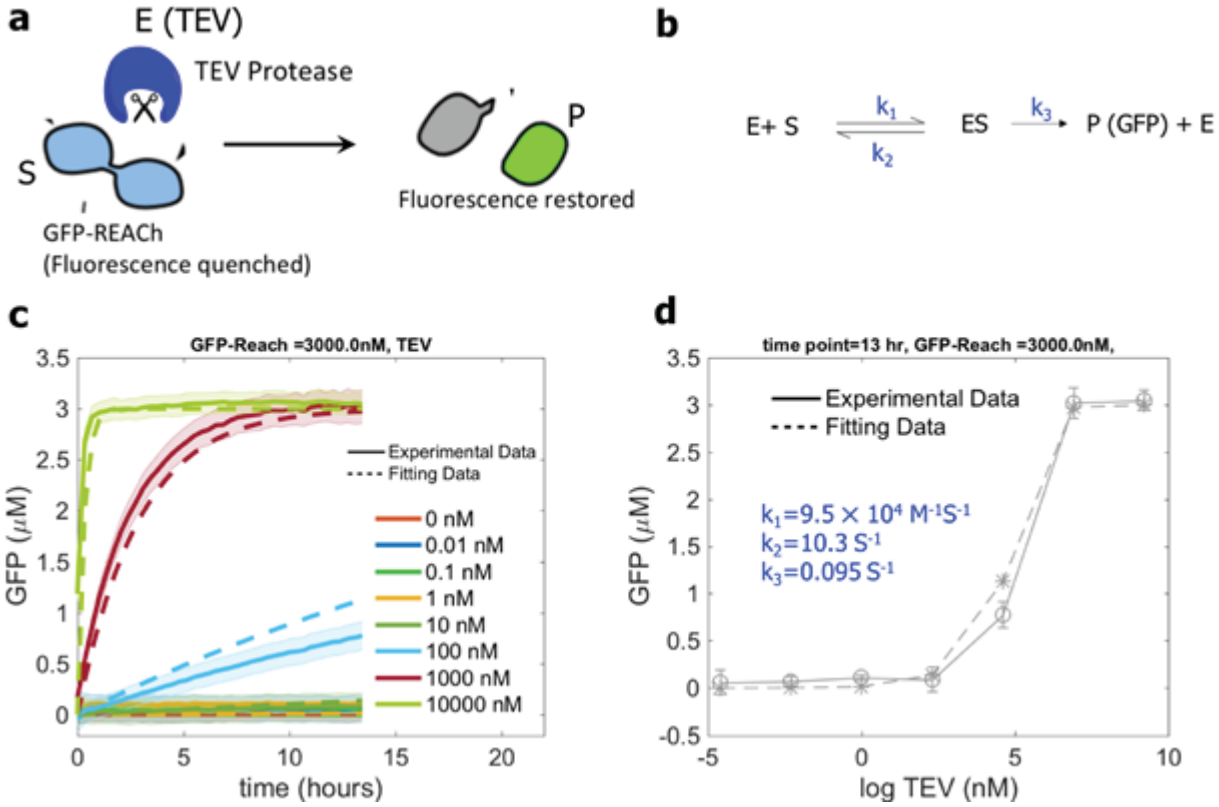


Figure 11: Modelling the FRET Based Readout Mechanism and Extracting the Reaction Parameters from the Experimental Data

*To detect the activity of a switch that uses protein-protein interaction induced by small molecule or light, FRET-based sensing mechanism is used where a recombinant intramolecular FRET construct consisting of GFP and REACh, fused by a recognition peptide sequence for TEV protease. TEV can cleave this sequence, causing GFP emission to increase. (a) Schematic diagram represents TEV and GFP-REACh interaction and the corresponding (b) chemical reaction equation. We derived ODE model from the chemical reaction equation and used it to extract parameters using least squares fitting of the experimental data. (c-d) Comparing the measured time domain response (solid lines) of (a) with the model data (dashed lines) at different concentrations of the TEV protease (E). (d) GFP response shown in (**c**) at t = 13 hours. With an increase in the TEV concentration, GFP expression increases and up to 3-fold increase in the output signal is observed.*

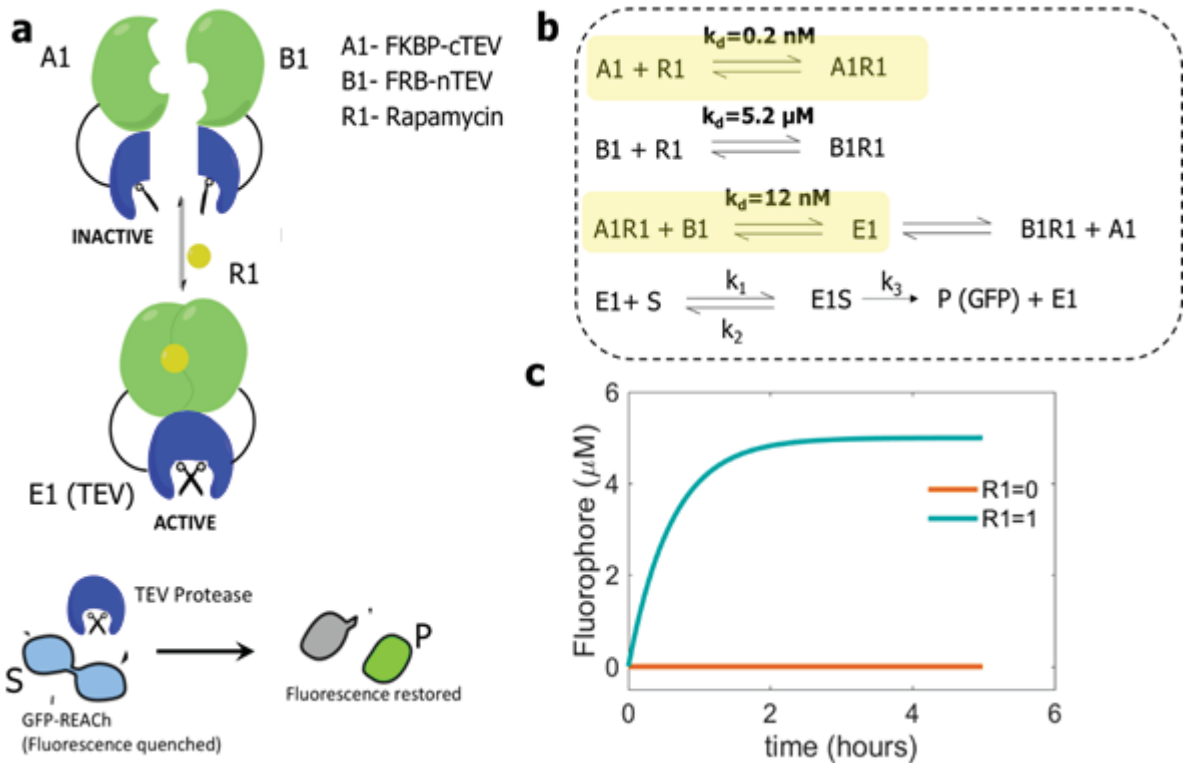


Figure 12: Modeling the Small Molecule Induced Switch

*(a) Schematic diagram represents the reaction network that involves the formation of the FKBP,rapamycin,FRB ternary complex which leads to the formation of TEV protease (E1) and the corresponding (b) chemical reaction equations. There are two possible pathways through which E1 complex can form. Because the interaction of FKBP to rapamycin is much stronger (k_d=0.2 nM) than the interaction of FRP to rapamycin (k_d=5.2 μM), only the highlighted pathway dominates in this reaction network. We derived ODE model from the chemical reaction equations and used it to simulate its response while using the reaction rates taken from the literature and determined through the data fitting. (**c**) In the absence of rapamycin, interaction between FKBP and FRB is almost negligible, and so the interaction between cTEV and nTEV fragments, resulting in zero fluorescence signal as the TEV protease cannot form. However, a high signal is detected when rapamycin is present. Here 0 and 1 correspond to without and with rapamycin respectively and k_d is the dissociation constant.*

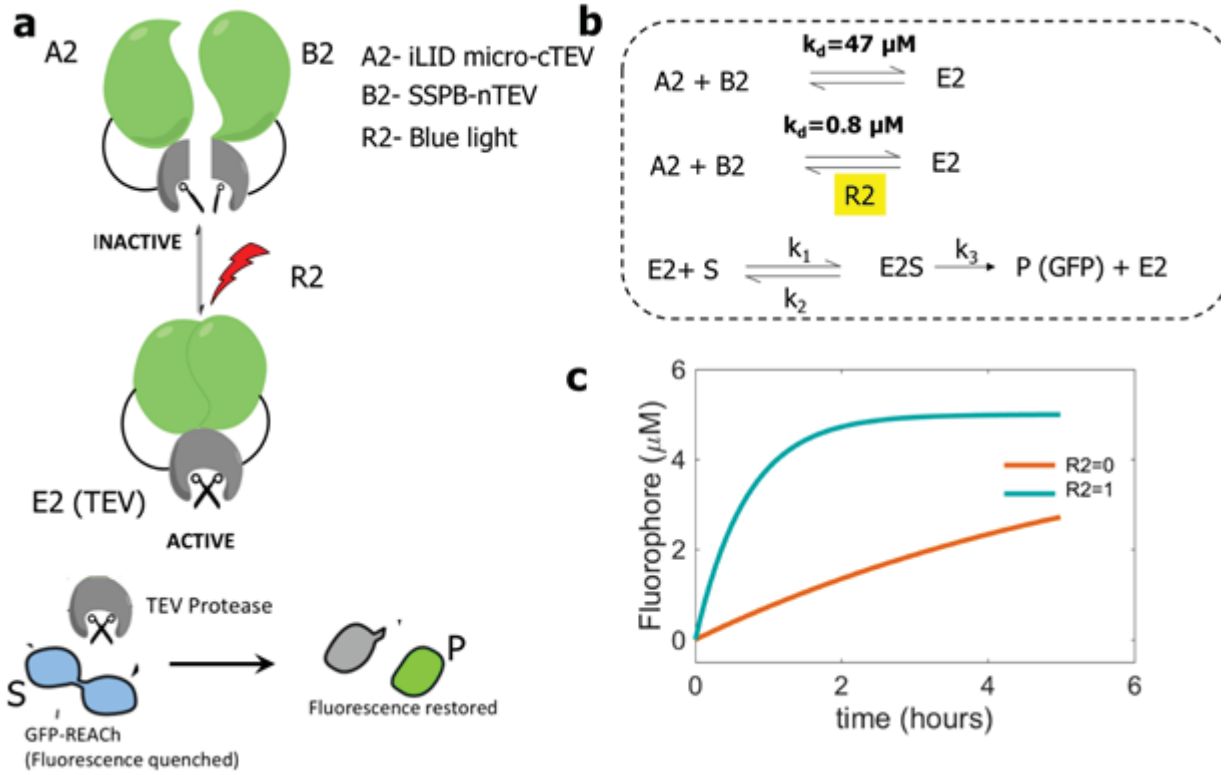


Figure 13: Modeling the Light-Induced Switch

(a) Schematic diagram represents the reaction network that involves the formation of the iLID micro and SSPB complex which leads to the formation of TEV protease and the corresponding (b) chemical reaction equations. We derived ODE model from the chemical reaction equations and used it to simulate its response while using the reaction rates taken from the literature and determined through the data fitting. (c) In the presence of blue light, the likelihood of the TEV formation increase by ~50 times leading to a higher GFP expression. Here 0 and 1 correspond to the absence and presence of the blue light respectively.

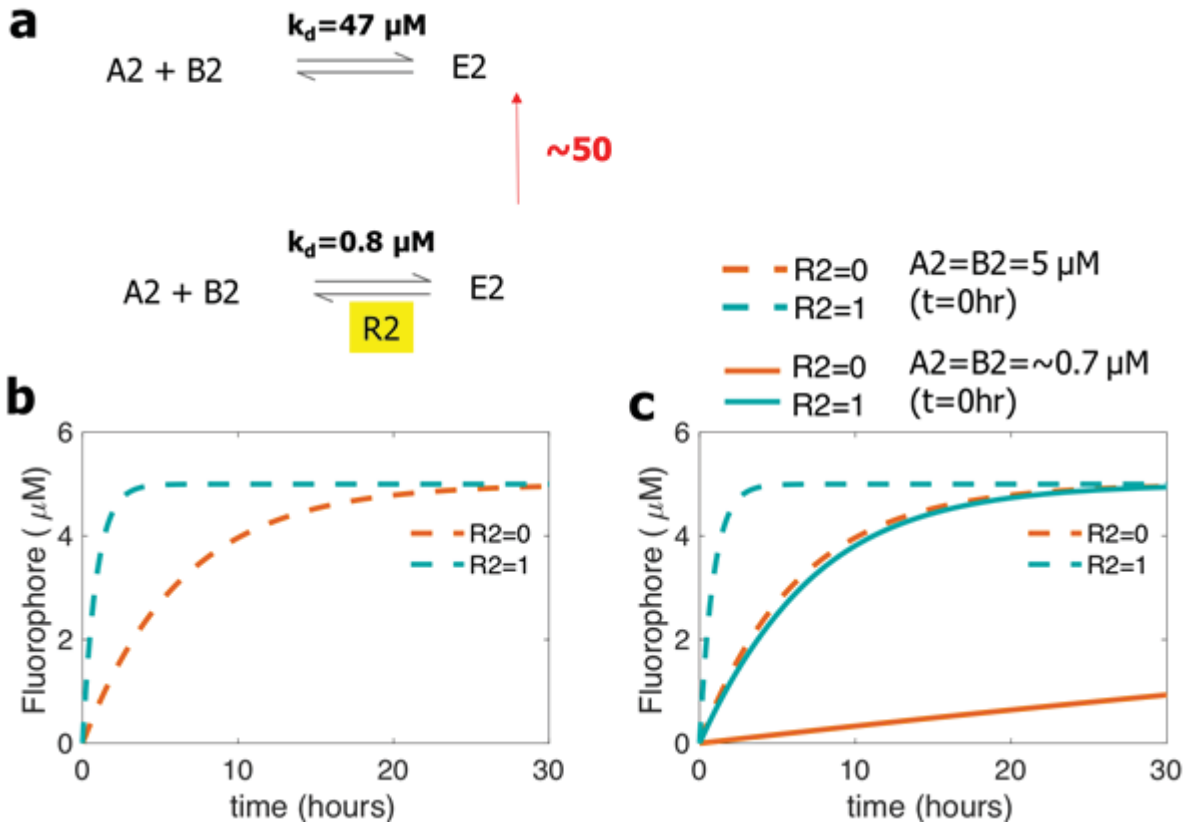


Figure 14: Improving the Dynamic Response of the Light-Induced Switch

(a) As per the current literature, the likelihood of the TEV complex formation can be increased by ~50 times when the blue light is present (k_d =0.8 μM) compared to when it is not (k_d =47 μM). (b) Because the TEV protease is still formed in the absence of the blue light, we observed a high fluorophore signal such that output activity of GFP is same in dark as in light. (c) Ideally we desire to have a very low signal in the dark compared to the light case. To achieve this, we optimized the initial concentrations of A2 and B2 using an optimizing algorithm, allowing to achieve a better response while considering the same reactions rates.

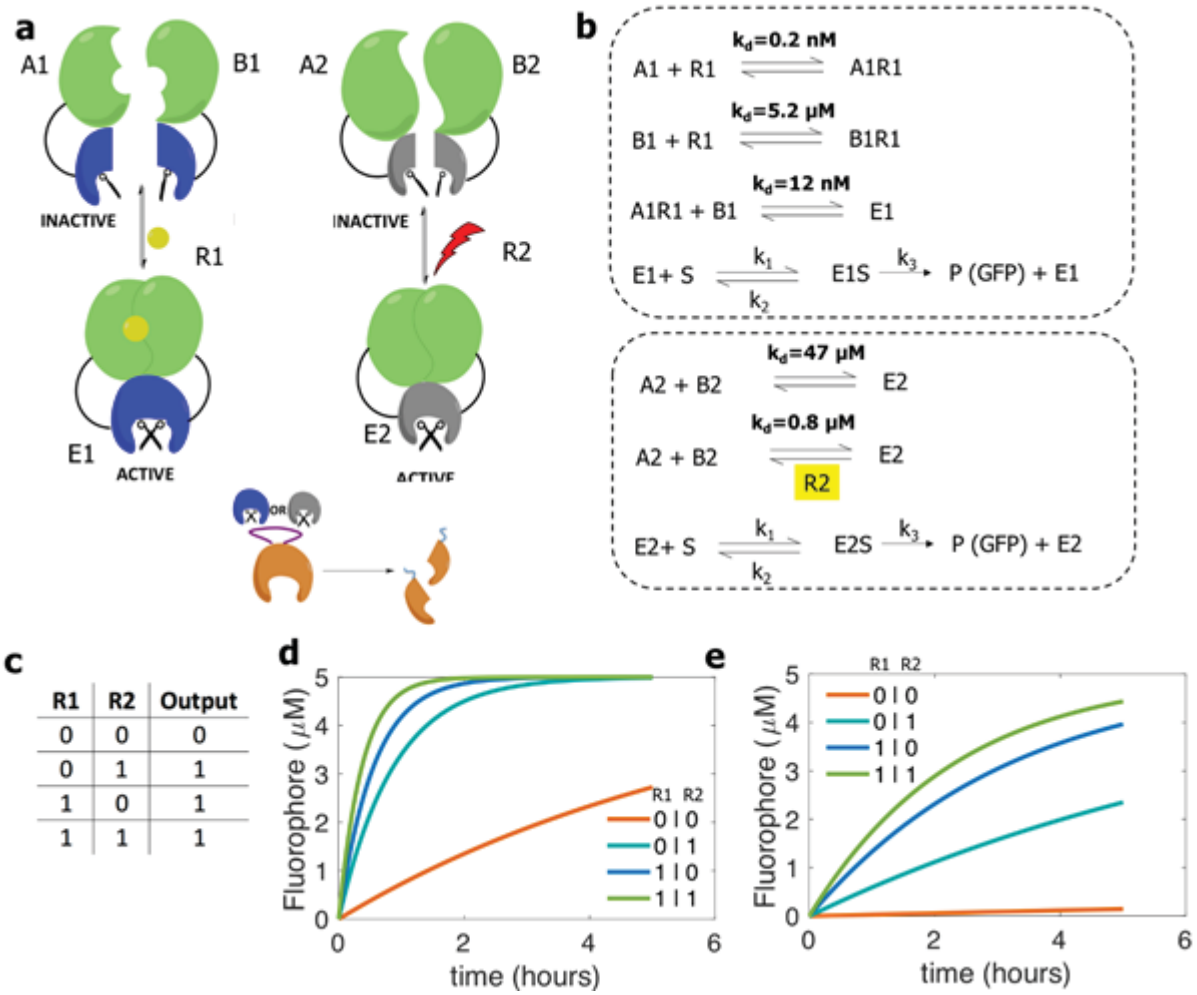


Figure 15: Modeling OR Gate

(a) Schematic diagram represents the reaction network of an OR gate that uses the small molecule and light-induced switches, and the corresponding (b) chemical reactions. Here R1 and R2 represent rapamycin and blue light respectively which are the inputs. (c) A high signal in terms of GFP expression results in when either or both the inputs are high demonstrating the operation of an OR gate. We derived ODE model from the chemical reaction equations and used it to simulate its response while using the reaction rates taken from the literature and determined through the data fitting. (d) When the initial concertation of A1, B1, A2 and B2 are 5 µM, a relatively high GFP expression is observed in the absence of both the inputs. Ideal output activity should be close to zero when R1 and R2 are zero. But because of E2 can form even when R2=0, a high fluorophore signal is observed. (e) Using the optimization algorithm we find that when the initial concertation of A2 and B2 are ~0.7µM, a relative high GFP expression is observed only when either or both the inputs are high.

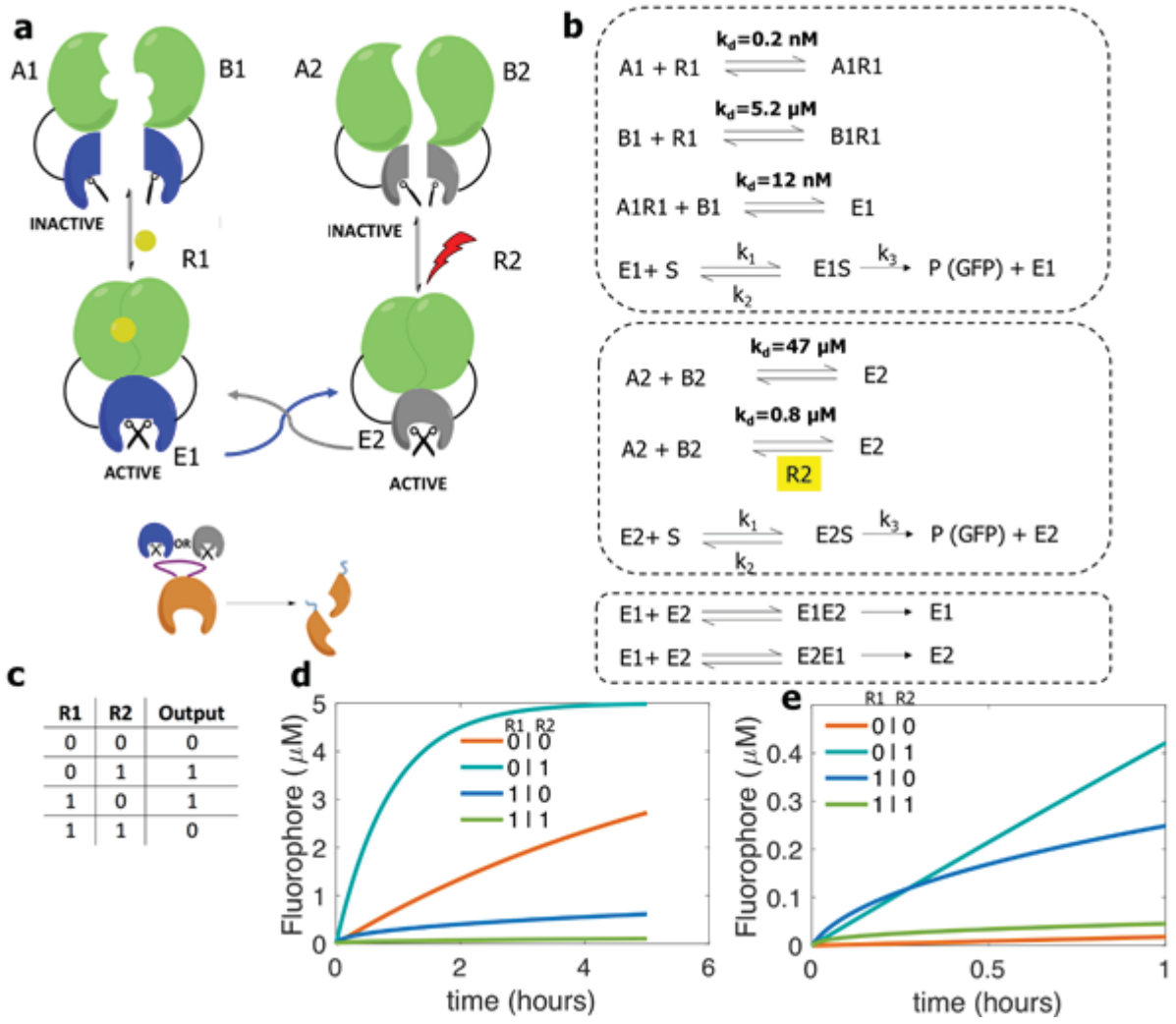


Figure 16: Modeling XOR Gate

(a) Schematic diagram represents the reaction network of a XOR gate that uses the small molecule and light induced switches and the corresponding (b) chemical reactions. Here R1 and R2 represents rapamycin and blue light respectively which are the inputs. (c) A high signal in terms of GFP expression results when only either inputs are high demonstrating the operation of an XOR gate. We derived ODE model from the chemical reaction equations and used it to simulate the response while using the reaction rates taken from the literature and determined through the data fitting. (d) When the initial concertation of A1, B1, A2 and B2 are 5 μM, a relatively low signal is observed when R1=1 and R2=0. Ideal output activity should be low when R1 and R2 are zero. But because E2 can form even when R2=0, E2 cleaves E1 such that there is no TEV protease left to detect, leading to a misleading output. (e) Using the optimization algorithm we find that when the initial concertation of A2 and B2 are ~0.7μM, a relative high GFP expression is observed when R1=1 and R2=0.

We designed and constructed an additional gate with **AND** integration using the same modules.

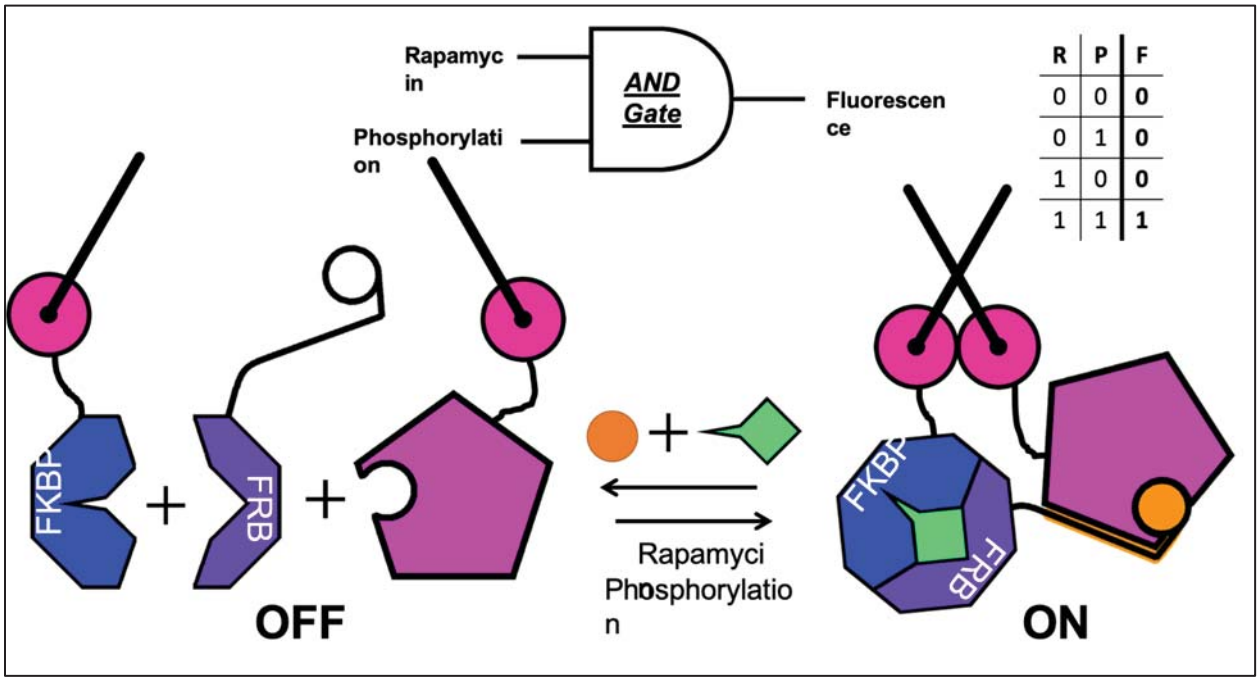


Figure 17: Overview Schematic of AND Gate

In AND protein logic gate, five components are necessary. Three components are proteins and two components are rapamycin and phosphorylation. Without the presence of both chemical inputs, the AND will not function. As shown in the schematic, the protein dimerization is formed between FRB-FKBP and SH2-pSH2 in the presence of rapamycin and phosphopeptide. This complex brings together the NTEV and CTEV fragments to form active TEV protease.

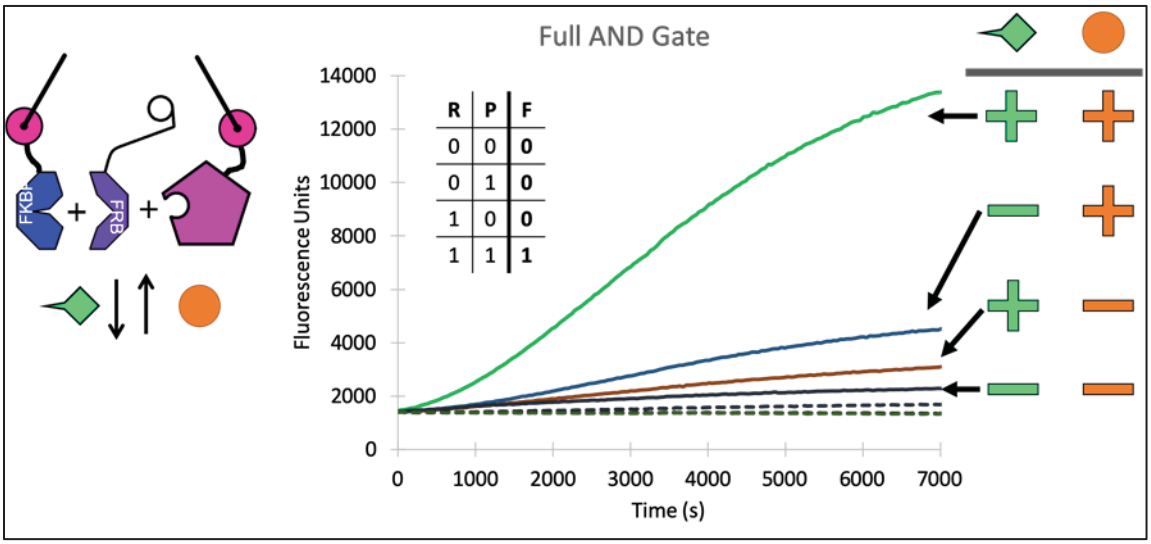


Figure 18: Experimental Realization of AND Gate

As **a general strategy to enhance the signal obtained from photoswitches**, we have developed a novel supramolecular approach which exploits the property of avidity, i.e., high local concentration to potentially generate non-linear signal amplification using the designed molecules:

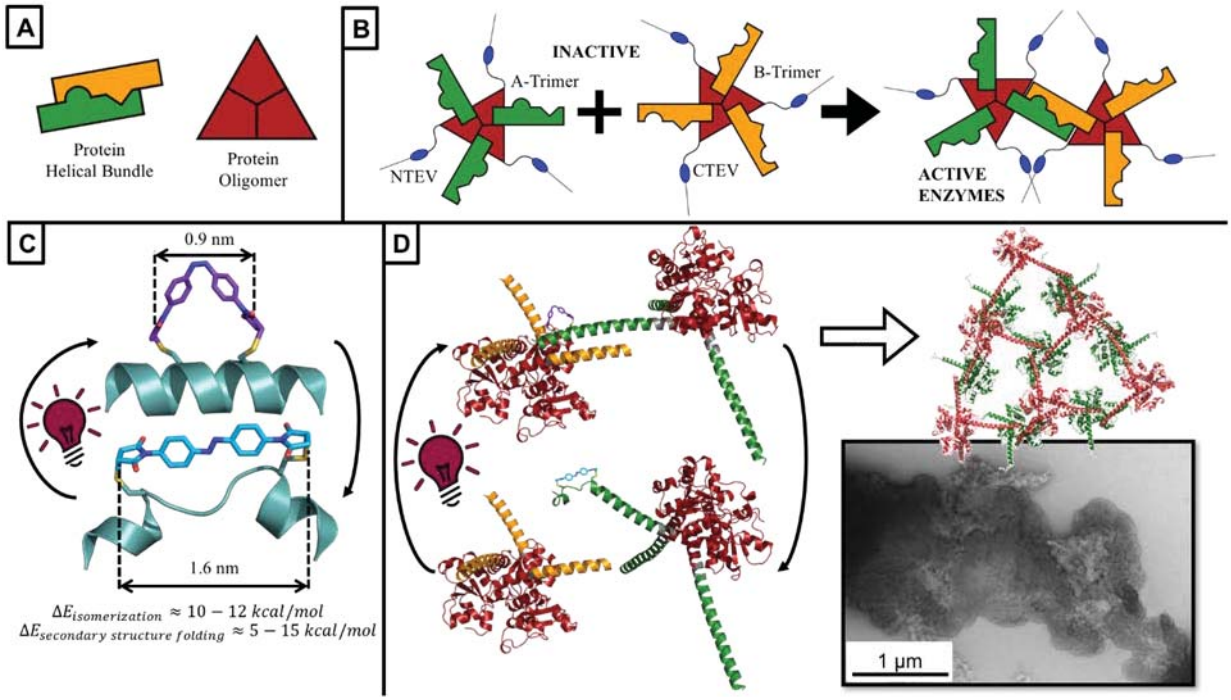


Figure 19: Enzyme Enhancement with Photoswitchable Assemblies

A) Utilizing an oligomerizing protein (the lp_1913 protein from Lactobacillus plantarum) and a heterodimeric binding domain (the coiled coil complex between MDB2 and p66alpha) (A) protein assemblies have been made through designed fusions. B) N-TEV and C-TEV protein sequences will be fused to these hetero-assembling proteins, enhancing N-TEV and C-TEV colocalization, binding, and activity. C) The photoswitching molecule azobenzene dimaleimide (ABDM) can be excited into its cis state with exposure to 370nm light. Previous work has demonstrated that secondary structural changes are possible with correct cysteine labeling with ABDM, followed by UV excitation. D) Labeling the coiled-coil binding domain with ABDM would permit UV control over protein assembly, colocalization, and enzymatic activity.

We are currently experimentally building these types of novel photoswitchable split enzymes.

3. CONCLUSIONS

In conclusion, our studies show that it is feasible to experimentally construct enzyme-based switches and develop computational modeling tools that can help analyze and optimize the parameters of designed enzyme-based circuitry. Using parameter sensitivity analyses, the design of future experiments will be guided and control theoretic properties of these circuits will be investigated. With the use of induced supramolecular clustering, we expect that the signal:background will also be increased. Overall, the combination of protein design and computational modeling will be further expanded to construct a general framework to rapidly, specifically and simultaneously detect chemical inputs and integrate them in sophisticated ways for biodetection and bioremediation of chemical and biological threats and other molecules of interest.

Analysis of costs: The costs were incurred in PI summer salary, personnel (1.5 FTE postdocs, 1 FTE graduate student, 1 Half-time research assistant), and the purchase of synthetic genes, laboratory chemicals and supplies for protein expression, purification and characterization.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

ACRONYM	DESCRIPTION
ABDM	azobenzene dimaleimide
AFRL	Air Force Research Laboratory
cTEV	C-Terminal part of tobacco etch virus
DoD	Department of Defense
E. coli	Escherichia coli
FKBP	FK506 binding protein
FRB	FKBP-rapamycin-binding
FRET	Förster resonance energy transfer
FTE	Full time engineer
GFP	Green fluorescent protein
iLID	Improved light-induced dimer
K_d	Dissociation constant
kDa	kilodalton
MBP	Maltose-binding protein
mg	milligram
nM	Nanomolar
nTEV	N-terminal part of tobacco etch virus
ODE	Ordinary Differential Equation
pSH2	Plasmodium falciparum specific helicase 2
REACh	Resonance Energy-Accepting Chromoprotein
SH2	Src homology 2
sspB	Stringent starvation protein B
TEV	Tobacco etch virus
UV	Ultra-Violet