UNIVERSITY OF CALIFORNIA

Santa Barbara

Engineer Single-Fluorescent Protein Biosensors Towards Anaerobic Fluorescence Imaging

A Thesis submitted in partial satisfaction of the

requirements for the degree Master of Science

in Chemical Engineering

by

Kang-Ching Fan

Committee in charge:

Professor Arnab Mukherjee, Chair

Professor Michelle O'Malley

Professor Siddharth Dey

March 2021

The thesis of Kang-Ching Fan is approved.

_____

Siddharth Dey

_____

Michelle O'Malley

_____

Arnab Mukherjee, Committee Chair

March 2021

Engineer Single-Fluorescent Protein Biosensors Towards Anaerobic Fluorescence Imaging

by

Kang-Ching Fan

ACKNOWLEDGEMENTS

I deeply acknowledge the patient mentoring and academic guidance from my advisor, Prof. Arnab Mukherjee, and for his enthusiasm in promoting the fearless creation of intellectual ideas in our lab. I acknowledge deep support from my close friends Chung-Ta, Chih-Cheng, and Tzu-Yu, for sustaining me with warm caring through the sleepless nights facing the challenges and difficulties in life and science. I acknowledge my lab members Harun, Jason, Nolan, Raymond, Emily, and Austin for their insightful inputs into my work and also providing opportunities to improve my scientific communication skill, also as great ears to share the ups-and-downs of daily lives as an international student. I'm grateful to my undergraduate apprentices Sumedha, Kevin, Charles, and Audrey for their tireless contribution to make this work possible. I'm particularly grateful for my research mentors Prof. Max Wilson, Prof. Michelle O'Malley, Prof. Siddharth Dey for supervising the scientific quality of the work with their science profession and passion over the past few years. Finally, I express my deepest gratitude to my parents, Ching-Hua and Zheng-Ming for encouraging me to fulfill my dreams and ambitions.

ABSTRACT

Engineer Single-Fluorescent Protein Biosensors Towards Anaerobic Fluorescence Imaging

by

Kang-Ching Fan

In this work, I describe the early attempts to engineering biosensors using a flavin-based fluorescent protein LOV. LOV proteins outperform fluorescent probes based on the green fluorescent protein (GFP) in anaerobic conditions and have the potential to serve as ligand-responsive allosteric reporters. Applications of LOV proteins to this end, however, have been limited due to a lack of understanding of engineerable sites within their protein structures. I systematically addressed this problem with the exhaustive library construction of domain-insertion chimeras using estrogen receptor ligand-binding domain (ERLBD) with iLOV. First, I extensively insert ERLBD into every single site of iLOV; the G480 site, which resides in the Hβ-Iβ loop within iLOV, retains 10% of the native iLOV fluorescence and exhibit a moderate increase in fluorescence upon ligand-binding ($\Delta F/F = 0.2$). I also performed independent theoretical prediction of engineerable sites based on a homology-based model, an ensemble allostery-based model, and a machine learning model for protein function classification; all of which reached consensus on the experimentally validated sites. Next, based on the identified engineerable sites within iLOV, I swapped the ERLBD with an ATP-binding protein to engineer an ATP sensor. However, direct insertion didn't yield a workable sensor; therefore, I introduced random short peptide linkers flanking the inserted domain. Furthermore, I

developed a potassium cyanide (KCN)-based assay for ATP sensor screening. Although my screening assay didn't allow the identification of a bona fide ATP sensor, I pointed out potential solutions regarding the control of cofactor availability, reduction of background fluorescence, and the use of an internal standard to achieve a streamlined screening process. The LOV-based biosensor is still in its early stage of development. My work presents a forefront framework to engineer LOV-based protein switches that may complement to GFP-based imaging in underexplored anaerobic biological systems.

TABLE OF CONTENTS

CHAPTER 1: BACKGROUND AND SIGNIFICANCE

Fluorescent proteins have enabled imaging of molecular-scale events required for signal transduction in biological systems [1]. In particular, the green fluorescent protein (GFP) and its other color variants have been widely adopted as a basic molecular biology tool to study cellular functions since GFP was discovered and optimized in the 1980s [2]. To be used as a "reporter", GFP is genetically coupled with endogenously derived proteins (*i.e.,* the sensing units) that participate in signal transduction pathways that allow quantifiable fluorescence change in response to changes in the states of the sensing unit. Such molecular "spies" play a major role in the elucidation of cellular events, which may help us engineer biological systems with applications such as the manufacturing specialty chemicals [3,4] or the detection of disease markers [5].

Fluorescence-based methods have proved to be indispensable in biological science because of their inherent high sensitivity of detection. Although GFP has long been considered the working horse in developing genetically-encodable biosensor, the strict requirement of molecular oxygen using GFP as the reporting unit sets the limit in the context of **anaerobic environments**, which can be mostly found in pathogenic anaerobic bacteria and fungi [6–8] or the microenvironment of highly dividing tumor cells [9]. To selectively label the molecule of interest *in vivo* for imaging purposes, several techniques including synthetic molecules with functional handles for site-specific labeling [10,11], incorporation of non-canonical amino acids that enables detection of chemical signals [12,13], and directed evolution of proteins that bind synthetic chromophores [14–16] seem, at the first glance, solve the problem of the oxygen requirement of GFP. However, despite having signals with a high signal-to-noise ratio using

the small molecule-based approaches, this advantage might be compromised due to low cell permeability and potential toxicity issues. Moreover, the dilution effect arising from cell proliferation may prohibit quantitative imaging for a long duration. **The lack of genetically encodable biosensors having oxygen-independent fluorescence response makes it challenging to study and engineer anaerobic organisms.**

We propose the use of **light-oxygen-voltage sensing proteins (LOV)** as a new molecular chassis for engineering genetically encodable biosensors for anaerobic imaging. LOV is an emerging class of fluorescent protein that uses flavin mononucleotide (FMN), an essential cofactor that ubiquitously appears in all kingdoms of lives, as the chromophore (**Figure 1**). LOV can be used for anaerobic imaging because unlike GFP, the chromophore maturation does not require molecular oxygen and is highly stable at extreme pH [17–19]. We thus **hypothesize** that using LOV protein as the reporting unit enables the construction of small molecule biosensors in anaerobic environments. The brief guidelines of our proposed aims are listed as follows:

Aim 1. Develop a generalizable molecular framework for engineering LOV-based fluorescent proteins. (Chapter 3)

The understanding of the basic structural-function relationship of proteins facilitates the engineering of metabolite biosensors. Existing GFP-based biosensors utilize protein backbone hotspots found serendipitously for protein engineering including protein insertion or dissection [20]. Since there is no existing attempt that tests the effect of manipulating the LOV backbone on its function, we suppose having a basic understanding of LOV may be a prerequisite for further applications. We investigate the allosteric effect of LOV proteins-

based domain-insertion chimeras using a systematic insertion approach. Independent computational techniques were also performed to justify the experimental results.

Aim 2. Build an ATP sensor as a proof of concept of modular domain-insertion to understand the energetics in anaerobes. (Chapter 4)

The cellular energetics of anaerobes is not well understood due to the limitations of enabling technology. We aim to develop an adenosine triphosphate (ATP) sensor based on our understanding of LOV protein as a proof of concept of our methodology in aim 1. We expect the use of time-lapse fluorescent imaging can be coupled to the use of the ATP sensor to test a hypothesis related to antibiotic evasion phenotype formation under a low oxygen environment.

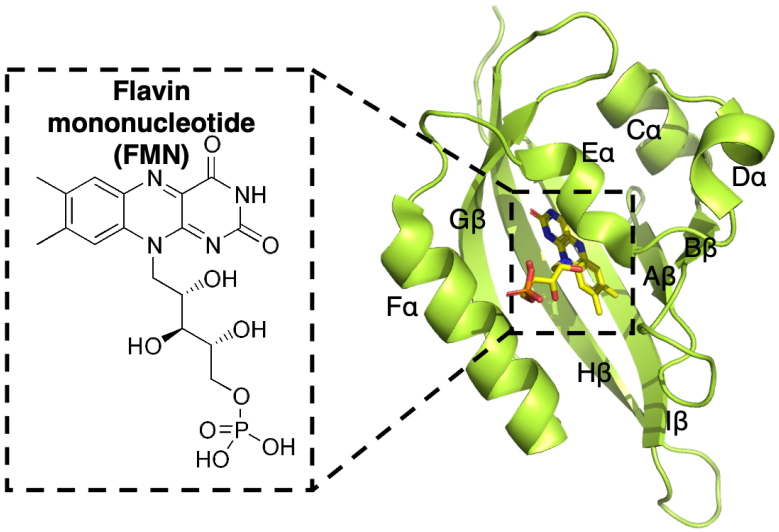

Figure 1. Schematics of iLOV (PDB: 4EES). LOV proteins utilize flavin mononucleotide (FMN) as the cofactor.

CHAPTER 2

MATERIALS AND METHODS

2.1 Chemicals, reagents, plasmids, and bacterial strains

2.1.1 Chemicals

The following chemicals were purchased from Fisher Scientific (Pittsburgh, PA): Lennox broth (LB), Agar, Disodium hydrogen phosphate (Na_2HPO_4), Potassium dihydrogen phosphate (KH_2PO_4), Potassium cyanide (KCN), Ammonium Chloride (NH_4Cl), Glucose, Magnesium Sulfate, Calcium Chloride, Casamino acids, Dimethyl sulfoxide (DMSO), B-PER bacterial extraction reagent, lysozyme. Antibiotics including ampicillin and kanamycin, and isopropyl β-D-1-thiogalactopyranoside (IPTG) were purchased from GoldBio (St. Louis, MO). (Z)-4-Hydroxytamoxifen was purchased from Tocris (Minneapolis, MN). Buffers and enzymes for Gibson assembly (HIFI master mix) and routine polymerase chain reaction (Q5 master mix) were purchased from New England BioLabs (Ipswich, MA). DNA oligos were synthesized from Genewiz (South Plainfield, NJ) or IDT (Coralville, IA).

2.1.2 Bacterial strains

E. coli DH5α (also known as NEB10) was used for cloning and plasmid propagation while E. coli BL21 (DE3) was used for protein expression. The genotype of each strain is shown in **Table 1**.

Table 1. The genotype corresponding to the bacterial strains used in this study.

Strain	Genotype
E. coli DH5α	F⁻ *endA1 gln V44 thi-1 recA1 relA1 gyrA96 deo R nupG purB20 φ80dlacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ⁻
E. coli BL21 (DE3)	F⁻ *ompT gal dcm lon hsd*S_B($r_B^- m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*⁺]$_{K\text{-}12}$(λ^S)

2.1.3 Plasmid design and construction

Plasmids were constructed using Gibson Assembly [21]. The gene sequences of iLOV-ERLBD chimeras were encoded in vector pQE80L-iLOV [9]. The gene of the human estrogen receptor ligand-binding domain (ERLBD, amino acid residue 302-552) was PCR amplified from pcDNA-HA-ER WT; pcDNA-HA-ER WT was a gift from Sarat Chandarlapaty (Addgene plasmid #49498; http://n2t.net/addgene:49498; RRID: Addgene_49498). A FRET-based ATP sensor, ATeam1.03, was a gift from Takeharu Nagai (Addgene plasmid #51960; http://n2t.net/addgene:51960;RRID:Addgene_51960 [22]). All plasmid design was aided by ApE and SnapGene Viewer.

2.2 Bioinformatic analysis

2.2.1 Identification of permissible domain-insertion sites from homology models

We identifed the surface-exposed amino acid residues by computing the solvent-accessible area (SAA) using Stride [23]. The positions with SAA larger than 30 Å^2 are considered to be surface-exposed. To identify evolutionarily non-conserved sites, the sequence alignment of over 400 iLOV homologs was fed into the MISTIC server [24]. MISTIC adopts the Kullback-Liebler (KL) conservation metric that accounts for the frequency of amino acids in a given position and the natural background frequency of the amino acids; KL less than 2 is considered

non-conserved. The crystal structure of iLOV (PDB: 4EES) was fed to the CMView program to generate a 2D contact map to identify tight loops. 3D molecular models were generated using molecular visualization software PyMOL (Schrodinger).

2.3 Molecular cloning techniques

All molecular cloning involved in this work follows the standard protocols of Gibson assembly. The inserted domain DNA sequence was PCR amplified using primers designed using NEBuilder (New England BioLabs) bearing complementary overhangs to the backbone domain insertion sites. Briefly, PCR was carried out in 25 μL reaction volume using 1 ng template DNA, 0.5 μM primers, and Q5 High-Fidelity 2X Master Mix (New England BioLabs). The PCR cycle consists of an initial denaturation at 98°C for 30 seconds followed by 30 cycles of 98°C for 10 s, annealing temperature for 20 seconds, and 72C for 20 s/kb. A final extension at 72°C for 2 min was used to complete the full-length synthesis. PCR amplicons were gel-purified with Monarch DNA Gel Extraction Kit (NEB). DNA fragments were assembled by mixing the insert and backbone fragments in a molar ratio of 2:1 followed by incubation at 50°C for 20 minutes. All plasmid constructs were transformed into E. coli DH5α using heat shock at 42˚C or electroporation at 1.8 kV using MicroPulser Electroporator (Biorad). Cells were plated onto LB agar plates supplemented with ampicillin or kanamycin overnight for selection. Plasmids were when isolated with PureYield Plasmid Miniprep System (Promega) and used to transform E. coli BL21 (DE3) for protein expression studies.

2.4 Protein expression and screening

2.4.1 Protein expression and screening for potential biosensors

DNA was transformed into chemically or electrocompetent *E. coli* BL21 (DE3) cells and plated onto LB agar containing 100 μg/mL ampicillin (Amp). Single colonies from the plates were used to inoculate 5 mL of LB broth supplemented with Amp (100 μg/mL) and grown overnight while being shaken at 37°C, 250 rpm. The following day, the overnight cultures were diluted 1:100 into 5 mL M9 minimal supplemented with 100 μg/mL Amp, 0.01% casamino acids, and 0.4% glucose as the carbon source. Cells were grown at 37°C with shaking until an optical density at 600 nm (OD_{600}) of 0.4-0.6 was reached. The cultures were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. Protein expression was continued for 5-6 hours at 37°C, followed by harvesting of the cells by centrifugation at 4,122 g for 10 minutes at 4°C. The cell pellets were washed with PBS twice (supplemented with ligand or vesicle control if necessary) and were resuspended into screening buffer (*i.e.*, PBS with ligands or vesicle control) before any *in vivo* fluorescence assays.

For *in vitro* assays, the PBS washed cell pellets were resuspended with 400 μL B-PER bacterial protein extraction reagent (Thermo Fisher) supplemented with lysozyme (200 μg/mL). Cells were lysed at room temperature for 30 minutes. The lysates were clarified by centrifugation at 17,000 g for 5 minutes. 80 μL of clarified lysates were pipetted into Costar 96-well black plate (Corning). Each aliquot was mixed with 80 μL of either Dimethyl sulfoxide (DMSO) or 10 μM (*Z*)-4-Hydroxytsmoxifen (4-HT) (dissolved in DMSO) to produce a final concentration of 5 μM 4-HT. The plate was centrifuged at 1,500 g for 10 minutes to remove bubbles. Fluorescence emission spectra were recorded between 470 nm

and 600 nm using a fluorescence spectrometer (Tecan Spark). Fluorescence intensities were normalized by the absorbance at 280 nm. Signal fold changes were calculated based on biological triplicates by dividing the integrated fluorescence signal with added 4-HT by the integrated fluorescence signal without added 4-HT.

2.4.2 Comparison between in vivo assay and the lysate-based screening

The initial attempt used the lysate-based method [25] to maximize the concentration of sensors available to 4-HT. However, *in vivo* assay is considered generally better than *in vitro* lysate-based screening. The lysis of cells might render the release of flavin from the native flavoproteins and can thus complicate the fluorescence analysis. Previous studies have shown that E. coli cells are permeable to 4-HT and β-estradiol [26,27]. By adding 4-HT and DMSO control in all the culture medium and the screening buffer, we measured the whole-cell fluorescence intensity (FI) normalized by cell density ($OD_{600\ nm}$) for each construct to identify domain-insertion hotspots with potential ligand-induced fluorescence change.

2.4.3 Screening criteria for ERLBD-iLOV chimeras

We simultaneously tested the basal level of fluorescence intensity, the fluorescence intensity after 4-HT is added, and the fold-change in fluorescence intensity. We defined a "positive hit" as one with fluorescence relative intensity fold-change ($\Delta F/F$) larger than twice the coefficient of variation (CV) of the iLOV fluorescence intensity. If the fluorescence intensity of a given construct remains at least 10% of that of iLOV but does not have significant change upon adding ligand, it is classified as "neutral". Any construct having less than twice the fluorescence intensity of the ER-LBD only background fluorescence is classified as "disruptive".

2.4.4 KCN assay for ATP sensor screening

The intracellular ATP level was quantified using commercial luminescent assay BacTiter-Glo (Promega) and a benchmark ATP sensor (ATeam1.03) [22]. Inoculate 8 μL overnight preculture into freshly made M9 media supplemented with Kanamycin followed by 5 hours growth at 37°C with shaking at 250 rpm; longer protein expression time is avoided to prevent bacterial cells from entering a stationary growth phase, during which the intracellular ATP level is lower than that during log phase [28]. For the KCN treatment test, 20 μL of 50 mM KCN solution or 20 μL PBS control was added into black 96-well plates followed by the addition of 80 μL cell cultures. The plates were sealed with air-permeable films to prevent water evaporation; the plates were then incubated at 37°C for 10 min before fluorescence measurements.

2.4.5 Fluorescence spectroscopy

Fluorescence spectroscopy was performed using optically clear, flat-bottom black 96-well plates (Corning). The Tecan Spark fluorometer was used for making 96-well plate measurements, where the excitation and emission slit widths were set at 5 nm and 20 nm respectively, and the gain was adjusted to 80. We reported all the fluorescence intensity (FI) data using the following equation:

$$\mathrm{FI}=\frac{\mathrm{FI}_{\mathrm{raw}}-\mathrm{FI}_{\mathrm{blank}}}{\mathrm{OD}_{\mathrm{raw}}-\mathrm{OD}_{\mathrm{blank}}} \quad (1)$$

where raw fluorescence intensities were corrected with blank measurements and divided by optical densities (OD) at 600 nm for normalization. In our KCN assay for ATP sensor screening, we chose the ODs from non-KCN treated control for normalization since KCN can cause an undesirable shift in ODs, making it difficult to quantify cell densities.

CHAPTER 3

DOMAIN-INSERTION ALLOSTERY OF LOV

3.1 Introduction

Domain-insertion is defined as the transfer of coding sequences of uniquely folded structures from one to another and was found naturally through protein evolution [29]. Such protein engineering strategy has been used in building protein switches with potential applications such as biosensors and optogenetic tools for gene editing [30–33]. A typical ligand-activated switch consists of a ligand-binding domain, which is inserted into an output domain whose function is modulated in response to ligand-binding. To facilitate the development of small-molecule biosensors, a modular switch platform is desirable [34]; a fusion topology that is suitable for sensing a particular ligand can be adapted to respond to any other kinds of inputs by changing the input domains with different cognate ligands.

We aim to build a generalizable correlation between protein local structure and domain-insertion permissibility using a light-oxygen-voltage protein (LOV) chassis derived from *Arabidopsis thaliana* (iLOV) [35]. A well-characterized human estrogen receptor ligand-binding domain (ERLBD) was chosen to be systematically inserted into every single sit of amino acid residue within the iLOV backbone. The binding of cognate ligands 4-hydroxytamoxifen (4-HT) or β-estradiol to ERLBD induces a large conformational change in ERLBD structure [31,36,37]. Using domain insertion, the ERLBD can transduce its change in conformation to iLOV and subsequently reduce or enhance the affinity of FMN binding to LOV, thus modulating the fluorescent intensity. The small size of LOV protein (10-15 kDa) compared to other

fluorescent proteins makes it an ideal candidate to rationally design protein chimeras for studying the correlation of structure and domain-insertion permissibility. Sites that are permissible to domain-insertion without compromising the overall protein stability are identified as "hotspots", where different ligand-binding domains may be inserted to build an array of biosensors with different cognate analytes. An ideal sensor should exhibit a low basal level of fluorescence intensity and high signal fold-change upon analyte binding. The aforementioned characteristics can be optimized by varying the length and composition of the peptide linkers flanking the inserted domain using random mutagenesis-based methods [38–41] (**Figure 2**).

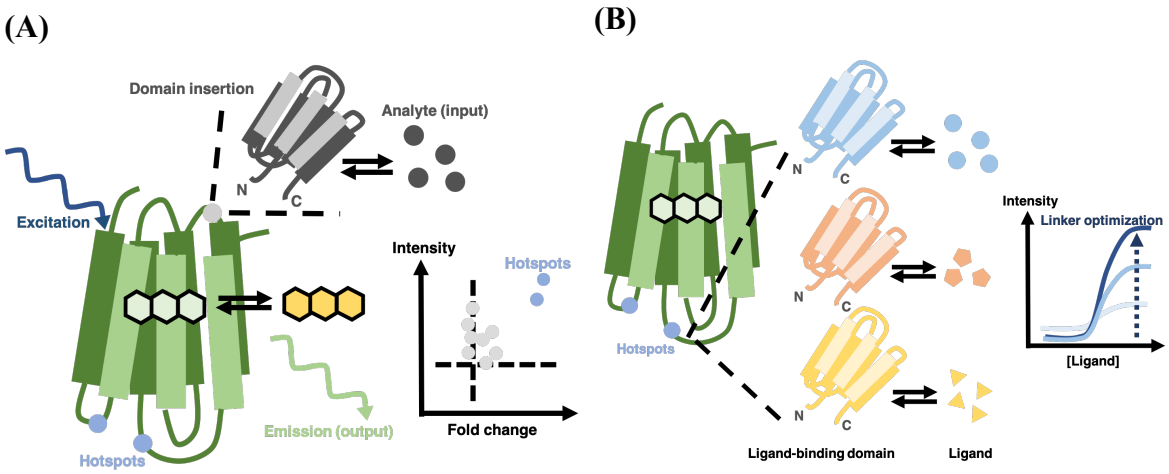

Figure 2. Schematics of testing domain insertion permissibility and allosteric response. (A) A ligand-binding domain (in gray) is systematically inserted into a flavin-based fluorescent protein, LOV (in green). The binding of ligand of the inserted domain triggers a conformational change of the LBD, which is then transduced into the LOV domain, changing the affinity of LOV to its chromophore, flavin. The sites inserted that render LOV to restore the fluorescence are identified as hotspots for domain insertion applications. (B) A generalizable molecular framework for generating metabolite sensors is shown accordingly.

3.2 Potential pitfalls and alternatives

3.2.1 Estrogen receptor ligand-binding domain may have limited solubility

We used the bacteria *Escherichia coli* (*E. coli*) as the expression system for biosensors. However, bacterial overexpression of human-derived proteins can sometimes lead to misfolding or inclusion body formation, which can make downstream purification or characterization difficult. The bacterial maltose-binding protein (MBP) [42] is an attractive protein expression accessory due to its excellent folding ability. If the ERLBD-based chimeras are insoluble due to the low solubility of ERLBD [43,44], we can perform a translational fusion of MBP to our protein chimera to facilitate downstream expression, purification, and analysis.

For future biosensor applications, if the analytes of interest are hard to transport into the cytoplasm of *E. coli* to bind the sensor proteins, a periplasmic export peptide (TorA) that is associated with the twin-arginine translocation (Tat) pathway can be genetically attached to the sensor proteins [45]. Campbell et al. has adopted such an approach to building genetically encoded calcium ion indicators with a variety of spectral properties [46].

3.2.2 Transposon-based mutagenesis for domain insertion library construction is not ideal in our study

Domain-insertion profiling with sequencing (DIP-seq) developed by Savage et al. is an efficient method for random domain insertion based on transposon (*vis*. jumping genes) [47]. Unfortunately, the intrinsic bias towards certain bases and the small size of the protein backbone for domain insertion in our study make transposon an undesirable tool to achieve our goal. Although an alternative method of insertional mutagenesis library construction suggests using oligo pool-based microarray technology may yield unbiased and comprehensive libraries [48], given the small size of LOV, such method is not economically

effective. To cover the entire space for investigating domain-insertion permissibility, we decided to do systematic cloning instead of such a library-based method. Such exhaustive covering of domain-insertion dataset also allows validation of existing and future independent theoretical methods for prediction validation or method improvement. We will report our results in **Chapter 3.3.3 to 3.3.5**.

3.2.3 Domain-insertion permissibility may provide insights into possible fission sites on LOV for constructing new circular permutation variants

Circular permutation of proteins (an approach that alters the amino acid connectivity but retains a similar 3D structure [20,49]) can potentially avoid the disruption of the inserted ligand-binding domain for developing biosensors. This method has been adopted as a general way to engineer biosensors for neurotransmitters such as γ-aminobutyric acid [41], acetylcholine [38], dopamine [50], norepinephrine [51], calcium [52], and for metabolite such as maltose [53], and nucleotides such as guanidine triphosphate (GTP) [54]. We hypothesize that the approach of ERLBD insertion can also help us identify possible fission sites for engineering circularly permuted LOV (cpLOV) [55]. If the ERLBD inserted constructs cannot fold well due to the large size of ERLBD compared to iLOV, we can choose the sites where ERLBD insertion still renders iLOV fluorescent and engineer cpLOV variants. Those variants can be inserted into a ligand-binding domain of interest based on an established protocol to build allosteric tunable biosensors [56].

3.3 Results and discussion

3.3.1 Domain insertion hotspots are rare

The initial sites on iLOV chose for domain-insertion are at the loop regions, which are presumably less prone to disruptive structural perturbation. In our domain-insertion site

notation, we indicate the amino acid "right after" the insertion site, which may be different from some other literature. We used the numbering system from iLOV's protein data bank structure (4EES). **Figure 3** reveals that domain-insertion permissible sites are rare (*i.e.*, 20 out of 110 constructs are either neutral or responsive). Loss of fluorescence after domain-insertion is an indication of protein unfolding, presumably due to the large distance between the N and the C terminus (~ 20 Å for the 4-HT bound state). Nonetheless, we found that position **G480** has a noticeable change in fluorescence intensity in response to 4-HT (ΔF/F = 20%). G480 may serve as a starting point for testing the domain-insertion modularity in the future.

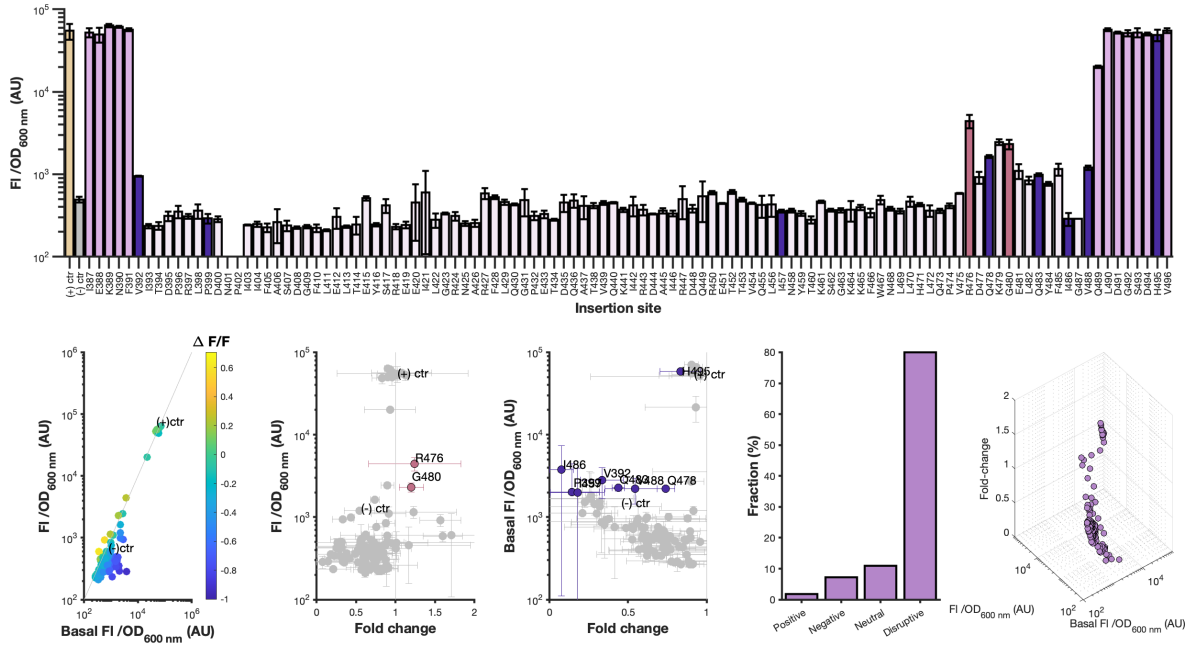

Figure 3. Domain insertion permissibility as revealed by the recovered fluorescence in the presence of ligand and the fluorescence intensity fold-change upon ligand-binding. The sites at the loop region of iLOV are chosen for inserting ER-LBD. Colors of the bar graph indicate the effect of domain-insertion: red indicates significant change fluorescence intensity upon ligand-binding; purple indicates the insertion has neutral effect on fluorescence; blue indicates disruptive effect which quenches the fluorescence. The positive control cells express unmodified iLOV fluorescent proteins, while the negative control cells express the ERLBD protein. Empty positions: data were not measured (e.g., N401 and P402).

3.3.2 The excitation fluorescence spectra may reveal the change in the chromophore-binding environment

To explore the mechanism of fluorescence change upon ligand binding, we speculated that the binding of 4-HT can either increase the affinity between FMN and LOV or enhance the fluorescent quantum yield. To differentiate between these two possibilities, we performed fluorescence excitation measurements for the cell-lysates. Restoration of FMN binding can be characterized by the presence of a shoulder at 495 nm and a decrease in the intensities at 375 nm and 400 nm relative to the largest peak at 450 nm. The peak at 375 nm has contributions from the intramolecular charge-transfer process and thus depends on the solvation environment [57] (**Figure 4**). To further elucidate the detailed mechanism behind the ligand-induced structural change, purification of final working sensors enables the acquisition of circular dichroism (CD) spectra to probe the integrity of the protein secondary structure [58].

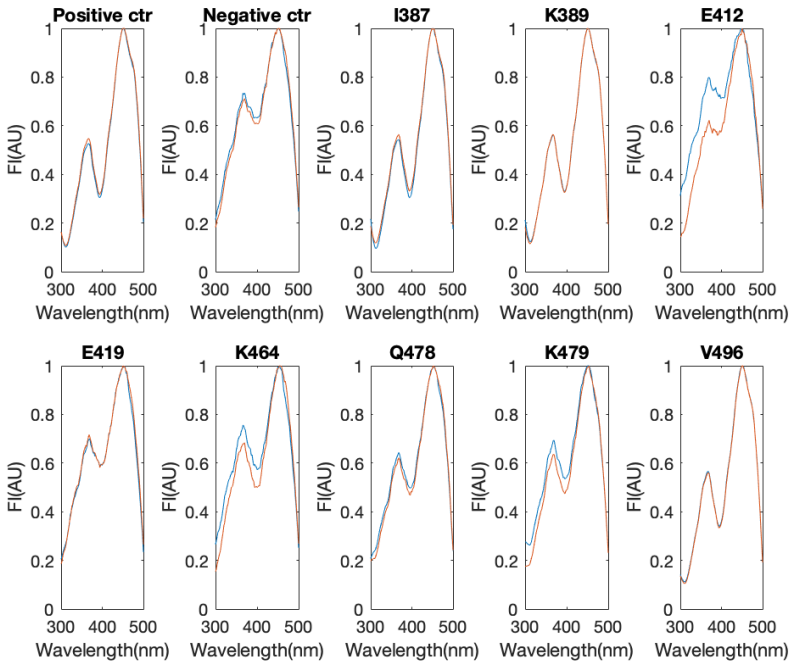

Figure 4. The normalized excitation spectra of some chosen ERLBD-iLOV chimeras reveal the folding ability of iLOV upon ligand-binding. Blue lines: DMSO vesicle treated control; orange lines: 4-HT treated lysates.

3.3.3 Identification of permissible domain-insertion sites from homology models

Surface-exposed, evolutionarily non-conserved tight loops (*i.e.,* loops that join proximal secondary structures) serve as promising candidates for domain-insertion. We argue that the insertion of a ligand-binding domain at such sites can create extrinsic disorder [59,60] that quench the LOV protein fluorescence in a ligand-free state. The ligand-bound state changes the energy landscape of the protein state ensemble that preferentially stabilize the fluorescent states [61,62].

Since the shape of the fluorescence excitation spectrum indicates the change in the local environment of the chromophore, changes in shape might infer an ordered-disordered transition. To explain the change in the shape of the excitation spectra, we adopted a pipeline for identifying permissible domain-insertion sites using a homology model approach described recently [56].

We predicted *loops L1 (Aβ-Bβ), L5 (Fα-Gβ), and L7 (Hβ-Iβ)* to be the most promising sites for domain-insertion allostery as derived from the pipeline (**Figure 5**). Our testing on the fluorescence excitation spectra of K464 (L6) and K479 (L7) showed that 4-HT binding changed the FMN-free state towards the FMN-bound state (**Figure 4**). However, K464 cannot fully restore its' fluorescence after ligand-binding based on our preliminary results (**Figure 3**). G480 (L7) showed ligand-dependent fluorescence change but is marginally fluorescent

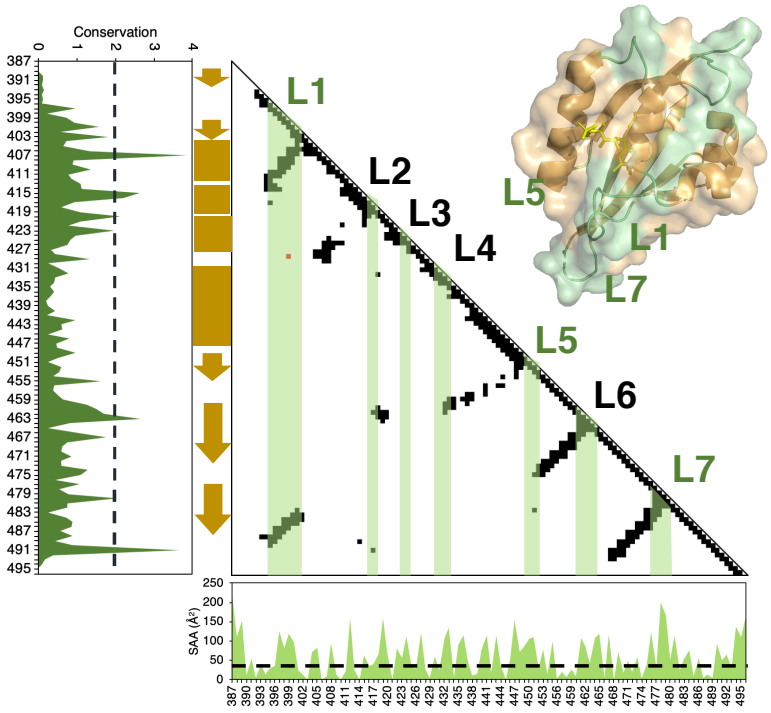

Figure 5. Prediction of domain-insertion hotspots using a conservation model. Contact map with threshold distance 7Å. Surface-exposed and evolutionarily non-conserved tight loops should be the most permissible for domain-insertion. The dashed lines indicate the threshold of the solvent accessible area (SAA) and the threshold of amino acid residue KL conservation, respectively. Lines extending perpendicularly from the diagonal indicate tight loops joining interacting secondary structures. Loop regions are shown as green bands, while secondary structures are shown as brown rectangles (helix) or arrows (sheets). The crystal structure of iLOV (*pdb: 4EES*) shows the domain-insertion hotspots (green).

(0.1-fold of iLOV). The conservation-based model may narrow the range of potential hits, but it does not guarantee if the protein is still functional after domain-insertion.

Coyote-Maestas et al. introduced the idea of differential permissibility based on their study on the latent allostery in inward rectifier potassium ion channel Kir2.1 [63]. They found that domain-insertion permissibility is correlated with dynamic properties (*i.e.*, normal modes of Gaussian network model [64]) but not correlate well with both static (e.g. B-value [65]) and conservation-related properties (e.g. Shenkin score [66]). A widespread theoretical framework called ensemble average model (EAM) [61] describes the thermodynamic interdependence of proteins structural sectors, whose stability may be perturbed by ligand binding. By inserting

distinct domains into Kir2.1, Coyote-Maestas et al. found that domain-insertion permissibility is surprisingly different between domains. They argue that "differential permissibility" may be a better metric for describing allosteric capacity since the emergence of allostery requires regions poised to undergo order/disorder transitions. Further test with different classes of proteins is required to validate this idea. Our work presents a good starting point toward the applicability of such protein allostery descriptors for LOV proteins. We hope to reconcile the theories of predicting allosteric hotspots with our systematic experimental approach using LOV. The domain-insertion fitness landscape from our results can also be used as a machine learning training set for protein engineering [67,68].

3.3.4 Ensemble modeling using COREX/BEST for the prediction of allosteric pivots

Ensemble allosteric model (EAM) [69] predicts that *intrinsically disordered* regions have the potential to establish allosteric properties and was experimentally verified [70]. The COREX/BEST [71] is a coarse-grained algorithm for calculating free energy using a surface-area-based method, which allows the prediction of conformational entropy and thus the propensity of local folding/unfolding. In iLOV, COREX/BEST predicts that the Hβ-Iβ loop (*i.e.,* the residues near residue number 480) possess an equally-like probability of folded/unfolded within the simulated ensemble, as can be seen from the stability ratio $K_f \sim 1$ (**Figure 6**). Thus, the Hβ-Iβ loop may be a hotspot for domain-insertion with allosteric coupling.

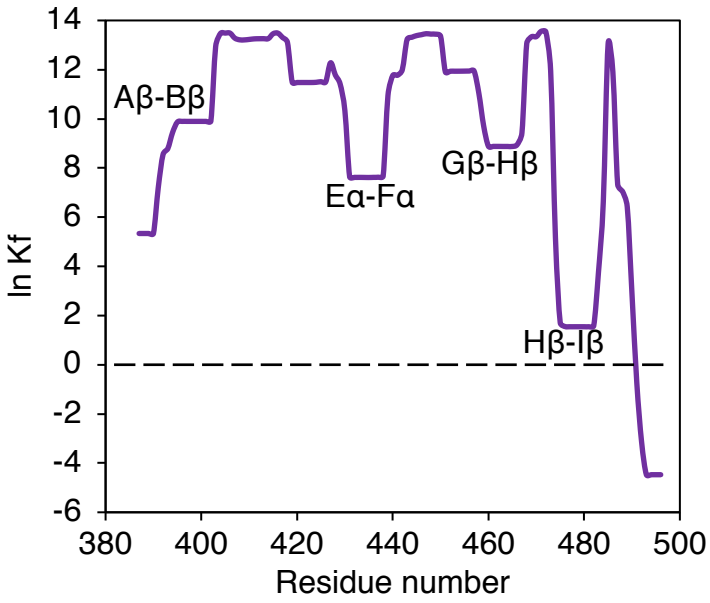

Figure 6. COREX/BEST ensemble allostery analysis. The intrinsically disordered regions are manifested by near one stability ratio (*i.e.*, ln $K_f \sim 0$). The Hβ-Iβ loop within iLOV is predicted to be intrinsically disordered and may be amenable to domain-insertion allostery.

3.3.5 A deep-learning protein classification algorithm facilitates the prediction of engineerable sites

DeeProtein is a deep neural network-based machine learning model for the identification of "engineerable" sites within protein structures [67]. Being a protein function classification model, DeeProtein was trained to assign gene ontology terms (GO) to a given protein's primary structure. By iteratively removing amino acid one at a time and perform the prediction of protein function analysis, the model reports the gain/loss of function "sensitivities" for each amino acid by comparing the likelihood of the protein being assigned a given GO term, before and after removing an amino acid.

We fed the primary structure of iLOV into the DeeProtein server. DeeProtein predicts that the analyzed protein structure has a high probability of ligand-binding (*i.e.*, GO: 0005488), which is consistent with the flavin-binding nature of iLOV. The GO sensitivities were then

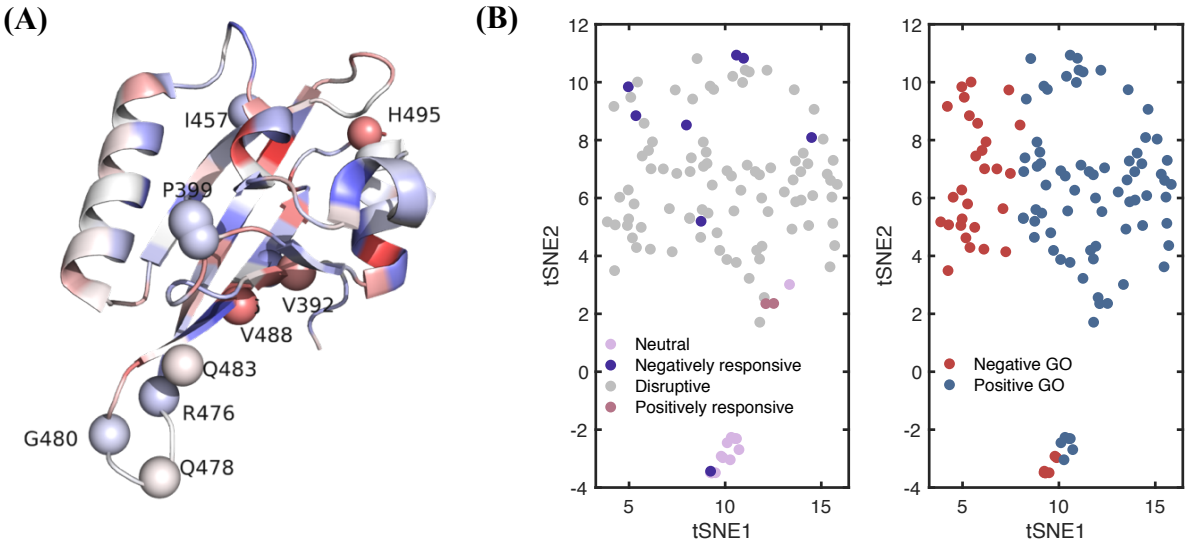

Figure 7. DeeProtein deep neural network-based prediction of engineerable sites within LOV protein. (A) The GO: 0005488 sensitivity corresponding to molecular binding, is mapped onto the crystal structure of iLOV. The responsive sites are marked by spheres; blue: positive sensitivity, red: negative sensitivity. (B) Domain-insertion chimera characteristics projection in reduced space (t-distributed stochastic neighbor embedding (t-SNE)).

mapped onto the iLOV crystal structure (PDB: 4EES). We further compared the experimentally determined fluorescence intensity fold-change and the fluorescence intensity in the presence of ligand from the ERLBD-iLOV fusion library with the predicted GO sensitivities. We performed t-distributed stochastic neighbor embedding (t-SNE) to reduce dimensionality of data for better visualization [72]. This technique facilitates classification since similar objects within the dataset form clusters while dissimilar objects are likely to be distant to each other. The "responsive" constructs (either positive or negative) in general have positive sensitivities, although the fluorescence response does not necessarily proportional to the magnitude of the sensitivity (**Figure 7**). This finding suggests DeeProtein may narrow the search space of domain-insertion sites, but better prediction of actual allosteric sites may require consideration of intrinsic dynamic properties such as normal modes [73].

3.3.6 Experimental validation of domain-insertion permissibility and allosteric response in LOV proteins

Recent attempts to construct fluorescent reporters based on the principle of bimolecular fluorescence complementation (BiFC) such as split-miniSOG-Jα [74] and split-CagFbFP [75] confirmed that Hβ-Iβ can tolerate insertion of flexible poly-Gly/Ser segments. Permissible sites in the loops Aβ-Bβ and Eα-Fα were also identified in CagFbFP potentially due to the high thermostability of CagFbFP [1976]. We argue, however, that the use of poly-Gly/Ser as the insertion permissibility probe could only serve as a naïve guide towards biosensor engineering, as both the "rigidity" and the "N- to C- separation" of the inserted domain dictates the overall folding stability of the protein chimera. For examples of existing biosensor applications, please refer to **Appendix B**.

Photoactive yellow protein (PYP) shares the same superclass, namely Per-ARNT-Sim (PAS) domain, as LOV proteins. It's interesting to see that a new class of fluorescent proteins engineered from PYP to bind synthetic fluorogenic chromophores – the fluorescence-activating and absorption shifting tag (FAST), allows domain-insertion at the corresponding Hβ-Iβ loop [77]. We believe future studies on the basic structural characteristics of PAS domains may help our understanding of developing LOV-based biosensors.

3.3.7 FMN availability affects the dynamic range of LOV-based biosensor

In this section, we analyze the effect of intracellular FMN level on the dynamic range of LOV-based biosensors. Different from traditional GFP variants, LOV proteins or any other "fluorogenic" proteins (*e.g.,* UnaG, iRFP) utilize cofactors in the solution to glowing; thus, the availability of cofactors should be considered when choosing or tuning an expression system for better signal sensitivities.

The fraction of sensor-FMN complex (α) is $\alpha = (1 + \frac{K_D}{[FMN]_{tot}})^{-1}$, where $[FMN]_{tot}$ is the total concentration of FMN in the cell, and K_D is the dissociation constant of FMN to LOV. Since the fluorescence intensity is proportional to the molecular brightness (B_+, the brightness of ligand-bound state; B_-, the brightness of ligand-free state) and the fraction it is in presence, the ratio between the "on" and the "off" state fluorescence intensities (*i.e.,* the dynamic range) can be described as:

$$\frac{I_+}{I_-} = \frac{B_+\alpha_+}{B_+\alpha_-} \propto \frac{1 + \frac{K_{D,-}}{[FMN]_{tot}}}{1 + \frac{K_{D,+}}{[FMN]_{tot}}} \quad (2)$$

It's straightforward to see that the dynamic range is approaching a maximum $\frac{B_+K_{D,-}}{B_+K_{D,+}}$ at low $[FMN]_{tot}$ (**Figure 8**); therefore, the background FMN level should be maintained at a low level. Unfortunately, the intracellular FMN level is constantly changing during flavoprotein overexpression. A previous study has shown overexpression of LOV protein can increase the intracellular level by 3-fold [78]. Our previous theoretical study for predicting the steady-state intracellular FMN level based on the bacterial FMN riboswitch model (unpublished) showed that the overexpression of flavoprotein makes the intracellular FMN level detrimentally low and will presumably lead to the metabolic burden that incurs cellular toxicity. Taken together, the conflict suggests that the FMN riboswitch itself cannot explain how FMN is regulated. More studies must be done to help understand how to control FMN concentration without causing detrimental metabolic burdens.

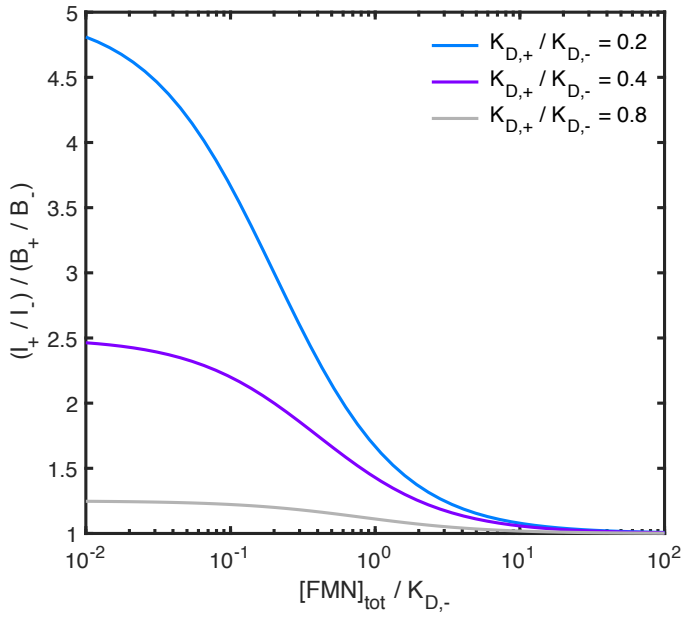

Figure 8. FMN level dictates the theoretical maximum fluorescence intensity fold-change of LOV-based biosensor. The normalized fluorescence intensity fold-change is plotted against the normalized total FMN concentration. Low concentration of FMN or tighter binding of FMN in the presence of ligand allows higher fluorescence intensity fold-change.

3.3.8 A proposed method to enhance the fluorescence contrast of LOV compared to the background

LOV-based fluorescent proteins are not without disadvantages, among which the most notorious would be low fluorescent intensity and autofluorescence in the background due to flavin, mostly flavin mononucleotide (FMN). To further enhance the fluorescence contrast between LOV-based fluorescent proteins and the dissociated flavin background, we reason that by enhancing the relative proportion of flavin adenine dinucleotide (FAD) to FMN, we can reach higher sensitivity of fluorescence due to intrinsically lower fluorescence quantum yield from FAD (*i.e.,* 0.03 from FAD compared to 0.26 from FMN). Note that although FMN is the natural cofactor of most LOV proteins, iLOV can bind FAD with a similar dissociation coefficient as FMN. Furthermore, the resulting FAD-bound iLOV has indistinguishable

fluorescent spectrum from FMN-bound iLOV due to unstacking of the adenine and isoalloxazine ring [79]; we can thus assume that the brightness of holo-iLOV is independent with flavin composition. Since the extinction coefficients for each flavin derivatives and LOV protein are similar (~12500 $mM^{-1}cm^{-1}$ at $\lambda = 450$ nm), we may define the fluorescence contrast (ψ) as:

$$\psi = \frac{\varphi_{\mathrm{LOV}}[\mathrm{holo-LOV}]}{\varphi_{\mathrm{FAD}}[\mathrm{FAD}] + \varphi_{\mathrm{FMN}}[\mathrm{FMN}]} \tag{3}$$

Consider the mass balance of LOV protein and flavin:

$$[\mathrm{apo-LOV}] + [\mathrm{holo-LOV}] = [\mathrm{LOV}]_{\mathrm{tot}} \tag{4}$$

$$[\mathrm{FMN}] + [\mathrm{FAD}] = [\mathrm{Flavin}]_{\mathrm{tot}} \tag{5}$$

The holo-protein concentration can be written as follows, where K_D is the dissociation constant:

$$[\mathrm{holo-LOV}] = \frac{[\mathrm{LOV}]_{\mathrm{tot}}}{1 + \frac{\mathrm{K_D}}{[\mathrm{Flavin}]_{\mathrm{tot}}}} \tag{6}$$

Now if we express the concentration ratio between FAD and FMN to be $\mathrm{r} = \frac{[\mathrm{FAD}]}{[\mathrm{FMN}]}$, [FMN] and [FAD] can thus be written as:

$$[\mathrm{FMN}] = \frac{[\mathrm{Flavin}]_{\mathrm{tot}}}{\mathrm{r}+1} \tag{7}$$

$$[\mathrm{FAD}] = \frac{r[\mathrm{Flavin}]_{\mathrm{tot}}}{\mathrm{r}+1} \tag{8}$$

Finally, the fluorescence contrast is therefore:

$$\psi = \frac{\varphi_{\mathrm{LOV}}(\mathrm{r}+1)[\mathrm{LOV}]_{\mathrm{tot}}}{(\varphi_{\mathrm{FAD}}\mathrm{r} + \varphi_{\mathrm{FMN}})(\mathrm{K_D} + [\mathrm{Flavin}]_{\mathrm{tot}})} \tag{9}$$

Assume $[\mathrm{LOV}]_{\mathrm{tot}} = 1.5\ \mathrm{mM}$. $\varphi_{\mathrm{LOV}} = 0.5, \varphi_{\mathrm{FAD}} = 0.03, \varphi_{\mathrm{FMN}} = 0.26$. It's straightforward to observe the larger the fraction of FAD, the higher the fluorescence contrast (*i.e.*, a

monotonic increasing function). This effect is more pronounced at lower flavin total concentration (**Figure 9**). To increase the FAD to FMN ratio, we proposed to co-express *flavin mononucleotide adenyltransferase (FMNAT)* with LOV, where FMNAT transforms riboflavin and FMN into FAD [80,81]. Further experimental validation of this idea is required in the future.

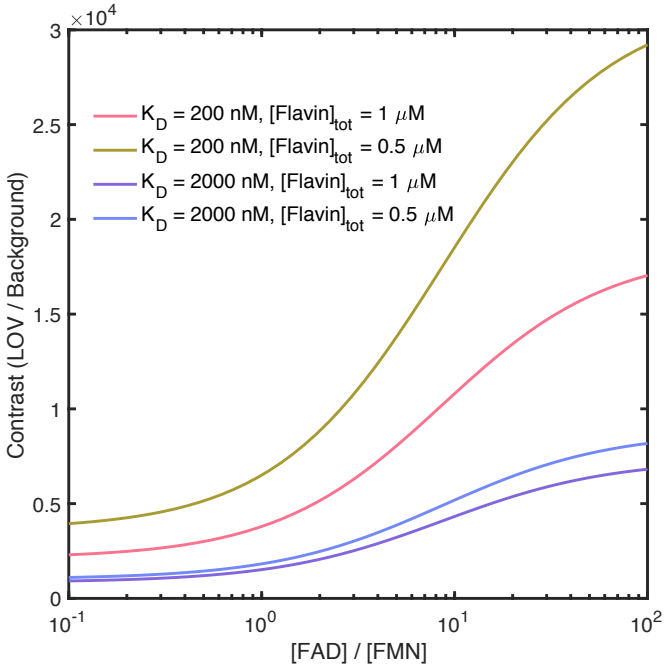

Figure 9. The relationship between LOV protein fluorescence contrast and FAD to FMN ratio. The simulation uses the following parameters: $[\mathrm{LOV}]_{\mathrm{tot}} = 1.5\ \mathrm{mM}$. $\varphi_{\mathrm{LOV}} = 0.5$, $\varphi_{\mathrm{FAD}} = 0.03$, $\varphi_{\mathrm{FMN}} = 0.26$.

Expression accessories are important to control the intracellular flavin content to maximize fluorescence contrast and the dynamic range of biosensors. The concept of fluorescent protein accessory may improve the availability of a non-covalently bound FP system. For example, a far-red fluorescent protein, smURFP, harnesses biliverdin as the fluorophore. Heme oxygenase-1 co-expression and addition of 5-aminolevulinic acid plus iron (ii) sulfate ensure not only the adequate availability of the fluorophore precursor but also the

elimination of undesirable fluorescent side product that may interfere with smURFP fluorescence [82].

3.3.9 Potential mechanisms behind a switch-like behavior of ligand-induced fluorescence change based on LOV protein scaffold

Once we have a protein construct that changes fluorescence intensity upon ligand-binding, we would like to study the mechanism behind such switch-like behavior. We speculate that potential mechanisms may involve changes in (1) The *affinity* between the flavin chromophore and LOV protein (i.e. the dissociation constant, K_D); (2) The *quantum yield* (ϕ) of the LOV protein that is associated with the local flavin-binding environment; (3) The *protein abundance* as affected by the ligand-induced thermodynamic stabilization; (4) *3D domain-swapping* derived from oligomerization of protein constructs to minimize distortion strain due to domain-insertion [83–85]. Further biochemical analysis can help us differentiate between those cases. For example, the dissociation constant of the flavin chromophore can be determined by stripping it out of the LOV protein backbone using a denaturation-renaturation process. Reconstitution of the apo-protein with flavin is then monitored with fluorescence measurement as previously described [86]. To determine the protein abundance, the use of *in vivo* fluorescence assay can be coupled with Western-blotting [87] to quantitatively detect changes in protein expression level within the cell.

3.4 Conclusions

In this chapter, we presented LOV as an emerging class of fluorescent protein to develop fundamental knowledge of engineerable sites within the protein structure. Extensive insertion of estrogen ligand-binding domain (ERLBD) into iLOV fluorescent protein yielding nearly 110 constructs enabled first experimental validation of domain-insertion permissibility and

allosteric response using LOV protein as a reporter scaffold. Theoretical approach including the traditional homology-based protein engineering pipeline, ensemble allosteric modeling, and deep-learning-based protein function classification model yielded consensus on engineerable sites within iLOV, but not achieved in the prediction within the single amino acid resolution. Such theoretical models may facilitate targeting potential secondary structures for domain-insertion; however, realistic domain-insertion permissibility and protein allostery also depend on parent protein stability and the context of inserted domains.

The Hβ-Iβ loop LOV protein is in general amenable to domain-insertion or dissection as is consistent with previous findings. Limited fluorescence signal and fold-change upon ligand-binding in our naïve domain-insertion library raised concerns in basic properties of LOV proteins including the availability of flavin cofactor and high fluorescence background due to dissociated flavin in the environments. We proposed conceptual assessments and possible solutions to improve the characteristics of LOV-based biosensors. Overall, we expect this work to encourage the continued development of LOV-based biosensors that enable both fundamental and applied biological applications.

CHAPTER 4

TOWARDS LOV-BASED ATP SENSOR

4.1 Introduction

Adenosine triphosphate (ATP) is a critically important metabolite involved in diverse biochemical reactions and reflects cellular energy levels. Several ATP sensors were developed to monitor ATP concentration with a spatial and temporal resolution at a single cell level to precisely understand how ATP controls cellular processes [88–91]. However, malfunctioned sensors due to immature fluorescent proteins can cause an undesirable shift in the reported signal during a change in cell growth rate [92], especially for fluorescence resonance energy transfer (FRET) type sensors. Moreover, the use of GFP-based ATP biosensors cannot faithfully report the ATP concentration in hypoxia. We hypothesize that using LOV as a scaffold for the development of an ATP sensor could help solve a long-standing challenge in bacterial infection treatment – bacterial persister formation.

Persister is a bacterial variant with a dormant phenotype that is associated with chronic infections and antibiotic treatment failure [93]. For example, *Staphylococcus aureus* is a common anaerobe that can cause osteomyelitis and endocarditis. To gain a better understanding of the mechanisms behind persister formation, several studies showed that persisters are metabolically inactive [94,95]; however, the persister physiology has largely been studied in aerobic environments, contradicting the fact that most persisters are anaerobes in nature. To observe the direct causal relationship between cell energetics and the probability of persister formation, we expect to engineer the first fluorescent reporter based on LOV

protein for visualizing ATP in low oxygen conditions (Anaerobically Visible ATP Sensor, **AnViAS**).

To engineer AnViAS, we used the domain-insertion approach – replace the ERLBD in the best performing ERLBD-iLOV construct with an ATP-binding domain – the *B. subtillis* F_0F_1 ATP synthase ε subunit. ATP binding triggers the bundling of the two α-helices at the C-terminal domain (**Figure 10**). Such ATP binding domain has been adopted to build a GFP-based ATP sensor making use of the conformational change ATP-binding incurs [90]. The affinity between ATP and the ε subunit can be tuned using site-directed mutagenesis to enhance the sensitivity in physiological relevant ATP concentrations. We may purify the overexpressed protein from *E. coli* and test the ATP sensing performance *in vitro*. To further enhance the dynamic range of the sensor, linker optimization can be performed as needed. To test the specificity toward ATP detection of our sensor, we may use chemical analogs similar in structure as ATP such as ADP, AMP, and deoxy-ATP to avoid false-positives when the sensor is used *in vivo*. To test the reversibility of ATP binding**,** we may incubate the ATP saturated AnViAS with ATP hydrolyzing enzymes such as apyrase [96] and measure the decrease in fluorescence intensity over time. To test the capability of AnViAS to sense ATP in anaerobic environments**,** we can also perform the aforementioned experiments for cells inoculated anaerobically in oxygen impermeable 96-well plates.

To test the real-time ATP detection capability of AnViAS in vivo, we may use respiration inhibitor 2-deoxyglucose (2DG) or potassium cyanide (KCN) to alter the intracellular ATP concentration. Cell growth can be continuously monitored with periodic measurement of fluorescence. For independent validation of intracellular ATP concentration, a commercial luciferase-based product may be used to be compared with AnViAS performance.

To measure the ATP concentration at the single-cell level and to validate the persister formation hypothesis using AnViAS, we can use time-lapse fluorescent microscopy coupled with a microfluidic platform previously described [97,98] to differentiate ATP concentration at the single-cell level. We hypothesize that persister formation in hypoxia is potentiated by the reduction in bacterial bioenergetics (*i.e.*, reduced level of intracellular ATP) associated with anaerobic adaptation (**Figure 11**). As an initial attempt, *E. coli* K-12 persister high persistence (hip) strain [99] may be used as the model system of this study.

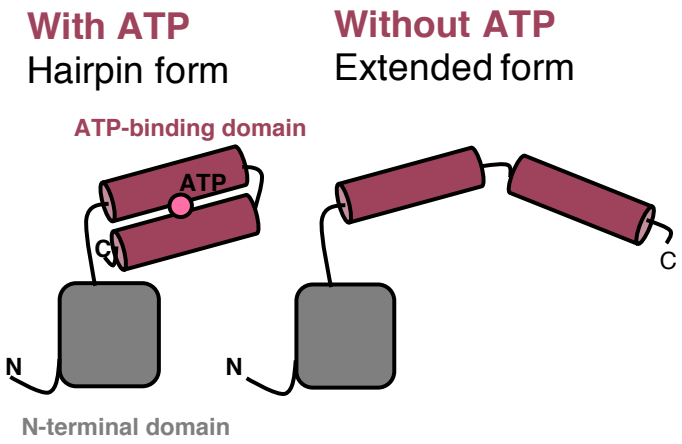

Figure 10. Schematics of Bacillus PS3 F_0F_1 ATP synthase ε subunit complexed with ATP. ATP binding triggers the bundling of the two α-helices. The C-terminal ATP-binding domain and the N terminal domain are in violet and in grey, respectively.

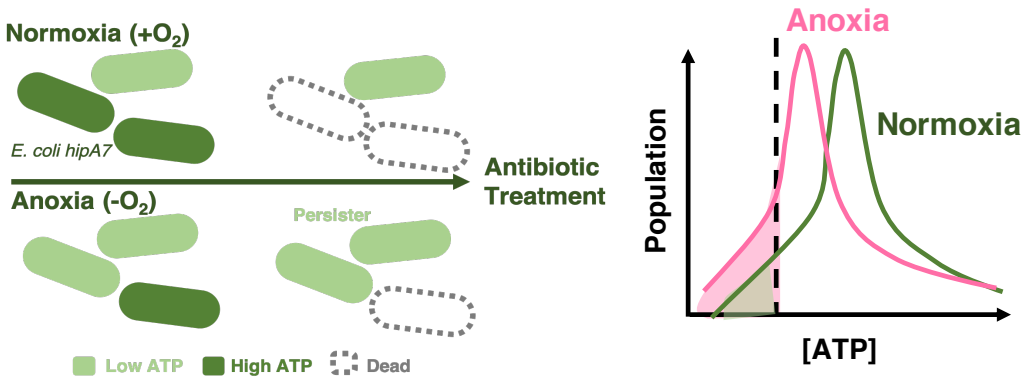

Figure 11. Schematics of the bacterial persister formation hypothesis. The fluorescence ATP indicator with time-lapse fluorescent microscopy can keep track of the viability and the energetic state of the cells. Bacterial persister formation is hypothesized to be more likely to appear from a dormant cell with lower energy level.

4.2 Potential pitfalls and alternatives

It is possible that *B. subtillis* F_0F_1 ATP synthase ε subunit may not provide the best sensitivity towards ATP detection. We may explore the use of alternative ATP binding proteins such as *Bacillus sp*. PS3 F_0F_1 ATP synthase ε subunit, which has stronger affinity towards ATP (K_D = 7.4 μM at 37°C) than *B. subtillis* F_0F_1 ATP synthase ε (K_D = 3.3 mM at 37°C) [92].

To avoid disrupting the LOV protein folding after domain-insertion, the separation between the -N and -C termini should be ideally ~ 10 Å. If the insertion of F_0F_1 ATP synthase ε leads to poor protein folding, a bacterial ATP-binding protein, GlnK1, can be used instead [88]. Although GlnK1 undergoes a large conformational change upon ATP-binding, the distance between the -N and -C termini remain in the order of 10 Å, thus making GlnK1 an ideal alternative to building ATP sensor using the domain-insertion approach.

4.3 Results and discussion

In **Chapter 3**, we experimentally identified loop Hβ-Iβ in iLOV as an amenable region for domain-insertion, among which construct G480 showed a ligand-induced change in fluorescence intensity. As our first attempt to develop an ATP sensor using the knowledge gained from **Chapter 3**, we swapped the ERLBD with B. subtillis F_0F_1 ATP synthase ε subunit and performed a KCN-based ATP sensor assay (See **2.4.4** for detailed protocols). Naïve swapping yielded non-fluorescent chimeras even with a high concentration of ligand (**Figure 12**). We noticed the large size of the N-terminal domain of our ATP-binding subunit may have possibly distorted the iLOV structure and subsequent unfolding events. To rescue the fluorescence of our chimeras, we decided to randomly introduce various lengths of linkers ranging from zero to four amino acids at each end of the inserted domain using Gibson

assembly, selected bright variants possessing enough folding stability for domain-insertion; finally, a negative selection of dim variants in ATP depleted environments was performed, indicating the switch-like property of protein allostery. To further reduce the size of our ATP sensor, a truncated version of the ATP-binding subunit lacking the N-terminal domain was also included in our linker-optimization process. We used the VST codon for our random mutagenesis to avoid incorporation of a stop codon and to make sure small flexible amino acids such as glycine and serine are included (**Figure 13**).

Either the intact or truncated libraries showed a span of fluorescence fold-change upon KCN treatment (**Figure 14**). Selected variants with a large fluorescence fold-change (ΔF/F) were then sequenced. Unfortunately, none of the interesting variants showed successful insertion of the ATP-binding domain; instead, unidentifiable short sequences were inserted, or just a native iLOV sequence was shown. We suspected that Gibson assembly might not be ideal for creating a random domain-insertion library since unpredictable complementary events might yield undesirable insertion chimeras. In the future, one may use multiple site-directed mutagenesis to avoid such problems.

Furthermore, another issue that may come into play is the reproducibility of our KCN assay. We found that KCN can cause an undesirable shift in either the fluorescence intensity or OD of the bacterial cultures expressing fluorescent proteins. By transcriptionally fusing a mCherry fluorescent protein to iLOV, mCherry may serve as an internal standard to correct for any fluorescence quenching events arising from KCN treatment (*i.e.,* express mCherry and iLOV under the control of the same promoter). Some previous studies used translational fusion [25,91,100], but due to interactions between two fused proteins, any unfolding events can potentially affect one another. To prevent such an issue, we decided to use transcriptional

fusion instead and reported the iLOV fluorescence signal by mCherry normalization. The fluorescence signals indeed remained stabilized after 10 min of KCN treatment (**Figure 15**). We believe such modification may help us identify bona fide ATP sensors with reproducible ATP sensing properties in the future.

Lastly, one potential issue that may confound the identification of allosteric chimera is that proximity of ATP to flavin fluorophore in the ligand-bound state might induce "fluorescence quenching", which was observed in FAD where an additional adenine moiety contributed to a 10-fold lower quantum yield compared to FMN [91]; similar fluorescence quenching effect was also seen in a GFP-based ATP sensor [101]. The ATP-bound state of our domain-insertion chimera is presumably higher in fluorescence than the ATP-unbound state because the ATP-bound state of the ATP ligand-binding domain is more compact. If the on-state ATP concentration exceeds the amount required for fluorescence quenching, we could lose those candidates with higher fluorescence fold-change when exposed to lower ATP concentrations.

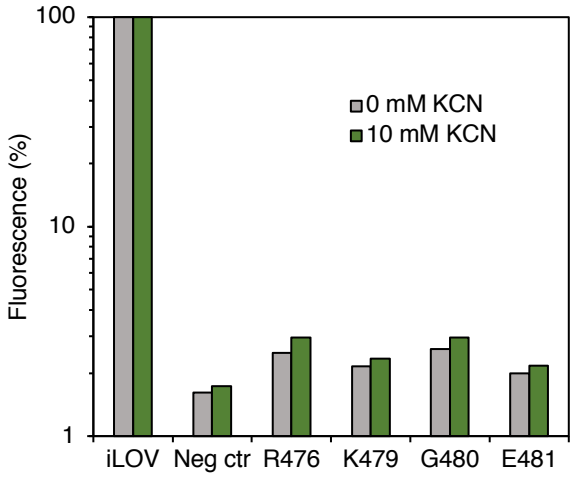

Figure 12. Naïve ATP-binding domain insertion into iLOV. Four different sites within the Hβ-Iβ loop of iLOV were selected for direct insertion of ATP-binding domain. Bacterial cells expressing only the ATP-binding domain serves as the negative control in our experiment.

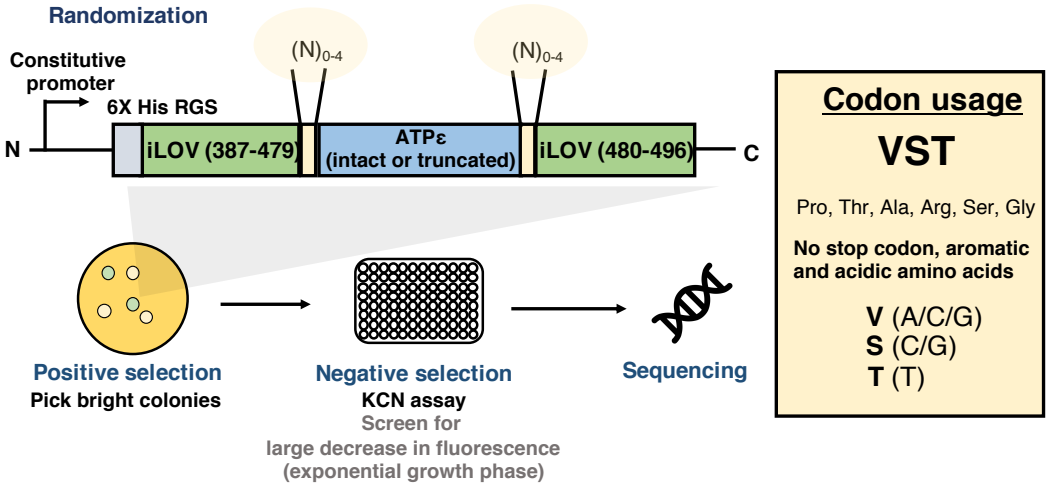

Figure 13. Workflow of ATP sensor screening. The ATP-binding domain (either with or without the N-terminal domain) with various lengths of linkers at each end was cloned into iLOV at the G480 position using Gibson assembly. Protein chimeras were expressed using a constitutive promoter to facilitate the screening process. A positive selection followed by a negative selection ensures variants with proper folding and switch-like behavior to be picked. A final DNA sequencing identifies potential ATP sensor candidates for further testing.

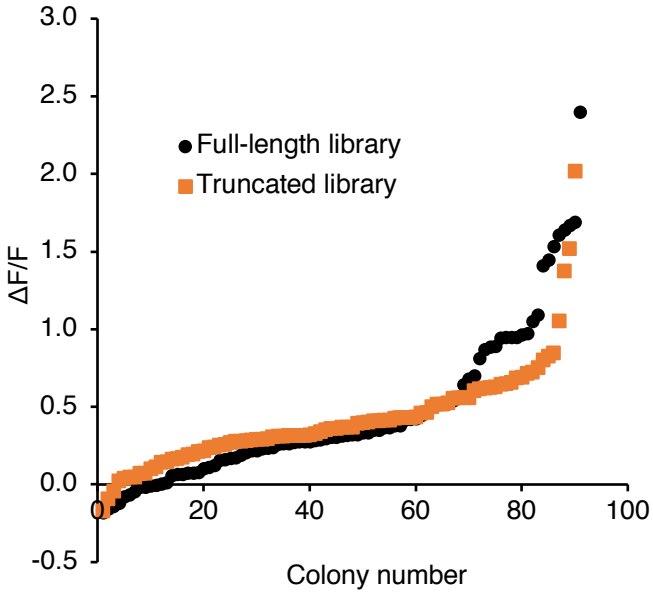

Figure 14. ATP sensor linker mutagenesis yields a wide span of ATP responsible variants. ΔF/F is defined as $(F_{PBS} - F_{KCN}) / F_{KCN}$, where F_{PBS} is the fluorescence intensity of the PBS treated groups, whereas F_{KCN} is the fluorescence intensity of the KCN treated groups.

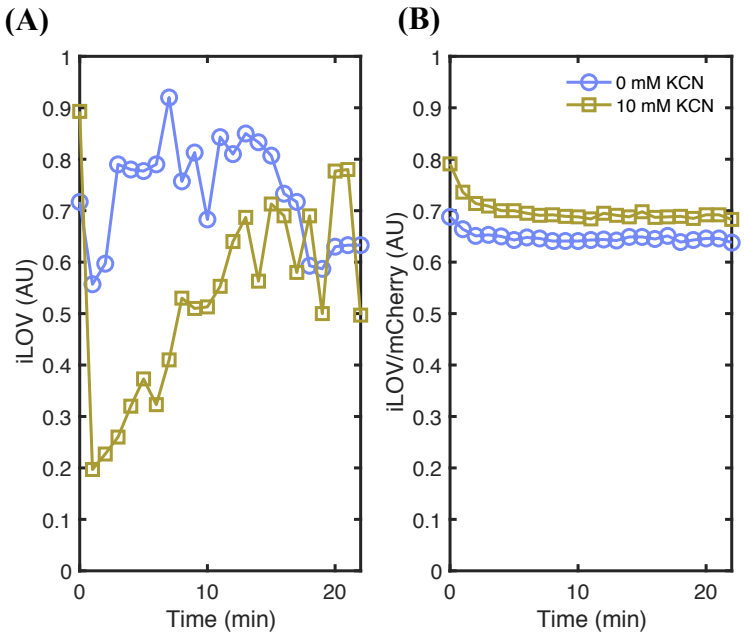

Figure 15. mCherry fluorescence as an internal standard for KCN-induced fluorescence drifting. (A) The fluorescence intensity time-series of iLOV alone suffers from undesirable shift during and after KCN treatment. (B) The fluorescence intensity normalization using mCherry makes it a robust signal reporter against KCN-induced shift.

4.4 Conclusions

In this chapter, we explored the possibility of building an anaerobically visible ATP sensor (AnViAS) using LOV protein as a reporter scaffold, with a long-term goal to use such sensor to investigate the relationship between bacterial cell energetics (*i.e.,* ATP levels) and the development of antibiotic-resistant phenotypes (*i.e.,* persisters). Acknowledging the domain-insertion hotspots of iLOV found from the previous chapter, we inserted *B. subtillis* F_0F_1 ATP synthase ε subunit into iLOV without additional peptide linkers. However, such naïve insertion didn't yield fluorescent constructs. To rescue fluorescence that possibly due to steric hindrance of the inserted domain, we relaxed the distance between the fused domains using Gibson assembly by introducing random short peptide linkers between them or truncate the N-terminal domain of *B. subtillis* F_0F_1 ATP synthase ε subunit irrelevant to ATP binding. We proposed a fluorescence screening workflow that includes a positive selection by

identification of bright colonies on plates, followed by a negative selection using KCN to reduce the ATP level. Such design may help pick up well-folded protein chimeras and those which exhibit ATP-dependent fluorescence response. Unfortunately, Gibson assembly yielded incomplete preparation of linker mutagenesis libraries. Moreover, the undesirable shift in fluorescence intensity during KCN treatment made it hard to identify genuine ATP responsive candidates. We transcriptionally fuse unmodified iLOV with a spectrally different fluorescence protein, mCherry. The mCherry fusion successfully stabilizes the iLOV fluorescence signal during KCN treatment. We believe such modification along with site-directed mutagenesis that introduces linkers between the two domains may help us identify bona fide LOV-based ATP sensors in the future.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

LOV fluorescent proteins represent a remarkable class of endogenous small molecule-dependent fluorescent units. In this thesis, we demonstrated the limitations of state-of-the-art techniques for predicting engineerable sites with an extensive experimental dataset of ligand-induced fluorescence response in iLOV protein. ERLBD was systematically inserted into nearly every site of iLOV (**Chapter 3**). The consensus Hβ-Iβ loop within iLOV is permissible to domain-insertion given the retained despite diminished brightness of iLOV. Among those protein chimeras, the G480 construct showed a ligand-dependent change in fluorescence, indicating successful mining of protein allostery hotspot in iLOV. Independent theoretical prediction based on traditional homology modeling, ensemble allostery model (EAM), and a protein function characterization machine learning model (DeeProtein) all justifies the existence of such a site within the Hβ-Iβ loop.

Given the small size of LOV proteins, maintaining the folding stability of the protein chimera is an integral part of the design of fluorescent biosensors. We anticipate that the need for changing the topology of domain-insertion is required in certain circumstances when the size of a ligand-binding domain (LBD) is significantly larger than the parent LOV protein; inserting a circularly permutated variant of LOV into the LBDs would be desirable. We believe a good start involves the development of bright circular permutated LOV proteins using thermostable parent LOV variants, as they are structurally resilient to protein fission and hence the construction of cpLOV. Since the Hβ-Iβ loop was verified to have the potential

of protein allostery, we could choose multiple fission sites within that specific loop and genetically connect the original N- and C- termini with suitable linkers in the future.

We tested our hypothesis of the found allosteric hotspot G480 within iLOV to build an ATP sensor by constructing a library of variants with different linkers flanking the inserted *B. subtillis* F_0F_1 ATP synthase ε subunit (**Chapter 4**). Our results demonstrated that domain-insertion permissible sites might not be transferrable between different inserted domains. Short peptide linkers prevent steric hindrance between the two domains and can potentially create an intrinsically disordered region with a favorable energy landscape for protein switches. In our experiments, however, the random linker mutagenesis enabled through Gibson assembly cannot prevent empty vector carryover and unexpected fragmentation of inserts due to sequence complementarity of inserted fragments to the backbone. A better way to circumvent the problem in the future is to apply multiple site-directed mutagenesis on a naïve insertion chimera, which ensures no unmodified native iLOV will be present in our biosensor screening.

We noticed in our screening of ATP sensors using the KCN can pose difficulties in identifying truly ligand-responsive chimeras; KCN treatment may induce cell apoptosis events that trigger fluorescent protein degradations and a change in flavin level, leading to an undesirable and even irreproducible shift in fluorescence signal. A translational or transcriptional fusion of benchmark fluorescent proteins that share limited spectral overlap with LOV proteins may serve as internal standards for signal normalization.

Although LOV fluorescent proteins (FP) have proved to be promising as a molecular biology reporter in studying anaerobic microbes, the development of biosensors using protein engineering is still in its infancy stage. A good understanding of the engineerable sites within

LOV would facilitate protein switch applications (**Chapter 3**). However, LOV suffers from limitations such as considerably lower brightness and higher background fluorescence compared to GFP-based counterparts [102]. A thermostable monomeric LOV variant may be a better starting point for the development of sensors because domain-insertion can unavoidably reduce the parent protein folding stability, leading to lower solubility and overall brightness. Further, a potential solution to the problem of flavin availability and flavin background fluorescence is described in **Chapter 3** where protein expression "accessories" can be theoretically introduced to control the intracellular flavin content, to maximize fluorescence contrast and dynamic range of biosensors.

Eukaryotes and prokaryotes utilize cyclic nucleotides as signaling molecules to regulate cellular processes. However, the role of such molecules in bacteria is still not well explored [103,104]. In bacteria, cyclic dimeric guanosine 3',5' monophosphate (c-diGMP), cyclic dimeric adenosine 3',5' monophosphate (c-diAMP), cyclic guanosine 3',5' monophosphate (cGMP), and cyclic adenosine 3',5' monophosphate (cAMP) was shown to be involved in virulence factor generation, biofilm formation or cell-to-cell communication (*i.e.*, quorum sensing). To better understand how bacterial pathogenesis works, enabling technology to detect cyclic nucleotides with the spatial and temporal resolution is thus in need. Although several genetically encodable biosensors to detect cyclic nucleotides have been developed [30,32,105,106], the protein maturation of such GFP-based sensors makes them not suitable to be used anaerobically, especially in the study of obligately anaerobic bacteria. We expect that with the knowledge we obtained in this research in mind, the use of the LOV protein scaffold will outperform existing tools for detecting cyclic nucleotides. For example, an Epac protein [107,108]

can be inserted to develop a cAMP sensor, while a PDE5α protein [109] can be inserted to develop a cGMP sensor.

In summary, our research poses important questions that need to be answered for LOV fluorescent proteins to be used as a routine workhorse for developing biosensors that work in anaerobic environments. Our attempts to engineer LOV fluorescent protein provides a framework to improve and apply LOVs in unmet needs in microbiological research, medicine, and biological industries.

REFERENCES

1. Greenwald, E. C., Mehta, S. & Zhang, J. Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. *Chem. Rev.* **118**, 11707–11794 (2018).
2. Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **373**, 663–664 (1995).
3. Kim, S. K. *et al.* A Genetically Encoded Biosensor for Monitoring Isoprene Production in Engineered Escherichia coli. *ACS Synth. Biol.* **7**, 2379–2390 (2018).
4. de los Santos, E. L. C., Meyerowitz, J. T., Mayo, S. L. & Murray, R. M. Engineering Transcriptional Regulator Effector Specificity Using Computational Design and In Vitro Rapid Prototyping: Developing a Vanillin Sensor. *ACS Synth. Biol.* **5**, 287–295 (2016).
5. Kanno, A., Yamanaka, Y., Hirano, H., Umezawa, Y. & Ozawa, T. Cyclic Luciferase for Real-Time Sensing of Caspase-3 Activities in Living Mammals. *Angew. Chemie Int. Ed.* **46**, 7595–7599 (2007).
6. Lobo, L. A., Smith, C. J. & Rocha, E. R. Flavin mononucleotide (FMN)-based fluorescent protein (FbFP) as reporter for gene expression in the anaerobe Bacteroides fragilis. *FEMS Microbiol. Lett.* **317**, 67–74 (2011).
7. Choi, C. H., DeGuzman, J. V, Lamont, R. J. & Yilmaz, Ö. Genetic Transformation of an Obligate Anaerobe, P. gingivalis for FMN-Green Fluorescent Protein Expression in Studying Host-Microbe Interaction. *PLoS One* **6**, e18499 (2011).
8. Tielker, D., Eichhof, I., Jaeger, K.-E. & Ernst, J. F. Flavin mononucleotide-based fluorescent protein as an oxygen-independent reporter in Candida albicans and Saccharomyces cerevisiae. *Eukaryot. Cell* **8**, 913–915 (2009).
9. Mukherjee, A., Walker, J., Weyant, K. B. & Schroeder, C. M. Characterization of Flavin-Based Fluorescent Proteins: An Emerging Class of Fluorescent Reporters. *PLoS One* **8**, e64753 (2013).
10. Los, G. V *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
11. Lukinavičius, G. *et al.* A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* **5**, 132 (2013).
12. Beatty, K. E. *et al.* Fluorescence Visualization of Newly Synthesized Proteins in Mammalian Cells. *Angew. Chemie Int. Ed.* **45**, 7364–7367 (2006).
13. Marchand, J. A. *et al.* Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019).
14. Plamont, M.-A. *et al.* Small fluorescence-activating and absorption-shifting tag for tunable protein imaging in vivo. *Proc. Natl. Acad. Sci.* **113**, 497 LP – 502 (2016).
15. Herwig, L. *et al.* Directed Evolution of a Bright Near-Infrared Fluorescent Rhodopsin Using a Synthetic Chromophore. *Cell Chem. Biol.* **24**, 415–425 (2017).
16. Sheng, W. *et al.* A Near-Infrared Photoswitchable Protein–Fluorophore Tag for No-Wash Live Cell Imaging. *Angew. Chemie Int. Ed.* **57**, 16083–16087 (2018).
17. Drepper, T. *et al.* Reporter proteins for in vivo fluorescence without oxygen. *Nat. Biotechnol.* **25**, 443 (2007).
18. Mukherjee, A. *et al.* Engineering and Characterization of New LOV-Based Fluorescent

Proteins from Chlamydomonas reinhardtii and Vaucheria frigida. *ACS Synth. Biol.* **4**, 371–377 (2015).
19. Nazarenko, V. V *et al.* A thermostable flavin-based fluorescent protein from Chloroflexus aggregans: a framework for ultra-high resolution structural studies. *Photochem. Photobiol. Sci.* (2019). doi:10.1039/C9PP00067D
20. Miyawaki, A. *et al.* Fluorescent indicators for Ca2+based on green fluorescent proteins and calmodulin. *Nature* **388**, 882 (1997).
21. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343 (2009).
22. Kotera, I., Iwasaki, T., Imamura, H., Noji, H. & Nagai, T. Reversible Dimerization of Aequorea victoria Fluorescent Proteins Increases the Dynamic Range of FRET-Based Indicators. *ACS Chem. Biol.* **5**, 215–222 (2010).
23. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–W502 (2004).
24. Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M. & Marino Buslje, C. MISTIC: mutual information server to infer coevolution. *Nucleic Acids Res.* **41**, W8–W14 (2013).
25. Dippel, A. B., Anderson, W. A., Evans, R. S., Deutsch, S. & Hammond, M. C. Chemiluminescent Biosensors for Detection of Second Messenger Cyclic di-GMP. *ACS Chem. Biol.* **13**, 1872–1879 (2018).
26. Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* **34**, 646 (2016).
27. Liang, R., Zhou, J. & Liu, J. Construction of a Bacterial Assay for Estrogen Detection Based on an Estrogen-Sensitive Intein. *Appl. Environ. Microbiol.* **77**, 2488 LP – 2495 (2011).
28. Tran, Q. H. & Unden, G. Changes in the proton potential and the cellular energetics of Escherichia coli during growth by aerobic and anaerobic respiration or by fermentation. *Eur. J. Biochem.* **251**, 538–543 (1998).
29. Aroul-Selvam, R., Hubbard, T. & Sasidharan, R. Domain Insertions in Protein Structures. *J. Mol. Biol.* **338**, 633–641 (2004).
30. Matsuda, S. *et al.* Generation of a cGMP Indicator with an Expanded Dynamic Range by Optimization of Amino Acid Linkers between a Fluorescent Protein and PDE5α. *ACS Sensors* **2**, 46–51 (2017).
31. Tucker, C. L. & Fields, S. A yeast sensor of ligand binding. *Nat. Biotechnol.* **19**, 1042–1046 (2001).
32. Kitaguchi, T., Oya, M., Wada, Y., Tsuboi, T. & Miyawaki, A. Extracellular calcium influx activates adenylate cyclase 1 and potentiates insulin secretion in MIN6 cells. *Biochem. J.* **450**, 365 LP – 373 (2013).
33. Bubeck, F. *et al.* Engineered anti-CRISPR proteins for optogenetic control of CRISPR–Cas9. *Nat. Methods* **15**, 924–927 (2018).
34. Nicholes, N. *et al.* Modular protein switches derived from antibody mimetic proteins. *Protein Eng. Des. Sel.* **29**, 77–85 (2015).
35. Chapman, S. *et al.* The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proc. Natl. Acad. Sci.* **105**, 20038 LP – 20043 (2008).
36. Shiau, A. K. *et al.* The Structural Basis of Estrogen Receptor/Coactivator Recognition

and the Antagonism of This Interaction by Tamoxifen. *Cell* **95**, 927–937 (1998).
37. Tanenbaum, D. M., Wang, Y., Williams, S. P. & Sigler, P. B. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc. Natl. Acad. Sci.* **95**, 5998 LP – 6003 (1998).
38. Jing, M. *et al.* A genetically encoded fluorescent acetylcholine indicator for in vitro and in vivo studies. *Nat. Biotechnol.* **36**, 726 (2018).
39. Fosque, B. F. *et al.* Labeling of active neural circuits in vivo with designed calcium integrators. *Science (80-.).* **347**, 755 LP – 760 (2015).
40. Zhao, Y., Shen, Y., Wen, Y. & Campbell, R. E. High-Performance Intensiometric Direct- and Inverse-Response Genetically Encoded Biosensors for Citrate. *ACS Cent. Sci.* **6**, 1441–1450 (2020).
41. Marvin, J. S. *et al.* A genetically encoded fluorescent sensor for in vivo imaging of GABA. *Nat. Methods* (2019). doi:10.1038/s41592-019-0471-2
42. Quiocho, F. A., Spurlino, J. C. & Rodseth, L. E. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997–1015 (1997).
43. Seielstad, D. A., Carlson, K. E., Katzenellenbogen, J. A., Kushner, P. J. & Greene, G. L. Molecular characterization by mass spectrometry of the human estrogen receptor ligand-binding domain expressed in Escherichia coli. *Mol. Endocrinol.* **9**, 647–658 (1995).
44. Nygaard, F. B. & Harlow, K. W. Heterologous expression of soluble, active proteins in Escherichia coli: the human estrogen receptor hormone-binding domain as paradigm. *Protein Expr. Purif.* **21**, 500–509 (2001).
45. Barrett, C. M. L., Ray, N., Thomas, J. D., Robinson, C. & Bolhuis, A. Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in Escherichia coli. *Biochem. Biophys. Res. Commun.* **304**, 279–284 (2003).
46. Zhao, Y. *et al.* An Expanded Palette of Genetically Encoded Ca^{2+} Indicators. *Science (80-.).* **333**, 1888 LP – 1891 (2011).
47. Nadler, D. C., Morgan, S.-A., Flamholz, A., Kortright, K. E. & Savage, D. F. Rapid construction of metabolite biosensors using domain-insertion profiling. *Nat. Commun.* **7**, 12266 (2016).
48. Coyote-Maestas, W., Nedrud, D., Okorafor, S., He, Y. & Schmidt, D. Targeted insertional mutagenesis libraries for deep domain insertion profiling. *Nucleic Acids Res.* **48**, e11–e11 (2020).
49. Yu, Y. & Lutz, S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol.* **29**, 18–25 (2011).
50. Patriarchi, T. *et al.* Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science (80-.).* **360**, (2018).
51. Feng, J. *et al.* A Genetically Encoded Fluorescent Sensor for Rapid and Specific In Vivo Detection of Norepinephrine. *Neuron* **102**, 745-761.e8 (2019).
52. Wang, Q., Shui, B., Kotlikoff, M. I. & Sondermann, H. Structural Basis for Calcium Sensing by GCaMP2. *Structure* **16**, 1817–1827 (2008).
53. Marvin, J. S., Schreiter, E. R., Echevarría, I. M. & Looger, L. L. A genetically encoded, high-signal-to-noise maltose sensor. *Proteins Struct. Funct. Bioinforma.* **79**, 3025–3036 (2011).
54. Bianchi-Smiraglia, A. *et al.* Internally ratiometric fluorescent sensors for evaluation of

intracellular GTP levels and distribution. *Nat. Methods* **14**, 1003 (2017).
55. Jones, A. M. *et al.* The Structure of a Thermophilic Kinase Shapes Fitness upon Random Circular Permutation. *ACS Synth. Biol.* **5**, 415–425 (2016).
56. Dagliyan, O., Dokholyan, N. V & Hahn, K. M. Engineering proteins for allosteric control by light or ligands. *Nat. Protoc.* **14**, 1863–1883 (2019).
57. Penzer, G. R. & Radda, G. K. The chemistry and biological function of isoalloxazines (flavines). *Q. Rev. Chem. Soc.* **21**, 43–65 (1967).
58. Munro, A. W., Kelly, S. M. & Price, N. C. Circular Dichroism Studies of Flavoproteins BT - Flavoprotein Protocols. in (eds. Chapman, S. K. & Reid, G. A.) 111–123 (Humana Press, 1999). doi:10.1385/1-59259-266-X:111
59. Dagliyan, O. *et al.* Engineering extrinsic disorder to control protein activity in living cells. *Science (80-.).* **354**, 1441 LP – 1444 (2016).
60. Karginov, A. V, Ding, F., Kota, P., Dokholyan, N. V & Hahn, K. M. Engineered allosteric activation of kinases in living cells. *Nat. Biotechnol.* **28**, 743 (2010).
61. Motlagh, H. N., Wrabl, J. O., Li, J. & Hilser, V. J. The ensemble nature of allostery. *Nature* **508**, 331 (2014).
62. Hilser, V. J. & Thompson, E. B. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci.* **104**, 8311 LP – 8315 (2007).
63. Coyote-Maestas, W., He, Y., Myers, C. L. & Schmidt, D. Domain insertion permissibility-guided engineering of allostery in ion channels. *Nat. Commun.* **10**, 290 (2019).
64. Li, H., Bahar, I., Chang, Y.-Y. & Yang, L.-W. iGNM 2.0: the Gaussian network model database for biomolecular structural dynamics. *Nucleic Acids Res.* **44**, D415–D422 (2015).
65. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891 (2007).
66. Shenkin, P. S., Erman, B. & Mastrandrea, L. D. Information-theoretical entropy as a measure of sequence variability. *Proteins Struct. Funct. Bioinforma.* **11**, 297–313 (1991).
67. Upmeier zu Belzen, J. *et al.* Leveraging implicit knowledge in neural networks for functional dissection and engineering of proteins. *Nat. Mach. Intell.* **1**, 225–235 (2019).
68. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* (2019). doi:10.1038/s41592-019-0496-6
69. Hilser, V. J., Wrabl, J. O. & Motlagh, H. N. Structural and Energetic Basis of Allostery. *Annu. Rev. Biophys.* **41**, 585–609 (2012).
70. Choi, J. H., Laurent, A. H., Hilser, V. J. & Ostermeier, M. Design of protein switches based on an ensemble model of allostery. *Nat. Commun.* **6**, 6968 (2015).
71. Vertrees, J., Barritt, P., Whitten, S. & Hilser, V. J. COREX/BEST server: a web browser-based program that calculates regional stability variations within protein structures. *Bioinformatics* **21**, 3318–3319 (2005).
72. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
73. Li, H., Chang, Y.-Y., Lee, J. Y., Bahar, I. & Yang, L.-W. DynOmics: dynamics of structural proteome and beyond. *Nucleic Acids Res.* **45**, W374–W380 (2017).
74. Boassa, D. *et al.* Split-miniSOG for Spatially Detecting Intracellular Protein-Protein Interactions by Correlated Light and Electron Microscopy. *Cell Chem. Biol.* **26**, 1407-

1416.e5 (2019).
75. Yudenko, A. *et al.* Rational Design of a Split Flavin-Based Fluorescent Reporter. *ACS Synth. Biol.* (2020). doi:10.1021/acssynbio.0c00454
76. Remeeva, A. *et al.* Effects of Proline Substitutions on the Thermostable LOV Domain from Chloroflexus aggregans. *Crystals* **10**, (2020).
77. Tebo, A. G. *et al.* Circularly Permuted Fluorogenic Proteins for the Design of Modular Biosensors. *ACS Chem. Biol.* **13**, 2392–2397 (2018).
78. Shu, X. *et al.* A Genetically Encoded Tag for Correlated Light and Electron Microscopy of Intact Cells, Tissues, and Organisms. *PLOS Biol.* **9**, e1001041 (2011).
79. Anderson, N. T., Weyant, K. B. & Mukherjee, A. Characterization of flavin binding in oxygen-independent fluorescent reporters. *AIChE J.* **66**, e17083 (2020).
80. Huerta, C., Grishin, N. V & Zhang, H. The "Super Mutant" of Yeast FMN Adenylyltransferase Enhances the Enzyme Turnover Rate by Attenuating Product Inhibition. *Biochemistry* **52**, 3615–3617 (2013).
81. Hou, Y. *et al.* Metabolic engineering of cofactor flavin adenine dinucleotide (FAD) synthesis and regeneration in Escherichia coli for production of α-keto acids. *Biotechnol. Bioeng.* **114**, 1928–1936 (2017).
82. Rodriguez, E. A. *et al.* A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. *Nat. Methods* **13**, 763–769 (2016).
83. Karchin, J. M., Ha, J.-H., Namitz, K. E., Cosgrove, M. S. & Loh, S. N. Small Molecule-Induced Domain Swapping as a Mechanism for Controlling Protein Function and Assembly. *Sci. Rep.* **7**, 44388 (2017).
84. Ha, J.-H., Karchin, J. M., Walker-Kopp, N., Castañeda, C. A. & Loh, S. N. Engineered Domain Swapping as an On/Off Switch for Protein Function. *Chem. Biol.* **22**, 1384–1393 (2015).
85. Nandwani, N. *et al.* A five-residue motif for the design of domain swapping in proteins. *Nat. Commun.* **10**, 452 (2019).
86. Arinkin, V. *et al.* Structure of a LOV protein in apo-state and implications for construction of LOV-based optical tools. *Sci. Rep.* **7**, 42971 (2017).
87. Heins, R. A., Choi, J. H., Sohka, T. & Ostermeier, M. In Vitro Recombination of Non-Homologous Genes Can Result in Gene Fusions that Confer a Switching Phenotype to Cells. *PLoS One* **6**, e27302 (2011).
88. Berg, J., Hung, Y. P. & Yellen, G. A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nat. Methods* **6**, 161–166 (2009).
89. Tantama, M., Martínez-François, J. R., Mongeon, R. & Yellen, G. Imaging energy status in live cells with a fluorescent biosensor of the intracellular ATP-to-ADP ratio. *Nat. Commun.* **4**, 2550 (2013).
90. Imamura, H. *et al.* Visualization of ATP levels inside single living cells with fluorescence resonance energy transfer-based genetically encoded indicators. *Proc. Natl. Acad. Sci.* **106**, 15651–15656 (2009).
91. Lobas, M. A. *et al.* A genetically encoded single-wavelength sensor for imaging cytosolic and cell surface ATP. *Nat. Commun.* **10**, 711 (2019).
92. Yaginuma, H. *et al.* Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Sci. Rep.* **4**, 6522 (2014).
93. Fisher, R. A., Gollan, B. & Helaine, S. Persistent bacterial infections and persister cells. *Nat. Rev. Microbiol.* **15**, 453 (2017).

94. Conlon, B. P. *et al.* Persister formation in Staphylococcus aureus is associated with ATP depletion. *Nat. Microbiol.* **1**, 16051 (2016).
95. Shan, Y. *et al.* ATP-Dependent Persister Formation in Escherichia coli. *MBio* **8**, e02267-16 (2017).
96. Davalos, D. *et al.* ATP mediates rapid microglial response to local brain injury in vivo. *Nat. Neurosci.* **8**, 752–758 (2005).
97. Mohan, R. *et al.* A microfluidic approach to study the effect of bacterial interactions on antimicrobial susceptibility in polymicrobial cultures. *RSC Adv.* **5**, 35211–35223 (2015).
98. Mohan, R. *et al.* A multiplexed microfluidic platform for rapid antibiotic susceptibility testing. *Biosens. Bioelectron.* **49**, 118–125 (2013).
99. Moyed, H. S. & Bertrand, K. P. hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. *J. Bacteriol.* **155**, 768 LP – 775 (1983).
100. Chesterfield, R. J. *et al.* Rational Design of Novel Fluorescent Enzyme Biosensors for Direct Detection of Strigolactones. *ACS Synth. Biol.* **9**, 2107–2118 (2020).
101. Hühner, J., Ingles-Prieto, Á., Neusüß, C., Lämmerhofer, M. & Janovjak, H. Quantification of riboflavin, flavin mononucleotide, and flavin adenine dinucleotide in mammalian model cells by CE with LED-induced fluorescence detection. *Electrophoresis* **36**, 518–525 (2015).
102. Ozbakir, H. F., Anderson, N. T., Fan, K.-C. & Mukherjee, A. Beyond the Green Fluorescent Protein: Biomolecular Reporters for Anaerobic and Deep-Tissue Imaging. *Bioconjug. Chem.* (2019). doi:10.1021/acs.bioconjchem.9b00688
103. Gomelsky, M. cAMP, c-di-GMP, c-di-AMP and now cGMP: bacteria use them all! *Mol. Microbiol.* **79**, 562–565 (2011).
104. Kalia, D. *et al.* Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signaling in bacteria and implications in pathogenesis. *Chem. Soc. Rev.* **42**, 305–341 (2013).
105. Christen, M. *et al.* Asymmetrical Distribution of the Second Messenger c-di-GMP upon Bacterial Cell Division. *Science (80-.).* **328**, 1295 LP – 1297 (2010).
106. Surdo, N. C. *et al.* FRET biosensor uncovers cAMP nano-domains at β-adrenergic targets that dictate precise tuning of cardiac contractility. *Nat. Commun.* **8**, 15031 (2017).
107. Rehmann, H. *et al.* Structure and regulation of the cAMP-binding domains of Epac2. *Nat. Struct. Biol.* **10**, 26–32 (2003).
108. de Rooij, J. *et al.* Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* **396**, 474–477 (1998).
109. Ho, Y. S., Burden, L. M. & Hurley, J. H. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J.* **19**, 5288–5299 (2000).

APPENDIX A

PROTEIN SEQUENCES

>Estrogen receptor ligand-binding domain (ERLBD)

KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHM
INWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDR
NQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSL
EEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHL
YSMKCKNVVPLYDLLLEMLDAHRLHAP

>iLOV

MIEKNFVITDPRLPDNPIIFASDGFLELTEYSREEILGRNARFLQGPETDQATVQKIRD
AIRDQRETTVQLINYTKSGKRFWNLLHLQPVRDQKGELQYFIGVQLDGSDHV

>mCherry

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTK
GGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVT
VTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEI
RQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYER
AEGRHSTGGMDELYK

APPENDIX B

COMMON LIGAND-BINDING PROTEINS IN BIOSENSORS

Ligand	Ligand-binding protein	N-C separation (Å)	
		Holo-form	Apo-form
Glucose	GGBP	Small	Large
4-Hydroxytamoxifen	Estrogen receptor ligand-binding domain	21	64
ATP	B. subtillis F_0F_1 ATP synthase ε subunit	25	~ 100
cAMP	Mouse Epac1	Large	Small
Ca^{2+}	M13 and CaM	< 15	> 25
cGMP	GAFa domain of mouse phosphodiesterase 5α	9.7	12