Engineering of photocleavable protein (PhoCl) and light-inducible nanocapsules

by

Xiaocen Lu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry
University of Alberta

Abstract

Optogenetics is a burgeoning range of biological techniques which involve the use of light and genetically encoded protein to control (actuators) or monitor (indicators) biochemical activities in living cells and tissues. In recent years, photo-controllable actuators engineered from photoreceptors have become widely used in various applications of chemical biology and physiology. In the optogenetic actuator toolkit, the key building block is a genetically encodable light-sensitive protein. Currently, most optogenetic actuators are engineered from naturally occurring photoreceptor protein from bacteria, fungi, and plants. Despite the tremendous progress in development and application of tools based on natural photoreceptors, the scope of physiological process that can be controlled remains limited by protein properties and photosensory mechanisms.

To expand the actuator toolkit, the photocleavable protein (PhoCl), which is engineered from a photoconvertible protein, was introduced by our group in 2017. Illumination of PhoCl with violet light produces a large empty barrel fragment and a small chromophore-containing peptide fragment that dissociate spontaneously. Although the first generation of PhoCl (PhoCl1) has been demonstrated to be a simple and versatile photocleavable linker in a range of cell physiology and materials chemistry applications, an important drawback of the original PhoCl1 was a relative slow rate of dissociation, with a half-time of about 500 s.

In this thesis, I describe my efforts in development of the second generation of PhoCl variants (PhoCl2) with improved rate and efficiency of dissociation. To better understand the structural changes associated with photoconversion and peptide dissociation, we determined the X-ray crystal structures of the green state, the red state, and the cleaved empty barrel of PhoCl1. Guided by the structural information, we further engineered PhoCl using a NanoLuc luciferase-based complementation assay for screening. Two variants of the second generation of PhoCl variants were developed via directed evolution. These variants were designated as PhoCl2c with higher dissociation contrast ratio and PhoCl2f with faster dissociation rate. Compared to the original PhoCl1, the PhoCl2 variants exhibited improved dissociation performances when characterized as purified proteins and in cell-based experiments. Furthermore, we describe the application of PhoCl2c for control of light-induced cell apoptosis in living cells. These data demonstrate that the new and improved variants, PhoCl2c and PhoCl2f, are useful for optogenetic control of subcellular localization and protein interactions.

I also described the development of photo-inducible nanoparticles by engineering protein hybrids composed of encapsulin and a light-sensitive protein. Encapsulin is a virus capsid-like natural nanocapsule with diameter of 25 nm to 32 nm, which is used by bacteria to isolate toxic enzymatic activities. I demonstrated in mammalian cell expression system that enhanced green fluorescent protein (EGFP) can be displayed on the surface of the nanocapsules and a co-expressed cargo can be auto-packaged into the capsules in a fully genetically encodable way. To use encapsulin as a light-inducible capsule, I designed different constructs by fusing encapsulin to PhoCl or blue light photoactivable CRY2-CIBN system. We expect to employ this nanocapsule system for optogenetic release and caging of proteins involved in the control of in cell physiology.

Preface

This thesis is an original work by Xiaocen Lu. No part of this thesis has been previously published.

Chapter 1 of this thesis has been published as Lu, X.; Shen, Y.; Campbell, R. E. Engineering Photosensory Modules of Non-Opsin-Based Optogenetic Actuators. *Int. J. Mol. Sci.* **21**, 6522 (2020). I was responsible for literature review and original draft preparation. Y.S. and R.E.C assisted on review and editing.

Chapter 2 and Chapter 3 of this thesis have been converted into a manuscript in preparation as Lu, X.; Wen, Y; Zhang, W.; Zhang, S.; Shen, Y; Lemieux M. J.; Campbell, R. E. "Engineering photocleavable protein (PhoCl) variants with improved rate and efficiency of dissociation". I was responsible for assembling all constructs, screening new variants, protein characterizations and cell-based experiments as well as the manuscript composition. Y.W. was responsible for protein crystallization, X-ray data collection and structure determination. W.Z. conceived the idea of light-induced cell death experiment and performed the initial test. S.Z. is performing computational simulation. Y.S. contributed to manuscript edits. M.J.L. and R.E.C. are the supervisory authors and assisted on review and editing.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Robert E. Campbell, for his continuous support and insightful guidance on my Ph.D. study throughout the past five years. Before I joined the Campbell group, I knew little about the molecular biology techniques. He taught me lots of knowledge and skill in this powerful area, which made me realized that the research I am doing is of great significant. Also, his personality and attitudes to research and students have a profound impact on my life.

Next, I would like to thank all the professors in my supervisory and examine committee: Dr. Frederick G. West, Dr. Todd L. Lowary, Dr. Sheref S. Mansy, as well as Dr. Kai Zhang from the Department of Biochemistry in University of Illinois at Urbana-Champaign School of Molecular and Cellular Biology for reviewing this thesis and attending my final oral exam.

I sincerely thank the other science faculty member: Mr. Gareth Lambkin in Department of Chemistry for technical training and assistance. I also would like to express my appreciation to my collaborators, as well as all of our group labmates and other friends in department who give me many useful advices, present and past. Particular thanks go to Wei, Yan, Eason, Sally, Rochelin, Shuce, Fang, Yong and Landon. Because of their generous help, I could finish my projects.

Last but not the least, I would like to thank my parent Junhua Liu and Desheng Lu, my husband Xiangyu Li for their continuous support and understanding during my study.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AsLOV2	*Avena sativa* phototropin 1 LOV2 domain
AmEnc	*Acarypchloris marina* MBIC11017 encapsulin
AtCRY1	*Arabidopsis thaliana* cryptochrome 1
avGFP	*Aequorea victoria* green fluorescent protein
αTAT	alpha-tubulin acetylase
BFM	biofunctional module
Bid	BH3 interacting-domain death agonist
BphP	bacterial phytochrome photoreceptor
BMC	bacterial microcompartment
BRET	bioluminescence resonance energy transfer
BV	biliverdin
Cas9	CRISPR associated protein 9
CBid	C-terminal domain of Bid
Cdc42	cell division control protein 42 homolog
CFEM	correlative fluorescence and electron microscopy
ChR2	channelrhodopsin-2
CIB1	cryptochrome-interacting basic-helix-loop-helix protein
CIBN	C-terminal truncated version of CIB1
CLP	cargo loading peptide
CRISPR	clustered regularly interspaced short palindromic repeats
CRY2	*Arabidopsis* cryptochrome 2

dclp	dichroic long pass
DMEM	Dulbecco’s modified Eagle medium
E. coli	*Escherichia coli*
EGFP	enhanced green fluorescent protein
ESI-MS	electrospray ionization mass spectrometry
em	emission
EM	electron microscopy
ePDZ	engineered PDZ domain
ex	excitation
FAD	flavin adenine dinucleotide
FBS	fetal bovine serum
Fd	ferredoxin
FKF1	flavin-binding kelch repeat F-box 1
FMN	flavin mononucleotide
FP	fluorescent protein
GEF	guanine nucleotide exchange factor
GFC	gel filtration chromatography
GFP	green fluorescent protein
HCC	hepatocellular carcinoma
HCV	hepatitis C virus
HEK293T	human embryonic kidney 293T cells
HHBSS	HEPES-buffered Hanks balanced salt solution
HOI	heme oxygenase 1

Hsp90	heat shock protein 90
IPTG	isopropyl β-D-1-thiogalactopyranoside
iRFP	near-infrared fluorescent protein
K_d	dissociation constant
kDa	kilodalton
LARIAT	light-activated reversible inhibition by assembled trap
LED	light emitting diode
LOV	light-oxygen-voltage-sensing
LOVTRAP	LOV2 trap and release of protein
MaB2Enc	*Methylocapsa acidiphila* B2 encapsulin
MAPK	mitogen-activated protein kinase
MBP	maltose binding protein
MD	molecular dynamics
MIB	malonic acid, imidazole and boric acid-buffered system
MP	multimeric protein domain of CaMKII
mRFP1	monomeric fluorescent protein 1
MxEnc	*Myxococcus xanthus* encapsulin
$NADP^+$	nicotinamide adenine dinucleotide phosphate
NBid	N-terminal domain of Bid
NES	nuclear exclusion sequence
NIR	near-infrared
NIR-GECO	near-infrared fluorescent genetically encoded Ca^{2+} indicator
NLS	nuclear localization sequence

NLuc	NanoLuc luciferase
NTA	nitrilotriacetic acid
nMag	negative Magnet
NMR	nuclear magnetic resonance
PA	photoactivatable
PAGE	polyacrylamide gel electrophoresis
PanX1	pannexin-1
PAS	Per-Arnt-Sim
PBS	phosphate-buffered saline
PCB	phycocyanobilin
PCR	polymerase chain reaction
PDB	protein data bank
PΦB	phytochromobilin
PIF	phytochrome interaction factor
PhoCl	photocleavable protein
phot	phototropin
PHR	photolyase homology region
PI3K	phosphatidylinositol 3-kinase
$PI(3,4,5)P_3$	phosphatidylinositol (3,4,5)-trisphosphate
pMag	positive Magnet
PEG	polyethylene glycol
PPI	protein-protein interaction
PTFP	phototransformable fluorescent protein

PhyB	*Arabidopsis thaliana* phytochrome B
Rac1	Ras-related C3 botulinum toxin substrate 1
RBS	ribosomal binding site
RFU	relative fluorescence units
RhoA	Ras homolog family member A
RLU	relative luminescence units
ROI	region of interest
SDS	sodium dodecyl sulfate
SEM	standard error of mean
SR	steroid receptor
STIM1	stromal interaction molecule 1
RMSD	root-mean-square deviation
TBS	Tris-buffered saline
TEV	tobacco etch virus
TmEnc	*Thermotoga maritima* encapsulin
TULIPs	tunable and light-controlled interacting protein tags
τ	time constant
$t_{1/2}$	half-life
VVD	fungal circadian clock photoreceptor Vivid
Δ	deletion
ε	extinction coefficient

Chapter 1 Engineering of optogenetic tools

1. 1 Overview and premise

Optogenetics, a set of technologies that involves the use of light and genetically encoded proteins to control cells and tissues, has enabled a tremendous number of biological breakthroughs in recent years[1–3]. Compared to conventional synthetic organic photoswitches, optogenetic tools enable precise targeting of specific cells or subcellular regions in a relatively non-invasive method. In addition, the numerous emerging optogenetic tools have brought new insights into modern neurobiology and cell physiology by enabling fast and reversible control of transient activities[4]. This combination of light and genetic engineering provides substantial advantages towards the understanding of many biological processes, such as uncovering the mechanism of complex processes (learning[5], sleep[6], addiction[7] and movement[8]) in brain and understanding the pathology of neurological disorders[9–11].

In the broad sense, optogenetic tools include two complementary groups: light-dependent controllable actuators and light-dependent optical indicators[12]. Optogenetic actuators are proteins that can be used to control a specific biological activity when illuminated[13], and optogenetic indicators are protein-based biosensors that transduce microenvironment changes into detectable optical readouts (**Figure 1.1**). The combined use of spectrally and biologically compatible pairs of actuator and indicator allow simultaneous manipulation and monitoring of biological processes[14–16]. Although the early optogenetics tools were mainly used in neuroscience, recent developments in this area are addressing a broader scope of studies in the areas of cell biology and physiology[12,17–19].

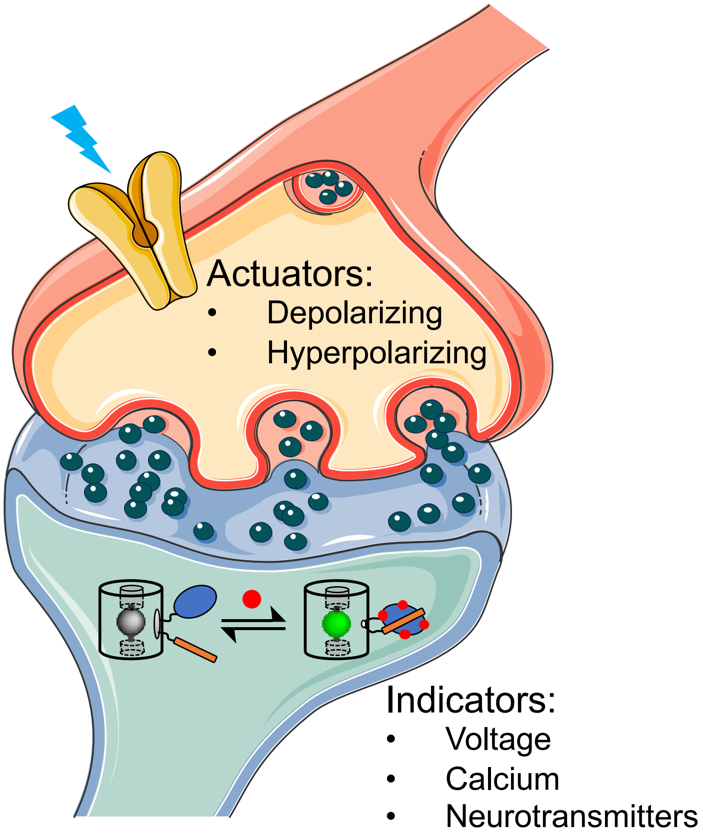


Figure 1.1 Schematic representation of optogenetic actuators and indicator to study neurotransmission. The light-inducible actuators can be used to control the neuron electrical activity by either depolarizing or hyperpolarizing the pre-synaptic neuron. Light-emitting indicators can monitor the changes of neurobiology in membrane potential, calcium ion (Ca^{2+}) signalling, and release of neurotransmitters. (Note: The shapes of synapses and membrane protein are from Servier Medical ART: SMART, https://smart.servier.com. This server is licensed under a Creative Commons Attribution 3.0 Unported License: https://creativecommons.org/licenses/by/3.0/.)

Optogenetic actuators typically consist of a combination of two functional moieties: a photosensory module and a biofunctional module[20]. The photosensory module generally contains a cofactor chromophore or a chromophore generated from autogenic post-translational modification of amino acids, which enables the molecule to absorb light. The absorbance of light typically leads to a photochemical reaction that is associated with a change in the protein structure or function. The biofunctional module is the part that acts as the bridge between the photosensory module and the biological system. The light-induced change in structure or function of the photosensory module changes the structure or function of the biofunctional module, and thus its interaction with the biological system is modulated.

Optogenetic indicators also consist of a photosensory module and a biofunctional module, but the directionality of the interaction between them is the opposite of that in optogenetic actuators. For actuators, the photosensory module absorbs light and undergoes photoreactions which lead to protein conformational changes which ultimately affect the activity of the biofunctional module. For indicators, the biofunctional module interacts with target molecules and alters the microenvironment of the chromophore in photosensory module. Its change is reflected in the fluorescence readout when exposed to excitation light.

At the protein structural level, the two modules can exist in a single protein domain, as in microbial opsin-based actuators (channelrhodopsin-2 [Ref. 21] and halorhodopsin[22]) and pH-sensitive fluorescent protein (FP)-based indicators (pHluorin[23] and pHuji[24]). It is more common for the two modules to be distinct domains which are genetically connected in a protein fusion, as for the genetically encoded Ca^{2+} indicators (cameleon[25] and GCaMP6 [Ref. 26]) and photoactivatable enzymes actuators (photoactivatable Rac1 [Ref. 27] and photoactivatable CRISPR-Cas9 [Ref. 19]). A variety of design strategies have inspired researchers to develop diversified optogenetic tools by using protein engineering via both rational design and directed evolution.

Within the current repertoire, most of the biofunctional modules are engineered from the target molecule-binding proteins[28–31], ion channels[32] or pumps[33]. Also, a colourful palette of FPs[34], or luminescent protein variants[35], has served as the photosensory module of numerous engineered genetically encoded indicators designed using various strategies. However, the light-sensitive protein moieties that serve as the photosensory modules of optogenetic actuators are limited to just several engineered naturally sourced photoreceptor proteins: animal and microbial rhodopsins; plant photochromes, cryptochromes, and phototropins; bacterial phytochromes; and coral photoactivatable FPs. According to their different light-induced responses, the previously reported light-sensitive domains can be divided into three categories[36] (**Figure 1.2**): (i) opsin-based light-activatable channels and pumps[37], (ii) proteins that undergo allosteric conformational changes with light stimulation[38,39], (iii) proteins that undergo light-induced dimerization[40,41].

Although the natural photoreceptor-based optogenetic actuators have many advantages and have been widely applied in biological research, the range of the available tools is still relatively small. There is growing demand for engineering novel light-sensitive protein variants with improved optical properties and light-induced responses to satisfy the needs of a wider variety of studies in the life sciences.

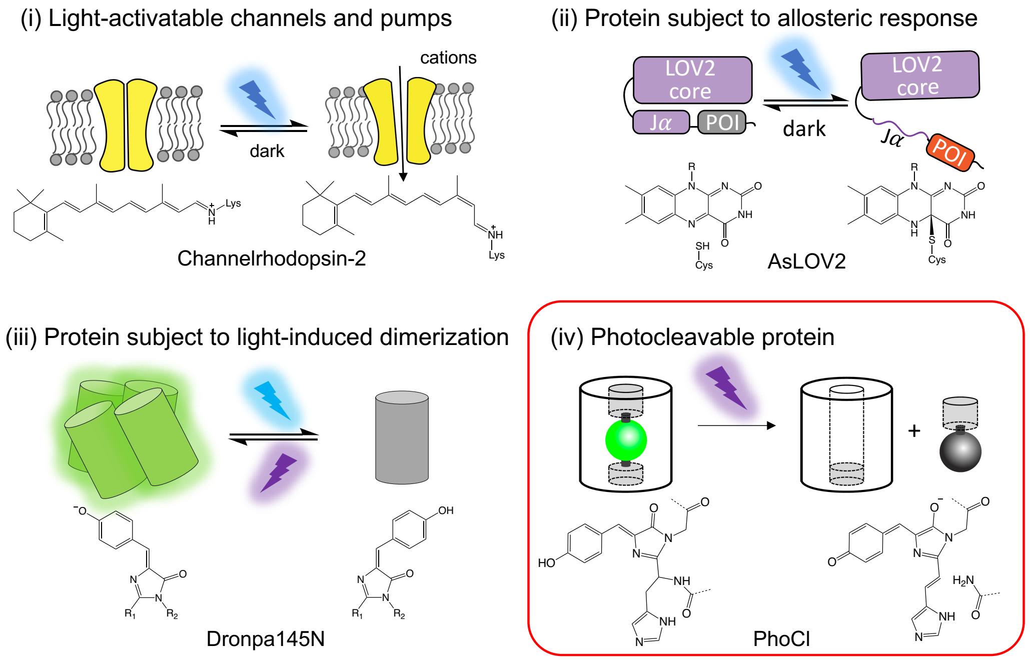


Figure 1.2 Four categories of the light-sensitive protein for actuators. The examples are channelrhodopsin-2 (ChR2) for (i); the LOV2 domain of *Avena Sativa* phototropin 1 (AsLOV2) for (ii); Dronpa145N for (iii); Photocleavable protein (PhoCl) for (iv).

In 2017 our group reported a further expansion of the optogenetics tool repertoire by introducing a fourth category of optogenetic tool (**Figure 1.2**): the photocleavable protein (PhoCl)[36], which was engineered from a photoconvertible FP. Illumination of PhoCl with violet light produces a small C-terminal fragment and a large “empty barrel” N-terminal fragment which can dissociate spontaneously. PhoCl was demonstrated to be a simple and versatile covalent caging of protein of interests that could be activated irreversibly by light. The first generation version of PhoCl was successfully employed for engineering of photoactivatable Cre recombinase, photoactivatable Gal4 transcription factor, and light-activation of protease-activatable ion channel pannexin-1 (PanX1). In work from other groups, PhoCl has been used as a photocleavable linker in photo-response biomaterials[42,43]. Although PhoCl has been applied in growing number of cell physiology and materials chemistry applications, a drawback of the original PhoCl was a relative slow rate of dissociation, with a half-time around 500 s.

For the research described in this thesis, I attempted to overcome the limitations of PhoCl by using protein engineering and structural-guided directed evolution. Having developed improve versions, I then applied PhoCl and another type of optogenetic tool (CRY2-CIB1) to engineer light-inducible nanocapsule systems by using a virus capsid-like protein encapsulin.

In this introductory chapter, I will review the relevant works in engineering of non-opsin-based photosensory domains, and their representative applications in cell biology and physiology. Photosensory proteins to be discussed include light-oxygen-voltage-sensing domain (LOV), cryptochrome (CRY), phytochrome (Phy and BphP), and photoactivatable FP.

1. 2 LOV domain

The light-oxygen-voltage-sensing domain (LOV domain) was first identified as the photosensory module in *Arabidopsis* plant phototropin 1 (Ref. 44,45). Since then, LOV domains have been found in a variety of bacterial, algal, fungal, and other plant species[46]. Structurally, these LOV domains belongs to a subclass of the Per-Arnt-Sim (PAS) superfamily which generally act as a molecular sensor in all kingdoms of life[47]. The small size of the LOV domain core (~ 110 amino acids) is composed of several α-helices and a five stranded antiparallel β-sheet that binds a flavin cofactor chromophore (FMN or FAD) (**Figure 1.3a**). The flavin chromophore absorbs blue light (wavelength, ~ 400 – 480 nm) resulting in the reversible formation of a covalent cysteinyl-flavin adduct in the hydrophobic core[45,48] (**Figure 1.3b**). The photoreduction reaction triggers a protein allosteric change through the highly variable N- or C-terminal extensions[49] which completes the transduction of the light into a biochemical signal[46,50]. This photocycle is thermally reversible in the dark with lifetimes of seconds to days depending on the particular LOV domain[51].

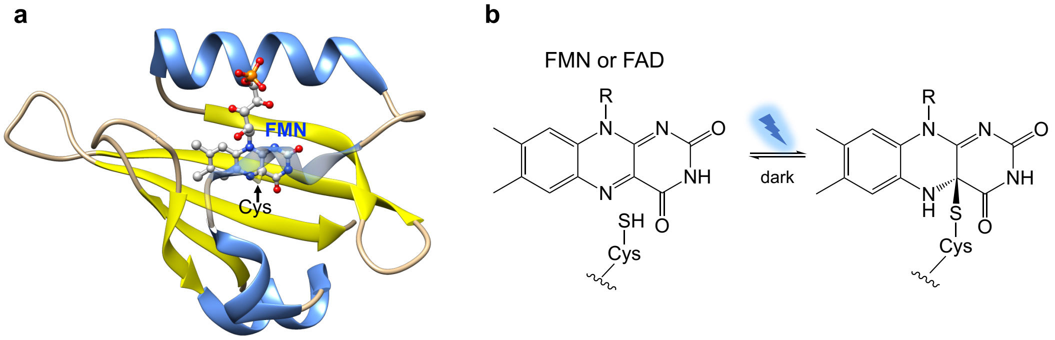


Figure 1.3 LOV domain structure and photochemistry. (**a**) Representative structure of LOV domain. The example is *Adiantum capillus-veneris* phy3-LOV2 domain in illuminated state (PDB: 1JNU). α-helices are represented in blue; β-strands are in yellow; FMN chromophore is represented with light grey ball and stick. (**b**) The photochemistry of flavin chromophore. Upon illumination, a thioether bond is formed between the flavin C4a position and the thiol of cysteine.

Recombinantly expressed LOV domain can incorporate an endogenous flavin cofactor found in various cell types. Thus, it is not necessary to add exogenous cofactor, which makes LOV-based tools applicable in a wide variety of different biological systems. In the optogenetic toolkit, the LOV domain has been utilized for allosteric control of the biofunctional module through two general strategies[52]: (i) light-activatable protein that drives a signalling pathway[27,39]; (ii) light-induced protein-protein interactions that manipulate the subcellular localization of a protein or a protein activity[53,54]. Considering the broad scope of the topic and the relevance to this thesis, I would like to specifically focus on the published examples of engineering of LOV-based actuators from the perspective of the photocycle, proteins with light-induced allosteric responses, and proteins with light-induced dimerization responses.

1. 2. 1 Engineering of LOV photocycles

The kinetics of LOV domain photocycle is dependent on the redox processes of the flavin chromophore[52,55] (**Figure 1.4**). Absorption of blue light promotes the dark ground state species LOV_{450} into a singlet excited state LOV*. LOV* rapidly undergoes intersystem crossing to form the triplet state LOV_{T715}. While in the triplet state, a proton and an electron are transferred from the conserved active site Cys in the domain peptide, creating a second triplet state $\text{LOV}_{\text{T Radical}}$. The radical specie recombines to form the Cys-FMN adduct state LOV_{390}. The adduct is thermally unstable in the dark, and reverts to the ground state LOV_{450} on a broad timescale that depends on the sequence of the particular LOV domain.

Compared to the other steps, the thermal dark recovery from the adduct to ground state is very slow in LOV domains, and is therefore the rate-limiting step in the photocycle. The half-lives of this process vary from a few seconds for *Arabidopsis* phot2 LOV2 (Ref. 56), to tens of seconds or several minutes in most plant and algal phototropin proteins[45,48], to several hours for bacterial and fungal LOV domains: *Bacillus subtilis* photosensor YtvA[57] and *Neurospora* Vivid (VVD)[58]. For *Arabidopsis* FKF1, the photoreduction reaction is irreversible, and the protein does not recover to the dark state after illumination[59]. Due to these differences, when developing new optogenetic tools it is important to choose LOV domain with appropriate decay kinetics according to the timescale of the biological applications. Suitable kinetics would maximize the efficiency of effector activation, and the rate of recycling the optogenetic tool for fast and reversible transient stimulations.

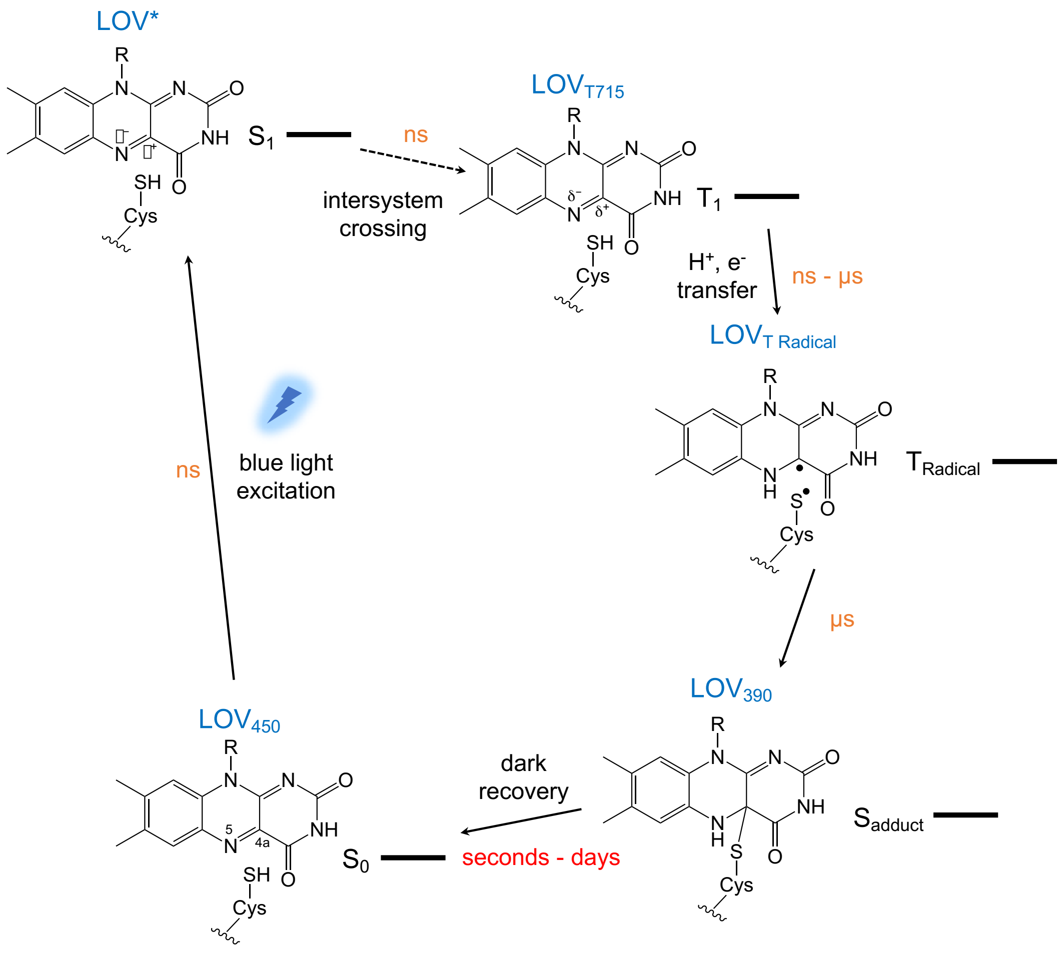
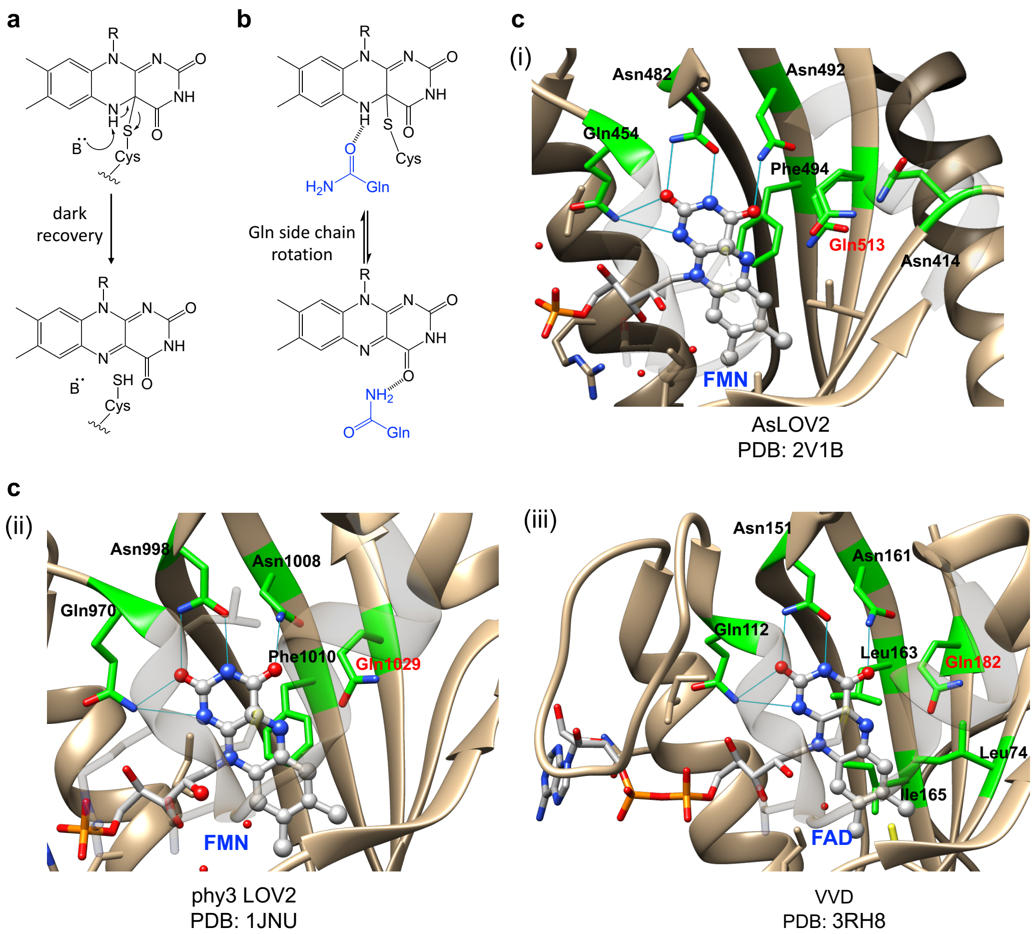


Figure 1.4 Photocycle and kinetics of LOV domain. The species labeled by numbers are defined by the wavelength of maximal absorption. The back bars beside the electronic energy level designations represent the relative energy levels.

Early efforts to engineer the LOV domain were focused on the use of rational design to optimize the dark recovery kinetics. Studies revealed that the Cys-FMN adduct cleavage is assisted by a base-catalyzed deprotonation of the flavin N5 (Ref. 60) (**Figure 1.5a**). The mechanism was found to involve a highly conserved glutamine residue[61] (**Figure 1.5b**) and surrounding side chains that interacted with the flavin and associated water molecules[62–64] (**Figure 1.5c**). Mutational analysis of the residues which interact with the flavin chromophore or vicinal residues near the binding pocket were performed to investigate the key mutations that can tune the photocycle[51,61,65–67].

Here I take the AsLOV2 (wild type: τ_{FMN} = 81 s)[51,65] as an example to review the mutations designed to alter the photocycle. There are three general strategies used in previous studies to alter the dark recovery rate: (i) modification of the electrostatic environment of the chromophore to accelerate or decelerate the deprotonation of the flavin N5 of the adduct (deprotonation factor); (ii) increase or decrease the solvent accessibility of the cofactor (solvent accessibility factor); (iii) alter the stability of the flavin through stacking or steric interactions (flavin stability factor). We will discuss these three factors one by one in the following sections.



AsLOV2
PDB: 2V1B

phy3 LOV2
PDB: 1JNU

VVD
PDB: 3RH8

Figure 1.5 Adduct cleavage mechanism and the conserved residues interacted with flavin in LOV variants. (**a**) Base-catalyzed mechanism for dark recovery. (**b**) Proposed side chain rotation and hydrogen bonding switch of the highly conserved Gln. (**c**) Conserved flavin binding pockets in LOV variants. The residues interacted with chromophore and surrounding water molecules are colored in green. Flavin chromophores (FMN or FAD) are represented in grey ball and stick. The examples are: *Avena Sativa* phototropin 1 LOV2 (AsLOV2; PDB ID: 2V1B) for (i); *Adiantum* phytochrome 3 (phy3; PDB ID: 1JNU) for (ii); *Neurospora* Vivid (VVD; PDB ID: 3RH8) for (iii).

Deprotonation factor

In the AsLOV2 domain, the flavin chromophore directly interacts electrostatically with the residues Gln454, Asn482, Asn492 and Gln513 ((i) in

Figure 1.5c). Mutations to the highly conserved glutamine residue Gln513 (**Figure 1.5b**) have a substantial effect on the recovery rate. The rate of recovery can be slowed down with hydrophobic substitutions Gln513Leu (τ_{FMN} = 1793 s)[61,67] and Gln513Ala (τ_{FMN} = 261 s)[65,67]. In contrast, substitution with a polar residue can provide different degrees of acceleration compared to the wild type. For example, the Gln513Asn has a recovery rate of τ_{FMN} = 43 s, Gln513His has a slightly faster rate (τ_{FMN} = 27 s), and Gln513Asp has a much faster rate (τ_{FMN} = 5 s). These results demonstrate that increasing the electronegativity of the side chain at position 513 facilitates the deprotonation of the adduct and speeds up the cleavage process[67].

Mutations of the other conserved residues Gln454, Asn482, Asn492 ((i) in **Figure 1.5c**) generally compromise the FMN binding or substantially lengthen the photocycle[67,68], due to the disruption of the electrostatic interactions with the chromophore.

Solvent accessibility factor

Other residues which do not directly interact with the chromophore can have effects on tuning the kinetics of the AsLOV2 photocycle. For example, computer simulations reveal that position Asn414 can engage in hydrogen bonding with the Gln513, and that the side chain of this residue undergoes a conformational change between the dark and light states[69,70] (**Figure 1.6**). This residue has indirect interactions with the chromophore and has a large effect on solvent accessibility to the chromophore that influences the deprotonation of N5 in adduct.

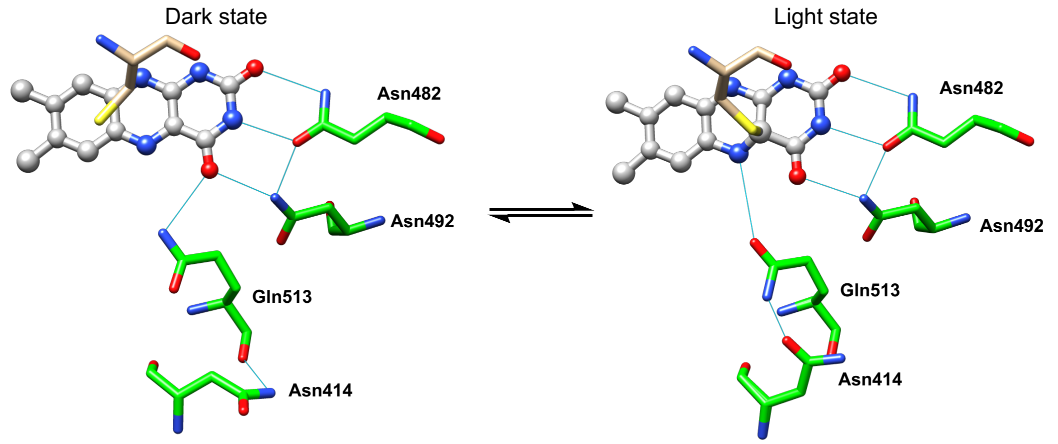


Figure 1.6 Hydrogen bonding networks and the conformational changes of side chains.

In 2014, Zayner *et al*. reported an extensive study on the effects of the Asn414 side chain on the photocycle. They found that the effects of Asn414 mutations are related both to the residue size and electrostatic properties. For hydrophobic residue, a larger sidechain at position 414 led to a longer recovery time (Asn414Leu: τ_{FMN} =1847 s > Asn414Ala: τ_{FMN} = 1427 s > Asn414Gly: τ_{FMN} = 615 s). It appears that the electrostatic factor has a greater impact than the steric factor, with the negative charged substitution Asn414Asp (τ_{FMN} = 69 s) having the fastest lifetime. These results suggested that Asn414 is involved in coordination of the water molecules that facilitates the deprotonation of N5 in adduct and base-catalyzed cleavage process.

Flavin stability factor

The oxidized flavin chromophore in the dark state can form a stronger π-π interaction with Phe494 than the reduced flavin chromophore in the light state ((i) in **Figure 1.5c**). Thus, mutations at position Phe494 generally lengthen the photocycle or compromise the flavin binding through the loss of favourable interactions with dark state form[65,67].

In 2009, Zoltowski *et al.* proposed the generality of two hydrophobic mutations (Ile427Val and Val416Ile) in different LOV domains which can alter the recovery rate. Introduction of these mutations was proposed to rearrange the steric packing around the binding site and affect the stability of the flavin. Mutation Ile427Val reduces the steric hindrance with the active Cys in the dark state (**Figure 1.7a**), and enlarges the solvent accessible space created upon adduct formation. In consequence this substitution increases the recovery rate via stabilizing the dark state[51,71]. In contrast, the mutation Val416Ile decreases the recovery rate by stabilizing the light state adduct through stabilizing the hydrophobic contacts[51] (**Figure 1.7b**). In summary, reducing the steric interactions around the packing can decrease the time of the photocycle, while increasing the steric interactions can lengthen the time of the photocycle.

In addition to the mutations discussed above, there are a number of other variants of note that have been reported to have significant effects on tuning the duration of the photocycle. Here, we summarize the lifetimes of all these previously reported AsLOV2 variants (**Table 1.1**).

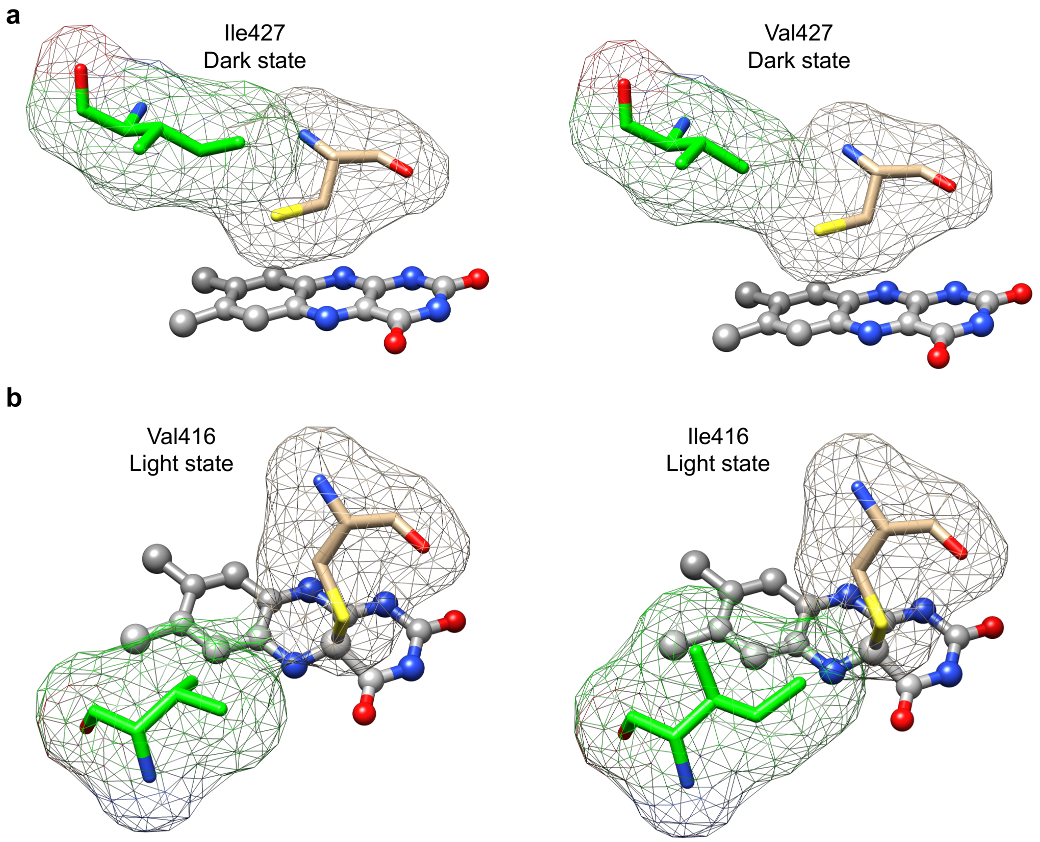


Figure 1.7 The effect of steric interactions. (**a**) Mutation Ile427Val in dark state. (**b**) Mutation Val416Ile in light state. The residues interactions are represented in meshed surface.

Table 1.1 Dark recovery kinetics of AsLOV2 variants.

Protein variants	**Kinetics: τ_{FMN} (s)**	**Reference**
WT	55	Kawano *et al.*, 2013 (Ref. 66)
	68.3	Nash *et al.*, 2008 (Ref. 61)
	80	Zayner *et al.*, 2012 (Ref. 65) Zayner and Sosnick., 2014 (Ref. 67)
	81	Zoltowski *et al.*, 2009 (Ref. 51)
Asn414Asp	69	Zayner and Sosnick., 2014
Asn414Gln	280	
Asn414Gly	615	
Asn414Ser	685	

Asn414Thr	892	
Asn414Ala	1427	
Asn414Leu	1847	
Asn414Val	≥ 12 h	Zayner *et al.*, 2012
Val416Ile	821	Zoltowski *et al.*, 2009
Val416Ile/Leu496Ile	1009	
Val416Thr	2.6	Kawano *et al.*, 2013
Val416Leu	4300	
Ile427Val	4	Kawano *et al.*, 2013
Phe434Leu	12	Zayner *et al.*, 2012
Cys450Val	NM	Zayner and Sosnick., 2014
Leu453Val	160	Zayner and Sosnick., 2014
Phe494Leu	206	Zayner *et al.*, 2012
Phe494Cys	282	Zayner and Sosnick., 2014
Phe494His	NM	
Gln513Asn	37	Nash *et al.*, 2008
Gln513Leu	1080	
Gln513Ala	261	Zayner *et al.*, 2012
Gln513Asp	5	Zayner and Sosnick., 2014
Gln513His	30	
Gln513Leu	1793	
Asn414Ala/Gln513His	2	Zayner and Sosnick., 2014
Asn414Leu/Gln513Ala	1900	
Asn414Ala/Gln513Ala	2081	

Note: Mutations that increase the recovery rate are represented in blue, and mutations that decrease the rate are represented in red.

1. 2. 2 Engineering of LOV domains with light-induced allosteric responses

In LOV-based optogenetic actuators, the photochemical reactions of the flavin chromophore are transduced to the biofunctional domain through protein conformational changes. Taking the AsLOV2 domain as a typical example, the formation of Cys-FMN adduct disrupts the β-sheet in the core of the domain and results in partial unfolding of the C-terminal Jα helix[72,73]. The activity of the biofunctional domain fused to the Jα helix can be modified by this light driven allosteric change.

Although biophysical studies have provided ample precedent for the light-induced conformational changes in the LOV domain, the Jα helix unfolding and conformational changes have not yet been observed in X-ray crystal structures[49]. This is likely due to the compact nature of crystal packing and the stabilizing effects of interprotein contacts. Hence, the extent of conformational change has primarily been evaluated by circular dichroism (CD) signal at 222 nm (θ_{222}). Using this measurement, the allosteric change parameter is defined as $\delta_{222} = (\theta_{222,\ \mathrm{dark}} - \theta_{222,\ \mathrm{light}}) / \theta_{222,\ \mathrm{dark}}$. The δ_{222} of wild type AsLOV2 is 0.30 units, and this parameter mostly reflects unfolding of the C-terminal helix and the mild change in the conformation of β-sheet[65,68].

Zayner *et al*. have reported the characterization of more than 100 variants of AsLOV2 with mutational analysis based on both photocycle kinetics and conformational changes[65,68] (**Figure 1.8a**). Their results reveal that there is no obvious correlation between these two properties (δ_{222} and τ_{FMN}), and mutations in the sensitive sites near the flavin which effect both δ_{222} and τ_{FMN} typically reduce the extent of the conformational change (Asn414Val, $\delta_{222} = 0.25$; Gln513Ala, $\delta_{222} = 0.18$). One possible mechanism to explain this observation is that upon light adduct formation the side chain rotamers of Asn414 and Gln513 (**Figure 1.6**) cause a bulge in the β-stand, which modulates the Jα helix docking affinity[62,69]. This proposed mechanism is supported by a recent molecular dynamics (MD) simulation. Luliano *et al*. used MD simulations to reveal a lever-like motion of Gln513 that induces changes in the hydrogen bonding network and initiates the allosteric pathways of conformational changes leading to Jα helix undocking[70].

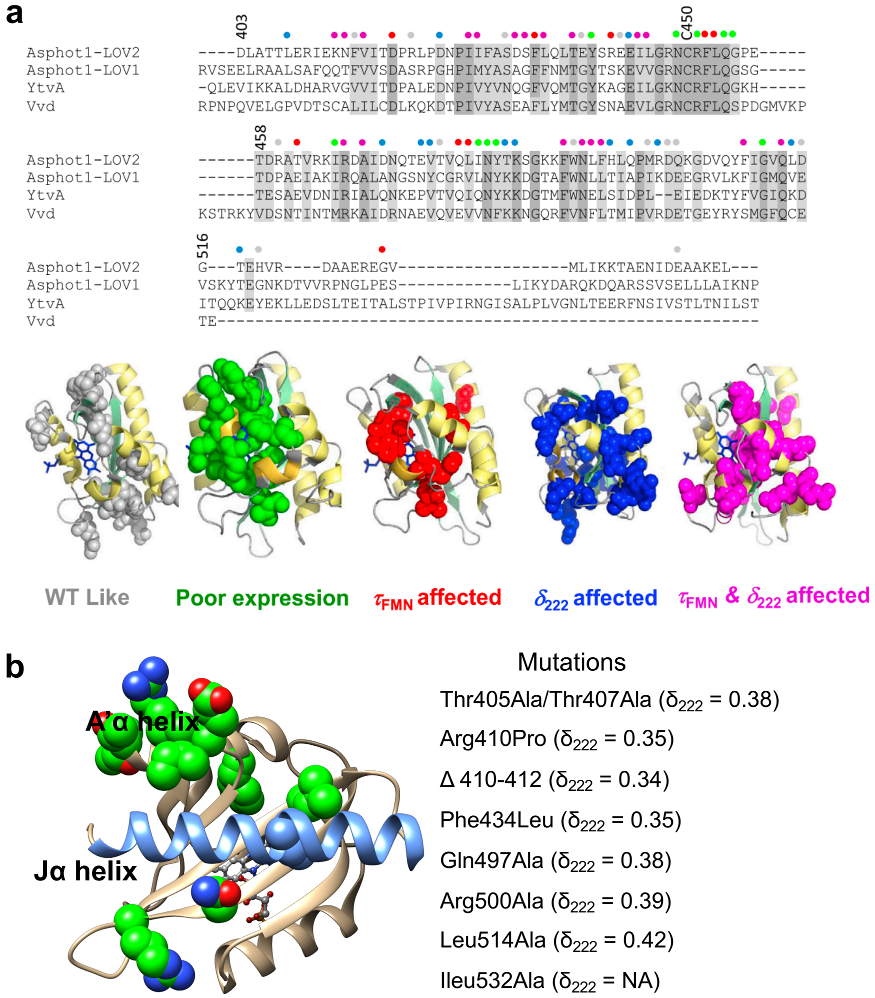


Figure 1.8 AsLOV2 mutations that affect lifetimes or conformational changes. (**a**) Sequence alignment of natural LOV domain variants and consequences of mutations mapped on AsLOV2. (**b**) Mutations that increase conformational change in LOV2. (*Note:* **Figure 1.8a** is reprint from Ref. 68 and used with permission of the publisher)

Based on the results from mutational analysis[68], the positions that contribute to larger conformational change are mostly located on the A' α helix near to the Jα helix and the interface between the Jα helix and the LOV core (**Figure 1.8b**). Substitutions to hydrophobic residues in this region could stabilize the Jα helix docking in the dark state[74]. For optogenetic applications, stability of Jα helix docking in the dark state can alter the affinity, and larger allosteric change could increase the dynamic change of light-dependent activation.

The mutations and mechanisms discussed above have been harnessed to optimize engineered AsLOV2-based optogenetic tools for applications in cell biology. For example, the mutations Thr406Ala, Thr407Ala, Val416Ile, Val529Asn and Ile532Ala were used to tune the equilibrium and kinetic parameters in Tunable, light-controlled interacting protein tags (TULIPs) reported by Strickland *et al.* in 2012 (Ref. 75) (**Figure 1.9**). In the TULIPs system, a biofunctional peptide was fused to the C-terminal of Jα helix and incorporated in helix docking by rational design, such that its function was caged in the dark state. "Uncaging" of the peptide upon illumination led to recruitment of the peptide binding protein (ePDZ), leading to induction of light-dependent protein translocation or other protein-protein interaction (**Figure 1.9a,b**).

For the TULIPs system, the dynamic range of the effector activity was modulated by a number of mutations (**Figure 1.9c**). LOV variants with mutation Val529Asn[76] (which decreases the affinity of Jα helix docking) was demonstrated to increase the binding to ePDZ both in dark and light states, resulting in lower caging of the biofunctional module. In contrast, variants with mutation Thr406Ala, Thre407Ala and Val532Asn[74] (which increase the affinity of Jα helix docking) decreased the recruitment of ePDZ in both states, exhibiting higher caging. This system could be further optimized by introduction of mutation Val416Ile, which led to an approximately 10-fold slower dark recovery rate. TULIPs system with various mutated LOV2 variants have been successfully applied for optical control of mitogen activated protein kinase (MAPK) activation[75].

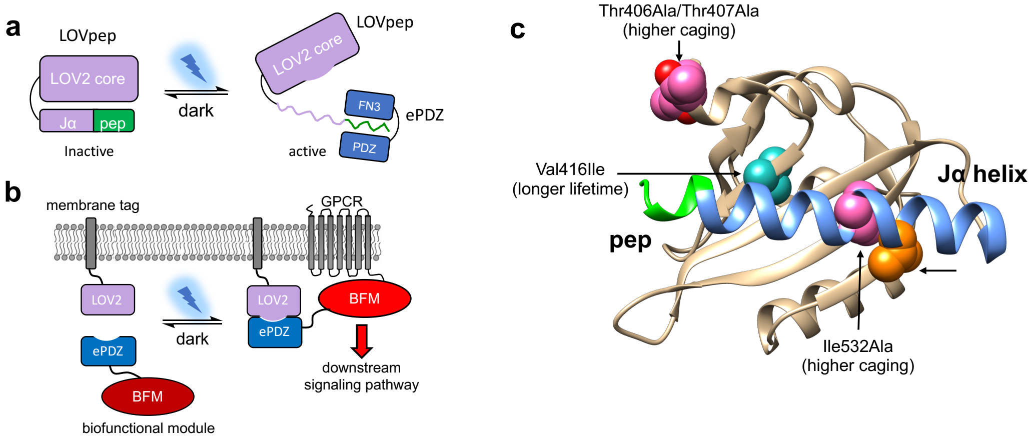


Figure 1.9 Engineering of the TULIP system. (**a**) Schematic of TULIPs design. (**b**) Schematic of optical control strategy by using TULIPs. "BFM" is short for biofunctional module. (**c**) Mutations used in TULIPs to tune equilibrium and kinetics.

Another important example of a LOV domain-based optogenetic tool is the photoinduced protein dissociation system designated LOVTRAP, reported by Wang *et al.* in 2016 (Ref. 53). To develop this system, a small protein named Zdark (Zdk), which selectively binds to the dark state of AsLOV2, was generated by mRNA display screening. Upon illumination, the binding is disrupted by the conformational change of the LOV2 domain, resulting in the uncaging of biofunctional domain by freeing the protein to move to its site of action. By incorporation of previously reported mutations (Ile427Thr, Val416Thr, Ile427Val, Val416Ile and Val416Leu), the time constant of the dark recovery could be tuned from a few seconds to hundreds of seconds. However, instead of only engineering of the AsLOV2 core domain, large differences in light-dark affinity could be achieved by the use of different Zdk variants. The LOVTRAP system, and its derivative Z-lock[77], have been used to mediate activation of different signaling protein that control cell-edge protrusion (Vav2, Rac1, RhoA, cofilin and αTAT) and represent a versatile approach for reversibly photo-caging and photo-uncaging.

Despite the many examples of optogenetic tool engineering using the AsLOV2 photosensory domain, rational improvements of the tools used in biological applications are limited by subtleties of allosteric regulation and can vary from case to case. Generally, development of the AsLOV2-based tool requires substantial empirical efforts for optimization of the LOV domain, linkers, and protein of interest.

1. 2. 3 Engineering of LOV domains that undergo light-induced dimerization

An alternative LOV-based optogenetic control strategy is to alter protein-protein interactions (PPI) via the light-induced protein dimerization of certain LOV variants. As an appropriate candidate, the fungal photoreceptor Vivid (VVD) from *Neurospora crassa* undergoes light-induced homodimerization or heterodimerization upon blue light illumination[58]. Although it uses FAD rather than FMN as a flavin chromophore cofactor, VVD shares very similar photochemistry to AsLOV2 and its photocycle can similarly be altered via rational design based on the strategies discussed above. Accordingly, the dark state lifetime of engineered VVD variants can vary from tens of seconds to several days[51].

For the wild type VVD, the light-induced conformational change propagates to the N-terminal cap (N-Cap) which induces protein homodimerization[50]. More specifically, the side chain rotation of the conserved Gln182 upon light adduct formation favors the hydrogen bonding with the reduced flavin and initiates the movement of the Cys71 side chain to interact with the β-stand where Gln182 is located[50,78,79] (**Figure 1.10a**). The rotation is propagated through rearrangement of the N-Cap, which exposes a hydrophobic cleft to facilitate homodimerization (**Figure 1.10b**). Overall, the formation of the homodimer is driven mostly by the conformational changes of the N-latch and N-hinge regions[78–80].

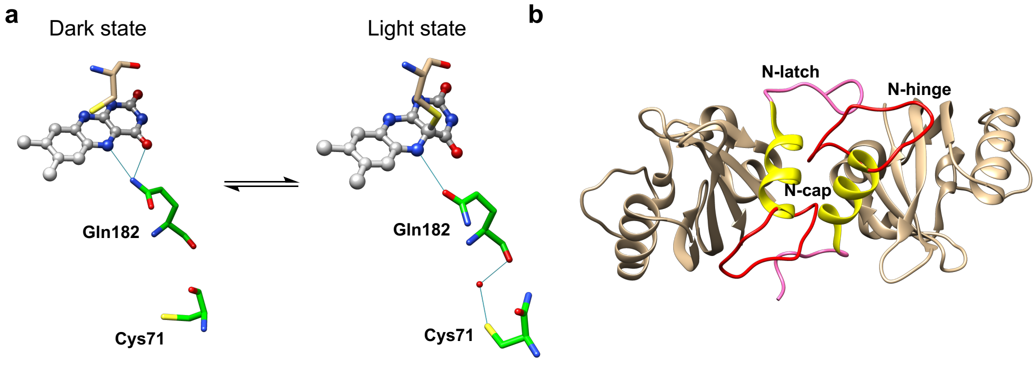


Figure 1.10 Light-induced side chain rotations and homodimerization. (**a**) Hydrogen bonding networks and the conformational changes of side chains. (**b**) VVD light state homodimer. The N-Cap helix is represented in yellow, N-hinge loop is represented in red and N-latch peptide is represented in hot pink.

Due to light-dependent homodimerization, VVD is a suitable scaffold for optogenetic control of PPIs. However, the relatively low affinity of the wild-type homodimer (K_d = 13 μM)[78] is not sufficient to induce robust light-dependent PPI in the context of live cells. To solve this problem, mutations that can increase the homodimer affinity were investigated[78,79]. Unfortunately, due to the highly specific nature of the interactions and the modest conformational change, most substitutions reduce the affinity or result in inactive variants that cannot undergo this light response. For example, variants with Cys71Ser or substitutions at Tyr40 are monomeric in both the dark and light states. The variant with Thr69Trp is dimeric in both the dark and light states[78,79]. To date, relatively few mutations, such as Cys71Val and Asn56Lys, have been reported to improve the affinity in the light state (but not the dark state)[78,79]. Compare to Cys, the Val residue side chain has two methyl groups that points to two directions, mimicking two rotamers of the Cys side chain. One methyl occupies the positions of thiol group in the dark state, and the other methyl mimics the thiol of Cys71 in the light state. The side chain of Val residue predisposes the VVD conformation towards the light state, therefore facilitates the dimer formation. Accordingly, the variant with Cys71Val mutation has an increased affinity (K_d = 0.4 μM)[78]. The Asn56Lys mutation is located on the helix of N-Cap and was found to form a salt bridge with the Asp68 residue on the second VVD domain and thereby stabilize the dimer[81]. The variant with these two mutations was reported to exhibit lower background signals and larger dynamic range than the wild type VVD in the "LightOn" gene expression system. In this system, the dimerization motif of a transcription factor is replaced with the light-inducible VVD[81].

Due to the fact that VVD undergoes a *homo*dimerization, this tool cannot selectively bring two different proteins together (i.e., *hetero*dimerization). To increase its utility, Kawano *et al*. developed two distinct types of VVD variants named pMag (positive) and nMag (negative)[54]. pMag was engineered by the introduction of positively charged residues (substitutions Ile52Arg and Met55Arg) in the N-Cap, and nMag was engineered by introduction of a negatively charged residue (substitution Ile52Asp) and a neutral mutation Met55Gly to prevent homodimerization. Unlike the original homodimeric VVD, the two variants engage in electrostatic interactions instead of primarily hydrophobic contacts and exhibit robust blue light-dependent heterodimerization (with a 21-fold change in bioluminescence complementary assay upon illumination). Moreover, the switch-off kinetics and dimerization efficiencies were optimized introduction of mutations that have been previously reported to alter the photocycle lifetime[51]. The Magnet toolkit was demonstrated to be useful for the optical control of phosphatidylinositol (PI) 3 – kinase (PI3K) for generation of PI(3,4,5)P_3. Overall, this system broadened the scope of optogenetic control by using VVD. In particular, light-induced heterodimerization provides a simple and versatile way to control the activity of split enzymes including Cre recombinase[82], CRISPR-Cas9 (Ref. 19) and T7 RNA polymerase[83] (**Figure 1.11**).

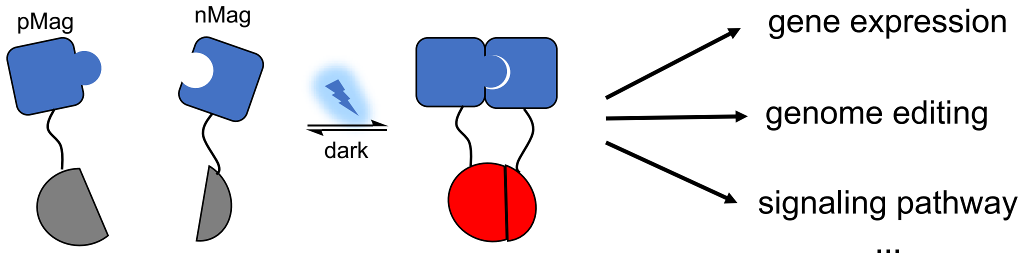


Figure 1.11 Schematic of Magnet system application strategy.

1. 3 Cryptochromes

Another type of photosensitive protein with a flavin cofactor is cryptochrome 2 (CRY2). CRY2 can simultaneously undergo light-induced reversible homo-oligomerization, and heterodimerization with its binding partner (CIB1), upon blue light illumination[84] (**Figure 1.12a**). Similar to the photochemistry of LOV2 domain, the conformational change of CRY2 is initiated by the photoreduction of its FAD chromophore (**Figure 1.12b**) that binds to the N-terminal photolyase homology region (PHR) of CRY2.

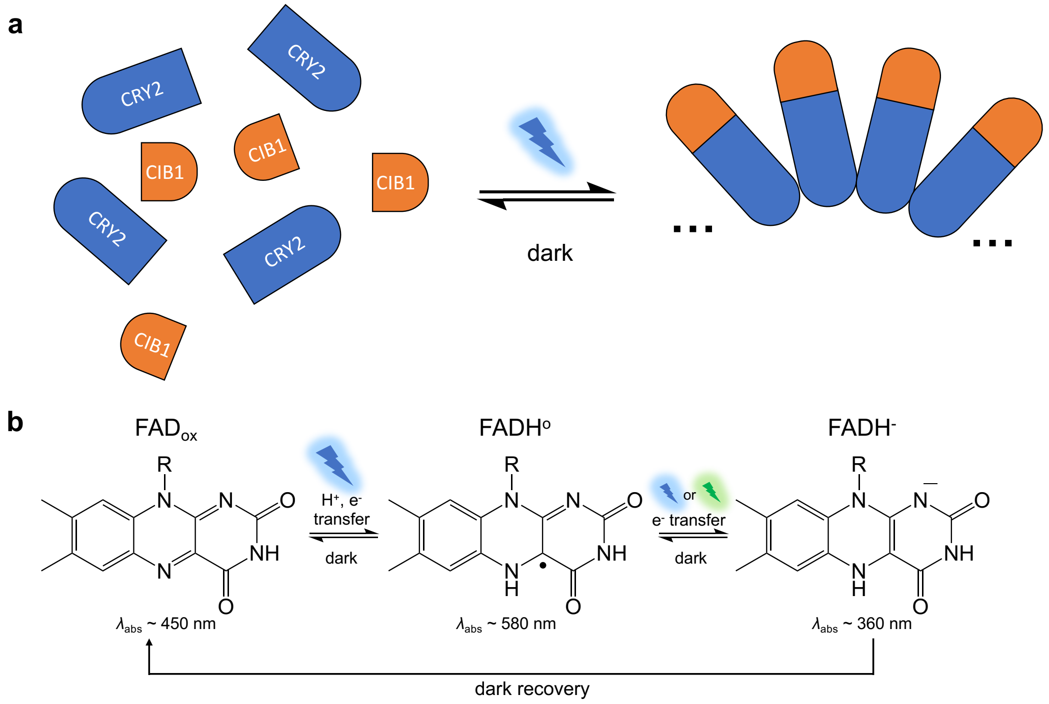


Figure 1.12 Schematic of CRY2-CIB1 system and its photochemistry. (**a**) Homo-oligomerization of CRY2 and heterodimerization between CRY2 and CIB1. (**b**) Photocycle of the FAD chromophore in CRY2.

The utility of the CRY2-CIB1 pair was first demonstrated through the optical control of transcriptional factor Gal4 and Cre recombinase in mammalian cells in 2010 by Kennedy *et al.*[85]. The photocycle of wild type *Arabidopsis* CRY2 is reversible within minutes, which makes it suitable for transient stimulation of many biological processes naturally. However, the wild type full length CRY2 and CIB1 had some limitations as a tool, including low expression level and punctate nuclear localization when expressed in mammalian cells. To address these problems, Kennedy *et al.* optimized this system by using only the PHR domain of CRY2 (CRY2PHR) and a CIB1 C-terminal truncated version (CIBN). They also mutated the predicted nuclear localization signal in CRY2 and CIB1. The resulting

CRY2PHR and CIBN variants expressed much better in cells and showed evenly distributed cytoplasm localization[85]. The optimized CRY2PHR-CIBN system has been applied widely in optogenetic manipulations of protein-protein interactions (PPIs) (**Figure 1.13a**), to control gene expression[85,86], genome editing[85,86], cellular signaling pathways[87,88], lipid metabolism[36], and organelle transport[89].

In 2016, Taslimi *et al.* reported a second generation CRY2-CIB system with smaller protein size and reduced interactions in the dark state[86]. The improved truncated variant CRY2(535) (residues 1-535) was demonstrated to have larger dynamic range and lower dark state self-association than CRY2PHR. They also attempted to identify a smaller binding domain derived from CIB1, but unfortunately efforts to truncate the domain substantially weakened the interaction with CRY2. To improve the versatility of the CRY2-CIB1 system, mutations adjacent to the flavin chromophore that alter the photocycle were identified via directed evolution based on a yeast two-hybrid screening. Compared to the dark recovery rate of wild type CRY2 ($t_{1/2}$ ~ 16 min)[90,91], the variant with long-cycling mutation Leu348Phe has a slower rate ($t_{1/2}$ ~ 24 min), and the variant with fast-cycling mutation Trp349Arg has a faster rate ($t_{1/2}$ ~ 2.5 min) in a cell membrane recruitment assay[85,86]. Guided by the crystal structure of the CRY2 homologue *Arabidopsis* cryptochrome 1 (AtCRY1), the mutation Leu348Phe is thought to stabilize the hydrophobic flavin binding pocket in the light state and therefore lengthens the photocycle. In contrast, Trp349Arg increased the solvent accessibility that facilitates dark thermal conversion, hence speeding up the recovery rate. These proposed explanations are consistent with the mechanisms used to tune the photocycles in LOV domains (as described in 1.2.1).

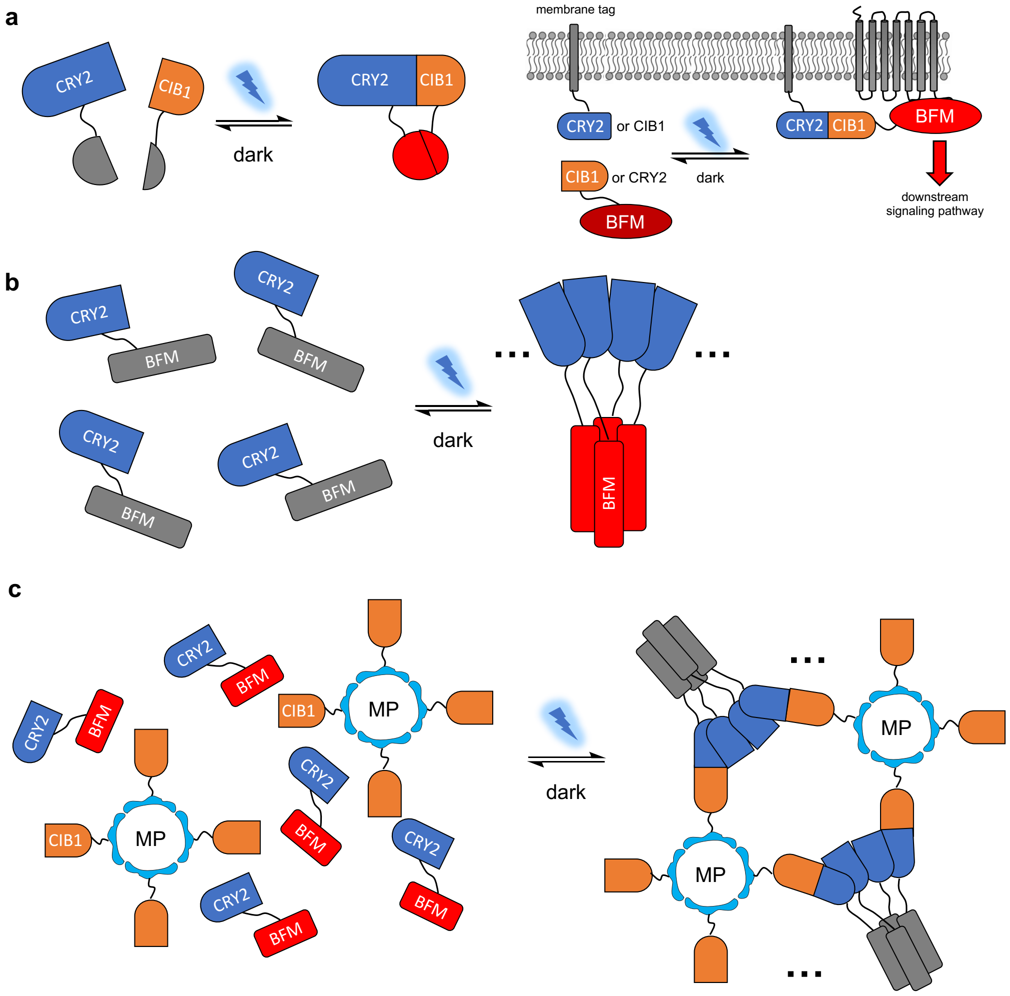


Figure 1.13 Schematics of the CRY2-based optogenetic tools design strategies. (**a**) Protein-protein interactions (PPIs) and protein translocations based on CRY2-CIB1 heterodimerization. (**b**) Activation of biofunctional modules (BFM) based on CRY2 homo-oligomerization. (**c**) Inhibition of biofunctional modules based on a combination of homo-oligomerization and heterodimerization.

In addition to heterodimerization, the light-dependent homo-oligomerization of CRY2 has been harnessed in optogenetic strategies to control a wide range of biological processes[92–95] (**Figure 1.13b**). In 2014，Taslimi *et al.* reported a new CRY2 variant name "CRY2olig" with enhanced homo-oligomerization[94]. The key mutation (Glu490Gly) of this improved variant was discovered through yeast two-hybrid screening. CRY2olig showed improved clustering rate with half-maximal clustering time in the range of tens of seconds after blue light illumination, while the dark dissociation of the larger CRY2olig clusters ($t_{1/2}$ ~ 23 min) is slower than the wildtype CRY2 ($t_{1/2}$ ~ 6 min)[93]. The utility of this light-dependent cluster formation was demonstrated by visualization of protein interactions, and transient inhibition of biofunctional modules through surrounding protein crowding effect. Another variant with improved cluster formation, designated "CRY2clust", was reported by Park *et al.* in 2017 (Ref. 95). The improved function was achieved by adding a 9 residue C-terminal extension following CRY2PHR domain. The improved clustering ability of CRY2clust was demonstrated in a Ca^{2+}-modulating optogenetic system (OptoSTIM1)[96] and optical control of protein kinase activity[97]. However, the mechanism by which the Glu490Gly mutation and C-terminal short peptide extension modulate the oligomerization remained unclear due to lack of structural knowledge of the CRY2 interaction interfaces, which has also limited rational efforts to engineer the CRY2 system.

To better understand the mechanisms of CRY2 interactions, Duan *et al.* proposed the charged residues that affect the heterodimerization and homo-oligomerization based on a model of CRY2 obtained by homology modeling[98]. They found that the CRY2-CIBI and CRY2-CRY2 interactions are dependent on different binding sites on CRY2 (**Figure 1.14**). The N-terminus (residues 2-6) of CRY2 is exposed to solvent, and is important to the heterodimerization to CIB1. The N-terminus which contains three Lys residues and one Asp residue such that it is highly charged, with a net positive charge at physiological pH. The deletion of the N-terminus or mutations of the charged residues to neutral residues significantly reduces the light-dependent binding affinity to CIB1 without impacting homo-oligomerization. In contrast, C-terminal residues tend to be crucial for CRY2-CRY2 interactions. The C-terminal region is predominantly negatively charged, with only a small region containing two positively charged residues Arg487 and Arg489. The mutational analysis indicated that the positively charged residues enhance the oligomerization, however, negatively charged residues counterbalanced this effect. The proposed deduction explained why the Glu490Gly substitution (discussed above)[94] improved the formation of CRY2 clusters. The variants with positively charged residues substitutions (Glu490Arg or Glu490His) exhibited even more improved clustering capacities than Glu490Gly. Based on these findings, the authors engineered two variants designated as CRY2high, with enhanced clustering containing the positively charged substitution Glu490Arg, and CRY2low with reduced oligomerization by using the negatively charged mutations CRY2(1-488)EED. The variants were demonstrated to enable different extents of extracellular signal-regulated kinase (ERK) activation by optogenetic manipulation. Despite the efforts in uncovering the mechanism of CRY2 clustering, the oligomerization interaction remains unclear and it seems increasingly likely that it involves multiple binding sites as well as stabilization by surfaces that are unchanged in both the light and dark states[95,98].

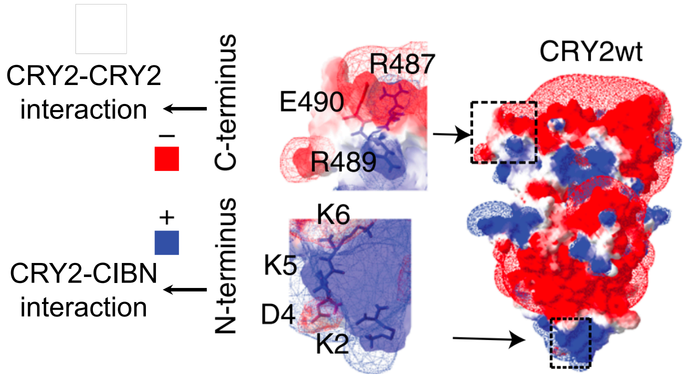


Figure 1.14 Surface charge of CRY2 wild type predicted by homology modeling. (*Note:* **Figure 1.14** is reprint from Ref. 98 and used with open access. This article is licensed under a Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0/.)

An alternative strategy to engineering of improved CRY2 variants, is to improve the cluster formation by combining both CRY2-CRY2 oligomerization and CRY2-CIB1 dimerization. This idea was firstly realized in a system named LARIAT (light-activated reversible inhibition by assembled trap) described by Lee *et al.* in 2014 (Ref. 18). In the LARIAT system, the CIB1 domain is fused to a multimeric protein (MP) (**Figure 1.13c**). Light-dependent homo-oligomerization and heterodimerization induce interactions among the MPs that facilitate cluster formation. This cluster formation leads to inactivation of the biofunctional module due to becoming effectively "trapped" in the clusters. The LARIAT system was demonstrated to be useful for reversible inhibition of biofunctional domains, including Vav2 which induces membrane protrusion, PI3 kinase which is involved in lipid metabolism, and microtubule polymerization in formation of mitotic spindle structure which is crucial in mitosis. A similar strategy has also been used to detect the protein interactions via visualization of the colocalization of the fluorescent clusters and their binding partners[94].

Overall, the engineering of CRY2 has been somewhat limited due to our current lack of understanding of the light-dependent interactions at the molecular level. There is no reported structure of the CRY2 so far. CRY2-based optogenetic tools have proven effective for providing a spatiotemporal control via a variety of design strategies.

1. 4 Phytochromes

Phytochromes are a class of photoreceptors found in plant, fungi, and bacteria which contain bilin chromophores and are sensitive to red and far-red light. The proteins in this superfamily consist a highly conserved photosensory region, and a diversified dimerization or signaling region[99,100]. The photosensory region typically is composed of three adjacent domains: a Per-Arnt-Sim (PAS) domain, a cGMP phosphodiesterase/adenylyl cyclase/FhlA (GAF) domain and a phytochrome-specific (PHY) domain. The chromophore binding occurs in the pocket of GAF domain, with the bilin cofactor covalently connected to a conserved Cys residue located in the GAF or PAS domain, depends on the particular species[100]. For the prototypical phytochrome, the photosensory domain exists in a stable red-light-absorbing conformation (Pr) in the dark state. Upon absorption of red light, Pr reversibly converts into a far-red-absorbing form (Pfr). Pfr can be converted back to Pr by illumination with far-red light or through a dark recovery[101]. The red-shifted excitation wavelengths absorbed by phytochromes (from 620 to 800 nm)[100] are mostly located in the near-infrared (NIR) window[102]. Within the NIR window, the absorption of water molecules and hemoglobin in blood is reduced, therefore light has the maximal depth of penetration in biological tissues[102]. In addition, red light and far-red light illumination minimizes the cross-talk with blue light excitation used for green fluorescent indicators such as GCaMP[26]. Due to their NIR absorbance, phytochromes represent an advantageous template for development into both optogenetic actuators[41,103,104] and fluorescent indicators[105–108]. Here we review the toolkit of phytochrome-based actuators derived from plant phytochromes and bacterial phytochromes.

1. 4. 1 Engineering of plant phytochrome-based actuators

The plant phytochromes are light-sensitive proteins responsible for regulating various plant biological processes, including growth, germination, flowering, and circadian rhythm[69,99,109]. However, only *Arabidopsis* phytochrome B (PhyB) has been employed as a photosensory module in optogenetic actuators. PhyB is a homodimeric protein composed of a N-terminal PAS-like domain, a central chromophore binding GAF domain, and a C-terminal PHY domain which undergoes light-dependent structural rearrangements[110–112] (**Figure 1.15a**). Upon red or far-red light illumination, the protein conformational change is initiated by the isomerization of phycocyanobilin (PCB) or phytochromobilin (PΦB) chromophore[113] (**Figure 1.15b**) which is endogenous in plant and cyanobacteria but is not present in mammalian cells.

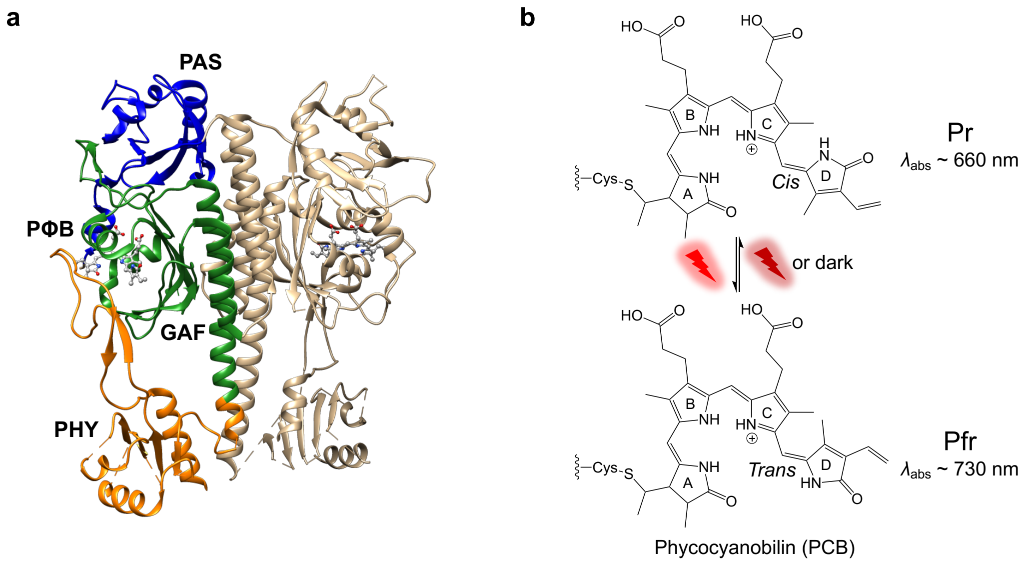


Figure 1.15 Structure of PhyB and its photochemistry. (**a**) Crystal structure of *Arabidopsis* PhyB photosensory region as Pr state dimer (PDB ID: 4OUR). PAS-like domain is represented in blue, GAF domain in green, and PHY domain in orange. The PΦB chromophore is represented in grey ball and stick. (**b**) Photoisomerization of PCB chromophore with red and far-red light.

With *cis* to *trans* isomerization, PhyB converts from an inactive Pr form into an active Pfr form, which in turn binds to the phytochrome-interacting factor (PIF). The PhyB-PIF pair was first demonstrated to be useful as an optogenetic tool in mammalian cells by Levskaya *et al.* in 2009 (Ref. 41). To utilize this protein pair as actuators, they firstly confirmed that PhyB could covalently bind to the exogenously added PCB chromophore in mammalian cells. They then optimized

PhyB (photosensory region with the C-terminal tandem PAS domains) and PIF (N-terminus of PIF6) by screening of variants using a protein translocation assay. The improved "Phy-PIF" pair exhibited reversible light-dependent association and dissociation both within a few seconds (τ_{on} = 1.3 s and τ_{off} = 4 s). Compared to other light-induced dimerization system such as VVD and the CRY2-CIBN pair, the controllability, fast kinetics of both association and dissociation, and red-shifted illumination are major advantages of this system. With these advantages, Phy-PIF pair has enabled spatial and temporal control with higher resolution in various cell biology applications via light-controllable PPI and protein translocation[41,114–118]. Examples include the optical control of Rac1 and Cdc42-mediacted G-protein signalling[41], gene expression by activation of transcriptional factors[114,116] and Cre-mediated DNA recombination in yeast cells[118].

Despite the unique advantages of Phy-PIF system, the major drawback of this system is that the PCB chromophore is only produced in photosynthetic organisms, thus needs to be added externally when using Phy-PIF in mammalian cells. In 2017 Uda *et al.* reported overcoming this limitation by reconstructing the PCB biosynthesis pathway in bacterial and mammalian cells[119]. They reported that endogenous heme can be converted into PCB by expressing a synthetic gene construct "PHFF" encoding four proteins: heme oxygenase (HO1), ferredoxin (Fd), Fd-NADP$^+$ reductase (Fnr) and phycocyanobilin:ferredoxin oxidoreductase (PcyA)[120–122]. The amount of PCB produced via biosynthesis was sufficient to reconstitute the light-dependent Phy-PIF interactions in cell assays[119].

Overall, the development of plant phytochrome-based photosensory module in optogenetic actuators is relatively limited to date. Although the PCB biosynthesis pathway has been transplanted into mammalian cells, this strategy is not ideal due to the extra steps required to incorporate the biosynthesis related genes into the targeted cells. The efficiency of PCB production can also be impacted by the metabolism of intermediates[119]. One potential solution to solve these problems, is engineering the chromophore binding pocket from binding PCB to binding biliverdin (endogenous in mammalian cells) via rational design based on PhyB crystal structure[123].

1. 4. 2 Engineering of bacterial phytochrome-based actuators

Similar to plant phytochromes, bacterial phytochrome photoreceptors (BphPs) undergo photoisomerization between Pr and Pfr forms upon illumination with red or far-red light. However, unlike the plant phytochromes, some bacterial phytochromes adopt the Pfr form, rather than the Pr form, in the ground state[124]. Another difference relative to plant phytochromes is that some eubacteria phytochromes use biliverdin IXα (BV) tetrapyrrole as the cofactor chromophore[100,125] (**Figure 1.16a**). As a product of heme catabolism, BV also exists ubiquitously in mammalian cells[126], therefore bacterial phytochromes are more feasible to be used as photosensory modules in optogenetic tools. Notably, bacterial phytochrome has been extensively engineered into various fluorescent reporters that have been utilized widely for cell imaging (such as iRFP variants[106,127]), a protein interactions reporter (Split[105,106]), and a genetically encoded calcium indicator (NIR-GECO[107,108]).

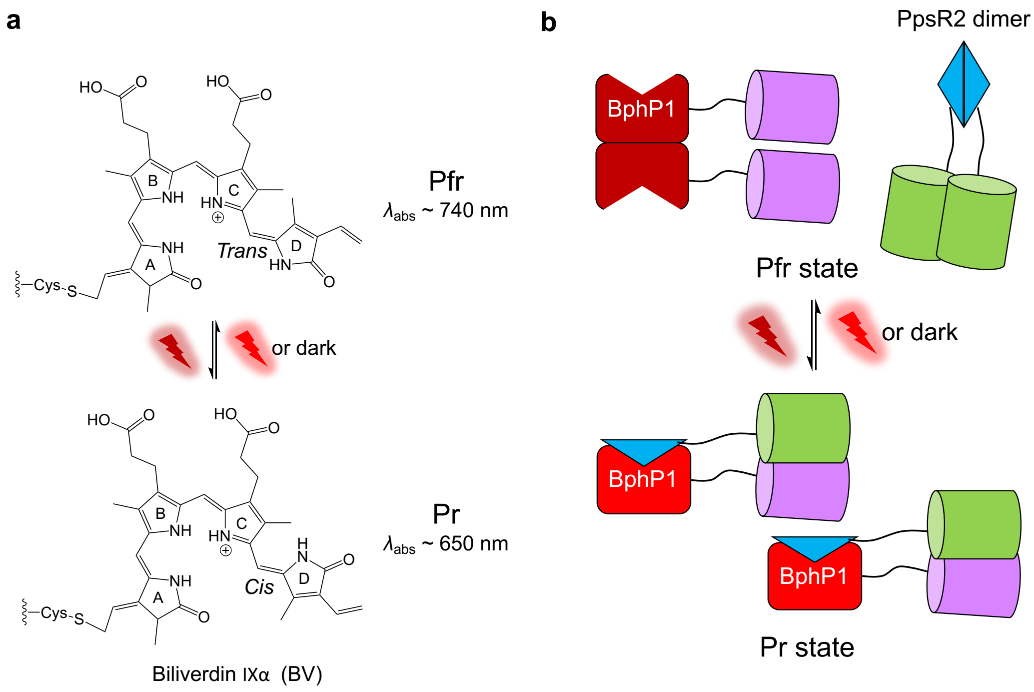


Figure 1.16 BV photochemistry in BphP1 and light-induced BphP1-PspR2 interaction.

The first example of a bacterial phytochrome-based optogenetic actuator, based on bacterial phytochrome BphP1 and a natural gene expression regulator PpsR2, was reported by Kaberniuk *et al.* in 2016 (Ref. 103). PpsR2 and the Pfr form of BphP1 and are both homodimers in dark state. Absorption of far-red light (~ 740 nm) by BV chromophore in BphP1 disrupts the homodimerization and induces heterodimerization between BphP1 and PspR2 (Ref. 103,128) (**Figure 1.16b**). In 2017, Redchuk *et al.* optimized PspR2 variant by identifying the minimal domain required for interaction which is termed as Q-PAS1 (Ref. 104). Using this optimized variant, light-dependent interactions were demonstrated via protein translocation assay in cells[103,104] and *in vivo* light-induced gene expression in mice upon illumination of far-red light[103]. However, one of the limitations of the BphP1-PspR2 system is the poor controllability of the dissociation step. Although the red light (~ 650 nm) illumination accelerates BphP1-PspR2 dissociation, the complete dissociation of the heterodimer only occurs through dark state relaxation. The relative slow dark recovery rate in cell experiments ($t_{1/2}$ ~ 900 s)[103] suggests there is further room for improvement.

1. 5 Fluorescent protein-based actuators

The discovery of the *Aequorea victoria* green fluorescent protein (avGFP)[129–131] ultimately launched a new era in cell biology studies. avGFP enables researchers to monitor biological processes in living cells and tissues by fusing FP to the protein of interest[132,133] or utilizing any of the numerous biosensors based on engineered FPs[25,34,134]. avGFP does not require an exogenous cofactor as the chromophore is generated from the autogenic post-translational modification of three amino acids with the polypeptide sequence (Ser65, Tyr66 and Gly67 in avGFP[134]). Taking avGFP as an example, it is generally believed that the mechanism of chromophore formation involves cyclization, oxidation, and dehydration steps[135] (**Figure 1.17a**). The mature chromophore is covalently bound with a central α-helix surrounded by a highly conserved β-barrel (**Figure 1.17b**).

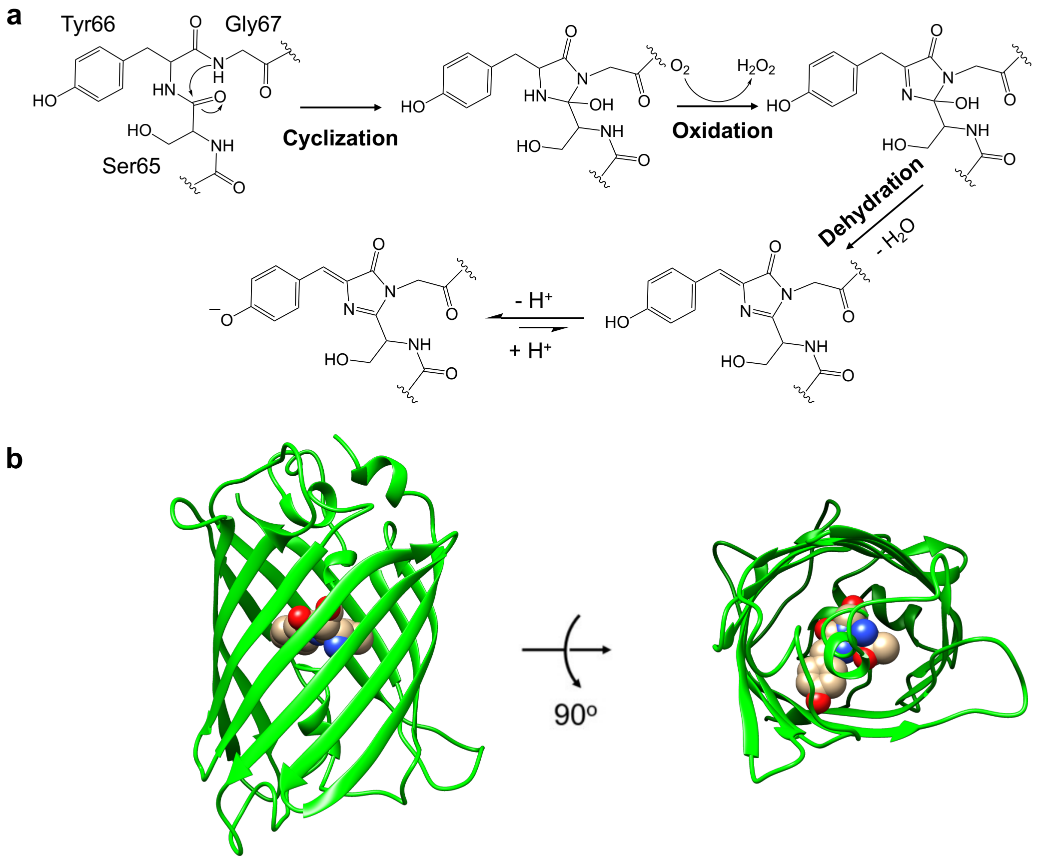


Figure 1.17 Green fluorescent protein (GFP). (**a**) Mechanism of GFP chromophore formation. (**b**) Three-dimensional structure of GFP (PDB ID: 1EMA).

1. 5. 1 Phototransformable FP

In addition to emitting fluorescence upon excitation, the chromophores in some FP variants also undergo photochemical reactions. These FPs are referred to as phototransformable FP (PTFPs)[136]. In contrast to the conventional FPs which exhibit constant fluorescence, the intensity or emission wavelength of PTFPs typically changes upon light illumination. Based on their different light-induced responses, the PTFPs can be divided into three main classes: (i) photoactivatable FP, (ii) photoconvertible FP and (iii) photoswitchable FP.

In a photoactivatable FP (such as PA-GFP[137], PA-mCherry[138] and PA-mRFP1 [Ref. 139]), the protein can be activated irreversibly from the low fluorescent state to high fluorescence state upon illumination with a specific wavelength of light. Photoactivation results from the light-induced (~ 405 nm) decarboxylation of the conserved Glu residue adjacent to the chromophore[140,141] (**Figure 1.18a**). For photoconvertible FP (such as EosFP[142], Kaede[143] and Dendra[144]), the mechanism of photoconversion involves polypeptide backbone cleavage and elongation of the chromophore π-conjugated system[145,146] (**Figure 1.18b**). In consequence the protein switches from green fluorescent state into red fluorescent state irreversibly. For photoswitchable FP (such as Dronpa[147], Dreiklang[148], and kindling FPs[149]), the reversible photoswitching is achieved through photoisomerization of the chromophore[150,151] (**Figure 1.18c**). The protein can be switched between the dark state and fluorescent state by illuminating with two distinct wavelengths of light. The photochemistry of chromophores in PTFPs has expanded the FP applications in important directions (e.g., super-resolution imaging), and has inspired researchers to engineer FPs into photosensory modules for optogenetic actuators.

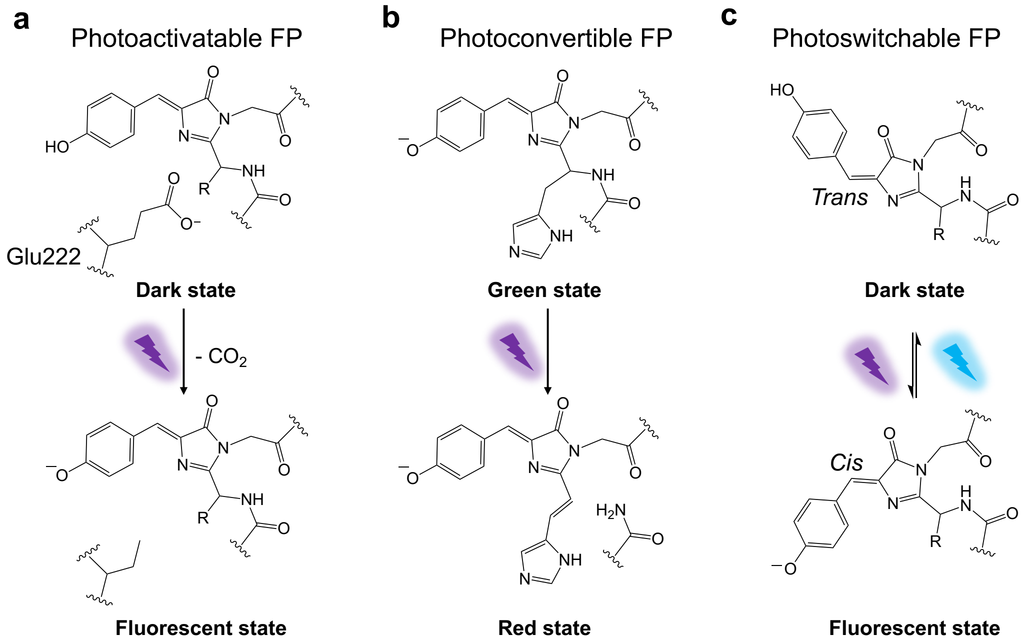


Figure 1.18 Representative photochemistry of phototransformable FPs (PTFPs).

1. 5. 2 Engineering of PTFPs-based actuators

The photochemistry of PTFPs is typically accompanied by a rearrangement of the chromophore pockets[151–153]. For example, in the photoswtichable FP Dronpa[147], the β-strands adjacent to chromophore were observed to be destabilized (by NMR analysis)[152] with cyan light illumination. That is, the *cis* to *trans* isomerization causes partial unfolding of the β-barrel and disturbs the interaction on cross-dimer interface in the tetrameric parent[152]. Taking advantage of this light-induced conformational change, in 2012 Zhou *et al.*[154] reported the optogenetic control of protein activity by using a previously reported Dronpa mutant (Dronpa145N)[155]. Illumination with cyan light induces the tetramer to monomer conversion of Dronpa145N, or the dimer to monomer conversion of the

Dronpa145K variant[154]. Although the affinity of Dronpa variants in multimeric form is relatively low, this system is suitable for "intramolecular oligomerization"-based applications by fusing two copies of Dronpa (**Figure 1.19**). The strategy was utilized for optical control of biofunctional domains such as Cdc42 GEF intersectin, HCV protease and various kinases[154,156].

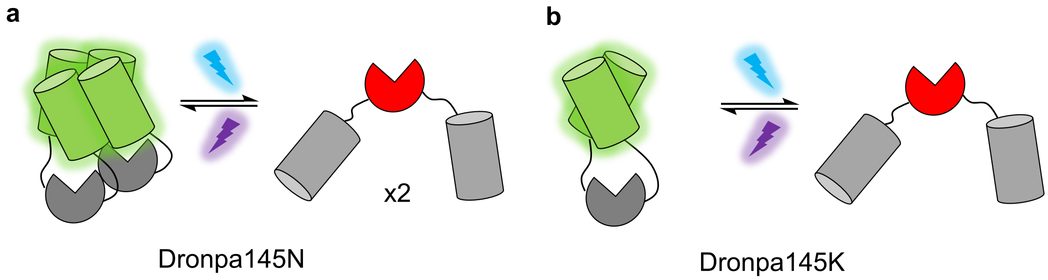


Figure 1.19 Schematic of the Dronpa-based optogenetic tool design strategy.

Another PTFP-based photosensory domain for actuators is the photocleavable protein (PhoCl)[36], engineered from a photoconvertible FP. PhoCl was developed in the Campbell lab and we will describe our work in engineering of first generation of PhoCl in more details in section 2.1.

1. 6 The scope of this thesis

Inspired by the precedent of past work on engineering of the photosensory domains used in optogenetic actuators (as we reviewed in this Chapter), we reasoned that our previously developed PhoCl[36] could likely be further improved by additional protein engineering and directed evolution. In this thesis, I describe my efforts to engineer second generation variants of PhoCl, demonstration of cell-based applications with improved PhoCl2 variants, and my efforts to explore a photo-inducible nanocapsule system via combining encapsulins with different photo-sensitive protein.

In Chapter 2, I describe my efforts to engineer the second generation of PhoCl variants (PhoCl2) with improved rate and efficiency of dissociation. Working with collaborators, we determined the X-ray crystal structures of the green state, the red state, and the cleaved empty barrel of PhoCl1. Using structure-guided engineering, two variants of the second generation of PhoCl variants were developed via directed evolution. These variants were designated as PhoCl2c with higher dissociation contrast ratio, and PhoCl2f for faster dissociation rate.

In Chapter 3, we demonstrate the utility of the improved PhoCl2 variants through protein characterizations. The kinetics and efficiencies of dissociation were compared between PhoCl2 variants and PhoCl1 by using spectroscopy, gel filtration chromatography (GFC) and SDS-PAGE analysis. The improvements in rate and extent of dissociation were further demonstrated in cells via protein translocation assays. We also explored the utility of PhoCl2 for optical control of cell apoptosis by engineering of a light-activatable pro-apoptotic protein Bid.

In Chapter 4, I describe my efforts to develop photo-inducible nanocapsule system via using encapsulins and photosensitive proteins. I expressed and purified the genetically encodable nanoparticles with cargo proteins in bacteria, and demonstrated the modification on both sides of capsules expressed in mammalian cells. I further attempted to achieve the photo-caging and photo-uncaging through fusing encapsulin with PhoCl or CRY2-CIBN pair.

In Chapter 5, we provide a summary of this thesis, as well as a brief discussion on the future directions regarding the development of PhoCl and photo-inducible encapsulins.

Chapter 2 Engineering of the second generation of PhoCl

2. 1 Introduction

Optogenetics is a burgeoning range of techniques that involves the use of light and genetically encoded proteins to enable spatial and temporal control of numerous biological processes in cell physiology[2,37,157]. The previously reported optogenetic actuators can be divided into three categories[36]: (i) opsin-based light-activatable channels and pumps[37], (ii) proteins that undergo allosteric conformational changes with light stimulation[38,39] (as described in section 1.2.2), and (iii) proteins that undergo light-induced dimerization[40,41,86,103] (as described in sections 1.2.3, 1.3, 1.4.1, 1.4.2 and 1.5.2). Although the advantages of these tools are widely recognized, the range of the available repertoire is still relatively small. There is a growing demand for tools with new capabilities to satisfy the needs of a wider variety of biological research applications.

In 2017 our group reported an expansion of this scope by introducing a fourth category of optogenetic tools: the photocleavable protein (PhoCl)[36], which is engineered from the green-to-red photoconvertible FP, mMaple[158]. This class of proteins contains a chromophore that derives from the tripeptide His-Tyr-Gly. The photoconversion reaction is a violet light-induced β-elimination reaction, which extends the conjugated system of the green chromophore to form a red chromophore[145,146] (as described in section 1.5.1). In mMaple, the bond cleavage of the polypeptide backbone forms a ~66 residues N-terminal fragment and a ~166 residues C-terminal fragment that remain associated.

To achieve the light-induced dissociation, a topological variant of mMaple was engineered via circular permutation by Dr. Wei Zhang. The original N- and C-terminus were joined with a flexible linker, and the new termini were introduced near the central helix which are close to the chromophore (**Figure 2.1a**). Upon violet light illumination, cleavage produces a ~10 residues small C-terminal peptide fragment and a ~220 residues large "empty barrel" N-terminal fragment which can dissociate spontaneously (**Figure 2.1b**). To improve chromophore formation efficiency and peptide dissociation rate, linker optimizations and several rounds of directed evolution were preformed, leading to the discovery of the first generation version PhoCl1.

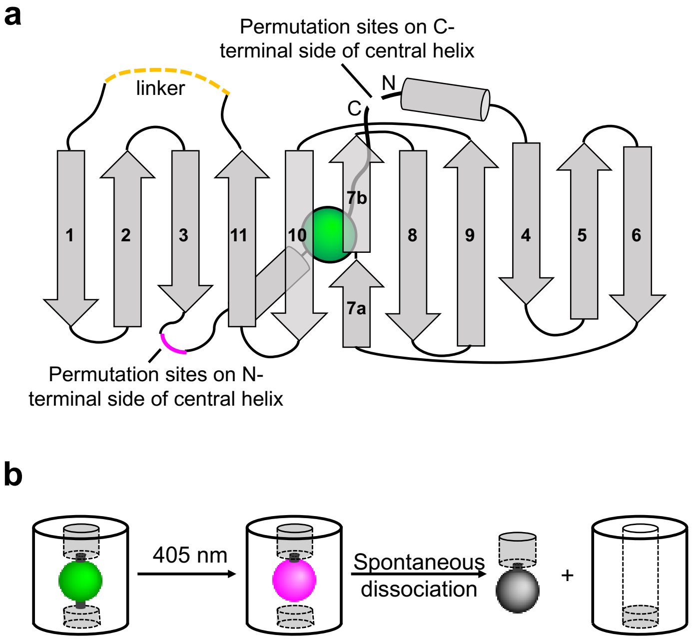


Figure 2.1 Schematics of PhoCl design strategy. (**a**) Engineering of PhoCl from mMaple via circular permutation. (**b**) Schematic of PhoCl photoconversion and dissociation. (*Note*: **Figure 2.1a** is reprint from Ref. 36 and used with author permission.)

The advantages of synthetic photocleavable linkers have been widely recognized in chemical biology studies[159–161]. Inspired by such work, PhoCl represents the first case of genetically encodable photocleavable tag that is composed of only natural amino acids. PhoCl enables a simple and versatile covalent caging of biofunctional module by fusing it to the protein of interest, such as Cre recombinase, Gal4 transcription factor, HCV protease and protease-activatable ion channel pannexin-1 (PanX1). Although PhoCl has been applied in several cell physiology[36] and materials chemistry applications[42,43], a drawback of the original PhoCl1 was a relative slow rate of dissociation, with a half-time around 500 s. We envisioned that PhoCl1 could be amenable to further improvement via protein engineering.

Photoconversion of PhoCl occurs when the green fluorescent chromophore is converted into a red fluorescent chromophore as the result of a photo-induced bond cleavage reaction. Following bond cleavage, the two fragments spontaneously dissociate and the red fluorescence is quenched by exposure of the chromophore to water molecules. Consequently, the original screening method was based on the disappearance rate of red fluorescence[36]. Although the variant (PhoCl1) identified by using this method has faster dissociation than the circular permutated prototype, the improvement was moderate. In addition, the original method is an indirect way to detect the physical separation of the cleaved two parts.

To overcome the limitations of the original screening method, and develop a further improved version of PhoCl, we employed a luciferase complementation assay as a novel high-throughput screening method to measure the dissociation more directly. Guided by the X-ray crystal structures of PhoCl, we have developed a second generation of PhoCl variants with improved dissociation rate or efficiency using protein engineering and directed evolution.

2. 2 Results and discussion

2. 2. 1 Crystal structures of PhoCl1

To obtain a better understanding of the structural changes associated with photoconversion and peptide dissociation, we collaborated with the group of Dr. M. Joanne Lemieux (Department of Biochemistry, University of Alberta) and determined the X-ray crystal structures of the green state of PhoCl1 (2.1 Å resolution), the red state (2.3 Å resolution), and the cleaved PhoCl1 empty barrel (2.8 Å resolution) (**Figure 2.2**).

The red state was produced by partial photoconversion of a crystal of the green state (**Figure 2.2b**). By illumination with 405 nm violet light (15 s light with LED array, 3.46 mW/mm^2), photoconverted PhoCl1 crystals changed from visibly yellowish to visibly orange-red. This orange-red color of the photoconverted PhoCl1 was stable for at least one month, and we did not observe any peptide dissociation in the context of the crystal, presumably due to the very high concentration of protein or steric constraints imposed by crystal packing.

The X-ray crystal structure of the red state revealed that only 2 of the 6 monomers in the asymmetric unit photoconverted from the green to the red state. Initial attempts to model each of the 6 monomers in the asymmetric unit of the photoconverted crystal as a mixture of green and red chromophore structures resulted in unsatisfactory fits to the experimental electron density. The best fit of the electron density was achieved by modeling 2 monomers as red state chromophores (chains A and B), and 4 monomers as green state chromophores (chains C, D, E, F). It remains unclear why 2 of 6 monomers were preferentially photoconverted upon illumination, rather than all 6 monomers experiencing equal photoconversion and each existing as a mixture of 33% red state and 67% green state. We speculate that highly efficient Förster resonance energy transfer (FRET) between the closely packed proteins in the crystal may have funneled energy into those monomers that were in the least favorable orientation to act as FRET donors and/or the most favorable orientation to act as FRET acceptors. These monomers may have been more likely to undergo photoconversion due to their higher probability of excitation or longer excited state lifetimes.

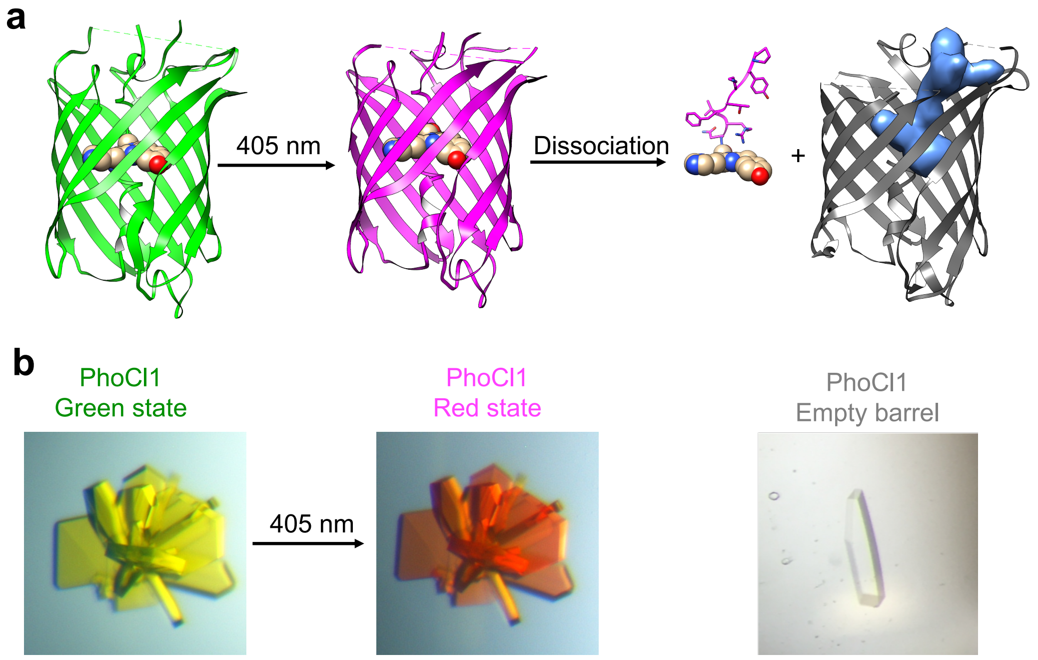


Figure 2.2 PhoCl1 protein structures and crystallization. (**a**) Representation of PhoCl1 structures. Green state of PhoCl1: green structure, the red state: magenta structure, and the cleaved PhoCl1 empty barrel: grey structure. Water-filled cavity was represented as a blue space-filling volume. Dissociated peptide fragment was shown as sticks for the peptide portion and spheres for the chromophore portion. (**b**) Visible color changes in a crystal of PhoCl1 upon illumination, and a transparent crystal of PhoCl1 empty barrel.

The chromophore of PhoCl is generated from the autogenic post-translational modification of His232, Tyr233, and Gly234, which are structurally aligned with Ser65, Tyr66, and Gly67, respectively, of *Aequorea victoria* green FP (avGFP)[134]. All residues are numbered consistently with the numbering used in the X-ray crystal structures. Upon illumination with violet light, the cleavage of the main chain of the protein at the amide bond connecting Phe231 to the His232-derived portion of the chromophore was observed in the red state (**Figure 2.3a**). Other than the changes in the vicinity of Phe231 and the chromophore, as shown here, the overall structures of the green and red states are essentially identical.

Photocleavage and dissociation leaves a water-filled cavity where the peptide fragment previously resided in the interior of the empty barrel structure. To purify the PhoCl empty barrel protein sample for crystallization, the intact PhoCl protein with His purification tag fusion was captured by Ni-NTA agarose and was photoconverted on the column with repeated violet light illumination, and the eluent containing the PhoCl empty barrel was collected. Compared to the red state, the conformational changes were observed in the determined crystal structure of the empty barrel (**Figure 2.3b,c** and **Figure 2.4**). The structure of the peptide fragment was not determined experimentally and is represented here for the sake of clarity.

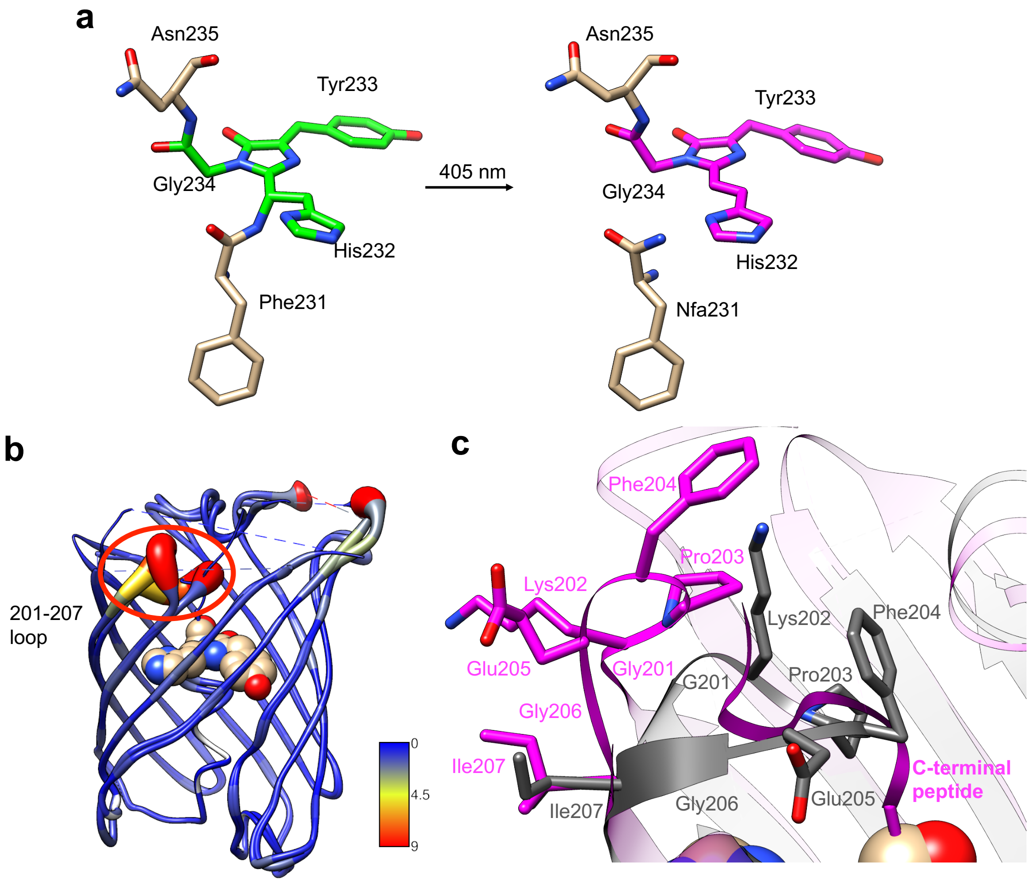


Figure 2.3 PhoCl chromophore structure change and protein conformational changes after illumination. (**a**) The experimentally-determined structure of the PhoCl1 chromophore in the green and red states. (**b**) Structure alignment of the PhoCl1 red state and empty barrel colored by RMSD values. The RMSD values were indicated by scale bar and coil thickness (thicker coil indicates higher RMSD). (**c**) Zoomed-in representation of the 201-207 loop in the aligned structures of the red state (magenta) and empty barrel (grey).

To show the conformational changes in structural alignment of the PhoCl1 red state and the empty barrel more clearly, the root-mean-square deviation (RMSD) of all residues (that is, the average RMSD for all atoms in each residue) between the two structures were calculated. The resulting RMSD values were used to color the cartoon coil alignment representation (using colour scale shown) and to adjust the coil diameter (thicker coil indicates higher RMSD) (**Figure 2.3b**). The representation reveals that the most substantial, and observable, conformational changes occur in the 201-207 loop (indicated by a red oval). In the empty barrel structure, the 201-207 loop has folded in towards the center of the barrel, relative to its position in the red state structure (**Figure 2.3c**).

Notably, if no protein rearrangement had occurred in the empty barrel following peptide dissociation, there should be perfect overlap between the peptide and chromophore and the cavity in the empty barrel. As shown in **Figure 2.4**, the two surfaces are highly overlapped, but the cavity is not identical to the space where the C-terminal peptide resides in the intact protein. The main difference results from folding of the 201-207 loop into the barrel. We speculate that this conformational change helps to stabilize the dissociated state by partially filling the void generated by dissociation of the small peptide fragment.

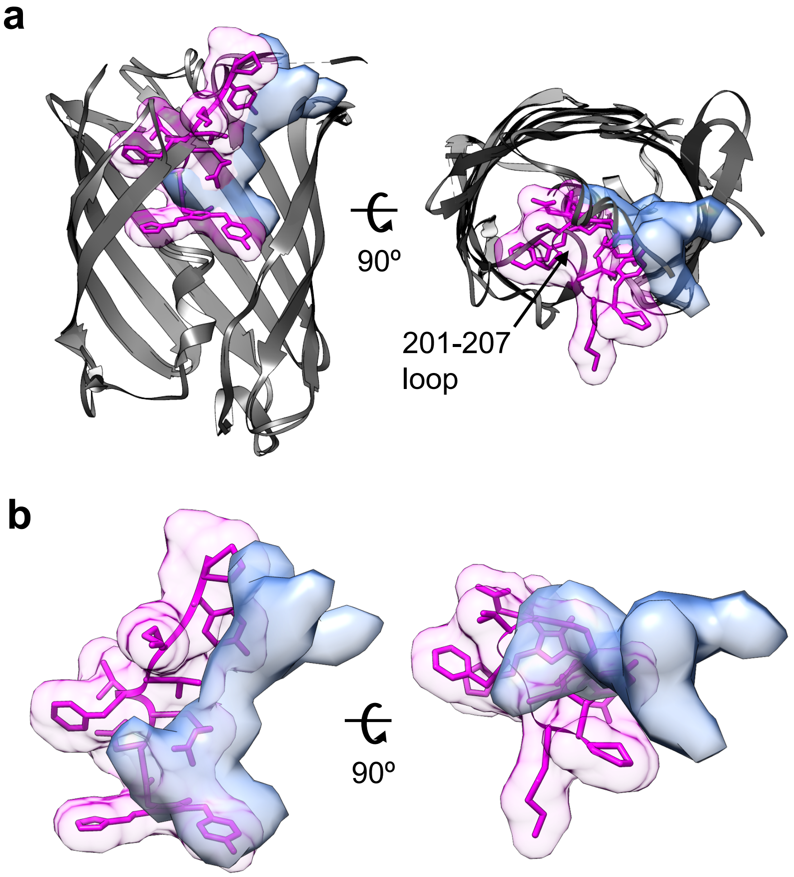


Figure 2.4 Additional information on protein conformational changes. (**a**) The superimposed PhoCl1 red state and empty barrel. The β-barrel of the red state structure is represented in grey, with the dissociable peptide and chromophore shown as magenta sticks surrounded by a semi-transparent magenta surface. The empty barrel structure is represented in grey with the water-filled cavity represented as semi-transparent light blue surface. (**b**) Zoomed-in representation of the superimposed dissociable peptide and chromophore in the red state and water-filled cavity in the empty barrel.

2. 2. 2 Luciferase-based complementation assays for screening

To overcome the drawback of slow dissociation, we attempted to achieve further improvement of PhoCl1 by protein engineering and structure-guided directed evolution. Here, I describe my efforts which led to the development of two second generation PhoCl variants, designated PhoCl2c (higher dissociation contrast ratio) and PhoCl2f (faster dissociation rate).

To screen libraries of PhoCl variants on the basis of dissociation rate and efficiency, we employed a NanoLuc luciferase[162]-based complementation assay (**Figure 2.5a**). PhoCl with long flexible linkers at both termini was incorporated into the NanoLuc Binary Technology (NanoBiT) system[163] (**Figure 2.5b**) which was firstly reported by Dixon *et al.* in 2016 (Ref. 163). NanoBiT luciferase uses the substrate furimazine[162] to produce high intensity luminescence (**Figure 2.5c**) only when the two split fragments are brought together to form the intact enzyme. The low affinity (K_D = 190 μM) of the NanoBiT fragments makes it a suitable split reporter for protein-protein interactions (PPIs) studies[163]. Dissociation of the intact PhoCl upon illumination can be monitored by the decrease in bioluminescence intensity resulting from separation of the two complementary fragments of NanoBiT: Large BiT (LgBiT; 18 kDa) and Small BiT (SmBiT; 11 amino acid peptide).

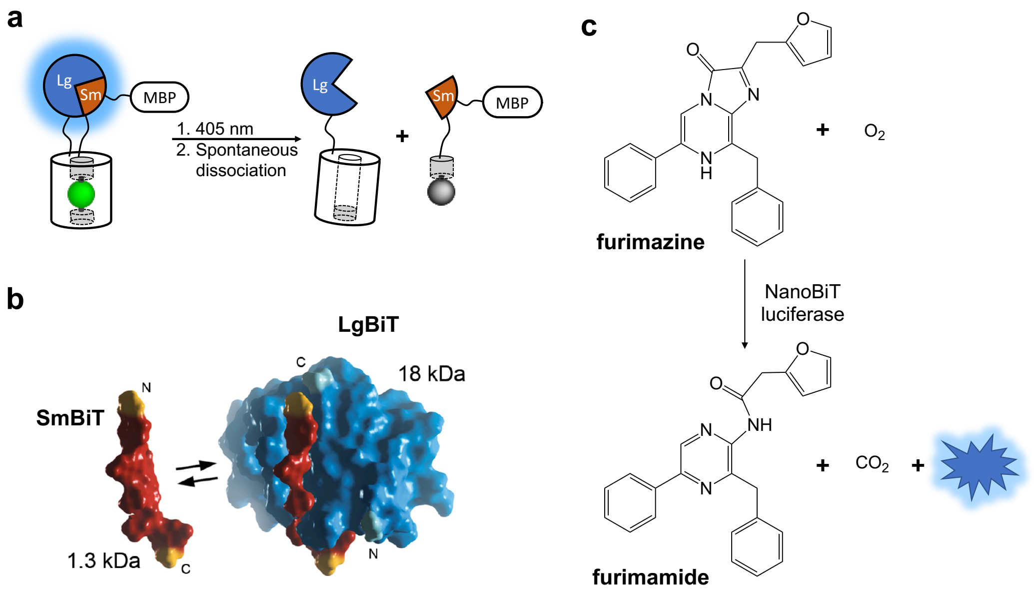


Figure 2.5 Luciferase-based complementary assay design. (**a**) Schematic of the PhoCl screening strategy design. (**b**) The two split complementary fragments of NanoBiT luciferase. LgBiT is presented in blue surface. SmBiT is presented in red surface and its N- and C-terminus are colored in yellow. (**c**) Reaction scheme of the furimazine substrate. (*Note:* **Figure 2.5b** is reprint from Ref. 162 (https://pubs.acs.org/doi/10.1021/acschembio.5b00753) and used with the permission of publisher. Further permissions related to this material excerpted should be directed to the ACS.)

As the proof-of-concept experiments, we tested several constructs containing split *NanoBiT* with or without *PhoCl* insertion (**Table 2.1**). Initially we repeated the original assay reported the literature[163]. Instead of incorporating PhoCl in the middle, LgBiT and SmBiT fragments are connected with a tobacco etch virus (TEV) protease substrate peptide (TEVs). After treatment the purified LgBiT-TEVs-SmBiT protein fusion (500 nM concentration) with TEV protease, the luminescence intensity decreases substantially (**Figure 2.6a**) which confirms that the low affinity between the two fragments of NanoBiT is outside the typical range for PPIs. We then tested the dissociation extent of our previously reported PhoCl (PhoCl1) with LgBiT-PhoCl1-SmBiT purified protein fusion. Bioluminescence signals of the protein samples, with and without illumination to induce photocleavage, were measured at 5 min after photoconversion. We found that PhoCl1 exhibited a moderate decrease in bioluminescence intensity upon illumination due to partially dissociation (**Figure 2.6b**), consistent with our expectations. To increase the solubility of the SmBiT fused to small peptide fragment of PhoCl, we fused the maltose binding protein (MBP) to the C-terminus of the construct. Addition of MBP to the small fragment produced a slightly greater bioluminescence decrease (emission at 460 nm) upon illumination. Accordingly, the construct *LgBiT-PhoCl-SmBiT-MBP* was employed in PhoCl library screening (**Figure 2.6c**).

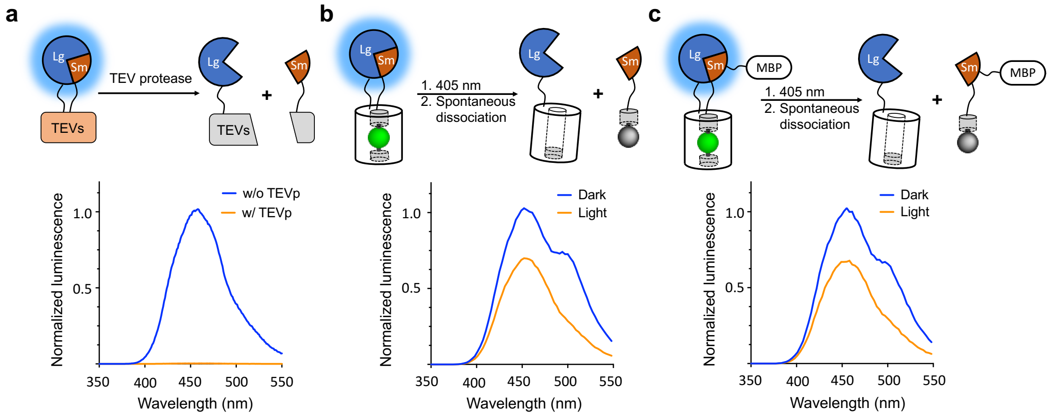


Figure 2.6 Bioluminescence spectra of the fusion proteins. (**a**) Luminescence emission spectra of the LgBiT-TEVs-SmBiT fusion protein, both with and without TEV protease (TEVp) treatment. (**b**) Luminescence emission spectra of the LgBiT-PhoCl1-SmBiT fusion protein, both with and without photoconversion. (**c**)

Luminescence emission spectra of the LgBiT-PhoCl1-SmBiT-MBP fusion protein, both with and without photoconversion.

Table 2.1 Summary table of dynamic ranges in luciferase complementary assay.

Fusion Protein	Emission (460 nm)	Emission (integration)
LgBiT-TEVs-SmBiT	-100%	-100%
LgBiT-PhoCl1-SmBiT	-32%	-42%
SmBiT-PhoCl1-LgBiT	-16%	-22%
SmBiT(86)*-PhoCl1-LgBiT	-12%	-22%
LgBiT-PhoCl1-SmBiT-MBP	-35%	-40%

Note: * SmBiT(86) is a peptide variant with high binding affinity to LgBiT (K_D = 700 pM)[163].

2. 2. 3 Development of PhoCl2 variants

Starting from the template encoding of PhoCl1, two error-prone PCR libraries and 32 site-directed mutagenesis libraries were screened (**Figure 2.7**) using the bioluminescence screening assay described in section 2.2.2.

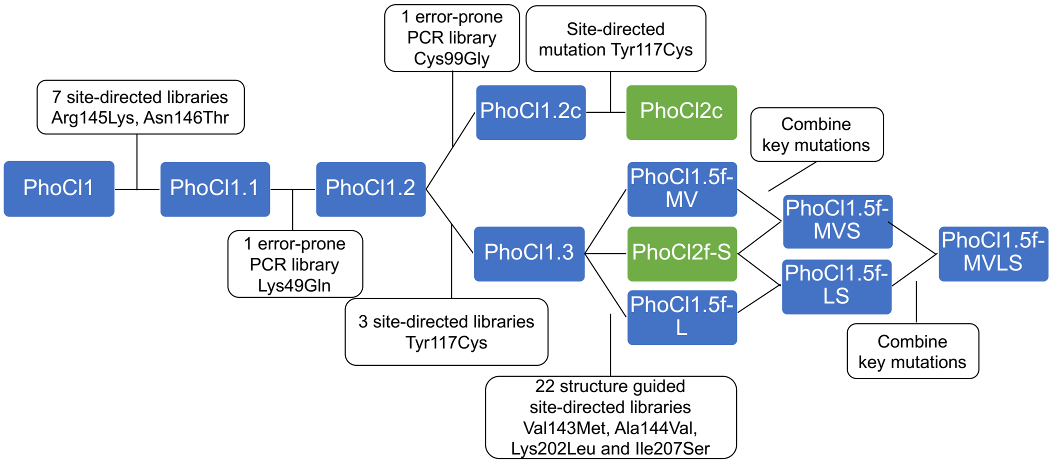


Figure 2.7 Flow chart of PhoCl evolution. The PhoCl variants in the screening process are represented in blue rectangles. The final PhoCl2 variants are presented in green rectangles. The library generation method, specific mutations discovered, and other key points, are described in white rectangles.

Guided by PhoCl1 protein structure of the red state, the surrounding residues that interact with the dissociable peptide were randomized by using site-directed mutagenesis libraries. With the exception of the residues that are hydrogen bonded to the chromophore directly, the residues involved in either hydrogen bonding or hydrophobic contacts with the peptide and chromophore (Ile3, Asp5, Phe7, Lys8, Phe11, Gly14, Tyr15, Ile35, Met37, Glu38, Gly39, Asp40, Phe42, Arg77, Asp78, Gly79, Val80, Met87, Tyr117, His118, Val120, His122, Leu138, Tyr139, Val143, Ala144, Arg145, Asn146, Met162, Asp177, Met178, Gly201, Lys202, Pro203, Phe204, Glu205, Gly206, Ile207, Gln208, Ile210, Thr229, Ala230, Thr239 and Lys240) (in **Table 2.2** and **Figure 2.8**), were randomised by site-directed mutagenesis to introduce a codon (NNK) encoding all 20 amino acids into the gene.

Table 2.2 Summary of the interactions of PhoCl1 dissociable peptide and chromophore in the red state structure.

Residues	**Hydrogen Bonding**	**Hydrophobic Interaction (2.90 Å – 3.90 Å)**
Chromophore	Trp17 (2.86), Arg19 (2.86, 2.44), Ser70 (2.98), Nfa231 (2.85)	Met87, His122, Ile124, Leu138, Tyr139, Glu140, Gln208, Ile210, Thr229, Ala230, Arg236
Asn235	Tyr15 (2.71), Asn180 (3.27), Val237 (3.30)	Ile35
Arg236	Tyr15 (3.00), Glu140 (3.17, 2.49), Tyr239 (2.92,2.68)	Phe7, Met74, His122, Ala142, Iey233, Phe238
Val237	Lys8 (2.96), Asn235 (3.30)	Phe11, Phe42, Pro203, Thr239
Phe238	NA	Ala142, Asn180, Gly201, Lys202, Pro203, Gly206, Gln208, Arg235
Thr239	Lys8 (2.62), Arg236 (2.68, 2.92), Tyr241 (3.12)	Pro203, Gly206, Val237
Lys240	Ala142 (3.26), Ala144 (3.32)	Val143, Phe204, Glu205, Gly206
Tyr241	His118 (2.60), Thr239 (3.12)	Ile3, Asp5, Pro4
Pro242	NA	NA

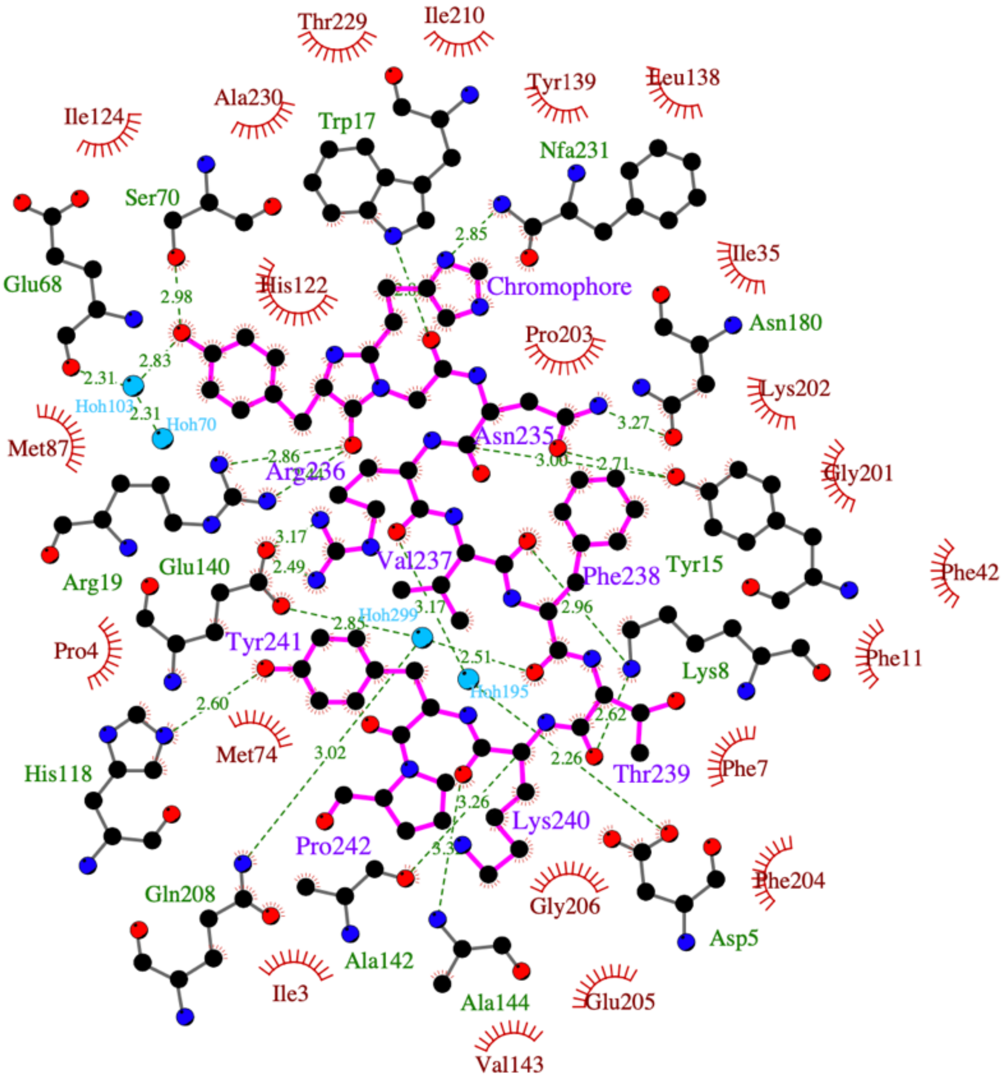
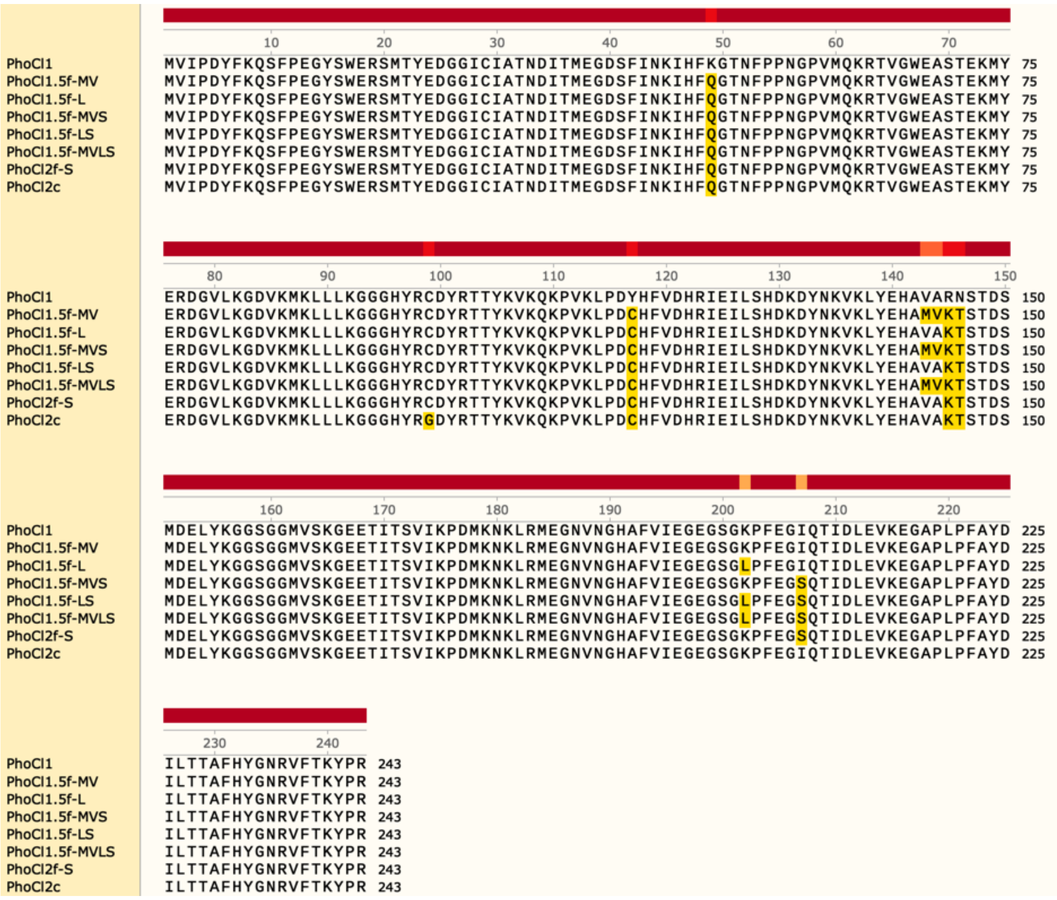


Figure 2.8 Schematic diagram of the dissociable peptide and chromophore in the PhoCl1 red state generated by LIGPLOT. The peptide and chromophore are represented with magenta-colored bonds. Hydrogen bonds are represented in green dashed lines and the hydrogen-bonded residues are represented with grey-colored bonds. Hydrophobic contacts are represented as red arcs with radiating lines.

Using the bioluminescence assay, gene libraries generated by error-prone PCR and site-directed mutagenesis were screened as bacterial lysates. Photoconversion was performed using 30 s of illumination with a 405 nm LED array. The top ~2% variants which exhibited large bioluminescence intensity decreases within 5 min after photoconversion were subjected to DNA sequencing and used as templates for the next round of evolution.

Based on the site-directed mutagenesis screening results, some specific mutations (for example Val143Met, Ala144Val, Lys202Leu and Ile207Ser in loop 201-207) led to significant faster dissociation kinetics (more details provided in section 3.2.1). However, the large majority of mutations caused decreased efficiency of chromophore formation (evident as reduced green fluorescence). Poor chromophore formation was considered to be undesirable because proteins that do not form the chromophore cannot be photoconverted and undergo dissociation.

A total of 34 libraries targeting 46 different positions in the protein were screened (**Figure 2.7**), ultimately leading to the identification of PhoCl2c variant (with Cys99Gly) with enhanced dissociation efficiency. Among the variants with faster dissociation, the variant designated PhoCl2f-S with key mutation Ile207Ser exhibited higher chromophore formation efficiency (will be described in section 3.2.1) and better expression in mammalian cells (will be described in section 3.2.2). Accordingly, this variant was re-named as the final PhoCl2f variant. We will describe the protein characterizations of the PhoCl2 variants with more details in section 3.2. The mutations in new variants are summarized in **Table 2.3** and represented in protein sequence alignment (**Figure 2.9**).



```
PhoCl1           MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFKGTNFPPNGPVMQKRTVGWEASTEKMY  75
PhoCl1.5f-MV     MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFQGTNFPPNGPVMQKRTVGWEASTEKMY  75
PhoCl1.5f-L      MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFQGTNFPPNGPVMQKRTVGWEASTEKMY  75
PhoCl1.5f-MVS    MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFQGTNFPPNGPVMQKRTVGWEASTEKMY  75
PhoCl1.5f-LS     MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFQGTNFPPNGPVMQKRTVGWEASTEKMY  75
PhoCl1.5f-MVLS   MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFQGTNFPPNGPVMQKRTVGWEASTEKMY  75
PhoCl2f-S        MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFQGTNFPPNGPVMQKRTVGWEASTEKMY  75
PhoCl2c          MVIPDYFKQSFPEGYSWERSMTYEDGGICIATNDITMEGDSFINKIHFQGTNFPPNGPVMQKRTVGWEASTEKMY  75

PhoCl1           ERDGVLKGDVKMKLLLKGGGHYRCDYRTTYKVKQKPVKLPDYHFVDHRIEILSHDKDYNKVKLYEHAVARNSTDS  150
PhoCl1.5f-MV     ERDGVLKGDVKMKLLLKGGGHYRCDYRTTYKVKQKPVKLPDCHFVDHRIEILSHDKDYNKVKLYEHAMVKTSTDS  150
PhoCl1.5f-L      ERDGVLKGDVKMKLLLKGGGHYRCDYRTTYKVKQKPVKLPDCHFVDHRIEILSHDKDYNKVKLYEHAVAKTSTDS  150
PhoCl1.5f-MVS    ERDGVLKGDVKMKLLLKGGGHYRCDYRTTYKVKQKPVKLPDCHFVDHRIEILSHDKDYNKVKLYEHAMVKTSTDS  150
PhoCl1.5f-LS     ERDGVLKGDVKMKLLLKGGGHYRCDYRTTYKVKQKPVKLPDCHFVDHRIEILSHDKDYNKVKLYEHAVAKTSTDS  150
PhoCl1.5f-MVLS   ERDGVLKGDVKMKLLLKGGGHYRCDYRTTYKVKQKPVKLPDCHFVDHRIEILSHDKDYNKVKLYEHAMVKTSTDS  150
PhoCl2f-S        ERDGVLKGDVKMKLLLKGGGHYRCDYRTTYKVKQKPVKLPDCHFVDHRIEILSHDKDYNKVKLYEHAVAKTSTDS  150
PhoCl2c          ERDGVLKGDVKMKLLLKGGGHYRGDYRTTYKVKQKPVKLPDCHFVDHRIEILSHDKDYNKVKLYEHAVAKTSTDS  150

PhoCl1           MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGKPFEGIQTIDLEVKEGAPLPFAYD  225
PhoCl1.5f-MV     MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGKPFEGIQTIDLEVKEGAPLPFAYD  225
PhoCl1.5f-L      MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGLPFEGIQTIDLEVKEGAPLPFAYD  225
PhoCl1.5f-MVS    MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGKPFEGSQTIDLEVKEGAPLPFAYD  225
PhoCl1.5f-LS     MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGLPFEGSQTIDLEVKEGAPLPFAYD  225
PhoCl1.5f-MVLS   MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGLPFEGSQTIDLEVKEGAPLPFAYD  225
PhoCl2f-S        MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGKPFEGSQTIDLEVKEGAPLPFAYD  225
PhoCl2c          MDELYKGGSGGMVSKGEETITSVIKPDMKNKLRMEGNVNGHAFVIEGEGSGKPFEGIQTIDLEVKEGAPLPFAYD  225

PhoCl1           ILTTAFHYGNRVFTKYPR  243
PhoCl1.5f-MV     ILTTAFHYGNRVFTKYPR  243
PhoCl1.5f-L      ILTTAFHYGNRVFTKYPR  243
PhoCl1.5f-MVS    ILTTAFHYGNRVFTKYPR  243
PhoCl1.5f-LS     ILTTAFHYGNRVFTKYPR  243
PhoCl1.5f-MVLS   ILTTAFHYGNRVFTKYPR  243
PhoCl2f-S        ILTTAFHYGNRVFTKYPR  243
PhoCl2c          ILTTAFHYGNRVFTKYPR  243
```

Figure 2.9 Sequence alignment of PhoCl variants. Alignment was performed on SnapGene software. The sequence of PhoCl1 was used as reference. The mutations in variants were highlighted in yellow-orange.

Table 2.3 Summary table of mutations in PhoCl2 variants.

Variants	**Mutations**
PhoCl1.5f-MV	Lys49Gln, Tyr117Cys, Val143Met, Ala144Val, Arg145Lys, Asn146Thr
PhoCl1.5f-L	Lys49Gln, Tyr117Cys, Arg145Lys, Asn146Thr, Lys202Leu
PhoCl1.5f-MVS	Lys49Gln, Tyr117Cys, Val143Met, Ala144Val, Arg145Lys, Asn146Thr, Ile207Ser
PhoCl1.5f-LS	Lys49Gln, Tyr117Cys, Arg145Lys, Asn146Thr, Lys202Leu, Ile207Ser

PhoCl1.5f-MVLS	Lys49Gln, Tyr117Cys, Val143Met, Ala144Val, Arg145Lys, Asn146Thr, Lys202Leu, Ile207Ser
PhoCl2f-S	Lys49Gln, Tyr117Cys, Arg145Lys, Asn146Thr, Ile207Ser
PhoCl2c	Lys49Gln, Cys99Gly, Tyr117Cys, Arg145Lys, Asn146Thr

2. 3 Conclusion

In this chapter, we describe the structure-guided engineering of the second generation of PhoCl variants (PhoCl2) with improved dissociation rate (PhoCl2f) and extent (PhoCl2c). The determined X-ray crystal structures of the PhoCl1 provided us with valuable insight into the conformational changes associated with light-dependent dissociation. By screening libraries in which residues of interest were randomized, a number of key mutations that affect the dissociation were identified. In ongoing work in our lab, the detailed dissociation process is being further elucidated via computational modeling and simulation study. An important conclusion from this work is that the luciferase-based complemention assay can be utilized for screening the dissociation of PhoCl. We expect that this high-throughput screening method based on protein-protein interactions (PPIs) could be used to engineer other improved photosensory domains in optogenetic actuators.

2. 4 Methods and materials

2. 4. 1 General molecular biology methods and materials

All synthetic DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). Plasmid constructions were performed with standard restriction enzyme cloning or In-fusion HD Cloning kit (Takara Bio USA) according to the manufacturer's protocol. Q5 high-fidelity DNA polymerase (New England BioLabs) was used for standard polymerase chain reaction (PCR) amplification, and Taq DNA polymerase (New England BioLabs) was used for error-prone PCR. Restriction enzymes and T4 ligase were purchased from Thermo Fisher Scientific. Site-directed mutagenesis was performed using QuikChange Lightening kit (Agilent). Transformation of bacterial cells with plasmid DNA was performed by electroporation using a MicroPulser Electroporator (Bio-Rad). GeneJET gel extraction kit and plasmid miniprep kit (Thermo Fisher Scientific) were used for DNA extraction. DNA sequencing was performed by the University of Alberta Molecular Biology Service Unit.

2. 4. 2 Plasmid construction

For constructing the PhoCl screening vector, the *NanoBiT* template was ordered as a synthetic gBlock from IDT. The *NanoBiT* gene sequence was back-translated from the amino acid sequence[163] with codon optimization for *Escherichia coli* expression. A gene encoding LgBiT-PhoCl-SmBiT containing 5' KpnI and 3' XbaI restriction sites at the ends of *PhoCl* was made by overlap extension PCR. Each terminus of *PhoCl* and split *NanoBiT* fragments were joined by linker sequences reported in a previous study[163]. The gene fragments encoding LgBiT-PhoCl-SmBiT (with 5' XhoI and 3' EcoRI restriction sites) and maltose binding protein (MBP) (with 5' EcoRI and 3' HindIII restriction sites) were digested and inserted between the XhoI and HindIII sites of pBAD/HisB vector by three-way ligation thus yielding the PhoCl screening construct of XhoI-*LgBiT*-KpnI-*PhoCl1*-XbaI-*SmBiT*-EcoRI-*MBP*-HindIII. The other constructs encoding the fusion proteins used in luciferase-based complementary assays (as described in section 2.2.2) were synthesized in a similar method but with different restriction sites of pBAD/HisB vector. These plasmids constructs are: XhoI-*LgBiT*-*TEVs*-*SmBiT*-HindIII; XhoI-*LgBiT*-*PhoCl1*-*SmBiT*-HindIII; XhoI-*SmBiT*-*PhoCl1*-*LgBiT*-HindIII; and XhoI-*SmBiT(86)*-*PhoCl1*-*LgBiT*-HindIII. Primer sequences are provided in **Table 2.4**. The protein sequence of TEV cut site used in this study is: ENLYFQS[164,165].

2. 4. 3 Protein purification

To purify protein for crystallization, the gene expressing PhoCl with a 6 × His tag at the C-terminal in pET-28a vector was used to transform *E. coli* strain BL21 (DE3) pLysS (Promega). A single colony was picked to inoculate a 5 mL starting culture that was allowed to grow overnight at 37 °C with 225 rpm before being diluted into large cultures (1 liter for PhoCl green state and red state; 10 liter for PhoCl empty barrel) of LB media supplemented with 100 µg/mL ampicillin. The cultures were allowed to grow at 37 °C to an OD_{600} of 0.5, and then induced with IPTG (0.1 mM) and grown overnight at 28 °C. The *E. coli* cells were harvested by centrifuge and lysed by sonication. Protein was purified by Ni-NTA chromatography (G-Biosciences). For PhoCl empty barrel purification, the intact His tag fused PhoCl protein was captured by Ni-NTA agarose and was photoconverted on the column with 10 × 15 s illumination under a 405 nm LED array (Loctite; 3.46 mW/mm^2), and the eluent containing the PhoCl empty barrel was collected. The protein crystallization, X-ray data collection and structure refinements were performed by our collaborator: Dr. Yurong Wen. The detailed information will be described in section Appendix.

To prepare the fusion proteins for luciferase-based complementary assays, the *E. coli* strain DH10B were transformed with the pBAD/HisB vector encoding the gene of interest. Cultures were induced with L-arabinose (0.2%) and allowed to grow 24 h at 30 °C. Protein were purified by Ni-NTA chromatography.

2. 4. 4 Luciferase-based complementary assays

The purified fusion protein (constructs described in section 2.2.2) with the concentration of 500 nM were applied in luciferase assays. For fusion protein LgBiT-TEVs-SmBiT, protein samples were treated with TEV protease before bioluminescence measurement. The hydrolysis reactions were performed with ProTEV Plus (Promega) at 30 °C for 4 hours according to the manufacturer's protocol. For the fusion protein with PhoCl insertion, samples were illuminated with 30 s violet light under LED flood array (Loctite) with an intensity of 3.46 mW/mm^2. At 5 min after photoconversion, the luminescence emission spectra of the protein, both with and without hydrolysis or photoconversion, were measured immediately after treatment with the luciferase substrate furimazine (working concentration: 10 μg/mL). The bioluminescence spectrums were acquired with a Synergy HTX plate reader (BioTek).

2. 4. 5 Mutagenesis and library screening

The *PhoCl* libraries were generated by error-prone PCR or site-directed mutagenesis with degenerate codons at the targeted sites. Primer sequences are provided in **Table 2.4**. *E. coli* strain DH10B (Thermo Fisher Scientific) electrocompetent cells were transformed with the gene libraries and grown on lysogeny broth (LB) agar plates supplemented with 100 μg/mL ampicillin and 0.02% L-arabinose. For libraries generated by error-prone PCR, approximately 10,000 colonies were screened in a given round. The colonies exhibiting the brightest green fluorescence (top ~ 2%), as determined using a custom-built colony screener[36] (for green channel: 470/40 nm excitation and 510/20 nm emission), were picked and cultured in 1 mL medium in 96-well deep block plates (Thermo Fisher Scientific). For libraries generated by randomization of one codon, 190 colonies (approximately six-fold the theoretical number of gene variants generated from an NNK codon, where N = A, G, C, T and K = G, T) were picked randomly and screened. For libraries generated by randomization of a combination of two codons, 570 green fluorescent colonies (approximately 50% of the theoretical number of gene variants) were picked and screened.

For each cultured variant, protein was extracted using 50 μL 0.5% of *n*-octyl-β-D-thioglucopyranoside in Tris buffered saline (TBS, pH 8) and then further diluted 1:4 in PBS buffer (pH 7.4). Photoconversion was performed in transparent microplates (Nunc-Thermo Fisher Scientific; Mfr.No. 265302) under LED flood array (3.46 mW/mm^2). Diluted samples (100 μL), with and without photoconversion, were tested on 96-well plates (Thermo Fisher Scientific) at 5 min after photoconversion. To each well was added 10 µL luciferase substrate furimazine in PBS solution (0.1 mg/mL), and the luminescence intensity (endpoint mode) was measured immediately using a Cytation 5 plate reader (BioTek). The top 2% variants with bright bioluminescence and large intensity decrease (> 60% decrease) were picked for sequencing and used as templates for the next evolution. A total of 34 libraries with 46 targeted amino acids were screened for the identification of the PhoCl2c and PhoCl2f variants.

Table 2.4 Primers used in Chapter 2.

Primers	**Sequence 5'-3'**
pET-28a-*PhoCl1*-6His construct	
F-NcoI-PhoCl1	ATATACCATGGTGATCCCTGACTACTTCAAGCAGAGCTTC
R-PhoCl1-linker-XhoI	GTAGCCTCGAGACCTCCACCTCCCCGTGGGTACTTGGTGAACACGC
pBAD/HisB-*PhoCl1* construct	
F-XhoI-PhoCl1	CCGAGCTCGAGTGTGATCCCTGACTAC
R-PhoCl1*-KpnI	CATCCGCCAAAACAGCCAAGCTTGGTACCTTACCGTGGGTACTTGGTGAACAC
pBAD/HisB-*LgBiT-PhoCl1-SmBiT-MBP* construct	
F-XhoI-LgBiT	GCAGGCTCGAGGATGGTCTTCACACTCGAAGATTTCGTTGGG
R-LgBiT-linker1-KpnI-PhoCl1 overlap PCR	GAAGTAGTCAGGGATCACGGTACCAGGTTCTCCGCCGCTAGAACCGCCTCCGCTACTCCC
F-linker1-Kpn1-PhoCl1 overlap PCR	GGCGGAGAACCTGGTACCGTGATCCCTGACTACTTCAAGCAGAGC
R1-PhoCl1-XbaI-linker2-SmBiT	CACGCCGCCACTGCTTCCGCCACCGGACGAACCTCTAGACCGTGGGTACTTGGTGAACAC
R2-linker2-SmBiT-EcoRI	CAGCTGAATTCCAGGATCTCTTCAAAAAGTCTGTATCCAGTCACGCCGCCACTGCTT
F-EcoRI-linker3-MBP	GTACGGAATTCGGGGGTGGAGGTTCAAAAATCGAAGAAGGTAAACTGGTAATCTGGAT
R-MBP*-HindIII	CAGCCAAGCTTTTAAGTCTGCGCGTCTTTCAGGG
pBAD/HisB-*LgBiT-TEVs-SmBiT* construct	
F-XhoI-LgBiT	GCAGGCTCGAGGATGGTCTTCACACTCGAAGATTTCGTTGGG
R1-LgBiT-linker1	GGTTCTCCGCCGCTAGAACCGCCTCCGCTACTCCCGCTGTTGATGGTTACTCGGAACAGC
R2-linker1-TEVs-linker2	GAACCATTATCACTCTGAAAATACAAGTTTTCTGTAGTAGGTTCTCCGCCGCTAGAACCG
R3-TEVs-linker2	AGCCGCCACTGCTTCCGCCACCGGACGAACCATTATCACTCTGAAAATACAAGTTTTCTG

R4-linker2-SmBiT	TTACAGGATCTCTTCAAAAAGTCTGTATCCAGTCACTCCACCCGAGCCGCCACTGCTTCC
R5-SmBiT-HindIII	TTACCAAGCTTTTACAGGATCTCTTCAAAAAGTCTGTATCCAG
pBAD/HisB-*LgBiT-PhoCl1-SmBiT* construct	
F-XhoI-LgBiT	GCAGGCTCGAGGATGGTCTTCACACTCGAAGATTTCGTTGGG
R-LgBiT-linker1-PhoCl1 overlap PCR	GAAGTAGTCAGGGATCACTGTAGTAGGTTCTCCGCCGCTAGAACCGCCTCCGCTACTCCC
F-linker1-PhoCl1 overlap PCR	GGCGGAGAACCTACTACAGTGATCCCTGACTACTTCAAGCAGAGC
R1-PhoCl1-linker2	CACGCCGCCACTGCTTCCGCCACCGGACGAACCATTATCCCGTGGGTACTTGGTGAACAC
R2-linker2-SmBiT-HindIII	CAGCTAAGCTTTTACAGGATCTCTTCAAAAAGTCTGTATCCAGTCACGCCGCCACTGCTT
pBAD/HisB-*SmBiT-PhoCl1-LgBiT* construct	
F1-XhoI-SmBiT-linker1	GGCGACTCGAGGGTGACTGGATACAGACTTTTTGAAGAGATCCTGGGGAGTAGCGGAGGC
F2-SmBiT-linker1	CTGGGGAGTAGCGGAGGCGGTTCTAGCGGCGGAGAACCTACTACA
F3-linker1-PhoCl1	GCGGCGGAGAACCTACTACAGTGATCCCTGACTACTTCAAGCAGAG
R-PhoCl1-linker2 overlap PCR	GCCGCCACTGCTTCCGCCACCGGACGAACCATTATCCCGTGGGTACTTGGTGAACACGC
F-linker2-LgBiT overlap PCR	GGTTCGTCCGGTGGCGGAAGCAGTGGCGGCGTCTTCACACTCGAAGATTTCGTTGGGGAC
R-LgBiT-HindIII	CAGCCAAGCTTTTAGCTGTTGATGGTTACTCGGAACAGC
pBAD/HisB-*SmBiT(86)-PhoCl1-LgBiT* construct	
F1-XhoI-SmBiT(86)-linker1	GGCGACTCGAGGGTGTCGGGTTGGCGGTTGTTCAAGAAAATCTCAGGGAGTAGCGGAGGC
F2-SmBiT(86)-linker1	CTCAGGGAGTAGCGGAGGCGGTTCTAGCGGCGGAGAACCTACTACA
R-LgBiT-HindIII	CAGCCAAGCTTTTAGCTGTTGATGGTTACTCGGAACAGC
Error-prone PCR for PhoCl libraries	
EP-F-KpnI-PhoCl	CGGAGAACCTGGTACCGTGATCCCT
EP-R-XbaI-PhoCl	CGGACGAACCTCTAGACCGTGGGTA
QuickChange primers for PhoCl libraries	
I3-antisense	TCTGCTTGAAGTAGTCAGGMNNCACGGTACCAGGTTCTCCG
D5-sense	GAACCTGGTACCGTGATCCCTNNKTACTTCAAGCAGAGCTTCCCC
F7-antisense	CCTCGGGGAAGCTCTGCTTMNNGTAGTCAGGGATCACGGTA
K8-antisense	GCCCTCGGGGAAGCTCTGMNNGAAGTAGTCAGGGATCAC
F11-antisense	CCAGCTGTAGCCCTCGGGMNNGCTCTGCTTGAAGTAGTC
G14 Y15-antisense	AGGTCATGCTGCGCTCCCAGCTMNNMNNCTCGGGGAAGCTCTGCTTGAAG
I35-antisense	CTGTCCCCCTCCATTGTMNNGTCGTTGGTGGCGATGC
M37 E38-antisense	TGATGAAGCTGTCCCCMNNMNNTGTGATGTCGTTGGTGGCGATGCAGATGCCG
G39 D40-antisense	CTGGAAGTGGATCTTGTTGATGAAGCTMNNMNNCTCCATTGTGATGTCGTTGGTGGCGAT
F42-antisense	CTTGAAGTGGATCTGGTTGATMNNGCTGTCCCCCTCCATTGTGAT
R77 D78-antisense	TCACGTCGCCCTTCAGCACGCCMNNMNNCTCGTACATCTTCTCGGTGCTG

G79 V80-antisense	CTTCATCTTCACGTCGCCCTTCAGMNNMNNGTCGCGCTCGTACATCTTCTCGGT
M87-sense	GCCCTTCAGCAGCAGCTTMNNCTTCACGTCGCCCTTCAG
Y117-antisense	CGGTGGTCCACGAAGTGMNNGTCGGGCAGCTTTACGG
Y117 H118-antisense	TCTCGATGCGGTGGTCCACGAAMNNMNNGTCGGGCAGCTTTACGGGCTTC
V120-antisense	TCTCGATGCGGTGGTCMNNGAAGTGGCAGTCGGGC
H122-antisense	GCTCAGGATCTCGATGCGMNNGTCCACGAAGTGGCAGTC
L138 Y139-antisense	GGAAGTCTTGGCCACGGCGTGCTCMNNMNNCTTCACCTTGTTGTAGTCCTTGTC
V143 A144-antisense	GTCCATGCTGTCGGTGGAAGTCTTMNNMNNGGCGTGCTCGTACAGCTTCACCTT
R145 N146-antisense	CTCGTCCATGCTGTCGGTGGAMNNMNNGGCCACGGCGTGCTCGTACAG
M162-sense	GGTGGCAGCGGTGGCNNKGTGAGCAAGGGCGAG
D177 M178-sense	GAGACCATTACAAGCGTGATCAAGCCTNNKNNKAAGAACAAGCTGCGCATGGAGGGCAAC
G201 K202-antisense	CGTCTGGATGCCCTCGAAGGGMNNMNNGCTGCCCTCGCCCTCGATCAC
P203 F204-antisense	ATCAATCGTCTGGATGCCCTCMNNMNNCTTGCCGCTGCCCTCGCCCTC
F204-antisense	CGTCTGGATGCCCTCMNNGGGCTTGCCGCTGCC
E205 G206-antisense	ACCTCCAAATCAATCGTCTGGATMNNMNNGAAGGGCTTGCCGCTGCCCTCGC
I207-antisense	CTCCAAATCAATCGTCTGMNNGCCCTCGAAGGGCTTGCC
I207 Q208-antisense	TCCTTCACCTCCAAATCAATCGTMNNMNNGCCCTCGAAGGGCTTGCCGCTGC
I210-antisense	CCTCCTTCACCTCCAAATCMNNCGTCTGGATGCCCTCGAAG
T229 A230-antisense	AACACGCGGTTGCCGTAGTGGAAMNNMNNGGTCAGGATGTCGTAGGCGAAGG
T239-antisense	CATTATCCCGTGGGTACTTMNNGAACACGCGGTTGCCGTAG
K240-antisense	ACCTCTAGACCGTGGGTAMNNGGTGAACACGCGGTTGCC

Chapter 3 PhoCl2 protein characterization and application

3. 1 Introduction

The photocleavable protein (PhoCl) reported by our group in 2017 (Ref. 36) represented a new category of photosensory domain for use in optogenetic actuators. Unlike other photosensory domains, PhoCl undergoes light-dependent irreversible dissociation via covalent bond cleavage. PhoCl is engineered from a circularly permutated version of a green-to-red photoconvertible fluorescent protein (FP)[158]. When illuminated with violet light, PhoCl undergoes bond cleavage at the green chromophore. This bond cleavage produces a large "empty barrel" N-terminal fragment, and a small C-terminal "chromophore peptide" fragment, which spontaneously dissociate from each other.

In our previous study, the spontaneous dissociation of the first generation version of PhoCl (PhoCl1) was validated and demonstrated via a variety of protein characterization assays[36]. Spectroscopic analysis of purified PhoCl revealed that photoconversion results in the transient formation of a red fluorescent state that is then quenched by water during the release of the C-terminal peptide fragment. Consequently, dissociation kinetics ($t_{1/2} \sim 500$ s) were determined by the loss of red fluorescence after photoconversion. The dissociation rate and extent of PhoCl1 has also been analyzed by time-lapse electrospray ionization mass spectrometry (ESI-MS). Analysis of the photocleavage of PhoCl1 as a fusion protein by western blotting and gel-filtration chromatography (GFC) provided additional evidence of spontaneous dissociation.

PhoCl1's utility and versatility as a photosensory domain has been demonstrated by the construction of optogenetic actuators that manipulate a number of biological processes in cells. Upon illumination, the trapped or caged biofuntional modules are released on a timescale of minutes, and the activated proteins could perform long-term actions until they are degraded in cells. In our previously reported study, PhoCl1 was utilized for optical control of biofunctional modules through three strategies (**Figure 3.1**), which will be discussed in more details.

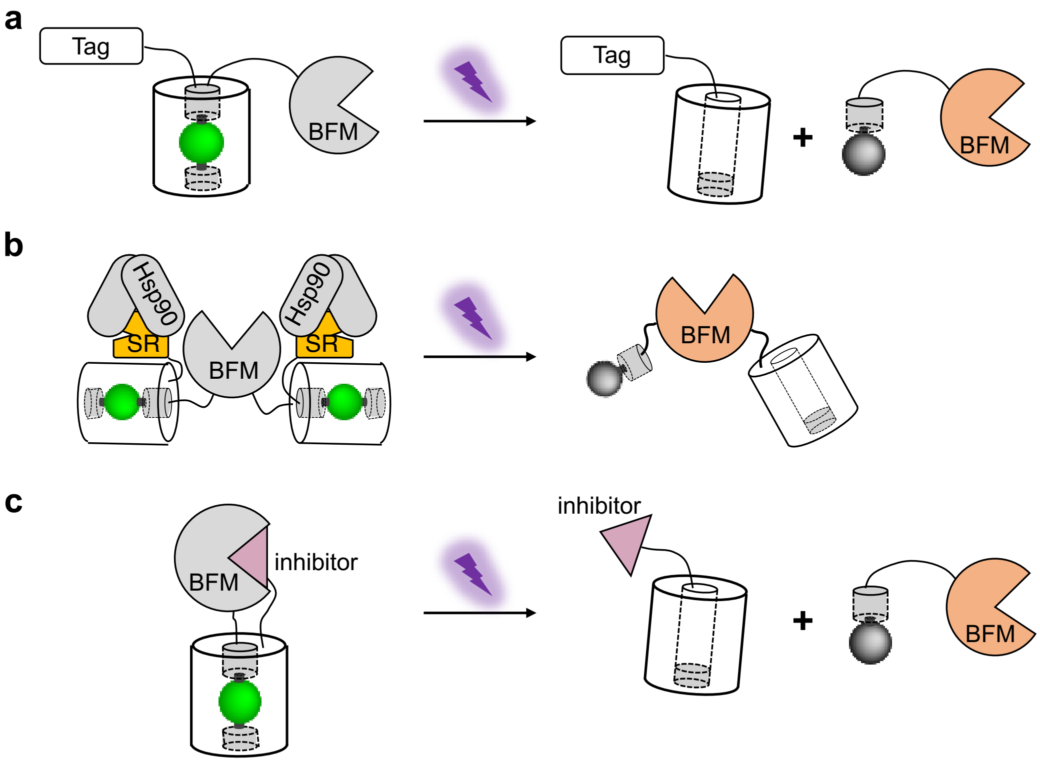


Figure 3.1 Schematic representations of PhoCl-based actuators design strategies. (**a**) Optical control of protein translations by photocleavage. "Tag" is short for localization tag, and "BFM" is short for biofunctional module. (**b**)

Optogenetic activation of the biofunctional module by releasing steric hindrance. (**c**) Optogenetic activation of the biofunctional module by releasing its inhibitor.

One strategy is to modulate the activity of a biofunctional module through manipulation of protein localization. In this approach, PhoCl and the biofunctional module are fused to a localization tag (such as nuclear exclusion sequence (NES)[166], nuclear localization sequence (NLS)[167], or membrane targeting motif (CAAX)[168]), such that the function of the biofunctional domain is caged in the dark state due to the inaccessibility to its target area. After transient illumination, subcellular protein translocation is manipulated due to detachment of the localization tag (**Figure 3.1a**). Following this strategy, optogenetic control of gene expression has been achieved by using PhoCl1 photocleavage to release a Gal4-VP16 transcription factor[169].

Another strategy is to design the PhoCl-mediated actuators based on steric controls. In previous work, we demonstrate blocking of the activity of biofunctional module by fusing to steroid receptor (SR) domains[170] (**Figure 3.1b**). Caging of the protein of interest likely involves interaction between SR domain with heat shock protein 90 (Hsp90), which leads to steric blocking, partial unfolding, and sequestering the biofunctional domain in the cytoplasm. Compared to the translocation strategy, the steric control strategy has enabled the activation of Gal4-VP16 transcription factor with much lower background signals. This PhoCl-SR-fusion also enabled the light-induced gene expression by photo-uncaging of Cre recombinase.

The third actuator design strategy is to utilize PhoCl as a photocleavable linker between the biofunctional domain and its proteinaceous inhibitor (**Figure 3.1c**). In this approach, the protein function would be inhibited in the dark state, until the inhibitor is released by photocleavage. Based on this strategy, our group reported the development of a photoactivatable HCV protease (HCVp) by inserting PhoCl between HCVp and its inhibitory peptide (I)[171]. We expect that this generic design could prove widely useful to engineer other photoactivatable enzymes.

In this Chapter, we describe our efforts to demonstrate the improved properties of the second generation versions of PhoCl (PhoCl2c and PhoCl2f). The kinetics and efficiency of dissociation were compared between PhoCl2 variants and PhoCl1 by using spectroscopy, gel filtration chromatography (GFC) and SDS-PAGE analysis. The dissociation rate and ratio were further demonstrated in cell-based experiments, by using our previous assay for manipulating subcellular protein localization through PhoCl-dependent cleavage of a NES. We also explored the utility of PhoCl2 for optical control of cell apoptosis by engineering of a light-activatable pro-apoptotic protein Bid. We expect that the PhoCl2-meidated optogenetic apoptosis could be useful for manipulation of cell life/death decisions with spatiotemporal precision, which has the therapeutic interest in cancer and neuropathology[172].

3. 2 Results and discussion

3. 2. 1 Characterization of the PhoCl2 variants

Dissociation kinetics were determined by loss of red fluorescence after photoconversion as described previously[36]. Kinetic measurements of the red fluorescence loss were monitored immediately following 15 s illumination with 405 nm light (LED array, 3.46 mW/mm^2) (**Figure 3.2**). Variants with mutations found within or near 201-207 loop (PhoCl1.5f-MV, PhoCl1.5f-L, and PhoCl2f-S) (**Figure 3.2**), were found to exhibit faster dissociation rates ($t_{1/2}$ = 76-140 s) than PhoCl1 ($t_{1/2}$ = 570 s) (**Figure 3.2a,c**).

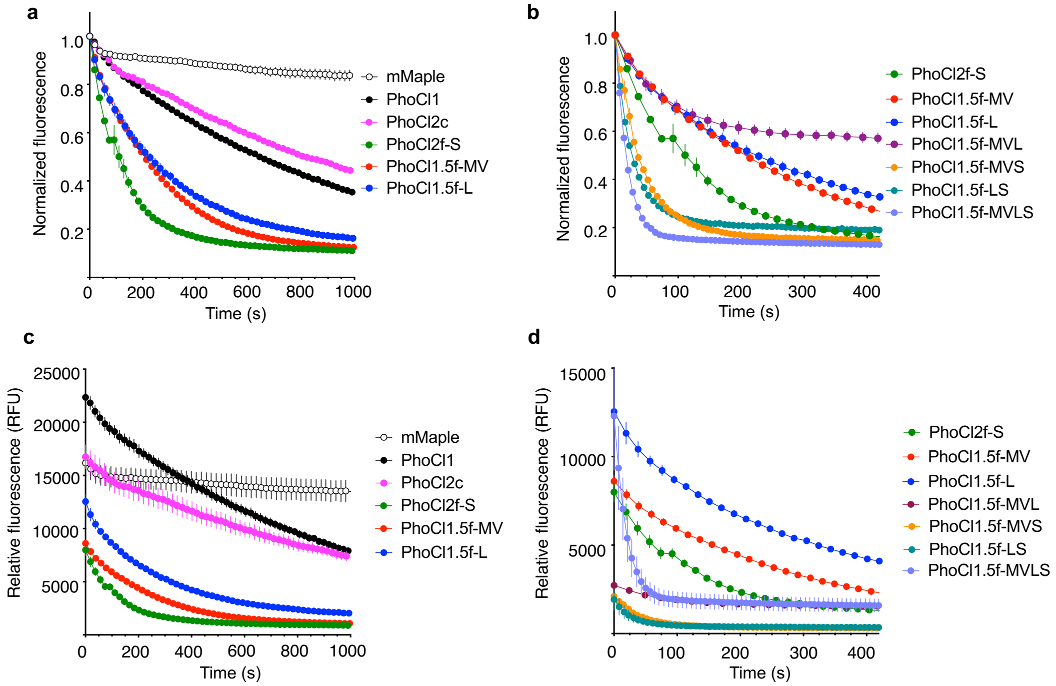


Figure 3.2 Dissociation kinetics of PhoCl variants. (**a**) Dissociation kinetics, as determined by loss of red fluorescence after photoconversion. The red fluorescence was monitored immediately following 15 s illumination with 405 nm light (LED array, 3.46 mW/mm^2). Values are mean ± SEM (n = 3 independent experiments).

One phase decay fit was used to determine the dissociation half time, $t_{1/2}$. *R*-squared values for fits range from 0.9830 to 0.9975. (**b**) Dissociation kinetics of PhoCl2f and PhoCl1.5f variants with the combinations of key mutations. (**c,d**) Loss of red fluorescence after photoconversion without normalization. These are the raw data shown in **a** and **b**. RFU, relative fluorescence units. Each protein was at a concentration of 500 nM, except for PhoCl1.5f-MVLS which was at 5 µM due to its poor chromophore formation.

We next attempted to combine key mutations to explore whether they would have additive beneficial effects on the rate of dissociation (**Figure 3.2b,d**). We found that the combination of Val143Met, Ala144Val and Lys202Leu (PhoCl1.5f-MVL) led to poor dissociation efficiency. The PhoCl1.5f-MVS variant (with Val143Met, Ala144Val and Ile207Ser), the PhoCl1.5f-LS variant (with Lys202Leu and Ile207Ser) and the PhoCl1.5f-MVLS variant (with all four key mutations) exhibited faster dissociation rates ($t_{1/2}$ = 15-30 s). However, these variants also exhibited dim brightness (**Table 3.1**) due to a decreased efficiency of protein folding. Due to its poor expression, we were unable to purify sufficient quantities of the PhoCl1.5f-MVLS variant to enable spectral characterization. Taking into account the factors of both dissociation rate and protein expression level, PhoCl2f-S variant with the mutation Ile207Ser was finally identified as the second generation version of PhoCl with faster dissociation rate ($t_{1/2}$ = 76 s).

Table 3.1 Properties of PhoCl variants.

Protein	Extinction Coefficients		Quantum Yield	Brightness	Dissociation
	ε (488 nm)	ε (405 nm)	488 nm	488 nm	Half Time (s)
mMaple	22	9.3	0.89	20	NA
PhoCl1	34	8.3	0.73	25	570
PhoCl2c	22	9.9	0.48	10	700
PhoCl2f-S	24	13	0.50	12	76
PhoCl1.5f-MV	27	11	0.53	15	130
PhoCl1.5f-L	25	12	0.45	11	140
PhoCl1.5f-MVS	25	13	0.29	7.3	30
PhoCl1.5f-LS	20	15	0.36	7.2	21
PhoCl1.5f-MVLS	NA	NA	NA	NA	14

Note: Extinction coefficient and brightness in $mM^{-1}cm^{-1}$. All measurements were performed with purified proteins in PBS solution (pH 7.4). EGFP extinction coefficient of 56 $mM^{-1}cm^{-1}$ (Ref. 173) and quantum yield of 0.6 (Ref. 174) at 488 nm were used as reference standards.

Spontaneous dissociation efficiency was determined using a gel filtration chromatography (GFC) assay, and the size of the eluted proteins in fractions of GFC eluent were confirmed by SDS-PAGE analysis (**Figure 3.3** and **Figure 3.4**). This analysis demonstrates that PhoCl2c-MBP fusion had higher dissociation efficiency with less remaining intact protein (7.9% intact protein) than the PhoCl1-MBP fusion (28.9% intact protein). Even with the molecular structure of PhoCl1 in hand, it remains unclear to us why PhoCl2c exhibits higher dissociation efficiency. The key mutation of PhoCl2c, Cys99Gly, is located on the opposite end of the barrel from the dissociable peptide.

The variants with faster dissociation (PhoCl1.5f and PhoCl2f variants) also exhibited higher dissociation efficiency than PhoCl1, which could be potentially be a consequence of increased extinction coefficients at 405 nm (**Figure 3.5** and **Table 3.1**) leading to more efficient photoconversion. The variant with the highest dissociation efficiency is PhoCl1.5f-MVS (7.3% intact protein), but the relatively poor chromophore formation of this variant will likely limit its potential application as an optogenetic tool.

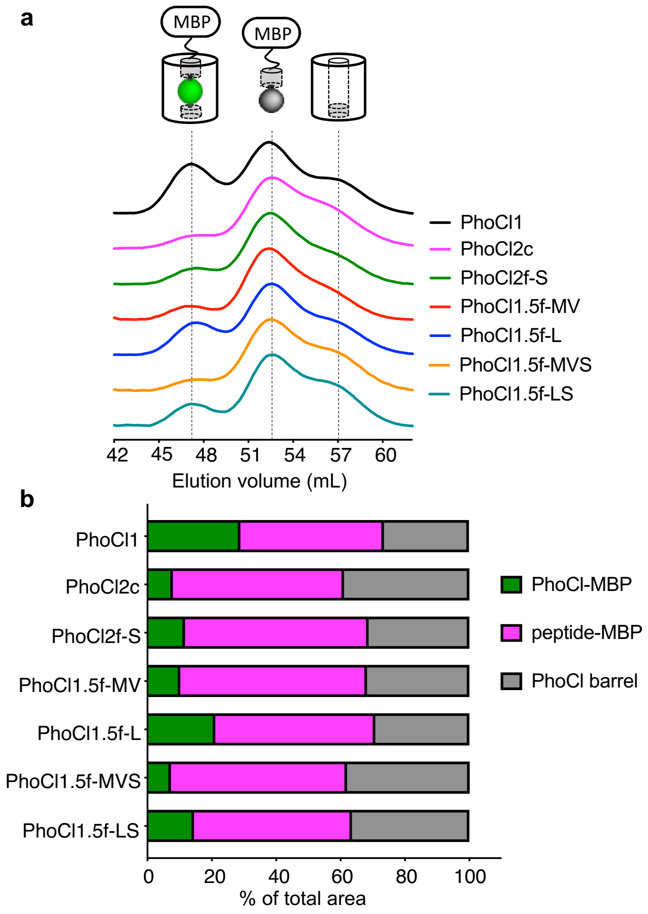


Figure 3.3 Spontaneous dissociation efficiencies of PhoCl-MBP fusion proteins. (**a**) Gel-filtration chromatography (GFC) analysis of maltose binding protein (MBP) fused PhoCl variants. All proteins were tested at the same

concentration (2 mg/mL) and illumination conditions (15 s LED array illumination, 3.46 mW/mm2). Partial photoconverted sample was loaded on the HiPrep 16/60 Sephacry S-100 column immediately after illumination. The intact PhoCl-MBP fusion (75 kDa) was eluted first from the column at a volume of ~ 47 mL. This was followed by the cleaved peptide-MBP fusion (45 kDa) at ~ 52 mL and then the cleaved empty barrel (30 kDa) at a volume of ~ 57 mL. (**b**) Summary of peak area ratios from the GFC analysis. Baseline correction, multi-peak fitting, and peak area integration were performed using the OriginLab software. A Gaussian fit model was used in peak fitting with three fixed widths curves corresponding to the three elution peaks. R-squared values for peak fits range from 0.9974 to 0.9993.

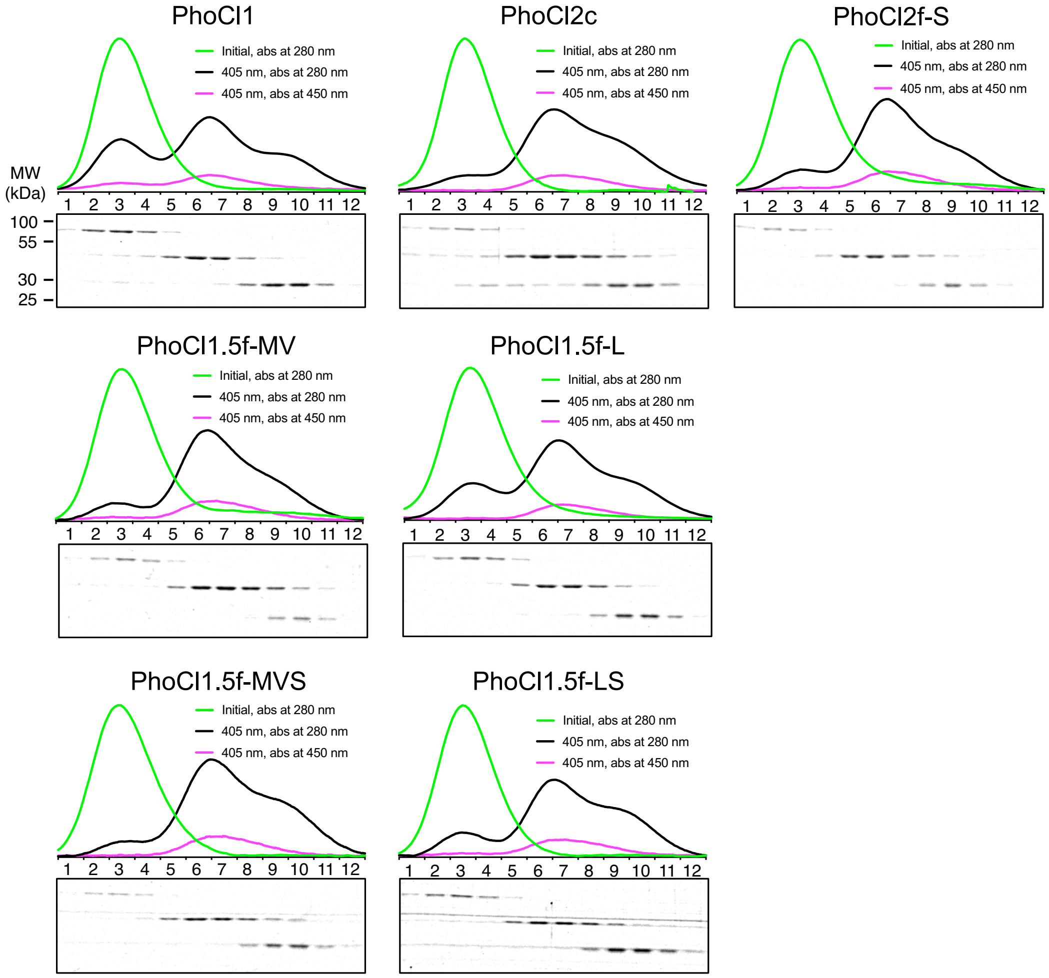


Figure 3.4 GFC and SDS-PAGE analysis of PhoCl-MBP fusions. The absorbance at 280 nm and 450 nm were detected. SDS-PAGE analysis of GFC fractions for partial photoconverted fusion protein is labeled by fraction numbers (12 × 1.5 mL, 43.5 - 61.5 mL elution volume).

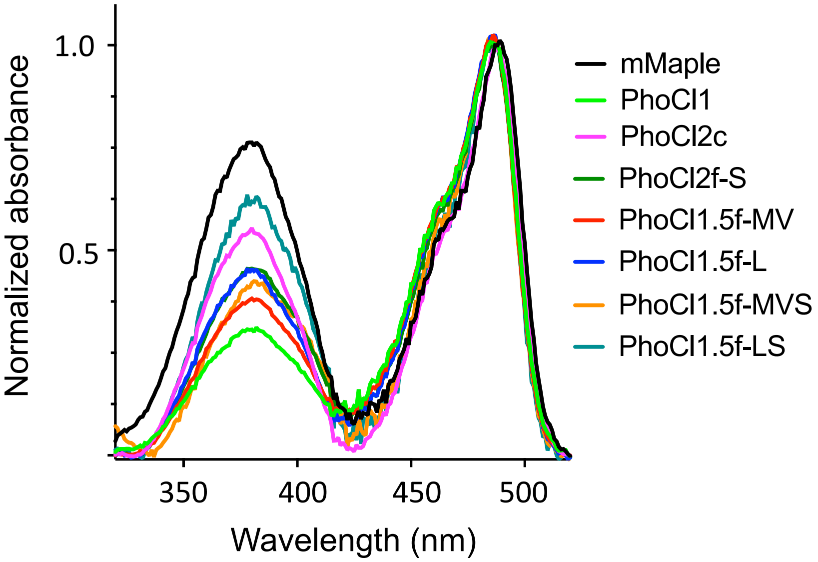


Figure 3.5 Absorbance spectra of mMaple and PhoCl variants. Compared to PhoCl1, increased extinction coefficients at 405 nm were observed with the improved variants.

The dissociation efficiency of PhoCl variants was also confirmed through the bioluminescence complementation assay (**Figure 3.6** and **Table 3.2**). A decrease of bioluminescence intensity was expected upon photocleavage. Photoconvertible protein mMaple inserted into NanoBiT was used as a negative control in this assay. However, even for this negative control we observed a decrease in bioluminescence (-26% at 460 nm). We rationalize this decreased bioluminescence as a result of the change in Bioluminescence Resonance Energy Transfer[175] (BRET) efficiency as the chromophore is converted from green to red.

More specifically, BRET to the green state results in sensitized emission that overlaps with the emission of NanoLuc itself. BRET to the red state results in sensitized emission at a longer wavelength peak that does not overlap with that of NanoLuc. This explanation is consistent with the emerging peak at 583 nm corresponding the mMaple red state (**Figure 3.6b**).

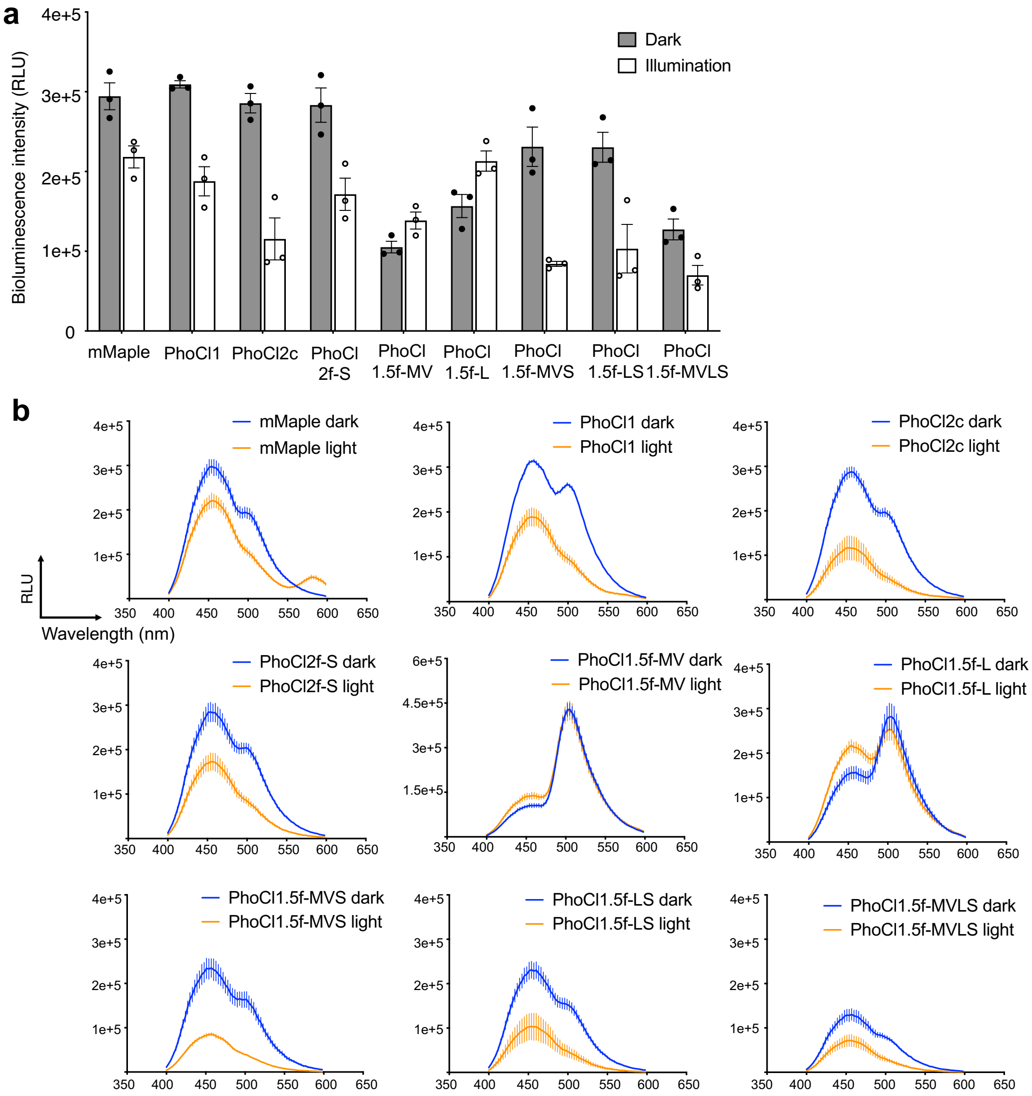


Figure 3.6 Bioluminescence complementation assay of PhoCl inserted NanoBiT. (**a**) Summary data of bioluminescence intensity (at 460 nm) on key

variants. Purified LgBiT-PhoCl-SmBiT-MBP fusion protein (500 nM; construct represented in **Figure 2.5a**) was applied in this assay. At 30 min after illumination (15 s with LED array), the luminescence emission spectra of the protein, both with and without photoconversion, were measured immediately after treatment with the luciferase substrate furimazine (final concentration: 10 µg/mL). Values are means ± SEM (*n* = 3 independent experiments). (**b**) Bioluminescence spectra of PhoCl inserted NanoBiT with (light) or without (dark) illumination. Measurement details were described in **a**.

Most PhoCl2 variants exhibited substantial decreases in bioluminescence intensity after illumination, consistent with our expectations (**Figure 3.6a**). PhoCl2c and PhoCl2f-MVS variants showed the largest decrease (~ 60%), which indicates that these two variants have the best dissociation efficiency, consistent with the GFC analysis. Based on the emission spectra of the dark state, it was apparent that the BRET efficiency varied between PhoCl variants (**Figure 3.6b** and **Table 3.2**). As with Fluorescence Resonance Energy Transfer (FRET)[176], BRET efficiency strongly depends on the interchromophore distance and orientation. With the same luciferase donor and linkers, the BRET ratio is also related to extinction coefficient and quantum yield of PhoCl acceptor. After photoconversion, the BRET signal should decrease or even disappear as a result of the photocleavage of the green chromophore. Compared to PhoCl1, most of the variants showed decreased BRET efficiencies (that is, higher ratios of NanoLuc emission at 460 nm to PhoCl emission at 505 nm), which can be explained by the diminished extinction coefficients and reduced quantum yields. For these variants with low BRET efficiency, bioluminescence intensity at 460 nm decreased upon illumination and photocleavage, as expected.

Table 3.2 Summary data of bioluminescence assay.

Protein	Emission (460 nm)	BRET ratio (dark)	BRET ratio (light)
mMaple	-26%	0.63	0.44
PhoCl1	-39%	0.82	0.47
PhoCl2c	-60%	0.66	0.40
PhoCl2f-S	-39%	0.69	0.45
PhoCl1.5f-MV	+32%	4.08	3.07
PhoCl1.5f-L	+36%	1.80	1.18
PhoCl1.5f-MVS	-64%	0.68	0.43
PhoCl1.5f-LS	-55%	0.63	0.41
PhoCl1.5f-MVLS	-45%	0.58	0.38

Note: The BRET ratio was determined by the use of the PhoCl (505 nm) to NanoBiT (460 nm) ratio (acceptor/donor ratio). All the data were the means from triplicate experiments for each variant.

However, for reasons that remain unclear to us, PhoCl1.5f-MV and PhoCl1.5f-L exhibited substantially higher BRET efficiency than other variants. Moreover, the luminescence of these two variants increased after photoconversion which is contrary to what we observed with the other variants. We speculate that Lys202Leu in PhoCl1.5f-L, and the combination of Val143Met and Ala144Val in PhoCl1.5f-MV, may have introduced stabilizing interactions between NanoLuc and PhoCl that resulted in a more favorable distance or orientation for increased BRET efficiency. The high BRET ratio still existed in partial converted protein fusion after illumination. Hence, the luminescence signal at 460 nm was observed to increase as a result of the reduced BRET efficiency.

Overall, the key mutations found from screening are clustered in close vicinity to where the C-terminal peptide emerges from the intact protein (**Figure 3.7**). It is reasonable to postulate that mutations in this region would affect the efficiency and rate of peptide dissociation. At this point the mechanisms by which these mutations are exerting their observed influence on peptide dissociation is unclear. Two possible mechanisms are: 1) Destabilization of the red state, leading to a decreased energy barrier for peptide dissociation and faster kinetics; and 2) stabilization of the empty barrel which would have no effect on the kinetics, but may lead to a greater overall efficiency of dissociation. The detailed information of how these mutations affect the peptide dissociation will be further elucidated by using computational simulation studies.

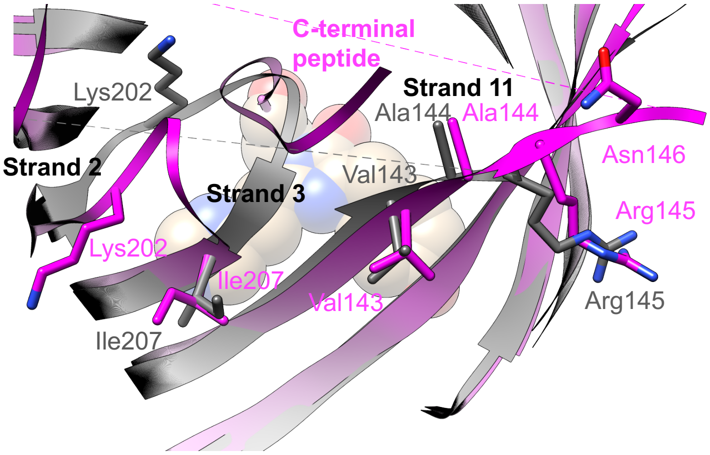


Figure 3.7 Key residues in or near the 201-207 loop that were identified during screening.

3. 2. 2 Optogenetic control of protein localization by PhoCl2 variants in HeLa cells

PhoCl dissociation rate and efficiency were verified in transiently transfected HeLa cells using our previously reported assay for manipulating subcellular protein localization through PhoCl-dependent cleavage of a nuclear exclusion sequence (NES)[36] (**Figure 3.8a**). Specifically, in this assay photocleavage produces redistribution of a red FP (mCherry) between the cytoplasm and nucleus. Relative to PhoCl1-mCherry, PhoCl2c-mCherry fusion exhibited a greater amount of nucleus-localized mCherry following photocleavage to remove the NES (**Figure 3.8b**). Following protein redistribution after photocleavage, the cytoplasm to nucleus intensity ratio for PhoCl2c-mCherry is lower than the ratio for PhoCl1-mCherry (**Figure 3.8c**).

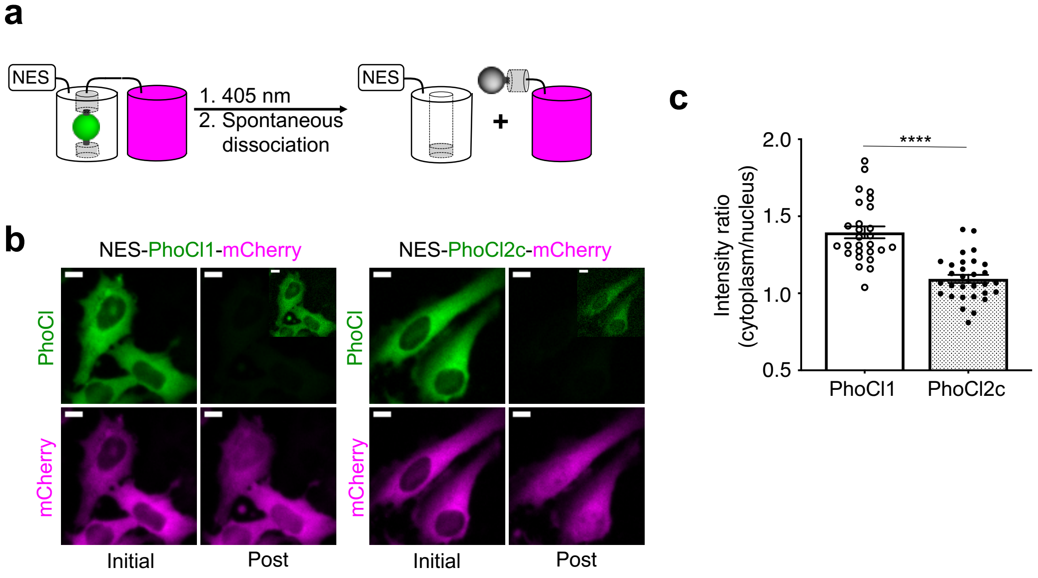


Figure 3.8 Optogenetic manipulation of protein translocation in HeLa cells. (**a**) Schematic of the NES-PhoCl-mCherry photocleavage. (**b**) Representative images of HeLa cells expressing NES-PhoCl-mCherry before and after (15 min) photoconversion. Conversion was performed with 10 s violet light pulses (395/40 nm, 2 mW/mm^2) every 15 s for 6 mins. Inset is the same image with 10× increased contrast. Scale bar, 10 µm. (**c**) Red fluorescence intensity ratios of cytoplasm to nucleus at 15 min after photoconversion. Ratios were calculated for single cells. Values are means ± SEM (n = 27 cells of PhoCl1, and n = 30 cells of PhoCl2c). ****$P < 0.0001$ by unpaired two-tailed t test (t (55) = 6.606).

The kinetics of photoconversion were determined by loss of green fluorescence, and dissociation was characterized by red fluorescence redistribution in HeLa cells expressing NES-PhoCl-mCherry (**Figure 3.9** and **Table 3.3**). PhoCl2f-S, as well as PhoCl1.5f-MV and PhoCl1.5f-L, exhibited faster mCherry redistribution and similar dissociation efficiency (**Figure 3.10**) compared to PhoCl1, which is consistent with our *in vitro* protein characterization results. However, three other 1.5f variants with the combinations of key mutations (PhoCl1.5f-MVS, PhoCl1.5f-LS and PhoCl1.5f-MVLS) were found to not undergo efficient dissociation in cells (refer to **Figure 3.10**), likely due to their poor efficiency of chromophore formation.

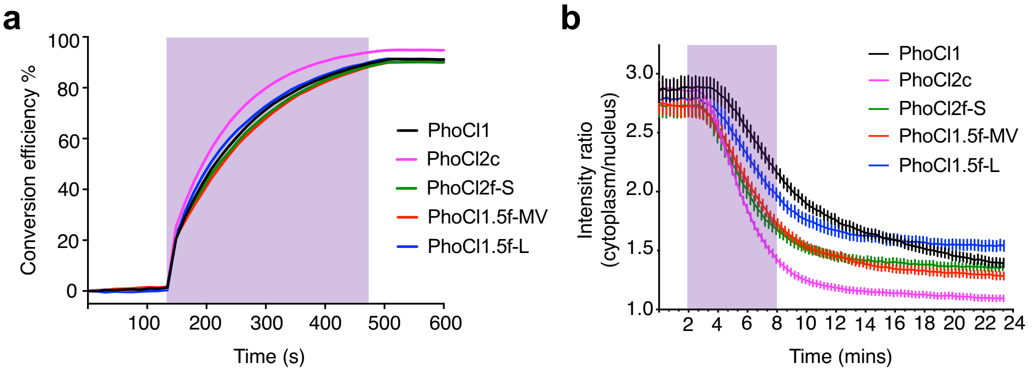


Figure 3.9 Dissociation kinetics and efficiencies of NES-PhoCl-mCherry fusion proteins in HeLa cells. (**a**) Photoconversion efficiencies of PhoCl variants, measured by loss of green fluorescence. (**b**) Red fluorescence intensity localization ratios of cytoplasm to nucleus versus time. Values are means ± SEM (*n* = 27 cells of PhoCl1 and PhoCl1.5f-L, and *n* = 30 cells of the other variants).

Table 3.3 Summary data of optogenetic manipulation of protein translocation assay with PhoCl variants.

Protein	**Photoconversion Half Time (s)**	**Dissociation Half Time (s)**
PhoCl1	75.5	241
PhoCl2c	62.7	114
PhoCl2f-S	78.1	135
PhoCl1.5f-MV	83.6	160
PhoCl1.5f-L	69.6	155

Note: Plateau followed by one phase decay fit was used in both analyses. For fit of photoconversion, *R*-squared values range from 0.9900 to 0.9951. For fit of dissociation, *R*-squared values range from 0.7006 to 0.9040.

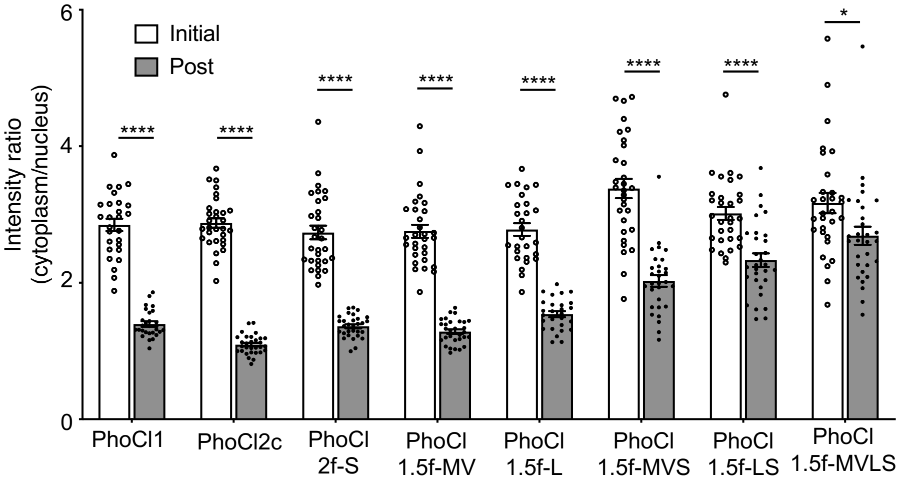


Figure 3.10 Red fluorescence intensity localization ratios (cytoplasm to nucleus) of NES-PhoCl-mCherry before and after photoconversion. Values are means ± SEM. The data is from the same experiments shown in **Figure 3.8a**. PhoCl1: ****$P < 0.000001$, t (52) = 14.85, n = 27 cells from 3 cultures; PhoCl2c: ****$P < 0.000001$, t (58) = 24.41, n = 30 cells from 3 cultures; PhoCl2f-S: ****P < 0.000001, t (58) = 13.09, n = 30 cells from 3 cultures; PhoCl1.5f-MV: ****P < 0.000001, t (58) = 14.47, n = 30 cells from 3 cultures; PhoCl1.5f-L: ****P < 0.000001, t (52) = 12.03, n = 27 cells from 3 cultures; PhoCl1.5f-MVS: ****P < 0.000001, t (58) = 8.187, n = 30 cells from 3 cultures; PhoCl1.5f-LS: ****P = 0.000006, t (58) = 4.997, n = 30 cells from 3 cultures; PhoCl1.5f-MVLS: *P = 0.020512, t (58) = 2.382, n = 30 cells from 3 cultures. Multiple t tests were used to analyze significant difference between group means.

It is interesting that in our cell experiments PhoCl2c-mCherry fusion also exhibited faster redistribution than PhoCl1-mCherry fusion in cell-based assays (**Figure 3.9b** and **Table 3.3**), which is at odds with our *in vitro* kinetics analysis (**Figure 3.2a** and **Table 3.1**). We suggest that this result can be explained by the higher dissociation efficiency of PhoCl2c. This result indicates that both the efficiency of photoconversion, and the kinetics of dissociation, are contributing to the effective rate of release in cells.

3. 2. 3 Optogenetic manipulation of cell apoptosis

We next attempted to achieve optogenetic manipulation of cell apoptosis through the use of photocleavage to activate BH3 interacting-domain death agonist (Bid). Bid is a pro-apoptotic protein belonged to Bcl-2 family. Following cleavage by caspase-8 during apoptosis, Bid translocates to the mitochondria where it induces an increase in outer mitochondria membrane permeability and cytochrome c release[177,178]. To achieve the optogenetic control of apoptosis, we engineered a photoactivatable Bid by inserting PhoCl into the middle of Bid at the position of the caspase-8 cut site (LQTDG). Photoinduced dissociation of C-terminal domain (CBid) from its autoinhibitory N-terminal domain (NBid) would induce the downstream apoptosis pathway (**Figure 3.11a**). The photo-induced apoptosis can be inhibited by the cell permeable pan caspase inhibitor Z-VAD-FMK[179]. Caspase-3 activity at the downstream is monitored via the NES-DEVD-mCardinal-NLS reporter. The cleavage of DEVD linker of the fusion protein by caspase-3 activation would result in the translocation of red FP mCardinal[180] into the nucleus (**Figure 3.11b**).

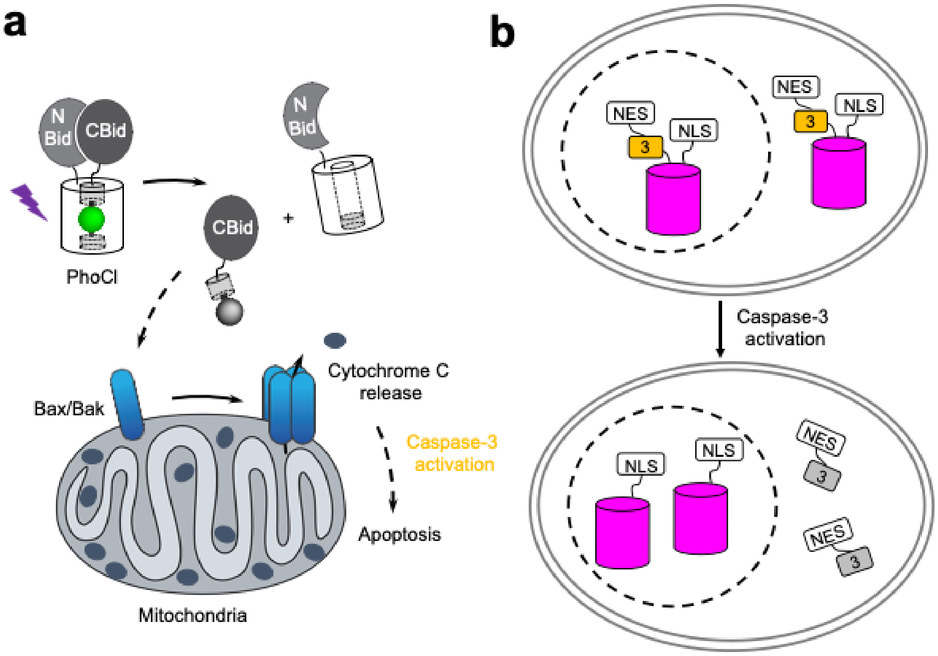


Figure 3.11 Schematics of optogenetic manipulation of cell apoptosis via PhoCl. (**a**) Optogenetic activation of apoptosis with a PhoCl-Bid construct. (**b**) Schematics of apoptosis reporter: NES-DEVD-mCardinal-NLS used in this experiment.

The mMaple insertion construct was used as negative control as it is not expected to dissociate following illumination. Caspase-3 activation was observed at around 30 min after conversion for the NBid-PhoCl2c-CBid and around 50 min after conversion for NBid-PhoCl1-CBid (**Figure 3.12** and **Figure 3.13**). Cell shrinkage and rounding followed by FP translocation were observed at the end of 2 hours of imaging photoconverted cells.

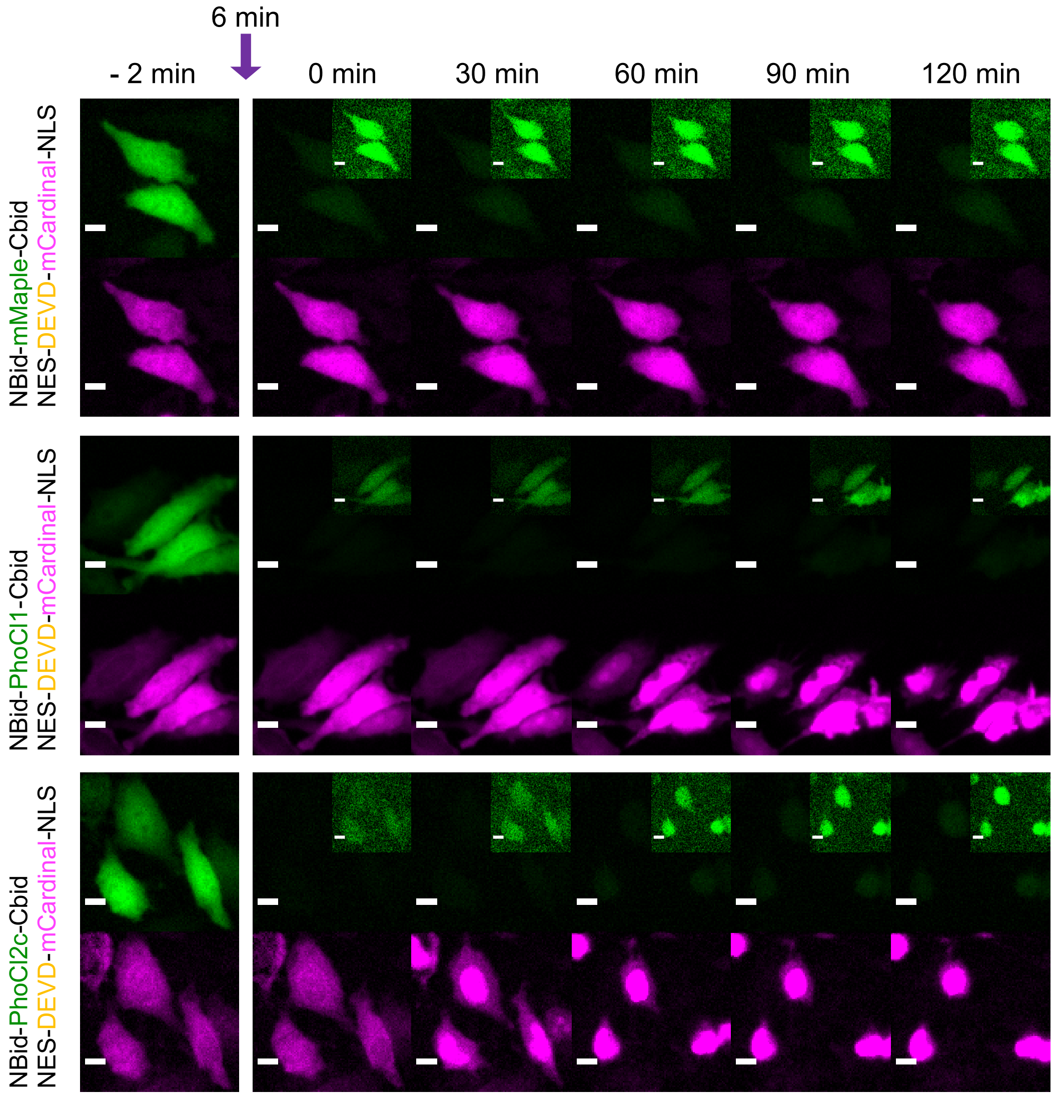


Figure 3.12 Transient transfected HeLa cells co-expressing NBid-mMaple-CBid or NBid-PhoCl-CBid with caspase-3 reporter. Cells were illuminated with 10 s violet light pulses (395/40 nm, 2 mW/mm^2) every 15 s for 6 mins, then imaged 2 hours after photoconversion. Inset is the same image with 10× increased contrast. Scale bar, 10 µm.

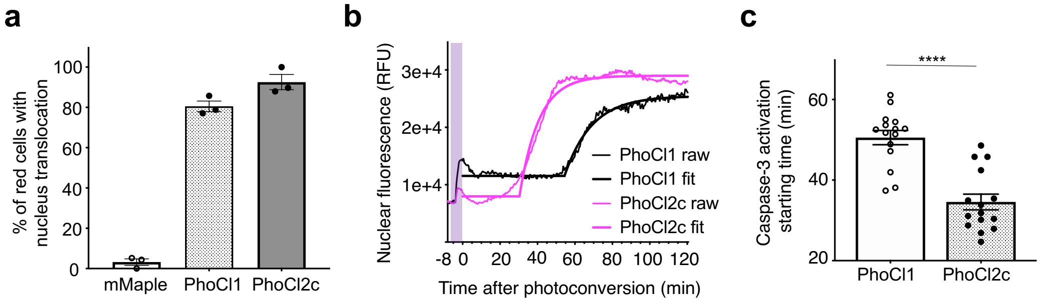


Figure 3.13 Demonstration of PhoCl-dependent induction of apoptosis using the caspase-3 translocation reporter. (**a**) Summary data of caspase-3 reporter translocation for cells expressing constructs described in **Figure 3.11**. Values are means ± SEM (n = 3 cell cultures of each variant. A total of 87 cells for mMaple, 89 cells for PhoCl1 and 113 cells for PhoCl2c). (**b**) Representative nuclear fluorescence intensity profiles of the HeLa cell represented in **Figure 3.12** co-expressing PhoCl inserted Bid and caspase-3 reporter. RFU, relative fluorescence units. Plateau followed by one phase decay fit was applied to each nuclear fluorescence profile. R-squared values range from 0.8238 to 0.9940. X0 is the time at which the decay begins in the fit. The caspase-3 activation time was characterized by X0 in the fit. The activation time X0 in each nuclear fluorescence profile was plotted in **c**. (**c**) Summary of caspase-3 activation starting time for cells in **Figure 3.12**. Values are mean ± SEM (n = 15 cells from 3 cultures for each variant). $****P < 0.0001$ by unpaired two-tailed t test ($t\,(28) = 6.043$).

To analyze the efficiency of photo-induced cell death, cells expressing mMaple inserted Bid or PhoCl inserted Bid with or without illumination were performed with DEAD cell viability assay (**Figure 3.14**). The dye in DEAD stain is ethidium homodimer-1 which is used to indicate loss of cell plasma membrane integrity. For cells expressing mMaple inserted Bid, there is no significant difference in cell viability with and without illumination. Illumination of cells expressing PhoCl1 inserted Bid resulted in significant increase of the cell death ratio, and a further enhanced death rate was observed with PhoCl2c due to its higher dissociation efficiency. Illumination of cells expressing the PhoCl-Bid construct without treatment of caspase inhibitor Z-VAD-FMK resulted in obvious cell shrinkage and rounding. The cell death followed by photoactivation can be inhibited effectively by treatment of Z-VAD-FMK pan caspase inhibitor.

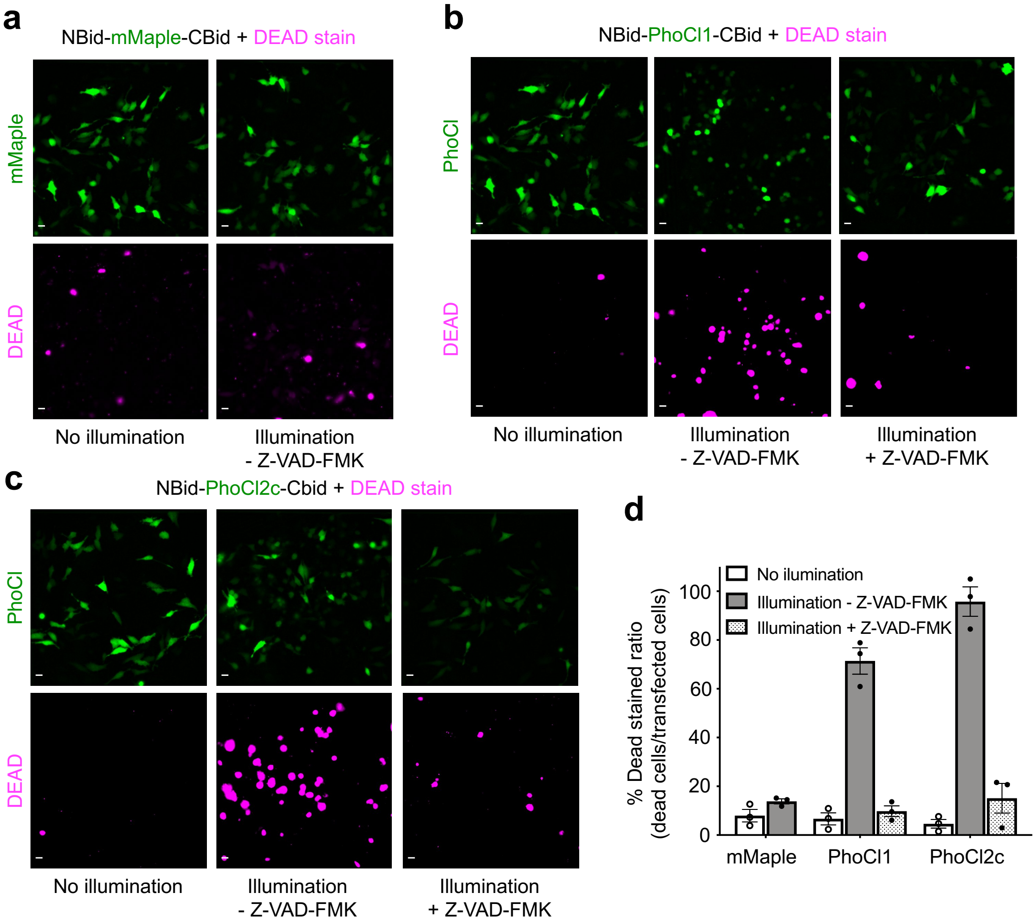


Figure 3.14 Additional information on demonstration of PhoCl-dependent induction of apoptosis using DEAD cell viability assays. (**a-c**) Representative cell images of HeLa cells expressing mMaple or PhoCl inserted Bid after DEAD stain. Scale bar, 20 µm. Photoconversion was performed with 30 s pulse light (LED array, 3.46 mW/mm^2). (**d**) Summary data of cell viability assay. Ratios are

calculated by determining the ratio of DEAD stained cells to green transfected cells. Values are means ± SEM (n = 3 cell cultures of each variant). A total of 321 cells for mMaple no illumination, 301 cells for mMaple illumination - Z-VAD-FMK, 316 cells for PhoCl1 no illumination, 328 cells for PhoCl1 illumination - Z-VAD-FMK, 326 cells for PhoCl1 illumination + Z-VAD-FMK, 325 cells for PhoCl2c no illumination, 337 cells for PhoCl2c illumination - Z-VAD-FMK and 318 cells for PhoCl2c illumination + Z-VAD-FMK.

3. 3 Conclusion

In this Chapter, we have described our efforts to characterize the performance of PhoCl2 variants as purified protein and expressed in cultured mammalian cells. The results demonstrate that the PhoCl2 variants exhibit faster and more efficient dissociation which enables more effective manipulation of protein localization and protein-protein interactions in living cells. In this work, we also explored the application of PhoCl for optical control of cell apoptosis which has potential therapeutic interest in the field of oncology. Although synthetic photocleavable linkers have been widely utilized in chemical biology studies, PhoCl is the first reported fully genetically encodable optogenetic system that shares this feature. Compared to the other reversible optogenetic technologies (such as LOV2 domain[53,77], CRY2-CIB1 [Ref. 89,181] and Dronpa145N[154,156]), PhoCl has no concentration dependence due to its irreversible photochemistry mechanism. After photoconversion by transient illumination, the caged biofunctional modules will be activated irreversibly and persist until they are degraded in cells. According to this feature, PhoCl is particular useful for optogenetic applications that require transient illumination for photoactivation and long-term action. The development of improved PhoCl variants expands the repertoire of photosensory domains that undergoes photocleavage with different kinetics and efficiencies. We anticipate that PhoCl2 variants will prove useful to in a wider variety of applications that requires greater dynamic range and faster photoactivation. For future work, we propose that we could further evolve a third generation variant with faster dissociation and higher chromophore formation efficiency via directed evolution by using the PhoCl1.5f-MVLS variant as the starting template.

3. 4 Methods and materials

3. 4. 1 General molecular biology methods and materials

The general molecular biology methods used in this chapter were same to those described in section 2.4.1. Primer sequences are provided in **Table 3.4**.

3. 4. 2 Protein purification and in vitro characterization

To prepare the proteins for *in vitro* characterizations of PhoCl variants, the *E. coli* strain DH10B were transformed with the pBAD/HisB vector encoding the gene of interest. Cultures were induced with L-arabinose (0.2%) and allowed to grow 24 h at 30 °C. Protein were purified by Ni-NTA chromatography. Molar extinction coefficients (ε) at 488 nm were determined by the alkali denaturation method[182] and then used as reference ε for the 405 nm. Quantum yields for PhoCl variants were measured using purified EGFP as the reference standard[174]. Photoconversion was performed with the 405 nm LED flood array. The absorbance spectra were acquired with a DU-800 UV-vis spectrophotometer (Beckman). The fluorescence spectra were acquired with a Safire2 plate reader (Tecan) and bioluminescence spectrums were acquired with a Cytation 5 plate reader (BioTek). The gel filtration was performed with a HiPrep 16/60 Sephacryl S-100 column on an AKTA chromatography system (GE Healthcare), and the size of the proteins in each eluted fraction were verified by SDS-PAGE.

3. 4. 3 PhoCl expression vectors and cell transfection

The previously reported pcDNA-NES-PhoCl-mCherry (Addgene #87690) plasmid was used as the backbone vector for the protein translocation experiments. A HindIII restriction site was introduced at the 3’ end of the NES sequence in the vector with the QuikChange Lightening kit. The DNA fragments encoding the PhoCl variants (with 5’ HindIII and 3’ KpnI restriction sites) were amplified via PCR and inserted into the appropriately digested NES-PhoCl-mCherry vector.

For the construction of PhoCl inserted Bid, human *Bid* template was ordered as a synthetic gBlock from IDT. The DNA sequence was backtranslated from protein sequence BH3-interacting domain death agonist isoform 2 (GenBank: NP_001187.1) with codon optimization for *Homo sapiens* expression. The gene encoding NBid-PhoCl-CBid, 5’ BamHI and 3’ KpnI restriction sites at the ends of *PhoCl*, was made using overlap extension PCR to connect three gene fragments (i.e., NBid, PhoCl, and CBid). The assembled gene (with 5’ NheI and 3’ XhoI restriction sites) was digested and inserted between the NheI and XhoI sites of pcDNA 3.1(+), thus yielding the PhoCl inserted Bid expression vector.

To construct the caspase-3 reporter, DNA encoding the NES of mitogen-activated protein kinase kinase (MAPKK)[166] and caspase-3 substrate (DEVD) was appended by PCR to the 5' end of the gene encoding mCardinal, and DNA encoding the 3 × NLS of SV40 (Ref. 183) was appended to the 3' end of fragment. The gene encoding NES-DEVD-mCardinal-NLS was then inserted between NheI and XhoI sites of pcDNA 3.1(+), thus yielding the caspase-3 reporter expression vector.

For expression of NES-PhoCl-mCherry, HeLa cells at 70% confluency in 35-mm cell culture dishes (Corning) were transfected with 2 μg plasmid DNA and 4 μL Turbofect (Thermo Fisher Scientific) according to the manufacturer's protocol. Imaging was performed at 24 h after transfection. For expression of PhoCl inserted Bid and caspase-3 reporter, HeLa cells were co-transfected with 1.75 μg NES-DEVD-mCardinal-NLS expression vectors and 0.25 μg PhoCl inserted Bid expression vectors mixed with 4 μL Turbofect. For expression of PhoCl inserted Bid for the DEAD viability assay, 1.75 μg pcDNA 3.1(+) empty vector and 0.25 μg PhoCl-Bid were used per transfection. The transfection was performed in 2 mL of serum-free DMEM for 2 h, after which the medium containing the transfection reagent and DNA mixture was replaced with 2 mL of complete medium supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich). Transfected cells were cultured 16 h before experiments.

3. 4. 4 Cell photoconversion and imaging conditions

Before imaging, the medium was changed to HEPES-buffered Hanks Balanced Salt Solution (HHBSS; 20 mM HEPES). Wide-field fluorescence imaging was performed using a Nikon Eclipse Ti-E epifluorescence microscope equipped with a 75-W Nikon xenon lamp and a Photometrics QuantEM 512SC camera. The NIS-Elements AR package software was used for automatic instrument control and image acquisition. Cells were imaged with a 20× air objective lens (NA 0.75; Nikon) using the following filter sets: PhoCl green (490/15 nm excitation: ex; 510 nm dclp dichroic and 525/50 nm emission: em), mCherry and DEAD stain (543/10 nm ex, 565 nm dclp dichroic and 620/60 nm em), photoconversion (395/40 nm ex and 425 nm dclp dichroic), and mCardinal (605/50 nm ex, 635 nm dclp dichroic and 670/50 nm em).

For imaging of NES-PhoCl-mCherry protein translocation, the images were acquired in both the PhoCl green and mCherry channel every 15 s for 2 min before photoconversion. Photoconversion was performed with 10 s pulse illumination (2 mW/mm^2) every 15 s for 6 min and each pulse of photoconversion light was followed by acquisition of PhoCl and mCherry fluorescence images. Imaging acquisition of green and red channels continued every 15 s for another 15 min after the photoconversion. Protein translocation was analyzed as the ratio of intensity of mCherry in the cytoplasm to the nucleus. Photoconversion efficiency was determined by the loss of green fluorescence intensity in the cytoplasm.

For imaging the caspase-3 activity triggered by photo-induced cell death with PhoCl inserted Bid, images were acquired in both the PhoCl green and mCardinal channel every 15 s for 2 min before photoconversion. Photoconversion was performed with 10 s pulse illumination (2 mW/mm^2) every 15 s for 6 min and each pulse conversion light was followed by acquisition of PhoCl and mCardinal fluorescence images. Imaging acquisition of green and red channels continued every 30 s for 2 h after the photoconversion. Caspase-3 activation was determined by the translocation of mCardinal from cytoplasm into nucleus.

For imaging of the DEAD stained HeLa cells with PhoCl inserted Bid expression, photoconversion was performed under the 405 nm LED flood array with 15 s illumination. Following photoconversion, the cells were incubated in HHBSS buffer (with or without Z-VAD-FMK) at room temperature for 30 min. The working concentration of Z-VAD-FMK (Sigma-Aldrich) was 20 μM. For analysis the cell death, the DEAD Cytotoxicity kit for mammalian cells (Invitrogen) was used to stain the dead cells with or without photoconversion in HHBSS buffer (with or without Z-VAD-FMK) for 30 min at room temperature according to the manufacturer's protocol. Images were acquired in white light, green fluorescence and red fluorescence channels. The extent of cell death was determined by the ratio of DEAD reagent stained cells to the green fluorescent transfected cells.

3. 4. 5 Statistical analysis

All data are expressed as mean ± SEM. Sample sizes (*n*), significant differences (*P*) and confidence of curve fitting (*R*-squared) are listed for each experiment. Results are reported as $P = P$ value, and t (degrees of freedom (df)) = t value. For all statistics, NS = $P \geq 0.05$, * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$ and **** = $P < 0.0001$. No samples were excluded from analysis and all experiments were reproducible. No randomization or blinding was used.

Table 3.4 Primers used in Chapter 3.

Primers	**Sequence 5’-3’**
QuickChange primers for the combinations of point mutations	
K202L-antisense	GATGCCCTCGAAGGGAAGCCCGCTGCCCTCGCCC
I207S-antisense	CTCCAAATCAATCGTCTGAGAGCCCTCGAAGGGCTTGCC
K202L I207S-antisense	CCTCCAAATCAATCGTCTGAGAGCCCTCGAAGGGAAGCCC
pBAD/HisB-*PhoCl-MBP* constructs	
F-XhoI-PhoCl	CCGAGCTCGAGTGTGATCCCTGACTAC
R-PhoCl-linker-KpnI	TTCATGGTACCTCCACCTCCCCGTGGGTACTTGGTGAACAC
pcDNA-*NES-PhoCl-mCherry* constructs	
quickchange-NES-HindIII-antisense	AGTAGTCAGGGATCACGCTAAGCTTGCCTCCCTGCTGCTCGTCC
F-HindIII-PhoCl	CTTAGAAGCTTAGCGTGATCCCTGACTACTTCAAG
R-PhoCl-KpnI	CCTCCGGTACCCCGTGGGTACTTGGTGAACACG
pcDNA-*NBid-PhoCl-CBid* constructs	
F-NheI-NBid	AGTCAGCTAGCGCCACCATGGATTGTGAGGTCAATAACGG
R-BamHI-NBid overlap PCR	GGATCCCTCATCGTAGCCTTCCCAC
F-NBid-BamHI-PhoCl overlap PCR	GTGGGAAGGCTACGATGAGGGATCCGTGATCCCTGACTACTTCAAGC
R-PhoCl-KpnI-CBid overlap PCR	GATCGGTTTCCATCGGTTTGAAGGGTACCCCGTGGGTACTTGGTGAAC
F-KpnI-CBid overlap PCR	GGTACCCTTCAAACCGATGGAAACCGATC
R-CBid*-XhoI	TGACTCTCGAGTTACTTGTTCAACCTGAGCAACCACG
F-BamHI-PhoCl	GTCAAGGATCCGTGATCCCTGACTACTTCAAGCAGAG
R-PhoCl-KpnI	CCTCCGGTACCCCGTGGGTACTTGGTGAACACG
pcDNA-*NBid-mMaple-CBid* construct	
F-BamHI-mMaple	AGTAGGAATCCGTGAGCAAGGGCGAGGAGACCA
R-mMaple-KpnI	GCACTGGTACCCTTGTACAGCTCGTCCATGCTG
pcDNA-*NES-DEVD-mCardinal-NLS* construct	
F1-NheI-NES	CACTGGCTAGCGCCACCATGAACCTGGTGGACCTGCAGAAGAAGCTGGAGG
F2-NES	GACCTGCAGAAGAAGCTGGAGGAGCTGGAGCTGGACGAGCAGCAGGGATCCGCCTCCGGC
F3-DEVD-mCardinal	CAGGGATCCGCCTCCGGCGATGAGGTGGATGGAGCCGTGAGCAAGGGCGAGGAG
R1-mCardinal-NLS	CGTTTTTTTTTTGGTCCGGAGCCCTTGTACAGCTCGTCCATGCC
R2-NLS-NLS	GGGGTCTACTTTGCGCTTCTTTTTTGGGTCAACTTTTCGTTTTTTTTTTGGTCCGGAGCC
R3-NLS*-XhoI	ACTGACTCGAGTTAGGTACCTACCTTGCGTTTTTTCTTGGGGTCTACTTTGCGCTTC

Chapter 4 Engineering of a light-inducible nanocapsule system

4. 1 Introduction

Cells utilize subcellular compartmentalization to regulate many different metabolic activities and establish physical boundaries for biological processes[184]. In eukaryotes, cells make use of membrane-bound organelles (such as mitochondria, lysosome and peroxisome) to create ion gradients and isolate specific biochemical reactions[185,186]. However, most prokaryote cells rely on the alternative strategy of using protein-based bacterial microcompartments (BMCs) (such as carboxysome and metabolosome)[187–189] (**Figure 4.1**).

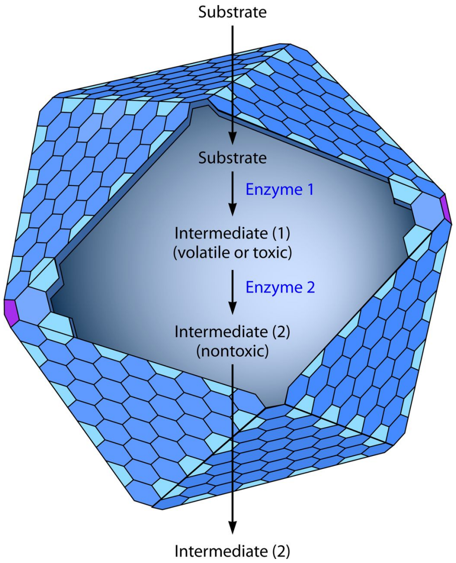


Figure 4.1 Schematics of the bacterial microcompartments. The protein shell of the microcompartment helps to sequester the toxic intermediates inside. (*Note*: **Figure 4.1** reprinted from Ref. 189 and used with permission of the publisher.)

These proteinaceous organelles typically share homologous shell proteins but have different cargo enzymes inside. The protein components of the compartment shells and their complement of enzymes are usually encoded within the same genetic loci or operon[188]. The semi-permeability of the protein shell allows small molecules, such as enzyme substrates, access to the cargo enzyme inside, and prevents the release of toxic intermediates and interference from other large functional molecules[190–192].

Among the many types of BMCs, encapsulins are virus capsid-like bacterial nanocompartments with diameters of 25 nm (Ref. 193) (**Figure 4.2a**) to 32 nm (Ref. 194). Encapsulins are typically used by bacteria to isolate toxic enzymatic activities[195–197]. The hollow icosahedral capsules are formed by 60 (in $T = 1$ capsules; T is triangulation number used to describe icosahedral symmetry[198]) or 180 (in $T = 3$ capsules) copies[193,194] of single shell protein which can be auto-assembled *in vivo* or *in vitro*[199].

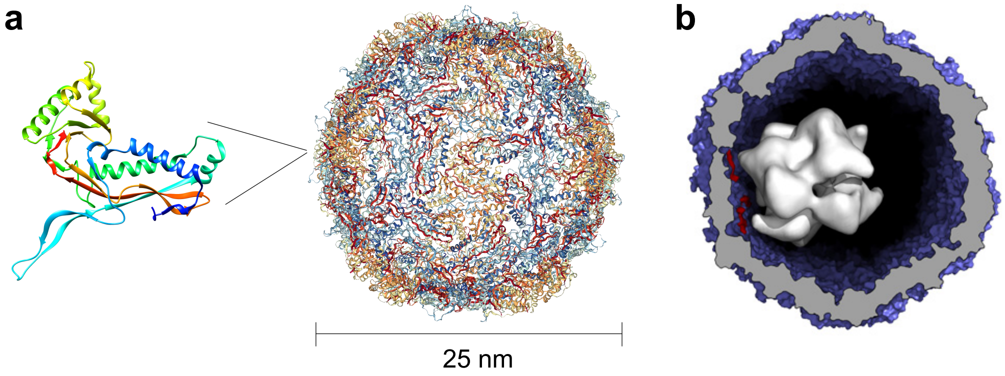


Figure 4.2 Structure of *Thermotoga maritima* encapsulin. (**a**) The 25 nm icosahedral capsule is formed from 60 identical subunits (PDB ID: 3DKT). (**b**)

Model of the cargo protein fitted into the capsule. (*Note*: **Figure 4.2b** is reprinted from Ref. 193 and used with permission of the publisher.)

As one of several bio-nanocompartments, encapsulin has several unique features. The N-terminus and C-terminus extend to the inner and outer faces of the capsule, respectively. In addition, cargo proteins can be packaged inside the capsule by genetic fusion of the cargo protein to a peptide that binds to the inner surface of the encapsulin shell protein[200,201] (**Figure 4.2b**). As the largest opening into encapsulin is 5 Å in dameter[193], there is limited ability of small substrates to enter or exit the capsule.

In recent years, encapsulin has been used in a wide variety of chemical biology applications that are unrelated to its natural roles. For example, in 2014 Moon *et al.* reported a versatile targeted drug delivery nanoplatform by displaying hepatocellular carcinoma (HCC) cell binding peptide (SP94-peptide) on the surface of engineered encapsulin[202]. Anticancer drugs and fluorescent probes were chemically attached to the nanocapsules and delivered to the target HCC cells.

Bacterial encapsulins have also been expressed as orthogonal compartments in eukaryotes including yeast and mammalian cells. In 2018, Lau *et al.* demonstrated that *Myxococcus xanthus* encapsulin can self-assemble to form capsules in yeast and the heterologous cargo proteins can be selectively packed inside[203]. Their study revealed that encapsulin compartments have the capacity to protect the cargo protein from degradation *in vivo*, co-localize split proteins inside, and allow for catalytic process to occur in the capsule interior[203]. Similarly, Sigmund *et al.*[204] utilized the *Myxococcus xanthus* encapsulin to engineer compartmentalized enzymatic reactions and size-constrained iron biomineralization in HEK293T cells. The iron-loaded capsules were demonstrated to have electron-rich cores and be excellent genetically encoded reporters for electron microscopy (EM)[204].

To expand the utility of this naturally sourced proteinaceous nanocapsules, in this Chapter I describe my efforts to engineer *Thermotoga maritima* encapsulin[193] into a light-inducible nanocapsule system. The encapsulin capsule was fused with enhanced green fluorescent protein (EGFP)[205] as a genetically encoded fluorescent nanocapsule marker in HeLa cells. To evolve the encapsulin for light-inducible control, I attempted to engineer the encapsulin shell via different strategies. These strategies include fusing the photocleavable protein (PhoCl)[36] to control the capsule assembly in a light-dependent manner, genetic insertion of a protease cleavage site at various position of the subunit, and using the LARIAT tool[18] (described in section 1.3) to develop an encapsulin-based liquid phase separation system. We expect that engineered encapsulin nanocapsules could be useful for live cell imaging and the development of novel actuators for optogenetic manipulation of protein function and cell physiology.

4. 2 Results and discussion

4. 2. 1 Engineering of a genetically encoded fluorescent nanocapsule

I first attempted to express the *Thermotoga maritima* encapsulin[193] (TmEnc) or *Myxococcus xanthus* encapsulin[194] (MxEnc) (**Figure 4.3a**) in *Escherichia coli* strain BL21 (DE3). The recombinant protein expression and the auto-assembly of the capsules were confirmed by 'pseudo-native' protein electrophoresis which is based on SDS-PAGE analysis of non-heated protein samples[206,207] (**Figure 4.3b**). The results revealed that only TmEnc variant can auto-assemble into potential capsules with large molecular weight that remained in the stacking gel (4%). The TmEnc oligomers species also showed resistance against trypsin hydrolysis to a certain extent. For MxEnc encapsulin, the oligomeric species was not detected with our protein expression and purification protocols for reasons that remain unclear to us. Consequently, I employed the TmEnc variant in our next studies for better capsules formation.

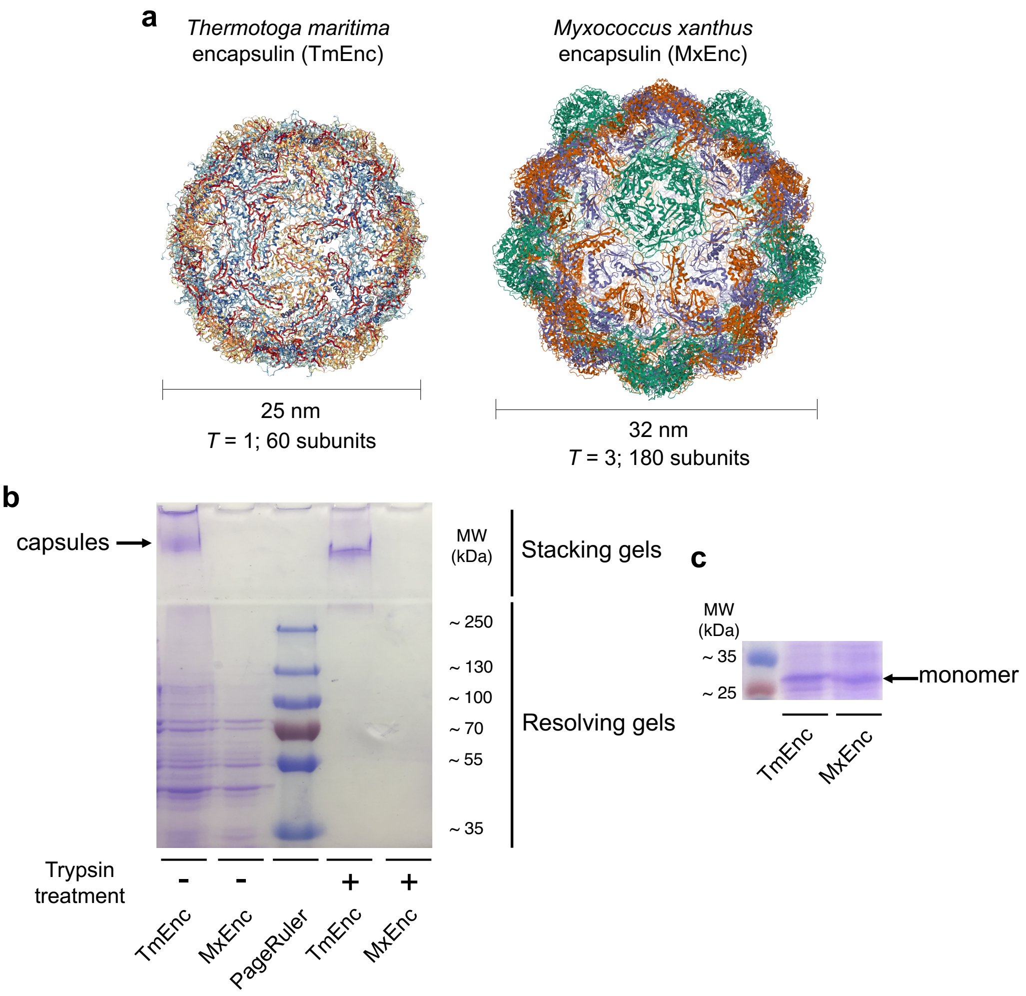


Figure 4.3 Encapsulin structures and recombinant protein expression in *E. coli*. (**a**) Assembly of *Thermotoga maritima* encapsulin (PDB ID: 3DKT) and *Myxococcus xanthus* encapsulin (PDB ID: 4PT2). (**b**) 'Pseudo-native' protein electrophoresis of the TmEnc and MxEnc, both with and without trypsin treatment. Non-heated protein samples were used in this assay. (**c**) SDS-PAGE analysis of the subunit of TmEnc and MxEnc. The molecular weight of monomer in both species is around 29 kDa.

To utilize the TmEnc as a genetically encoded nanocapsule in mammalian cells, I next tested protein expression and self-assembly of EGFP-fused encapsulin capsules in HeLa cells. Based on the protein structure of the *Thermotoga maritima* encapsulin (PDB ID: 3DKT)[193], we found that its N-terminus and C-terminus extend to the inner and outer surface of the capsule respectively. I reasoned that proteins, such as EGFP, that are genetically fused to the N-terminus can be effectively caged within the inner volume of encapsuling capsules. However, upon expression of EGFP-TmEnc fusion protein, I found that the protein capsules did not form (**Figure 4.4a**), as revealed by the green fluorescence being distributed evenly in the cytoplasm (**Figure 4.4b**). It is likely that inner volume of encapsulin is too small to accommodate a stoichiometric number of EGFP proteins (60 copies). In contrast, a C-terminal fusion protein would be displayed on the outside of the capsules (**Figure 4.4c**). Upon expression of the TmEnc-EGFP fusion protein, I found that the green fluorescence was localized to small, freely-diffusing, fluorescent “dots” in the cytoplasm (**Figure 4.4d**). These fluorescent small dots, which I interpret as the correctly assembled nanocapsule with 60 EGFP protein attached to its surface, can be used to observe the expression and package of the capsules via fluorescent microscopy.

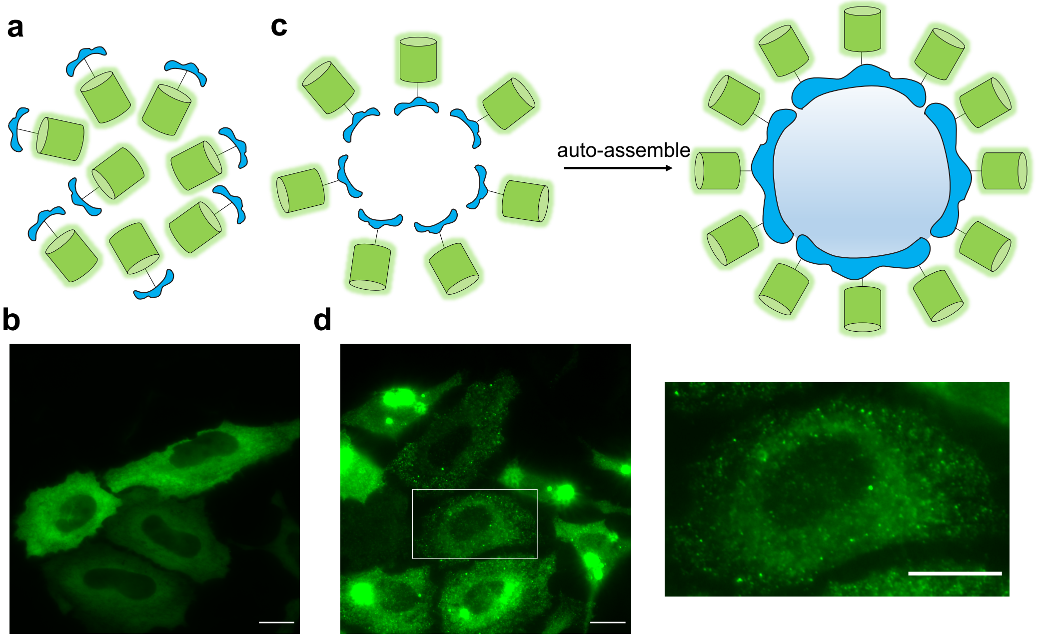
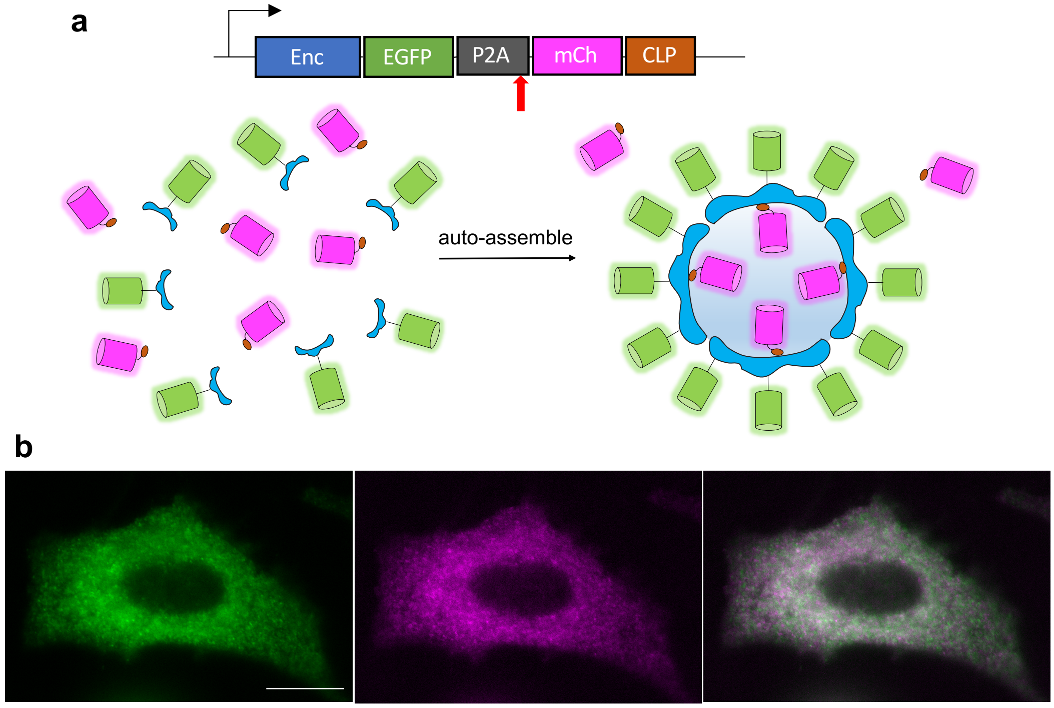


Figure 4.4 Expression of the engineered fluorescent encapsulin in HeLa cells. (**a**) Representation of EGFP fusion to the N-terminus of encapsulin. (**b**) Transient transfected HeLa cells expressing EGFP-TmEnc fusion protein. Scale bar, 10 μm. (**c**) Representation of EGFP fusion to the C-terminus of encapsulin. The expressed TmEnc-EGFP fusion protein can auto-assemble into fluorescent capsules in cells. (**d**) Transient transfected HeLa cells expressing TmEnc-EGFP fusion protein. The right panel is the zoomed-in region of interest (ROI), marked by white box in the left panel. Scale bar, 10 μm.

To load a heterologous cargo protein inside nanocapsules, a P2A bicistronic expression construct[208] was employed for co-expression of fluorescent encapsulin and the cargo protein (**Figure 4.5a**). In this study, I adopted mCherry fluorescent protein (FP) as a foreign cargo. mCherry was genetically fused with a minimal cargo loading peptide (CLP) identified at the C-terminal of a native cargo protein[201]. The result revealed that the TmEnc-EGFP fusion protein has the similar distribution pattern and localization as the presumed mCherry cargo, supporting the idea that the mCherry cargo can be auto-loaded into the capsules (**Figure 4.5b**).



TmEnc-EGFP mCherry-CLP co-localization

Figure 4.5 Co-expression of cargo protein and encapsulin. (**a**) A P2A bicistronic expression construct encoding the shell protein encapsulin (Enc) with EGFP as a C-terminal fusion, as well as mCherry fused with a cargo loading peptide (CLP). The P2A peptide is self-cleaved during translation. The cleavage site is marked with red arrow. The encapsulin subunit and the mCherry cargo can auto-assemble into capsules. (**b**) Transient transfected HeLa cells co-expressing TmEnc-EGFP fusion protein and mCherry-CLP fusion protein by using the construct described in **a**. Scale bar, 10 µm.

I also tested the expression, assembly, and cargo loading capacity of the *Myxococcus xanthus* encapsulin (MxEnc) as well as four encapsulin variants from other bacterial species[192] by using a construct analogous to the one described above (in **Figure 4.5a**). The CLP peptides for these encapsulin variants have not been determined experimentally. Hence, instead of using CLP peptide, the foreign mCherry FP was fused to their corresponding native cargo proteins. I found that these encapsulins from other species cannot form stable cargo-loaded nanocapsules, as revealed by the observation of few fluorescent dots in cell imaging. We speculate that the large size of the cargo fusion protein may introduce steric constraints at the inner surface of the capsules which impedes the capsule auto-assembly.

4. 2. 2 Engineering of light-closable and light-openable nanocapsules

To develop *Thermotoga maritima* encapsulin (TmEnc) into a light-closable capsule, we envisioned fusing the photocleavable protein (PhoCl) to the N-terminus of encapsulin. The N-terminal fusion would prevent complete capsule formation before photocleavage. Upon illumination with violet light, the cleavage of PhoCl would release the steric blocking effect and allow the auto-assembly of the nanocapsules. However, I found that encapsulin with only PhoCl fused to the N-terminus already exhibited small fluorescent dots before photoactivation, probably due to the partial formation of capsules or the oligomeric tendency of PhoCl at high local protein concentrations (**Figure 4.6a**). To solve this problem, we turned to our previously reported approach of protein caging by fusion to steroid receptor (SR) domains (described in section 3.1 and **Figure 3.1b**). I fused the SR domain to the N-terminus of the encapsulin via PhoCl, which would effectively serve as a photocleavable linker. The interaction of the SR with heat-shock protein 90 (Hsp90) leads to large steric hinderance, which would prevent the formation of capsules as well as PhoCl oligomers. This SR-PhoCl fusion strategy appeared to enable optogenetic activation of capsule assembly (**Figure 4.6b**) and cargo protein loading (**Figure 4.7**). However, further characterization (such as analysis of capsule diffusion rate, as well as correlative fluorescence and electron microscopy (CFEM)) will be required to demonstrate that the small fluorescent dots correspond to single nanocapsules.

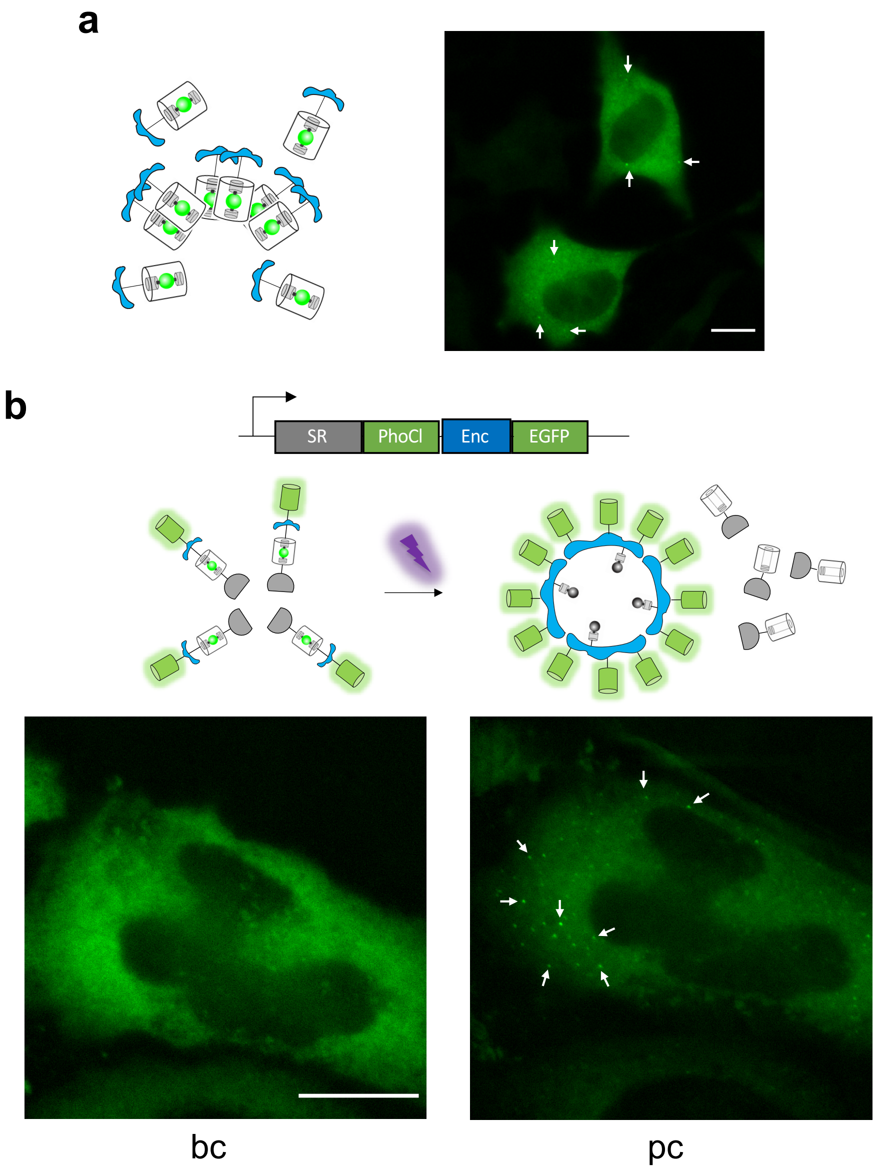


Figure 4.6 Optogenetic manipulation of nanocapsule assembly. (**a**) Schematic representation of PhoCl-TmEnc fusion protein, and transient transfected HeLa cells expressing PhoCl-TmEnc fusion protein. Several of the small fluorescent dots are marked with white arrows. Scale bar, 10 µm. (**b**) Schematic representation of SR-PhoCl fusion strategy for optical control of capsules formation, and the transient transfected HeLa cells expressing the construct. EGFP is displayed on the outer surface of encapsulin to indicate the expression and package of capsules. The formed small fluorescent dots after photoconversion are marked with white arrows. Photoconversion was performed with 5 s violet light pulses (395/40 nm, 2

mW/mm^2) every 1 min for 30 mins. bc, before-conversion; pc, post-conversion. Scale bar, 10 µm.

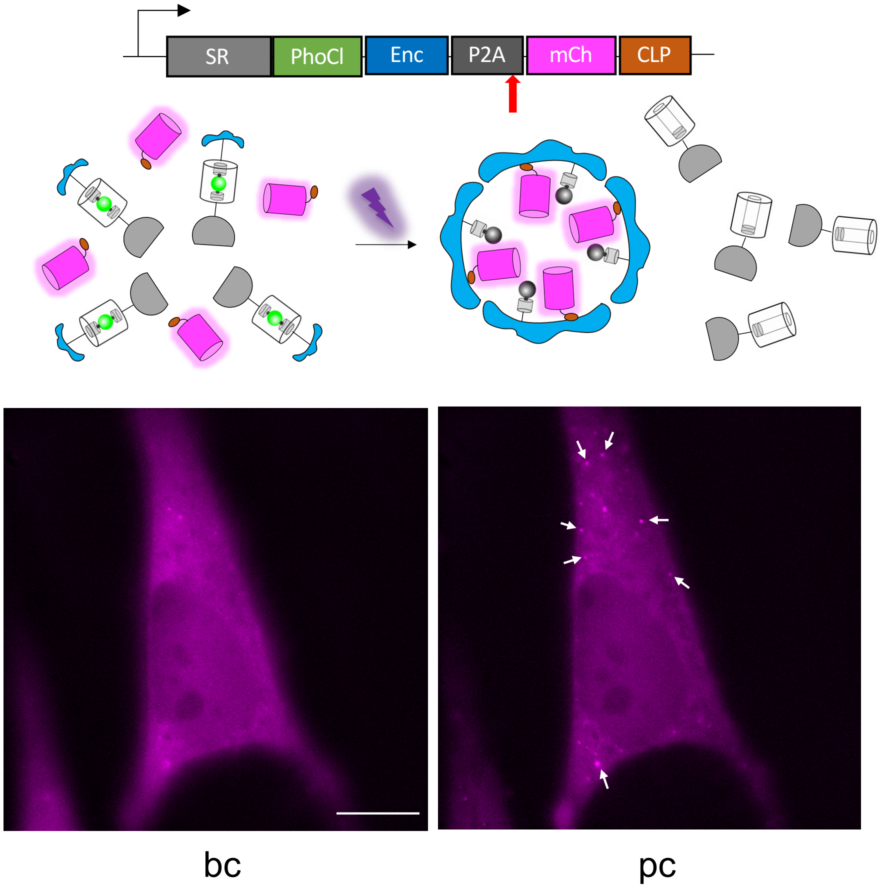


Figure 4.7 Optogenetic manipulation of cargo protein loading inside capsules. Schematic representation of the design strategy, and the transient transfected HeLa cells expressing the construct. Several of the small fluorescent dots formed after photoconversion are marked with white arrows. Photoconversion was performed with 5 s violet light pulses (395/40 nm, 2 mW/mm^2) every 1 min for 30 mins. bc, before-conversion; pc, post-conversion. Scale bar, 10 µm.

To engineer encapsulin into a light-openable cage, we proposed two design strategies to utilize PhoCl to break the polypeptide backbone of encapsulin and destabilize the capsule structures. The structural alignments of the two other encapsulin variants with the well characterized TmEnc suggested that TmEnc subunit is tolerant of several loop insertions on the capsule exterior[192] (**Figure 4.8**). Inspired by this finding, our first strategy was to insert protease substrate peptide into these insertion sites and manipulate the cleavage of the substrate by using our previously reported photoactivatable protease[36] (**Figure 4.9a**).

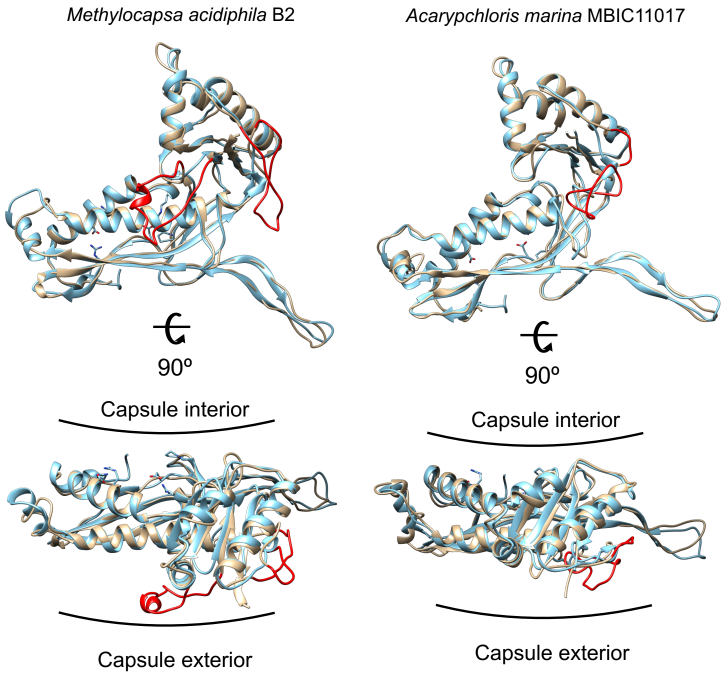


Figure 4.8 Alignment of other encapsulin variants with *Thermotoga maritima* encapsulin (TmEnc). TmEnc (PDB ID: 3DKT) is represented in beige; *Methylocapsa acidiphila* B2 encapsulin (MaB2Enc) and *Acarypchloris marina* MBIC11017 (AmEnc) encapsulin are represented in blue; loop insertions are highlighted in red. The PDB files of MaB2Enc and AmEnc homology models are unpublished data provided by our collaborator Dr. Tobias W. Giessen.

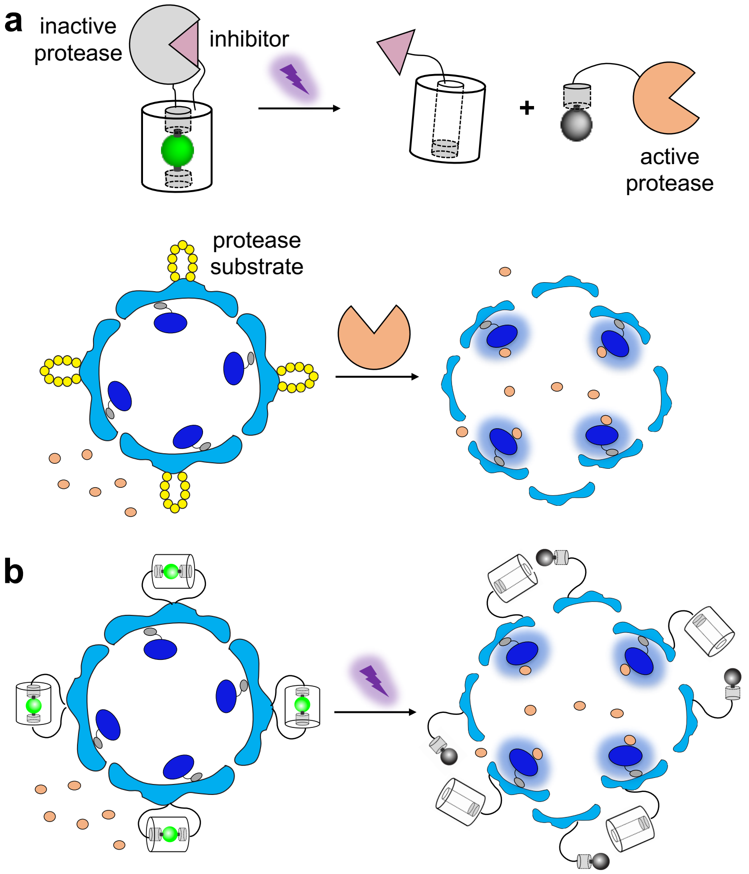


Figure 4.9 Schematic representation of proposed strategies for light-openable capsules. (**a**) Photo-uncaging of the cargo protein, by using the PhoCl-mediated photoactivatable protease to break encapsulin polypeptide backbone. (**b**) Optogenetic manipulation of capsule opening, by destabilizing of the capsules via PhoCl photocleavage.

As a preliminary proof-of-concept experiment, I inserted tobacco etch virus (TEV) protease cleavage sites (TEVs) at four positions of TmEnc separately (the positions are where the loops insertions are located on, represented in **Figure 4.8**). This effort led to four engineered TmEnc variants: TmEnc(A1), TmEnc(A2), TmEnc(M1) and TmEnc(M2). To assay the opening of the capsule, I preloaded it with co-expressed NanoLuc (NLuc) luciferase[162] cargo fused to the cargo loading peptide (CLP) (**Figure 4.10a**). Treatment of the purified loaded capsules with TEV protease (TEVp) resulted in an increase in bioluminescence (**Figure 4.10b**). This result suggests that the cleavage of encapsulin backbone is sufficient to disrupt the capsule structure such that small molecules, such as the NanoLuc substrate, furimazine (represented in **Figure 2.5c**), are able to enter the capsule.

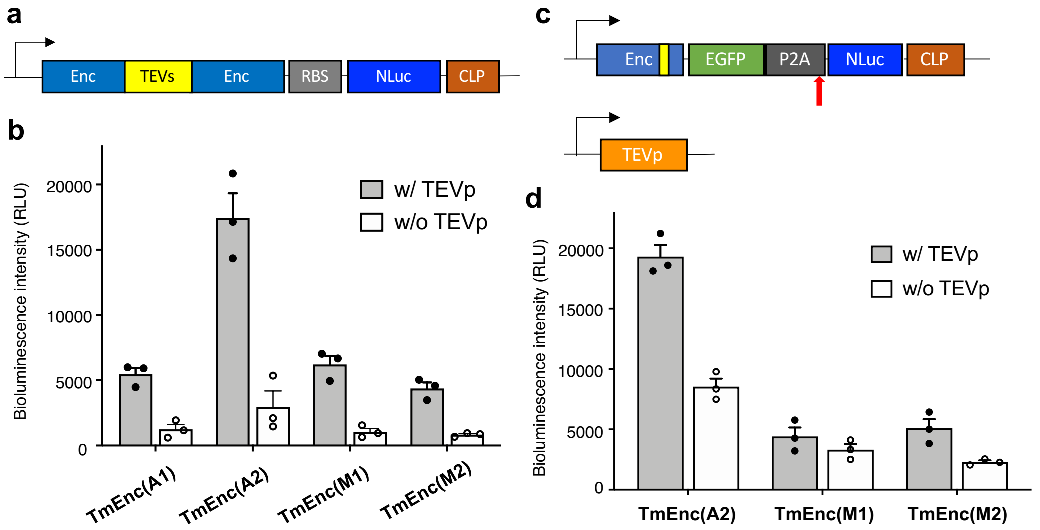


Figure 4.10 Bioluminescence assays of TmEnc variants with inserted TEV protease substrate sequences. (**a**) Genetic construct encoding the TEVs inserted encapsulin and cargo NanoLuc luciferase. The construct is used for expression of purified capsules in *E. coli*. RBS, ribosomal binding site. (**b**) Bioluminescence assay of purified engineered capsules, both with and without treatment of TEV protease. RLU, relative luminescence units. Values are means ± SEM (*n* = 3 independent experiments). (**c**) Genetic constructs used in HeLa cells for co-expressing the cargo-loaded capsules with TEV protease. (**d**) Bioluminescence assay of the engineered capsules expressed in HeLa cells. Cell suspensions were used in this assay. For the samples that were not transfected with TEVp expressing vector, the empty pcDNA 3.1(+) vector was co-transfected as control. Values are means ± SEM (*n* = 3 independent experiments).

I next co-expressed the cargo-loaded capsules with TEVp in mammalian cells (**Figure 4.10c**). TmEnc(A2), TmEnc(M1), and TmEnc(M2), (but not TmEnc(A1)), exhibited small fluorescent dots (potential capsules) when visualized by fluorescence microscopy. However, the TEV-dependent fold change in luminescence in the live cells assays (**Figure 4.10d**) were smaller than the changes observed with purified nanocapsules. We speculated that there are two possible reasons for this reduced contrast ratio. One is that the loading efficiency of luciferase cargo protein is not 100%, and there always be some unencapsulated NanoLuc enzyme inside the cells. In addition, some of the capsules in cells may be only partially formed, leaving large ‘holes’ in the capsule that allow free diffusion of substrate into the interior. Further optimizations of this strategy may be required to achieve highly effective release of the cargo protein.

To develop encapsulin into a light-openable capsule, our second strategy was to genetically insert PhoCl directly into the encapsulin subunit polypeptide. I attempted to introduce PhoCl into the three different positions of TmEnc, which correspond to the insertion sites used in the TmEnc(A2), TmEnc(M1) and TmEnc(M2) variants. Unfortunately, fluorescence live cell imaging exhibited substantial protein aggregations of these variants when expressed in HeLa cells, leading me to conclude that none of these PhoCl inserted encapsulins were properly forming the expected capsule structure.

4. 2. 3 Engineering of an encapsulin-based LARIAT system

As an alternative strategy to engineering encapsulin protein for optogenetic applications, I envisioned assembling the nanocapsules into clusters through the use of the LARIAT system[18] (described in section 1.3). This encapsulin-based LARIAT system was initially designed by replacing the multimeric proteins (MPs) with encapsulins. I fused TmEnc to CIBN, which can undergo blue light-induced heterodimerization with cryptochrome 2 (CRY2). Upon illumination, CRY2 homo-oligomerization and heterodimerization with CIBN leads to further interactions among encapsulin capsules that induces cluster formation (**Figure 4.11a**).

I first tested whether the light-dependent CRY2-CIBN interaction could be used to induce protein clustering. The CRY2-mCherry fusion protein and TmEnc-EGFP-CIBN fusion protein were co-expressed in HeLa cells. Initial experiments revealed the nuclear localization of CRY2 and CIBN as expected. Blue light illumination induced the formation of red fluorescent puncta, which reversibly dissembled in dark and reassembled with repeated illumination (**Figure 4.11b,e**). I observed no light-induced cluster formation in cells expressing CRY2-mCherry fusion protein alone (**Figure 4.11c,f**). The punctuate nuclear localizations were observed in cells expressing TmEnc-EGFP-CIBN fusion protein (**Figure 4.11d**), probably due to the aggregation of capsules at high protein concentrations in nucleus.

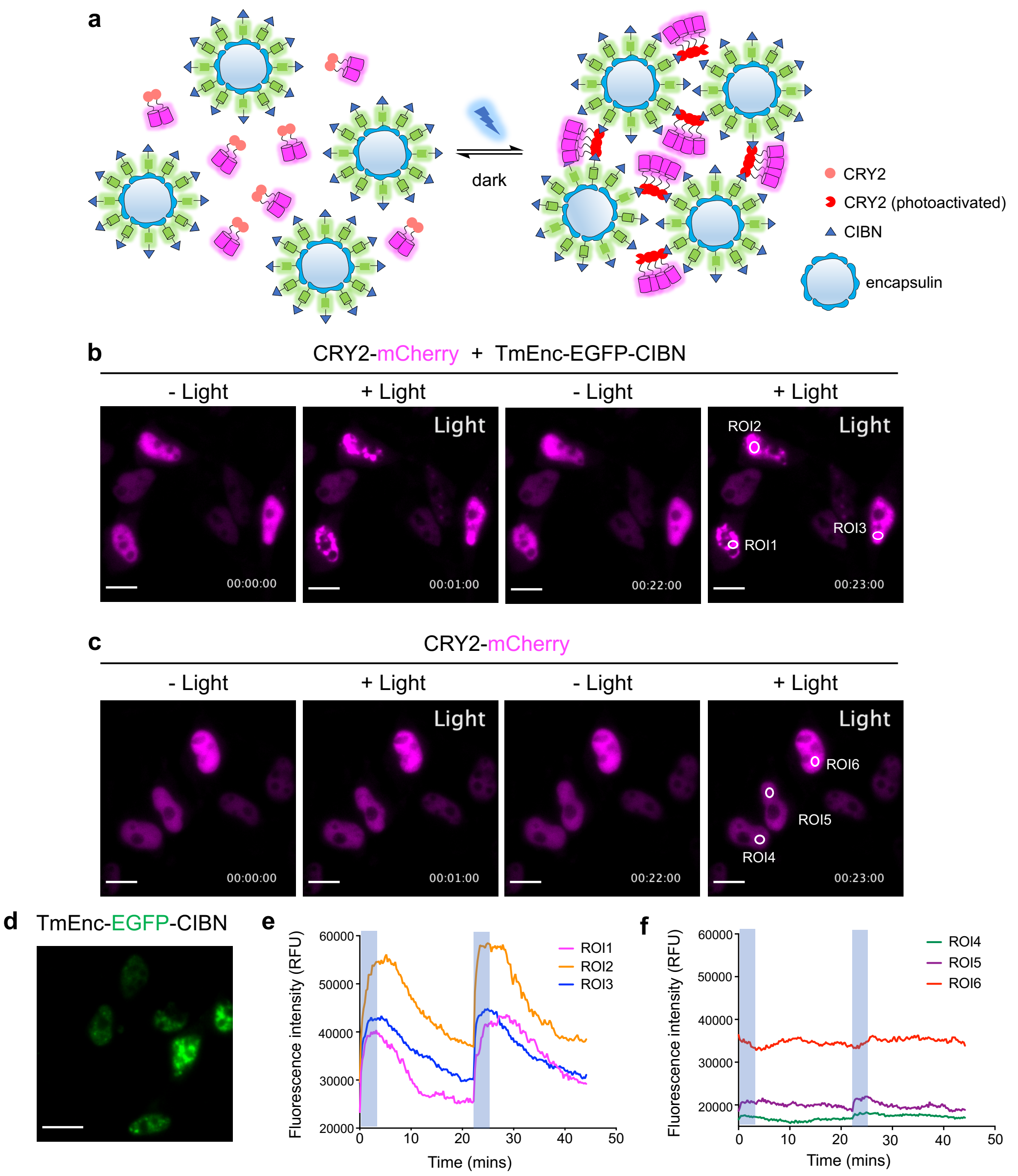


Figure 4.11 Encapsulin-based LARIAT system. (**a**) Schematic representation of encapsulin-based LARIAT design strategy. (**b**) Representative images of HeLa cells co-expressing CRY2-mCherry and TmEnc-EGFP-CIBN fusion protein. Cells were illuminated with blue light twice at a 20 min interval. Illumination was performed with 300 ms blue light pulses (490/15 nm, 0.5 mW/mm^2) every 15 s for 2 mins. Scale bar, 10 µm. Time lapse label is display in HH:MM:SS. (**c**) HeLa cells expressing CRY2-mCherry fusion protein alone. Illumination was performed in the method described in **b**. Scale bar, 10 µm. (**d**) HeLa cells expressing TmEnc-EGFP-CIBN fusion protein alone. Scale bar, 10 µm. (**e**) Fluorescence intensity profiles of

the light-induced clusters represented in **b**. RFU, relative fluorescence units. ROI, region of interest. (**f**) Fluorescence intensity profiles of the ROI represented in **c**.

I next attempted to mutate the residues that have previously been predicted to cause nuclear localization of CRY2 and CIBN[85] (**Figure 4.12**). As expected, the variants exhibited cytoplasmic distribution, in the context of TmEnc-EGFP-CIBN fusion protein and CRY2-mCherry, when expressed in mammalian cells (**Figure 4.13**). However, compared to the wild-type CRY2 and CIBN (**Figure 4.11b**), the mutants exhibited reduced cluster size and cluster formation efficiency upon illumination (**Figure 4.13a**). To improve protein clustering, we will attempt to optimize this system through several strategies in our further studies. These strategies include enhancing the expression level of TmEmc-EGFP-CIBN fusion protein, as well as testing the CRY2PHR and CIB1 variants[85].

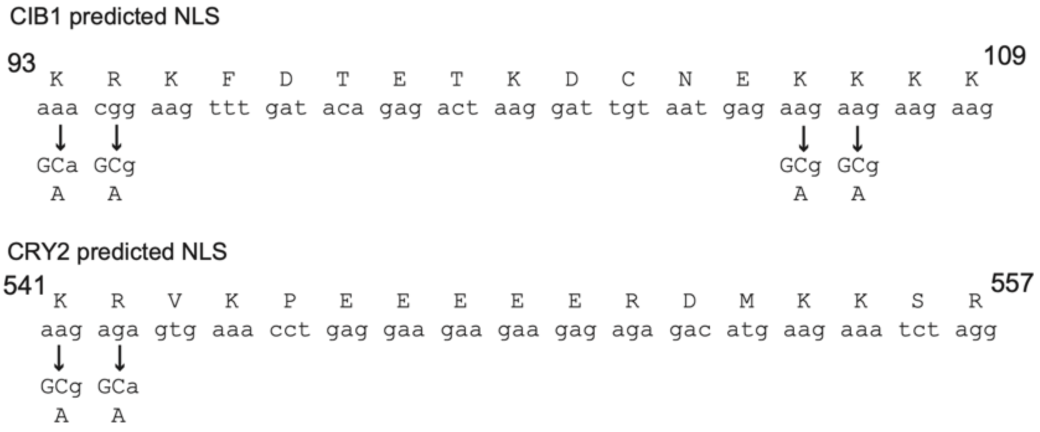


Figure 4.12 Mutations of the predicted nuclear localization sequence in CRY2 and CIB1. The number indicate amino acid residue position. NLS, nuclear localization sequence. (*Note*: **Figure 4.12** reprinted from Ref. 85 and used with permission of the publisher.)

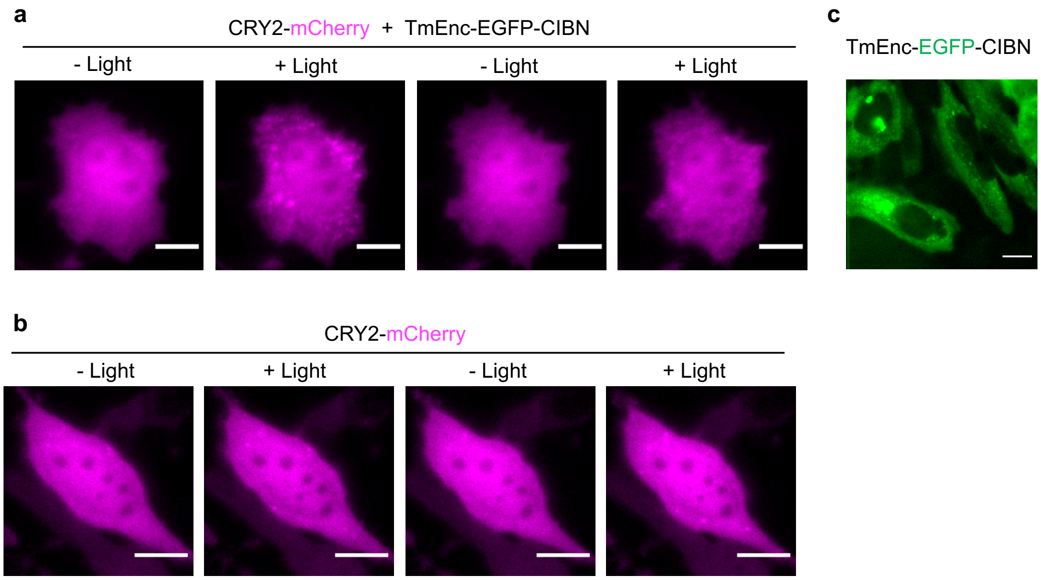


Figure 4.13 Encapsulin-based LARIAT system with mutated CRY2 and CIBN variants. (**a**) Representative images of HeLa cells co-expressing the mutated CRY2-mCherry and TmEnc-EGFP-CIBN fusion protein. Cells were illuminated with blue light twice at a 20 min interval. Illumination was performed with 300 ms blue light pulses (490/15 nm, 0.5 mW/mm^2) every 15 s for 2 mins. Scale bar, 10 µm. (**b**) HeLa cells expressing the mutated CRY2-mCherry fusion protein alone. Illumination was performed in the method described in **a**. Scale bar, 10 µm. (**c**) HeLa cells expressing the mutated TmEnc-EGFP-CIBN fusion protein alone. Scale bar, 10 µm.

4. 3 Conclusion

By utilizing protein engineering and rational protein design, I have made progress towards developing a bacterial encapsulin-based nanocapsule platform for use in optogenetic tools. My preliminary results demonstrate that *Thermotoga maritima* encapsulin can be engineered to be a genetically encoded fluorescent nanocapsule marker by displaying EGFP on the outer surface. I have engineered a photo-closable encapsulin shell by using PhoCl-SR fusion strategy, and a protease-openable capsule by genetic insertion of a TEV protease cleavage site at various positions. I further attempted to achieve optogenetic protein clustering by combining encapsulin with the CRY2-CIBN pair. Despite the many promising results in development of optogenetic tools based on encapsulin, the unambiguous evidence for complete capsule formation in mammalian cells, and the question of whether the small fluorescent particles represent single nanocapsules, remain to be demonstrated. We expect that the encapsulin-based nanocapsule system will be further optimized in future work, and ultimately prove its utility in optical control of biological processes.

4. 4 Methods and materials

4. 4. 1 General molecular biology methods and materials

The general molecular biology methods used in this chapter were same to those described in section 2.4.1. Primer sequences are provided in **Table 4.1**.

4. 4. 2 Protein purification and characterization

The gene expressing TmEnc, MxEnc in pET-28a vector, as well as TEVs inserted TmEnc and NLuc cargo protein in the pET-28a bicistronic vector were used to transform *E. coli* strain BL21 (DE3) pLysS (Promega). A single colony was picked to inoculate a 5 mL starting culture that was allowed to grow overnight at 37 °C with 225 rpm before being diluted into 1 liter large culture of LB media supplemented with 50 μg/mL ampicillin. The cultures were allowed to grow at 37

°C to an OD_{600} of 0.5, and then induced with IPTG (1 mM) and grown 4 h at 37 °C. The *E. coli* cells were harvested by centrifuge, the pellets were immediately frozen and stored in - 20 °C.

Encapsulin protein was purified according to the previously reported procedure[192]. Pellets were thawed and resuspended in 15 mL PBS (pH = 7.4) buffer, DNase I (Thermo Fisher Scientific; final concentration: 1 μg/mL) was added and cells were incubated for 20 min on ice, then the cell suspensions were lysed by sonication. Cell debris was subsequently removed through centrifugation (8000 rpm, 40 min, 4°C). The cleared supernatant containing encapsulin was treated with 0.3 g NaCl and 1.5 g of PEG-8000 was added, followed by incubation on ice for 30 min. The precipitate was collected through centrifugation (8000 rpm, 25 min, 4°C), then suspended in 10 mL PBS buffer and filtered using a 0.2 μm syringe filter. The samples were then subjected to gel filtration with a HiPrep 16/60 Sephacryl S-500 column on an AKTA chromatography system (GE Healthcare).

For the purified TmEnc and MxEnc capsules, trypsin hydrolysis reactions were performed with trypsin protease (Thermo Fisher Scientific) according to the manufacturer's protocol, followed by pseudo-native protein electrophoresis analysis. For the purified NLuc cargo protein loaded capsules, protein samples were treated with TEV protease before bioluminescence measurement. The hydrolysis reactions were performed with ProTEV Plus (Promega) at 30 °C for 4 hours. The luminescence emission intensity (at 460 nm) of the protein, both with and without hydrolysis, were measured immediately after treatment with the luciferase substrate furimazine (working concentration: 10 μg/mL). The bioluminescence intensity were acquired with a Synergy HTX plate reader (BioTek).

4. 4. 3 Construction of encapsulin expression vectors

To construct the vectors expressed in *E. coli*, *Thermotoga maritima* encapsulin (TmEnc) and *Myxococcus xanthus* encapsulin (MxEnc) templates were ordered as synthetic gBlocks from IDT. The DNA fragments encoding the encapsulin variants were amplified via PCR and inserted into the digested pET-28a vector at NcoI restriction site by using Gibson Assembly Master Mix (NEB). For the construction of bicistronic vector, the gene encoding NLuc-CLP-RBS (with 5' NcoI and 3' EcoRI restriction sites) and TmEnc (with 5' EcoRI and 3' XhoI restriction sites) were digested and inserted between the NcoI and XhoI sites of pET-28a vector by three-way ligation thus yielding the construct of NcoI-*NLuc*-*CLP*-RBS-EcoRI-*TmEnc*-XhoI. The protein sequence of cargo loading peptide (CLP) used in this study is: GGSENTGGDLGIRKL[201], and the DNA sequence of the ribosomal binding site (RBS) is: TTTGTTTAACTTTAAGAAGGAGA[209]. The TEV protease cut sites were introduced into various positions by using site-directed mutagenesis.

For mammalian cells expression, the pcDNA 3.1(+) vector was used as the backbone vector. The DNA fragments encoding the EGFP-TmEnc or TmEnc-EGFP fusion proteins were made using overlap extension PCR to connect the *EGFP* and *TmEnc* gene fragments. The assembled genes (with 5' NheI and 3' XhoI restriction sites) were digested and inserted between the NheI and XhoI sites of pcDNA 3.1(+), thus yielding the EGFP labelled encapsulin expression vectors. For co-expressing the TmEnc and the cargo protein in mammalian cells, the gene encoding TmEnc-EGFP-P2A (with 5’ NheI and 3’ KpnI restriction sites), and mCherry-CLP or NLuc-CLP (with 5’ KpnI and 3’ XhoI restriction sites) were digested and inserted between the NheI and XhoI sites of pcDNA 3.1(+), thus yielding the P2A bicistronic construct. The protein sequence of P2A peptide is: ATNFSLLKQAGDVEENPGP[208].

The previously reported pCAG-ERT2-PhoCl-Cre-PhoCl-ERT2 (Addgene #87694) plasmid was used as the backbone vector for the construction of SR-PhoCl fused encapsulins. The DNA fragments encoding the TmEnc-EGFP or TmEnc-P2A-mCherry-CLP (with 5’ NheI and 3’ NotI restriction sites) were amplified via PCR and inserted into the appropriately digested pCAG vector.

To construct the plasmids used in encapsulin-based LARIAT system, DNA encoding the CRY2 (with 5’ NheI and 3’ KpnI restriction sites) and the mCherry (with 5’ KpnI and 3’ XhoI restriction sites) were digested and inserted between the NheI and XhoI sites of pcDNA 3.1(+), thus yielding CRY2-mCherry expression vector. And the gene fragments encoding the TmEnc-EGFP (with 5’ NheI and 3’ Kpn2I restriction sites) and the CIBN (with 5’ Kpn2I and 3’ XhoI restriction sites) were inserted into pcDNA 3.1(+), thus yielding TmEnc-EGFP-CIBN expression vector. Mutating the predicted nuclear localization residues in CRY2 and CIBN were performed by using site-directed mutagenesis.

4. 4. 4 Cell photoconversion and imaging conditions

The general imaging conditions used in this chapter were same to those described in section 3.4.4.

For imaging of PhoCl-mediated photo-closable encapsulins, photoconversion was performed with 5 s pulse violet light illumination (395/40 nm nm, 2 mW/mm^2) every 1 min for 30 min, and each pulse of photoconversion light was followed by acquisition of EGFP or mCherry fluorescence images. Cells are imaged with an Axiovert 200M inverted microscope (Zeiss), which was equipped with a 75 W xenon-arc lamp, 63× objective lens (NA 1.40, oil; Zeiss) and an ORCA-Flash4.0 CMOS camera (Hamamatsu). The MetaMorph software was used for automatic instrument control and image acquisition.

For imaging of the light-induced protein clustering in encapsulin-LARIAT system, the images were acquired in mCherry channel before photoconversion. Photoconversion was performed with 300 ms pulse blue light illumination (490/15 nm, 0.5 mW/mm^2) every 15 s for 2 min and each pulse of photoconversion light was followed by acquisition of mCherry fluorescence images. Imaging acquisition of red channel continued every 15 s for another 20 min after the photoconversion, then a repeated round of photoconversion and imaging acquisition was performed in the same method. Cells were imaged with an Nikon Eclipse Ti-E epifluorescence microscope equipped with a 75-W Nikon xenon lamp, 20× objective lens (NA 0.75, air; Nikon) and a Photometrics QuantEM 512SC camera. The NIS-Elements AR package software was used for automatic instrument control and image acquisition. Image analysis was performed using ImageJ software (open source).

Table 4.1 Primers used in Chapter 4.

Primers	**Sequence 5’-3’**
pET-28a-*TmEnc* construct	
F-TmEnc	GTTTAACTTTAAGAAGGAGATATACCATGGAGTTTCTGAAGCGC
R-TmEnc	CTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGAACTTAAGCAGGATCAGAGC
pET-28a-*MxEnc* construct	
F-MxEnc	GTTTAACTTTAAGAAGGAGATATACCATGCCAGATTTTCTGGGTCAC
R-MxEnc	CTTCCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGCGACGCTCAGTAGCACC
pET-28a-*NLuc-CLP*-RBS-*TmEnc* construct	
F-NcoI-NLuc	GATATACCATGGTCTTCACACTCGAAGATTTCG
R1-NLuc-CLP	TTACGGATACCAAGATCACCGCCGGTATTTTCAGACCCGCCCGCCAGAATGCGTTCGCAC
R2-CLP-RBS	GTTAAACAAATCTAGAGGTCACAACTTACGGATACCAAGATCACC
R3-RBS-EcoRI	CTCCATGAATTCTCCTTCTTAAAGTTAAACAAATCTAGAGGTCACAAC
F-EcoRI-TmEnc	AAGGAGAATTCATGGAGTTTCTGAAGCGCTCCTTTG
R-TmEnc-XhoI	TGGTGCTCGAGGAACTTAAGCAGGATCAGAGCCTCGG
Quickchange primers for TEVs insertions	
TmEnc(A1)-sense	GCGTAAAGGGCTTACTTAGTGAAAACTTGTATTTTCAGAGTTTCGAAGAACGTAAGATTG
TmEnc(A2)-sense	TACTTAGTTTCGAAGAACGTGAAAACTTGTATTTTCAGAGTAAGATTGAATGTGGTTCCA
TmEnc(M1)-sense	TCGAAGAACGTAAGATTGAAGAAAACTTGTATTTTCAGAGTTGTGGTTCCACGCCGAAAG
TmEnc(M2)-sense	TTACAACTCCTCGCATTGAAGAAAACTTGTATTTTCAGAGTGACGCCTTGGTAGTAAGCG
pcDNA-*EGFP-TmEnc* construct	
F-NheI-EGFP	AAGCTGGCTAGCGCCACCATGGTGAGCAAGGGCGAGGAG
R-EGFP-TmEnc overlap PCR	GCTTCAGAAACTCGCTTCCACCCTTGTACAGCTCGTCCATGCC
F-EGFP-TmEnc overlap PCR	GAGCTGTACAAGGGTGGAAGCGAGTTTCTGAAGCGCTCCTTTGC
R-TmEnc-XhoI	TCTAGACTCGAGTTAGAACTTAAGCAGGATCAGAGCCTCG
pcDNA-*TmEnc-EGFP* construct	
F-NheI-TmEnc	AAGCTGGCTAGCGCCACCATGGAGTTTCTGAAGCGCTCC
R-TmEnc-EGFP overlap PCR	CCCTTGCTCACGCTTCCACCGAACTTAAGCAGGATCAGAGCCTCG
F-TmEnc-EGFP overlap PCR	CCTGCTTAAGTTCGGTGGAAGCGTGAGCAAGGGCGAGGAGC
R-EGFP-XhoI	TCTAGACTCGAGTTACTTGTACAGCTCGTCCATGCC
pcDNA-*TmEnc-EGFP-P2A-mCherry-CLP* construct	
F-NheI-TmEnc	AAGCTGGCTAGCGCCACCATGGAGTTTCTGAAGCGCTCC
R1-TmEnc-EGFP-P2A	TACGTCACCAGCTTGTTTGAGGAGTGAGAAATTTGTTGCCTTGTACAGCTCGTCCATGCC
R2-P2A-KpnI	GCTCACGGTACCTGGTCCTGGGTTTTCTTCTACGTCACCAGCTTGTTTGAGGAG
F-KpnI-mCherry	GGACCAGGTACCGTGAGCAAGGGCGAGGAGG

R1-mCherry-CLP	TACGTCACCAGCTTGTTTGAGGAGTGAGAAATTTGTTGCCTTGTACAGCTCGTCCATGCC
R2-CLP-XhoI	ATAGGCAGCCTGCACCTGAGGAGTGCGGCCGCTCACAACTTACGGATACCAAGATCACCG
pcDNA-*TmEnc-EGFP-P2A-NLuc-CLP* construct	
F-KpnI-NLuc	GGACCAGGTACCATGGTCTTCACACTCGAAGATTTCGTTG
R1-mCherry-CLP	CTTACGGATACCAAGATCACCGCCGGTATTTTCAGACCCGCCCGCCAGAATGCGTTCGCAC
R2-CLP-XhoI	TCTAGACTCGAGTCACAACTTACGGATACCAAGATCACCGCCG
pCAG-*ERT2-PhoCl-TmEnc-EGFP* construct	
F-NheI-TmEnc	CCACGGGCTAGCGGAGGCAGCATGGAGTTTCTGAAGCGCTCC
R-EGFP-NotI	AGGAGTGCGGCCGCTATCACTTGTACAGCTCGTCCATGCC
pCAG-*ERT2-PhoCl-TmEnc-P2A-mCherry-CLP* construct	
F-NheI-TmEnc	CCACGGGCTAGCGGAGGCAGCATGGAGTTTCTGAAGCGCTCC
R-TmEnc-XhoI	TCGCCCTTGCTCACGCTCTCGAGGAACTTAAGCAGGATCAGAGCCTCG
F-XhoI-P2A	AGGCTCTGATCCTGCTTAAGTTCCTCGAGGCAACAAATTTCTCACTCCTCAAACAAGCTG
R-CLP-NotI	ATAGGCAGCCTGCACCTGAGGAGTGCGGCCGCTCACAACTTACGGATACCAAGATCACCG
pcDNA-*CRY2-mCherry* construct	
F-NheI-CRY2	TATAGGGAGACCCAAGCTGGCTAGCGCCACCATGAAGATGGACAAAAAGACTATAGTTTG
R-CRY2-KpnI	CCTCCGGTACCTTTGCAACCATTTTTTCCCAAACTTGTAG
F-KpnI-linker-mCherry	CTACAAGTTTGGGAAAAAATGGTTGCAAAGGTACCGGAGGCGGGGGCAGCG
R-mCherry-XhoI	AGCGGGTTTAAACGGGCCCTCTAGACTCGAGTTACTTGTACAGCTCGTCCATGCC
pcDNA-*CRY2-mCherry* construct	
F-NheI-CRY2	TATAGGGAGACCCAAGCTGGCTAGCGCCACCATGAAGATGGACAAAAAGACTATAGTTTG
R-CRY2-KpnI	CCTCCGGTACCTTTGCAACCATTTTTTCCCAAACTTGTAG
F-KpnI-linker-mCherry	CTACAAGTTTGGGAAAAAATGGTTGCAAAGGTACCGGAGGCGGGGGCAGCG
R-mCherry-XhoI	AGCGGGTTTAAACGGGCCCTCTAGACTCGAGTTACTTGTACAGCTCGTCCATGCC
pcDNA-*TmEnc-EGFP-CIBN* construct	
F-NheI-TmEnc	CTCACTATAGGGAGACCCAAGCTGGCTAGCGCCACCATGGAGTTTC
R-EGFP-Kpn2I	CCTCCTCCGGACTTGTACAGCTCGTCCATGCCGAG
F-Kpn2I-CIBN	TGGACGAGCTGTACAAGTCCGGAGGAGGCGGTAGCATGAATGGAGCTATAGGAGGTGACC
R-CIBN-XhoI	GTTTAAACGGGCCCTCTAGACTCGAGTTAAATATAATCCGTTTTCTCCAATTCCTTCGTC
Quickchange primers for NLS mutation	
CRY2 NLS mut-antisense	CTTCCTCAGGTTTCACTGCCGCTGACCCGTTGTAACGAACAGCCGAAG
CIBN NLS mut1-antisense	CCTTAGTCTCTGTATCAAACTTCGCTGCCTTGAAATTTCCAGTCCCAAGCG
CIBN NLS mut2-antisense	TCATCTCTGTTCATCGTCATCTTCTTCGCCGCCTCATTACAATCCTTAGTCTCTGTATC

Chapter 5 Conclusions and future directions

5. 1 Summary of the thesis

The development of optogenetic actuators has resulted in a broad repertoire of genetically encodable tools with diversity in terms of mechanism, spatiotemporal scale, and applicability to the control of various biological processes. These tools provide researchers with a powerful toolkit to expand our understanding of cell biology and physiology, and may one day find use as therapeutic therapies in clinical applications[210–214]. Despite the tremendous progress in the development and application of tools, optogenetic technology remains a burgeoning field. There is growing demand for new tools with new capabilities, including improved optical properties and light-induced responses, to satisfy the needs of a wider variety of studies in the life sciences.

To expand the actuator toolkit, we previously introduced the first generation of photocleavable protein (PhoCl1), which is engineered from a green-to-red photoconvertible fluorescent protein (FP)[36]. Compared to the other photoreceptor proteins used in optogenetics, PhoCl has some unique beneficial properties, such as robust heterologous expression, self-sufficient chromophore formation[215], and a uniquely irreversible photocleavage mechanism[146,216]. PhoCl1 has been demonstrated to be a simple and versatile photocleavable linker for regulation of protein translocation, gene expression and enzyme activities[36].

To further improve PhoCl1 for faster and more efficient dissociation, in Chapter 2 I presented my efforts to develop novel variants of PhoCl. Working with collaborators, the protein X-ray crystal structures of PhoCl1 (green, red, and empty barrel states) were determined in order to obtain a better understanding of the structural changes associated with the dissociation process. To screen libraries based on dissociation, I developed a high-throughput screening method by using a luciferase complementation assay. This work led to the identification of the second generation of PhoCl variants: PhoCl2c with improved dissociation efficiency, and PhoCl2f with faster dissociation rate. In Chapter 3, I demonstrated the improved dissociation performances of PhoCl2 variants when characterized as purified proteins and in cell-based experiments. I further explored the application of PhoCl2c for optogenetic apoptosis in living cells. These data demonstrate the utility of PhoCl2 variants for the development of novel and improved PhoCl-based optogenetic tools.

In Chapter4, I described my efforts to explore the utility of encapsulin, a naturally sourced bacterial protein-based nanocompartment, for optogenetic applications. This proteinaceous capsules can pack foreign cargo proteins inside and be auto-assembled in cells, making it a promising building block to be used in optogenetic tools. By combining the use of encapsulin and different photosensory domains, I have engineered a photo-closable encapsulin shell through using photocleavage of PhoCl, as well as an optogenetic protein clustering system by using the blue light-activatable CRY2-CIBN pair.

In summary, this thesis described the development, characterization and application of PhoCl2 variants and encapsulin. In the following sections, I will propose our future directions for these studies.

5. 2 Future directions

5. 2. 1 Development of novel screening strategies and applications for PhoCl

In Chapter 2, I described a novel high-through screening method based on a protein-fragment complementation assay. However, the screened library size still remained limited (with ~ 10,000 variants in total) by using this strategy. A limiting factor is that the protein concentration is too high in *E. coli* expression system, such that the NanoBiT luciferase remains associated inside the bacteria after photocleavage. Unfortunately, this high concentration means that the screening strategy cannot be performed directly in *E. coli* colonies on agar plates. Rather, the colonies have to be picked, cultured, and lysed manually before measurements. To address this problem for more high-throughput screening, the protein expression level in *E. coli* would need to be further optimized to balance both aspects of dissociation affinity and detectable luminescence signals.

Another promising screening strategy would be to utilize the mRNA display[217–219] or cDNA display[220–222] technologies. This type of *in vitro* display method results in the association of the translated proteins with their gene sequence information. The nucleotide sequence of a single mRNA-PhoCl fusion protein molecule that dissociates faster, or needs lower light intensity for conversion, could be identified from large size library. Currently, such an mRNA display strategy is being used to screen for improved PhoCl variants by our collaborator Professor Donald B. Arnold at the University of Southern California.

In addition to high-throughput screening strategies, the range of possible applications of PhoCl should be further explored. In principle, the barrel structure of PhoCl could allow the inhibition of small peptide or the active site of enzymes directly by blocking their mode of action through a steric effect. Such a PhoCl-mediated strategy could be used to covalently trap or cage cell signaling peptide (such as translocation signal peptide and enzyme inhibitor) and protein enzymes in the dark and trigger the activity upon light illumination by releasing the steric constraint.

5. 2. 2 Development of encapsulin-based nanocapsule platform

For encapsulin-based nanocapsules, future efforts should focus on demonstration of the complete capsule formation and monodisperse particles distribution in mammalian cells by using a variety of characterization assays. These assays could include western blot, native mass spectrometry, and correlative fluorescence and electron microscopy. After these further demonstrations, the advantages and utility of encapsulins could be explored as compartments used in optogenetic tools to regulate a wide variety of metabolic and physiologic processes.

In addition to be used as a nanocompartment, I have demonstrated that the oligomerization of encaspulin subunits can be used in the encapsulin-based LARIAT system. In this system, the complete capsule formation is not strictly required, which may expand the versatility of encapsulins. This strategy could be further demonstrated for its utility for optogenetic probing and controlling of liquid-liquid phase separation[223–225]. Furthermore, the cargo protein loading capacity of encapsulin would bring new insights into this area by enabling the manipulation of biochemical reactions with spatiotemporal precision in a simple and versatile approach.

5. 3 Concluding remarks

In the research work presented in this thesis, I have described my effort to develop the second generation of PhoCl variants and the bacterial encapsulin-based light-controllable nanocapsules. First, by using protein engineering and structural-guided directed evolution, we successfully engineered two novel variants PhoCl2c (with improved dissociation efficiency) and PhoCl2f (with faster dissociation rate). We demonstrated the improvement of PhoCl2 variants *in vitro* and in cell-based experiments. Compared to PhoCl1, the PhoCl2c variant has been demonstrated to have higher dissociation efficiency in optogenetic apoptosis application. Although the protein structure of the novel PhoCl2 variants have not been solved yet, the determined X-ray crystal structures of PhoCl1 presented in this study expand our understanding of the dissociation process at the molecular level and will guide even further protein engineering studies. Second, we explored the utility of bacterial encapsulin nanocapsules in mammalian cell expression system. We have described our efforts to engineer the light-controllable encapsulin shell through various design strategies. Although the further characterizations are required to demonstrate the capsule integrity in cells, we propose that the encapsulin-based optogenetic tools has great potential in imaging cell dynamics, regulation the activities of functional protein, and controlling of protein phase separation.

Appendix: Protein crystallization and structure determination of PhoCl1

The protein crystallization, X-ray data collection and structure refinement were performed by our collaborator: **Dr. Yurong Wen** from Dr. M. Joanne Lemieux's lab in Department of Biochemistry, University of Alberta. The detailed method and data collection are described below.

A. 1 Crystallization and data collection

The green state of PhoCl1 and PhoCl1 empty barrel recombinant proteins were further purified with size exclusion chromatography using SD200 column (GE Healthcare) and buffer exchanged to 20 mM Tris pH 7.4, 150 mM NaCl. Initial crystallization trails were carried out with 384-well plate via sitting drop vapor diffusion against commercially available sparse matrix screens (Hampton Research and Molecular Dimensions) at room temperature. The green state of PhoCl was crystallized in 0.1 M MIB buffer pH 6.0, 25% w/v PEG 1500. Crystals of the red state of PhoCl1 was generated from the PhoCl1 green crystals through photoconversion with the violet light (15 s illumination with 405 nm LED array, 3.46 mW/mm^2). The PhoCl1 empty barrel crystals were observed in 0.056 M $NaH_2PO_4 \cdot H2O$, 1.344 M K_2HPO_4, pH 8.2. All the crystals were cryoprotected with the reservoir condition in supplement with 15-20% glycerol and flash frozen in liquid nitrogen. X-ray diffraction dataset were collected at Canadian Light Source

CMCF-ID beamline. All the datasets were processed and scaled with XDS package[226]; the data collection details and statistics are summarized in **Table A.1**.

A. 2 Structure determination and refinement

The structure of the green state of PhoCl1 was solved by molecular replacement using mTFP1 (PDB ID: 2HQK) as search model[227,228]. The red state of PhoCl1 and PhoCl empty barrel structure determination were carried out using the solved the green state of PhoCl1 as template. The manual model building and refinement were performed with Coot[229] and PHENIX[230] refine. The CR8 and IEY were used as the chromophore ligands for the green state and the red state of PhoCl1 respectively. In the PhoCl red crystal, the chromophore turned out to be partially converted, therefore two of the six monomers in the asymmetric unit with high occupancy for IEY were modeled in red form, and the other monomer was built in green form.

The PhoCl green form was solved to 2.10 Å with a R_{Work} factor of 0.1824 and R_{Free} factor of 0.2162 in $P2_12_12_1$ spacegroup. The Photoconversion of the PhoCl green crystal induced the packing of the crystal and the PhoCl red model was refined to 2.30 Å with a R_{Work} factor of 0.2186 and R_{Free} factor of 0.2638 in P1 spacegroup. The PhoCl empty barrel was determined to 2.82 Å in $P2_1$ spacegroup with anisotropical diffraction; the dataset was processed with the Diffraction Anisotropy Server[231] and refined with a R_{Work} factor of 0.2587 and R_{Free} factor of 0.3029.

All the structures showed favorable stereochemistry and exhibited good distribution of dihedral angles on a Ramachandran plot. Detailed refinement statistics are summarized in **Table A.1**. All the structure figures were generated using the molecular visualization PyMOL package[232].

Table A.1 X-ray data collection and refinement statistics.

Crystal	**PhoCl1 green state**	**PhoCl1 red state**	**PhoCl1 emptyBarrel**
Data collection			
Spacegroup	$P2_12_12_1$	P1	$P2_1$
a, b, c (Å)	60.36, 112.76, 144.85	38.93, 72.49, 126.55	46.4, 119.6, 65.3
α, β, γ (°)	90.0, 90.0, 90	92.5, 97.3, 92.5	90, 107.7, 90
Resolution (Å)	42.88-2.10 (2.18-2.10)	38.57-2.30 (2.38-2.30)	43.11-2.82 (2.92-2.82)
*R*merge	0.066 (0.709)	0.042 (0.533)	0.073 (0.760)
*R*meas	0.078 (0.835)	0.059 (0.734)	0.089 (0.936)
Multiplicity	3.4 (3.5)	1.8 (1.7)	2.7 (2.7)
CC(1/2)	0.997 (0.885)	0.998 (0.855)	0.997 (0.811)
CC*	0.999 (0.969)	1 (0.96)	0.999 (0.946)
I/σ(I)	9.73 (1.72)	9.60 (1.20)	8.54 (1.28)
Completeness (%)	96.8 (97.7)	95.9 (94.9)	92.9 (94.5)
Wilson B-factor (Å^2)	25.64	28.86	43.91
Refinement			
Total Reflections	194926 (31020)	103548 (16288)	44693 (7133)
Unique Reflections	57043 (9153)	58440 (9388)	16400 (2688)
R_{Work}/R_{Free}	0.1824/0.2162	0.2186/0.2638	0.2587/0.3029
Number of atoms:			
Protein	6715	10075	5031
Ligands	100	148	-
Water	410	488	-
Average B-factor (Å^2)	34.07	35.53	38.70
Protein ADP (Å^2)	34.35	35.59	38.71
Ligands (Å^2)	22.13	29.24	-
Water	32.51	31.77	-

Ramachandran plot:			
Favored/Allowed (%)	97.70/2.20	94.0/5.4	91.0/8.7
Root-Mean-Square-Deviation:			
Bond lengths (Å)	0.012	0.004	0.005
Bond Angle (°)	1.50	1.0	0.82

Note: Statistics for the highest resolution shell are shown in parentheses.

Bibliography

1. Kim, C. K., Adhikari, A. & Deisseroth, K. Integration of optogenetics with complementary methodologies in systems neuroscience. *Nat. Rev. Neurosci.* **18**, 222–235 (2017).
2. Fenno, L., Yizhar, O. & Deisseroth, K. The Development and Application of Optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
3. Deisseroth, K. *et al.* Next-generation optical technologies for illuminating genetically targeted brain circuits. *J. Neurosci.* **26**, 10380–10386 (2006).
4. Liu, Q. & Tucker, C. L. Engineering genetically-encoded tools for optogenetic control of protein activity. *Curr. Opin. Chem. Biol.* **40**, 17–23 (2017).
5. Schroll, C. *et al.* Light-Induced Activation of Distinct Modulatory Neurons Triggers Appetitive or Aversive Learning in Drosophila Larvae. *Curr. Biol.* **16**, 1741–1747 (2006).
6. Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. & De Lecea, L. Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* **450**, 420–424 (2007).
7. Witten, I. B. *et al.* Cholinergic interneurons control local circuit activity and cocaine conditioning. *Science.* **330**, 1677–1681 (2010).
8. Proville, R. D. *et al.* Cerebellum involvement in cortical sensorimotor circuits for the control of voluntary movements. *Nat. Neurosci.* **17**, 1233–1239 (2014).
9. Tye, K. M. & Deisseroth, K. Optogenetic investigation of neural circuits

underlying brain disease in animal models. *Nat. Rev. Neurosci.* **13**, 251–266 (2012).

10. Steinbeck, J. A. *et al.* Optogenetics enables functional analysis of human embryonic stem cell-derived grafts in a Parkinson's disease model. *Nat. Biotechnol.* **33**, 204–209 (2015).

11. Yamamoto, K. *et al.* Chronic Optogenetic Activation Augments Aβ Pathology in a Mouse Model of Alzheimer Disease. *Cell Rep.* **11**, 859–865 (2015).

12. Dugué, G. P., Akemann, W. & Knöpfel, T. A comprehensive concept of optogenetics. *Prog. Brain Res.* **196**, 1–28 (2012).

13. Guru, A., Post, R. J., Ho, Y. Y. & Warden, M. R. Making sense of optogenetics. *Int. J. Neuropsychopharmacol.* **18**, 1–8 (2015).

14. Guo, Z. V., Hart, A. C. & Ramanathan, S. Optical interrogation of neural circuits in Caenorhabditis elegans. *Nat. Methods* **6**, 891–896 (2009).

15. Akerboom, J. *et al.* Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front. Mol. Neurosci.* **6**, 1–29 (2013).

16. Packer, A. M., Russell, L. E., Dalgleish, H. W. P. & Häusser, M. Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo. *Nat. Methods* **12**, 140–146 (2015).

17. Ding, Y. *et al.* Ratiometric biosensors based on dimerization-dependent fluorescent protein exchange. *Nat. Methods* **12**, 195–198 (2015).

18. Lee, S. *et al.* Reversible protein inactivation by optogenetic trapping in cells.

Nat. Methods **11**, 633–636 (2014).

19. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33**, 755–760 (2015).

20. Zhang, W. Expanding the optogenetic toolkit with a photocleavable protein and a near infrared pH indicator. (University of Alberta, 2017).

21. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13940–13945 (2003).

22. Gradinaru, V., Thompson, K. R. & Deisseroth, K. eNpHR: A Natronomonas halorhodopsin enhanced for optogenetic applications. *Brain Cell Biol.* **36**, 129–139 (2008).

23. Gero Miesenbock, D. A. D. A. & J. E. R. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent protein. *Nature* **394**, 192 (1998).

24. Shen, Y., Rosendale, M., Campbell, R. E. & Perrais, D. pHuji, a pH-sensitive red fluorescent protein for imaging of exo- and endocytosis. *J. Cell Biol.* **207**, 419–432 (2014).

25. Allen, G. J. *et al.* Cameleon calcium indicator reports cytoplasmic calcium dynamics in Arabidopsis guard cells. *Plant J.* **19**, 735–747 (1999).

26. Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).

27. Wu, Y. I. *et al.* A genetically encoded photoactivatable Rac controls the

motility of living cells. *Nature* **461**, 104–108 (2009).

28. Bayley, P. M., Findlay, W. A. & Martin, S. R. Target recognition by calmodulin: Dissecting the kinetics and affinity of interaction using short peptide sequences. *Protein Sci.* **5**, 1215–1228 (1996).

29. Marvin, J. S. *et al.* An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat. Methods* **10**, 162–170 (2013).

30. Sato, M., Ozawa, T., Inukai, K., Asano, T. & Umezawa, Y. Fluorescent indicators for imaging protein phosphorylation in single living cells. *Nat. Biotechnol.* **20**, 287–294 (2002).

31. Sun, F. *et al.* A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. *Cell* **174**, 481-496.e19 (2018).

32. Lundby, A., Mutoh, H., Dimitrov, D., Akemann, W. & Knöpfel, T. Engineering of a genetically encodable fluorescent voltage sensor exploiting fast Ci-VSP voltage-sensing movements. *PLoS One* **3**, 1–5 (2008).

33. Kralj, J. M., Hochbaum, D. R., Douglass, A. D. & Cohen, A. E. Electrical spiking in Escherichia coli probed with a fluorescent voltage-indicating protein. *Science.* **333**, 345–348 (2011).

34. Zhao, Y. *et al.* An expanded palette of genetically encoded Ca2+ indicators. *Science.* **333**, 1888–1891 (2011).

35. Suzuki, K. *et al.* Five colour variants of bright luminescent protein for real-time multicolour bioimaging. *Nat. Commun.* **7**, 1–10 (2016).

36. Zhang, W. *et al.* Optogenetic control with a photocleavable protein, Phocl.

Nat. Methods **14**, 391–394 (2017).

37. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nat. Neurosci.* **18**, 1213–1225 (2015).
38. Peter, E., Dick, B. & Baeurle, S. A. Mechanism of signal transduction of the LOV2-Jα photosensor from Avena sativa. *Nat. Commun.* **1**, (2010).
39. Lungu, O. I. *et al.* Designing Photoswitchable Peptides Using the AsLOV2 Domain. *Chem. Biol.* **19**, 507–517 (2012).
40. Liu, H. *et al.* Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis. *Science.* **322**, 1535–1539 (2008).
41. Levskaya, A., Weiner, O. D., Lim, W. A. & Voigt, C. A. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997–1001 (2009).
42. Shadish, J. A., Strange, A. C. & Deforest, C. A. Genetically Encoded Photocleavable Linkers for Patterned Protein Release from Biomaterials. *J. Am. Chem. Soc.* **141**, 15619–15625 (2019).
43. Xiang, D. *et al.* Hydrogels With Tunable Mechanical Properties Based on Photocleavable Proteins. *Front. Chem.* **8**, 1–9 (2020).
44. Huala, E. *et al.* Arabidopsis NPH1: A protein kinase with a putative redox-sensing domain. *Science.* **278**, 2120–2123 (1997).
45. Salomon, M., Christie, J. M., Knieb, E., Lempert, U. & Briggs, W. R. Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin. *Biochemistry* **39**, 9401–9410 (2000).

46. Crosson, S., Rajagopal, S. & Moffat, K. The LOV domain family: Photoresponsive signaling modules coupled to diverse output domains. *Biochemistry* **42**, 2–10 (2003).

47. Taylor, B. L. & Zhulin, I. B. PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. *Microbiol. Mol. Biol. Rev.* **63**, 479–506 (1999).

48. Swartz, T. E. *et al.* The Photocycle of a Flavin-binding Domain of the Blue Light Photoreceptor Phototropin. *J. Biol. Chem.* **276**, 36493–36500 (2001).

49. Halavaty, A. S. & Moffat, K. N- and C-terminal flanking regions modulate light-induced signal transduction in the LOV2 domain of the blue light sensor phototropin 1 from Avena sativa. *Biochemistry* **46**, 14001–14009 (2007).

50. Zoltowski, B. D. *et al.* Conformational switching in the fungal light sensor vivid. *Science.* **316**, 1054–1057 (2007).

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

52. Pudasaini, A., El-Arab, K. K. & Zoltowski, B. D. LOV-based optogenetic devices: Light-driven modules to impart photoregulated control of cellular signaling. *Front. Mol. Biosci.* **2**, 1–15 (2015).

53. Wang, H. *et al.* LOVTRAP: An optogenetic system for photoinduced protein dissociation. *Nat. Methods* **13**, 755–758 (2016).

54. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat. Commun.* **6**, (2015).

55. Liscum, E., Hodgson, D. W. & Campbell, T. J. Blue Light Signaling through the Cryptochromes and Phototropins. So That's What the Blues Is All about. *Plant Physiol.* **133**, 1429–1436 (2003).

56. Kasahara, M. *et al.* Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from Arabidopsis, rice, and Chlamydomonas reinhardtii. *Plant Physiol.* **129**, 762–773 (2002).

57. Möglich, A. & Moffat, K. Structural Basis for Light-dependent Signaling in the Dimeric LOV Domain of the Photosensor YtvA. *J. Mol. Biol.* **373**, 112–126 (2007).

58. Schwerdtfeger, C. & Linden, H. VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation. *EMBO J.* **22**, 4846–4855 (2003).

59. Imaizumi, T., Tran, H. G., Swartz, T. E., Briggs, W. R. & Kay, S. A. FKF1 is essential for photoperiodic-specific light signalling in Arabidopsis. *Nature* **426**, 302–306 (2003).

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

61. Nash, A. I., Ko, W. H., Harper, S. M. & Gardner, K. H. A conserved glutamine plays a central role in LOV domain signal transmission and Its duration. *Biochemistry* **47**, 13842–13849 (2008).

62. Song, S. H. *et al.* Modulating LOV domain photodynamics with a residue

alteration outside the chromophore binding site. *Biochemistry* **50**, 2411–2423 (2011).

63. Chan, R. H. & Bogomolni, R. A. Structural water cluster as a possible proton acceptor in the adduct decay reaction of oat phototropin 1 LOV2 domain. *J. Phys. Chem. B* **116**, 10609–10616 (2012).

64. Freddolino, P. L., Gardner, K. H. & Schulten, K. Signaling mechanisms of LOV domains: New insights from molecular dynamics studies. *Photochem. Photobiol. Sci.* **12**, 1158–1170 (2013).

65. Zayner, J. P., Antoniou, C. & Sosnick, T. R. The amino-terminal helix modulates light-activated conformational changes in AsLOV2. *J. Mol. Biol.* **419**, 61–74 (2012).

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

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

68. Zayner, J. P., Antoniou, C., French, A. R., Hause, R. J. & Sosnick, T. R. Investigating models of protein function and allostery with a widespread mutational analysis of a light-activated protein. *Biophys. J.* **105**, 1027–1036 (2013).

69. Möglich, A., Yang, X., Ayers, R. A. & Moffat, K. Structure and Function of Plant Photoreceptors. *Annu. Rev. Plant Biol.* **61**, 21–47 (2010).

70. Iuliano, J. N. *et al.* Unraveling the Mechanism of a LOV Domain

Optogenetic Sensor: A Glutamine Lever Induces Unfolding of the Jα Helix. *bioRxiv* 2020.01.29.925040 (2020). doi:10.1101/2020.01.29.925040

71. Christie, J. M. *et al.* Steric interactions stabilize the signaling state of the LOV2 domain of phototropin 1. *Biochemistry* **46**, 9310–9319 (2007).

72. Harper, S. M., Neil, L. C. & Gardner, K. H. Structural basis of a phototropin light switch. *Science.* **301**, 1541–1544 (2003).

73. Harper, S. M., Christie, J. M. & Gardner, K. H. Disruption of the LOV-Jα helix interaction activates phototropin kinase activity. *Biochemistry* **43**, 16184–16192 (2004).

74. Strickland, D. *et al.* Rationally improving LOV domainâ€"based photoswitches. *Nat. Methods* **7**, 623–626 (2010).

75. Strickland, D. *et al.* TULIPs: Tunable, light-controlled interacting protein tags for cell biology. *Nat. Methods* **9**, 379–384 (2012).

76. Yao, X., Rosen, M. K. & Gardner, K. H. Estimation of the available free energy in a LOV2-Jα photoswitch. *Nat. Chem. Biol.* **4**, 491–497 (2008).

77. Stone, O. J. *et al.* Optogenetic control of cofilin and αTAT in living cells using Z-lock. *Nat. Chem. Biol.* **15**, 1183–1190 (2019).

78. Zoltowski, B. D. & Crane, B. R. Light activation of the LOV protein vivid generates a rapidly exchanging dimer. *Biochemistry* **47**, 7012–7019 (2008).

79. Vaidya, A. T., Chen, C. H., Dunlap, J. C., Loros, J. J. & Crane, B. R. Structure of a light-activated LOV protein dimer that regulates transcription. *Sci. Signal.* **4**, 1–8 (2011).

80. Lamb, J. S. *et al.* Illuminating Solution Responses of a LOV Domain Protein

with Photocoupled Small-Angle X-Ray Scattering. *J. Mol. Biol.* **393**, 909–919 (2009).

81. Wang, X., Chen, X. & Yang, Y. Spatiotemporal control of gene expression by a light-switchable transgene system. *Nat. Methods* **9**, 266–269 (2012).

82. Kawano, F., Okazaki, R., Yazawa, M. & Sato, M. A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nat. Chem. Biol.* **12**, 1059–1064 (2016).

83. Baumschlager, A., Aoki, S. K. & Khammash, M. Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control. *ACS Synth. Biol.* **6**, 2157–2167 (2017).

84. Che, D. L., Duan, L., Zhang, K. & Cui, B. The Dual Characteristics of Light-Induced Cryptochrome 2, Homo-oligomerization and Heterodimerization, for Optogenetic Manipulation in Mammalian Cells. *ACS Synth. Biol.* **4**, 1124–1135 (2015).

85. Kennedy, M. J. *et al.* Rapid blue-light-mediated induction of protein interactions in living cells. *Nat. Methods* **7**, 973–975 (2010).

86. Taslimi, A. *et al.* Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nat. Chem. Biol.* **12**, 425–430 (2016).

87. Zhang, K. *et al.* Light-mediated kinetic control reveals the temporal effect of the Raf/MEK/ERK pathway in PC12 cell neurite outgrowth. *PLoS One* **9**, (2014).

88. Katsura, Y. *et al.* An optogenetic system for interrogating the temporal

dynamics of Akt. *Sci. Rep.* **5**, 1–10 (2015).

89. Duan, L. *et al.* Optogenetic control of molecular motors and organelle distributions in cells. *Chem. Biol.* **22**, 671–682 (2015).

90. Engelhard, C. *et al.* Cellular metabolites enhance the light sensitivity of arabidopsis cryptochrome through alternate electron transfer pathwaysc. *Plant Cell* **26**, 4519–4531 (2014).

91. Herbel, V. *et al.* Lifetimes of Arabidopsis cryptochrome signaling states in vivo. *Plant J.* **74**, 583–592 (2013).

92. Chang, K. Y. *et al.* Light-inducible receptor tyrosine kinases that regulate neurotrophin signalling. *Nat. Commun.* **5**, (2014).

93. Bugaj, L. J., Choksi, A. T., Mesuda, C. K., Kane, R. S. & Schaffer, D. V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nat. Methods* **10**, 249–252 (2013).

94. Taslimi, A. *et al.* An optimized optogenetic clustering tool for probing protein interaction and function. *Nat. Commun.* **5**, (2014).

95. Park, H. *et al.* Optogenetic protein clustering through fluorescent protein tagging and extension of CRY2. *Nat. Commun.* **8**, 30 (2017).

96. Kyung, T. *et al.* Optogenetic control of endogenous Ca 2+ channels in vivo. *Nat. Biotechnol.* **33**, 1092–1096 (2015).

97. Wend, S. *et al.* Optogenetic control of protein kinase activity in mammalian cells. *ACS Synth. Biol.* **3**, 280–285 (2014).

98. Duan, L. *et al.* Understanding CRY2 interactions for optical control of intracellular signaling. *Nat. Commun.* **8**, 4–13 (2017).

99. Rockwell, N. C., Su, Y.-S. & Lagarias, J. C. Phytochrome Structure and Signaling Mechanisms. *Annu. Rev. Plant Biol.* **57**, 837–858 (2006).

100. Sharrock, R. A. The phytochrome red/far-red photoreceptor superfamily. *Genome Biol.* **9**, (2008).

101. Sweere, U. *et al.* Interaction of the response regulator ARR4 with phytochrome B in modulating red light signaling. *Science.* **294**, 1108–1111 (2001).

102. Smith, A. M., Mancini, M. C. & Nie, S. Bioimaging: Second window for in vivo imaging. *Nat. Nanotechnol.* **4**, 710–711 (2009).

103. Kaberniuk, A., Shemetov, A. A. & Verkhusha, V. V. A bacterial phytochrome-based optogenetic system controllable with near-infrared light. *Nat. Methods* **13**, 591–597 (2016).

104. Redchuk, T. A., Omelina, E. S., Chernov, K. G. & Verkhusha, V. V. Near-infrared optogenetic pair for protein regulation and spectral multiplexing. *Nat. Chem. Biol.* **13**, 633–639 (2017).

105. Filonov, G. S. & Verkhusha, V. V. A near-infrared bifc reporter for in vivo imaging of protein-protein interactions. *Chem. Biol.* **20**, 1078–1086 (2013).

106. Shcherbakova, D. M. *et al.* Bright monomeric near-infrared fluorescent proteins as tags and biosensors for multiscale imaging. *Nat. Commun.* **7**, (2016).

107. Qian, Y. *et al.* A genetically encoded near-infrared fluorescent calcium ion indicator. *Nat. Methods* **16**, (2019).

108. Qian, Y. *et al.* Improved genetically encoded near-infrared fluorescent

calcium ion indicators for in vivo imaging. *bioRxiv* 2020.04.08.032433 (2020). doi:10.1101/2020.04.08.032433

109. Anders, K. & Essen, L. O. The family of phytochrome-like photoreceptors: Diverse, complex and multi-colored, but very useful. *Curr. Opin. Struct. Biol.* **35**, 7–16 (2015).

110. Burgie, E. S. & Vierstra, R. D. Phytochromes: An atomic perspective on photoactivation and signaling. *Plant Cell* **26**, 568–4583 (2014).

111. Takala, H. *et al.* Signal amplification and transduction in phytochrome photosensors. *Nature* **509**, 245–248 (2014).

112. Nagano, S. From photon to signal in phytochromes: similarities and differences between prokaryotic and plant phytochromes. *J. Plant Res.* **129**, 123–135 (2016).

113. Dasgupta, J., Frontiera, R. R., Taylor, K. C., Lagarias, J. C. & Mathies, R. A. Ultrafast excited-state isomerization in phytochrome revealed by femtosecond stimulated Raman spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1784–1789 (2009).

114. Beyer, H. M. *et al.* Red Light-Regulated Reversible Nuclear Localization of Proteins in Mammalian Cells and Zebrafish. *ACS Synth. Biol.* **4**, 951–958 (2015).

115. Buckley, C. E. *et al.* Reversible Optogenetic Control of Subcellular Protein Localization in a Live Vertebrate Embryo. *Dev. Cell* **36**, 117–126 (2016).

116. Müller, K. *et al.* A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. *Nucleic Acids Res.* **41**, (2013).

117. Toettcher, J. E., Weiner, O. D. & Lim, W. A. Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. *Cell* **155**, 1422–1434 (2013).

118. Hochrein, L., Mitchell, L. A., Schulz, K., Messerschmidt, K. & Mueller-Roeber, B. L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. *Nat. Commun.* **9**, 1–10 (2018).

119. Uda, Y. *et al.* Efficient synthesis of phycocyanobilin in mammalian cells for optogenetic control of cell signaling. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 11962–11967 (2017).

120. Gambetta, G. A. & Lagarias, J. C. Genetic engineering of phytochrome biosynthesis in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10566–10571 (2001).

121. Mukougawa, K., Kanamoto, H., Kobayashi, T., Yokota, A. & Kohchi, T. Metabolic engineering to produce phytochromes with phytochromobilin, phycocyanobilin, or phycoerythrobilin chromophore in Escherichia coli. *FEBS Lett.* **580**, 1333–1338 (2006).

122. Müller, K. *et al.* Synthesis of phycocyanobilin in mammalian cells. *Chem. Commun.* **49**, 8970–8972 (2013).

123. Burgie, E. S., Bussell, A. N., Walker, J. M., Dubiel, K. & Vierstra, R. D. Crystal structure of the photosensing module from a red/far-red light-absorbing plant phytochrome. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10179–10184 (2014).

124. Rottwinkel, G., Oberpichler, I. & Lamparter, T. Bathy phytochromes in

rhizobial soil bacteria. *J. Bacteriol.* **192**, 5124–5133 (2010).

125. Bhoo, S. H., Davis, S. J., Walker, J., Karniol, B. & Vierstra, R. D. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature* **414**, 776–779 (2001).

126. Seyfried, H., Klicpera, M., Leithner, C. & Penner, E. Bilirubinstoffwechsel; I. Physiologie. *Wien. Klin. Wochenschr.* **88**, 477–482 (1976).

127. Shcherbakova, D. M. & Verkhusha, V. V. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nat. Methods* **10**, 751–754 (2013).

128. Papiz, M. Z., Bellini, D., Evans, K., Grossmann, J. G. & Fordham-Skelton, T. Light-induced complex formation of bacteriophytochrome RpBphP1 and gene repressor RpPpsR2 probed by SAXS. *FEBS J.* **286**, 4261–4277 (2019).

129. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan,Aequorea. *J. Cell. Comp. Physiol.* **59**, 223–239 (1962).

130. Morise, H., Shimomura, O., Johnson, F. H. & Winant, J. Intermolecular energy transfer in the bioluminescent system of aequorea. *Biochemistry* **13**, 2656–2662 (1974).

131. Shimomura, O. Structure of the chromophore of Aequorea green fluorescent protein. *FEBS Lett.* **104**, 220–222 (1979).

132. Kanda, T., Sullivan, K. F. & Wahl, G. M. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.* **8**, 377–385 (1998).

133. Lippincott-Schwartz, J., Snapp, E. & Kemvorthy, A. Studying protein

dynamics in living cells. *Nat. Rev. Mol. Cell Biol.* **2**, 444–456 (2001).

134. Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).

135. Rosenow, M. A., Huffman, H. A., Phail, M. E. & Wachter, R. M. The Crystal Structure of the Y66L Variant of Green Fluorescent Protein Supports a Cyclization-Oxidation-Dehydration Mechanism for Chromophore Maturation. *Biochemistry* **43**, 4464–4472 (2004).

136. Adam, V., Berardozzi, R., Byrdin, M. & Bourgeois, D. Phototransformable fluorescent proteins: Future challenges. *Curr. Opin. Chem. Biol.* **20**, 92–102 (2014).

137. Patterson, G. H. & Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science.* **297**, 1873–1877 (2002).

138. Subach, F. V. *et al.* Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat. Methods* **6**, 153–159 (2009).

139. Verkhusha, V. V. & Sorkin, A. Conversion of the monomeric red fluorescent protein into a photoactivatable probe. *Chem. Biol.* **12**, 279–285 (2005).

140. Van Thor, J. J., Gensch, T., Hellingwerf, K. J. & Johnson, L. N. Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222. *Nat. Struct. Biol.* **9**, 37–41 (2002).

141. Bell, A. F., Stoner-Ma, D., Wachter, R. M. & Tonge, P. J. Light-driven decarboxylation of wild-type green fluorescent protein. *J. Am. Chem. Soc.* **125**, 6919–6926 (2003).

142. Nienhaus, K., Nienhaus, G. U., Wiedenmann, J. & Nar, H. Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP. *Proc. Natl. Acad. Sci.* **102**, 9156–9159 (2005).

143. Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H. & Miyawaki, A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12651–12656 (2002).

144. Gurskaya, N. G. *et al.* Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat. Biotechnol.* **24**, 461–465 (2006).

145. Mizuno, H. *et al.* Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein. *Mol. Cell* **12**, 1051–1058 (2003).

146. Tsutsui, H. *et al.* The E1 Mechanism in Photo-Induced β-Elimination Reactions for Green-to-Red Conversion of Fluorescent Proteins. *Chem. Biol.* **16**, 1140–1147 (2009).

147. Ando, R., Mizuno, H. & Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science.* **306**, 1370–1373 (2004).

148. Brakemann, T. *et al.* A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. *Nat. Biotechnol.* **29**, 942–950 (2011).

149. Chudakov, D. M. *et al.* Kindling fluorescent proteins for precise in vivo photolabeling. *Nat. Biotechnol.* **21**, 191–194 (2003).

150. Andresen, M. *et al.* Structure and mechanism of the reversible photoswitch

of a fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13070–13074 (2005).

151. Andresen, M. *et al.* Structural basis for reversible photoswitching in Dronpa. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13005–13009 (2007).

152. Mizuno, H. *et al.* Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9227–9232 (2008).

153. Adam, V. *et al.* Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18343–18348 (2008).

154. Zhou, X. X., Chung, H. K., Lam, A. J. & Lin, M. Z. Optical Control of Protein Activity by Fluorescent Protein Domains. *Science.* **338**, 810–814 (2012).

155. Mizuno, H. *et al.* Higher resolution in localization microscopy by slower switching of a photochromic protein. *Photochem. Photobiol. Sci.* **9**, 239–248 (2010).

156. Zhou, X. X., Fan, L. Z., Li, P., Shen, K. & Lin, M. Z. Optical control of cell signaling by single-chain photoswitchable kinases. *Science.* **355**, 836–842 (2017).

157. Rost, B. R., Schneider-Warme, F., Schmitz, D. & Hegemann, P. Optogenetic Tools for Subcellular Applications in Neuroscience. *Neuron* **96**, 572–603 (2017).

158. McEvoy, A. L. *et al.* mMaple: A Photoconvertible Fluorescent Protein for

Use in Multiple Imaging Modalities. *PLoS One* **7**, (2012).

159. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3**, 769–772 (2007).

160. Lee, H. M., Larson, D. R. & Lawrence, D. S. Illuminating the chemistry of life: Design, synthesis, and applications of 'caged' and related photoresponsive compounds. *ACS Chem. Biol.* **4**, 409–427 (2009).

161. Klán, P. *et al.* Photoremovable protecting groups in chemistry and biology: Reaction mechanisms and efficacy. *Chem. Rev.* **113**, 119–191 (2013).

162. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).

163. Dixon, A. S. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **11**, 400–408 (2016).

164. Boulware, K. T., Jabaiah, A. & Daugherty, P. S. Evolutionary optimization of peptide substrates for proteases that exhibit rapid hydrolysis kinetics. *Biotechnol. Bioeng.* **106**, 339–346 (2010).

165. Kostallas, G., Löfdahl, P. Å. & Samuelson, P. Substrate profiling of tobacco Etch virus protease using a novel Fluorescence-Assisted whole-cell assay. *PLoS One* **6**, (2011).

166. Henderson, B. R. & Eleftheriou, A. A comparison of the activity, sequence specificity, and CRM1-dependence of different nuclear export signals. *Exp.*

Cell Res. **256**, 213–224 (2000).

167. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.* **5**, (2014).

168. Choy, E. *et al.* Endomembrane trafficking of ras: The CAAX motif targets proteins to the ER and Golgi. *Cell* **98**, 69–80 (1999).

169. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563–564 (1988).

170. Picard, D. Posttranslational regulation of proteins by fusions to steroid-binding domains. *Methods Enzymol.* **327**, 385–401 (2000).

171. Kügler, J. *et al.* High affinity peptide inhibitors of the hepatitis C virus NS3-4A protease refractory to common resistant mutants. *J. Biol. Chem.* **287**, 39224–39232 (2012).

172. Hughes, R. M. *et al.* Optogenetic Apoptosis: Light-Triggered Cell Death. *Angew. Chemie - Int. Ed.* **54**, 12064–12068 (2015).

173. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909 (2005).

174. Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **373**, 663–664 (1995).

175. Xu, Y., Piston, D. W. & Johnson, C. H. A bioluminescence resonance energy transfer (BRET) system: Application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 151–156 (1999).

176. Campbell, R. E. Fluorescent-Protein-Based Biosensors: Modulation of

Energy Transfer as a Design Principle. *Anal. Chem.* **81**, 5972–5979 (2009).

177. Esposti, M. D. The roles of Bid. *Apoptosis* **7**, 433–440 (2002).
178. Billen, L. P., Shamas-Din, A. & Andrews, D. W. Bid: A Bax-like BH3 protein. *Oncogene* **27**, S93–S104 (2008).
179. Slee, E. A. *et al.* Benzyloxycarbonyl-Val-Ala-Asp (OMe) fluoromethylketone (Z-VAD.FMK) inhibits apoptosis by blocking the processing of CPP32. *Biochem. J.* **315**, 21–24 (1996).
180. Chu, J. *et al.* Non-invasive intravital imaging of cellular differentiation with a bright red-excitable fluorescent protein. *Nat. Methods* **11**, 572–578 (2014).
181. Idevall-Hagren, O., Dickson, E. J., Hille, B., Toomre, D. K. & De Camilli, P. Optogenetic control of phosphoinositide metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).
182. Gross, L. A., Baird, G. S., Hoffman, R. C., Baldridge, K. K. & Tsien, R. Y. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11990–11995 (2000).
183. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499–509 (1984).
184. Diekmann, Y. & Pereira-Leal, J. B. Evolution of intracellular compartmentalization. *Biochem. J.* **449**, 319–331 (2012).
185. Martin, W. Evolutionary origins of metabolic compartmentalization in eukaryotes. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 847–855 (2010).
186. Schrader, M., Godinho, L. F., Costello, J. L. & Islinger, M. The different

facets of organelle interplay-An overview of organelle interactions. *Front. Cell Dev. Biol.* **3**, 1–22 (2015).

187. Cornejo, E., Abreu, N. & Komeili, A. Compartmentalization and organelle formation in bacteria. *Curr. Opin. Cell Biol.* **26**, 132–138 (2014).

188. Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. *PLoS Comput. Biol.* **10**, (2014).

189. Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse Bacterial Microcompartment Organelles. *Microbiol. Mol. Biol. Rev.* **78**, 438–468 (2014).

190. Pfeifer, F. Distribution, formation and regulation of gas vesicles. *Nat. Rev. Microbiol.* **10**, 705–715 (2012).

191. Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A. & Polka, J. K. The Bacterial Carbon-Fixing Organelle Is Formed by Shell Envelopment of Preassembled Cargo. *PLoS One* **8**, 1–13 (2013).

192. Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat. Microbiol.* **2**, 1–11 (2017).

193. Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939–947 (2008).

194. McHugh, C. A. *et al.* A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J.* **33**, 1896–1911 (2014).

195. Valdés-Stauber, N. & Scherer, S. Isolation and characterization of Linocin

M18, a bacteriocin produced by Brevibacterium linens. *Appl. Environ. Microbiol.* **60**, 3809–3814 (1994).

196. Rosenkrands, I. *et al.* Identification and characterization of a 29-kilodalton protein from Mycobacterium tuberculosis culture filtrate recognized by mouse memory effector cells. *Infect. Immun.* **66**, 2728–2735 (1998).

197. Hicks, P. M., Rinker, K. D., Baker, J. R. & Kelly, R. M. Homomultimeric protease in the hyperthermophilic bacterium Thermotoga maritima has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria. *FEBS Lett.* **440**, 393–398 (1998).

198. Johnson, J. E. & Speir, J. A. Quasi-equivalent viruses: A paradigm for protein assemblies. *J. Mol. Biol.* **269**, 665–675 (1997).

199. Rahmanpour, R. & Bugg, T. D. H. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS J.* **280**, 2097–2104 (2013).

200. Corchero, J. L. & Cedano, J. Self-assembling, protein-based intracellular bacterial organelles: Emerging vehicles for encapsulating, targeting and delivering therapeutical cargoes. *Microb. Cell Fact.* **10**, 1–8 (2011).

201. Cassidy-Amstutz, C. *et al.* Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **55**, 3461–3468 (2016).

202. Moon, H., Lee, J., Min, J. & Kang, S. Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplatform. *Biomacromolecules* **15**, 3794–3801 (2014).

203. Lau, Y. H., Giessen, T. W., Altenburg, W. J. & Silver, P. A. Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nat. Commun.* **9**, (2018).

204. Sigmund, F. *et al.* Bacterial encapsulins as orthogonal compartments for mammalian cell engineering. *Nat. Commun.* **9**, (2018).

205. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).

206. Baird, G. S., Zacharias, D. A. & Tsien, R. Y. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11984–11989 (2000).

207. Yanushevich, Y. G. *et al.* A strategy for the generation of non-aggregating mutants of Anthozoa fluorescent proteins. *FEBS Lett.* **511**, 11–14 (2002).

208. Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6**, 1–8 (2011).

209. Olins, P. O. & Rangwala, S. H. A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in Escherichia coli. *J. Biol. Chem.* **264**, 16973–16976 (1989).

210. Busskamp, V., Picaud, S., Sahel, J. A. & Roska, B. Optogenetic therapy for retinitis pigmentosa. *Gene Ther.* **19**, 169–175 (2012).

211. Toyama, Y., Miyawaki, A., Nakamura, M. & Jinzaki, M. *Make Life Visible*. (Springer Singapore, 2020). doi:10.1007/978-981-13-7908-6

212. Towne, C. & Thompson, K. R. Overview on Research and Clinical

Applications of Optogenetics. *Curr. Protoc. Pharmacol.* **75**, 11.19.1-11.19.21 (2016).

213. El Koussy, M. M. & Jadavji, N. M. Therapeutic Potential of Optogenetic Treatment for Individuals with Multiple Sclerosis. *J. Young Investig.* **33**, 77–82 (2017).

214. Joshi, J., Rubart, M. & Zhu, W. Optogenetics: Background, Methodological Advances and Potential Applications for Cardiovascular Research and Medicine. *Front. Bioeng. Biotechnol.* **7**, (2020).

215. Davidson, M. W. & Campbell, R. E. Engineered fluorescent proteins: Innovations and applications. *Nat. Methods* **6**, 713–717 (2009).

216. Nienhaus, K., Nienhaus, G. U., Wiedenmann, J. & Nar, H. Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9156–9159 (2005).

217. Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3750–3755 (2001).

218. Gold, L. mRNA display: Diversity matters during in vitro selection. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4825–4826 (2001).

219. Takahashi, T. T., Austin, R. J. & Roberts, R. W. mRNA display: Ligand discovery, interaction analysis and beyond. *Trends Biochem. Sci.* **28**, 159–165 (2003).

220. Yamaguchi, J. *et al.* cDNA display: A novel screening method for functional

disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. *Nucleic Acids Res.* **37**, (2009).

221. Mochizuki, Y., Suzuki, T., Fujimoto, K. & Nemoto, N. A versatile puromycin-linker using cnvK for high-throughput in vitro selection by cDNA display. *J. Biotechnol.* **212**, 174–180 (2015).

222. Naimuddin, M. & Kubo, T. A High Performance Platform Based on cDNA Display for Efficient Synthesis of Protein Fusions and Accelerated Directed Evolution. *ACS Comb. Sci.* **18**, 117–129 (2016).

223. Watanabe, T. *et al.* Genetic visualization of protein interactions harnessing liquid phase transitions. *Sci. Rep.* **7**, 1–13 (2017).

224. Zhang, Q. *et al.* Visualizing Dynamics of Cell Signaling In Vivo with a Phase Separation-Based Kinase Reporter. *Mol. Cell* **69**, 334-346.e4 (2018).

225. Tang, L. Optogenetic tools light up phase separation. *Nat. Methods* **16**, 139–139 (2019).

226. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 133–144 (2010).

227. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 32–41 (2006).

228. Ai, H. W., Henderson, J. N., Remington, S. J. & Campbell, R. E. Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: Structural characterization and applications in fluorescence imaging. *Biochem. J.* **400**, 531–540 (2006).

229. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

230. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

231. Strong, M. *et al.* Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8060–8065 (2006).

232. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).