Designing Chemical- and Light-Activated Protein Switches for Regulating Peptide Functions

by

Jiaqi Shen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemistry)
in the University of Michigan
2024

Doctoral Committee:

Professor Wenjing Wang, Chair
Professor Jennifer Bridwell-Rabb
Professor Kristina Hakansson
Professor Anna Mapp

Jiaqi Shen

shenjiaq@umich.edu

ORCID iD: 0000-0002-7435-4106

Dedication

To my parents, Suhua Cui and Guilong Shen.

Acknowledgements

I express my sincere gratitude to all those who have guided and helped me in the past twenty years of the pursuit of knowledge. I used to imagine that I would have many to say and to express at this moment, but words hardly drop onto paper to pull out these emotions.

My first acknowledgement is to my parents. I still clearly remember the scene of my parents sending me to schools with me sitting on the backseat on those two old bikes. I would not forget the efforts they made to provide educational resource for me. I am mostly grateful to the warm family atmosphere and the resilience on life.

I want to acknowledge my high school chemistry teacher, Minghua Zhang, who inspired my early enthusiasm on chemistry. I want to acknowledge the several professors I met during undergraduate study. Dr. Wenjun Lu made great impact on my values towards science. “Modern chemistry rooted from alchemy”, he said, “has always been a subject to pursue the Midas touch and immortality”. Dr. Yuhong Xu, my undergraduate thesis advisor, helped me so much to explore my real research interest. Dr. John Wolfe, the advisor of my research internship, offered me opportunities to conduct research at the University of Michigan and brought me to where I am now.

Thank you to Kayla, or Dr. Kroning as an updated version, my desk buddy, coffee buddy, lab mate, and my good friend in graduate school. I appreciate your patience for my silly jokes and my weird statements. The time we had together in lab has been the diamonds scattering on a beach,

shining under light irradiation. However, I still firmly hold my claim that coffee is soy milk and tea is vegetable soup.

Thank you to Shenyuan, my close friend for almost ten years now. Thank you for all the talks we had and all the suggestions you gave me. Thank you to Yimin. Thank you for the suggestions and help for my job search and the accompany during this time. I am awaiting the more time and new experience we can share.

Thank you to Gwen, my mentee for four years. Teaching you and watching your progress is such an accomplishment for me in graduate school. You are very talented, and I believe you will have a very successful graduate school career.

Thank you to all the Wang lab members. It has been a joy to work in this lab with all of you. I appreciate everyone's tolerance on my absurdness although I do like to see people's reactions when trying to digest my ridiculous statements.

Thank you to my committee, Dr. Mapp, Dr. Hakansson, and Dr. Bridwell-Rabb for your help and suggestions on my research. I would always remember the feeling when you announced me to be a Ph.D. candidate.

And the best advisor, Dr. Wenjing Wang. Thank you for providing me with such a great research experience and the strongest support on my career. I truly had limited knowledge on biology when I rotated in the lab, and it is you that taught me everything hand-in-hand. It has been a fun scientific journey for me, starting from learning what bases pairs and PCR are, and your understanding and patience has always been there. I was an arguably bad presenter and I remember that your comment on one of my candidacy slides was "bad". Five years later, in conferences and job interviews, the comment I got is that I gave a really nice and clear talk. I am now worried that I used up all my luck to have found your lab. I am grateful to the myself five years ago, who sent

a second email contacting you when you thought the first one was a missent email. I will remember the days we sit on the ground and set up cell culture hood together. I will remember the time we chat about science, universe and life. Your success is my highest pride.

What's past is prologue. Thank you to those have come that I got. Thank you to those have gone that I lost. Thank you to all the hardship and struggles. Thank you to the younger myself that brought me here.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Peptides serve important biological functions including neuromodulation, hormonal regulation, cell signaling, protein localization, and enzyme inhibition. The ability to modulate peptide functions with precision is invaluable in biological research. Genetic tools have offered precise control over biological systems with cell-type specificity, and chemogenetic and optogenetic techniques have expanded this control, providing high temporal and spatial resolution through small molecules and light as signaling inputs. While these methods have been extensively applied to regulate protein functions, their application in controlling peptide functions is less explored.

This thesis describes the engineering of chemogenetic and optogenetic protein domains for regulating peptide functions. These domains offer versatile control over various peptides, modulating biological processes through orthogonal signal inputs from small molecules and light. A directed evolution platform for optimizing these domains is also introduced.

For chemogenetic control over peptides, we developed the chemically activated protein domains (CAPs) for controlling the accessibility of both the N- and C-terminal portions of functional peptides. CAPs were developed through directed evolution of an FK506 binding protein (FKBP). By fusing a peptide to one or both CAPs, the peptide's function is blocked until a small molecule displaces them from the FKBP ligand binding site. CAPs are generally applicable to a range of short peptides, including a protease cleavage site (TEVcs), a dimerization-inducing heptapeptide (SsrA), a nuclear localization signal peptide (NLS), and an opioid peptide

(enkephalin), with a chemical dependence up to 156-fold. We show that the CAPs system can be utilized in cell cultures and multiple organs in living animals.

The second light, oxygen, voltage sensing domain from *Avena sativa* phototropin 1 (AsLOV2) has been widely applied to modulate the activity of various peptides by light. However, due to geometry restrictions, AsLOV2 is not applicable for peptides whose functions requires fusion-free N-terminus. We re-engineered AsLOV2 using circular permutation strategy to generate cpLOV. This modification allows modulation of the C-terminal accessibility of functional peptides while leaving the N-terminus unfused. Using the same strategy as CAPs and showcased by TEVcs, functional peptides can be fused to both AsLOV2 and cpLOV tandemly to reduce the basal activity and tune the dynamic range.

To further optimize these chemical- and light-switchable protein domains, we established an efficient yeast surface based directed evolution platform. This platform simultaneously exhibits activation and leakage signals on the same yeast cell, enabling further optimization of CAPs' caging efficiency. The improved CAPs were then applied to regulate three neuropeptides: enkephalin, pituitary adenylate cyclase-activating polypeptide (PACAP), and α-melanocyte-stimulating hormone (α-MSH), showcasing their broad applicability in modulating peptide functions.

Potential future work includes the optimization of the developed switchable protein domains, expanding the scope of using CAPs to modulate other neuropeptides, and development of orthogonal switchable protein domains.

This thesis contributes significantly to the field of peptide function modulation, offering novel chemogenetic and optogenetic tools and methodologies that have profound implications for biological research.

Chapter 1 Genetically Encoded Protein Switches for Modulating Cellular Processes

Chapter 1 is partially adapted from: Shen, J., Zhou, G. and Wang, W., Chemogenetic Tools in Focus: Proximity, Conformation, and Sterics. *Chemistry-Methods.* (under review)

Part of the figures are adapted from the article above and are produced by Guanwei Zhou, a Ph.D. student in the Wenjing Wang lab, University of Michigan.

Genetic tools have revolutionized the ways to investigate biological systems at molecular level. Genetic encoding enables selective expression of the tools in a subset of cell populations to investigate their functional roles in various physiological processes and behavioral studies in animal models. Since biological systems and cellular processes are highly dynamic, it is important to have temporal and spatial control of the genetic tools delivered. Based on the controlling method, genetic tools can be classified into two categories: chemogenetic and optogenetic tools.

Chemogenetic tools enable modulating biological processes through activation or inhibition of specific protein targets by using small molecules. Optogenetic tools utilize light as the user defined signal to modulate these processes. By introducing protein domains that can sense small molecules and light, precise control over cellular activities can be achieved, allowing scientists to further study biological systems.

Here in this chapter, I summarize the designs of chemogenetic and optogenetic tools, as well as the knowledge gap this dissertation aim to bridge in this field. I mainly discuss about chemogenetic tools and introduce some typical optogenetic tool design, both categorized based on

the mechanisms of chemogenetic and optogenetic domains. Genetic tools that incorporate unnatural amino acids into protein domains are not involved here in this thesis.

1.1 Chemogenetic tool design principles

Chemogenetic tools are genetic systems controlled by chemicals. Chemogenetic tools allow precise temporal manipulation of various cellular processes and have been widely applied to interrogate the complex biological system. Since the first chemogenetic tool reported in 1993 using a small molecule to dimerize an engineered T-cell antigen receptor (TCR)[1], many new chemogenetic domains have been developed for regulating cell signaling, protein transportation, protein stability, gene expression, neuronal activity, etc. The wide range of chemogenetic tools share some general working mechanisms: chemically induced and disrupted proximity, chemically induced conformational change and allosteric control, and chemically induced steric unblocking. Here in this section, design strategies on chemogenetic tools will be discussed, categorized by the three different mechanisms above.

1.1.1 Proximity control

Chemically induced proximity (CIP) utilizes small molecules to induce the formation of artificial protein complexes to manipulate the protein of interest (POI) (**Figure 1-1**). CIP provides a powerful mechanism for regulating biological processes, mainly in two ways: 1) locating POIs to functional locations (on switch) or away from functional locations (off switch); and 2) reconstituting split proteins or inducing the interaction of POIs.

Chemically disrupted proximity (CDP), the inverse process of CIP, has also been developed to provide complementary design strategies (**Figure 1-1**). In CIP- and CDP-based chemogenetic tools, the applications and outputs are solely dependent on the POIs' function rather than on the CIP or CDP protein, providing a generalizable chemogenetic tool design strategy. Therefore, CIP and CDP have been widely applied to control cell signaling, protein degradation, protein transportation, and gene expression.

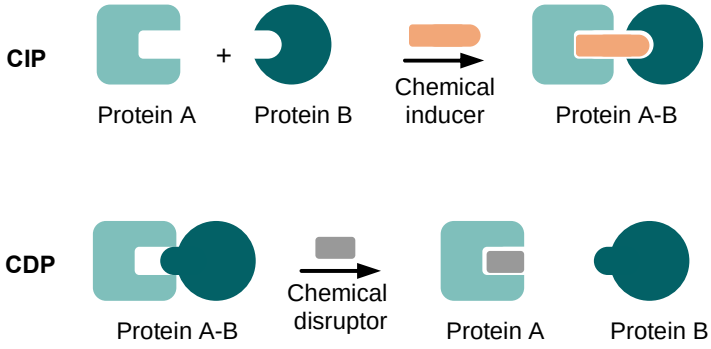


Figure 1-1 CIP system (top) involves a chemical inducer for bringing two proteins in proximity; a CDP system (bottom) involves a chemical disrupter for dissociating a protein pair.

1.1.1.1 Available chemogenetic protein domains for proximity control

The discovery of CIP began with three small molecules: Cyclosporin A (CsA), FK506, and rapamycin. CsA dimerizes CyP and calcineurin[2,3]; FK506 dimerizes FK506 binding protein 12 (FKBP12, here after referred as FKBP) and calcineurin[3,4]; rapamycin dimerizes FKBP and the mammalian target of rapamycin (mTOR)[4,5] (**Table 1-1**). The minimum binding domain of mTOR was further determined as the FKBP-rapamycin binding domain (FRB)[6]. The three small molecules act as "molecular glues", where they bind to the first protein (CsA to CyP, FK506 and rapamycin to FKBP), forming a new binding surface, and further interact with the second one (CsA and FK506 to calcineurin, rapamycin to FRB) (**Figure 1-2**, **Table 1-1**). FKBP is one of the most broadly used domains for chemogenetic tool design. Consequently, several new ligands for

FKBP have been developed. A synthetic ligand for FKBP (SLF) was synthesized to tightly bind to FKBP[7]. To abolish the binding towards endogenous FKBP, the "bump-and-hole" strategy has been applied on FKBP and SLF, so that the new ligand SLF* (with the bump) specifically binds to the FKBP(F36V) mutant (with the hole)[8]. The optimized version of SLF* is called shield-1[9]. The same strategy was applied to FRB and rapamycin, leading to the development of rapamycin derivatives that specifically bind to a triple mutant FRB*(K2095P, T2098L, W2101F)[10,11]. Recently, chemically induced trimerization has also been developed[12] in which rapamycin induces the trimerization of FKBP and two split FRB fragments.

Table 1-1 Summary of molecular glue induced CIP pairs introduced in this section.

Chemical Inducer	Protein A	Protein B
CsA	CyP	Calcineurin
FK506	FKBP	Calcineurin
Rapamycin	FKBP	mTOR, FRB
ABA	Pyl	ABI1
GA3	GID1	GAI
ABT-737	BCL-xL	AZ1
MTX	VHH	NanoCLAMP8
CBD	Nanobody CA-14	Nanobody DB-21
FPP	MBP	AR
Auxin	OsTIR1	IAA17
Danoprevir	NS3a	DNCR
Grazoprevir	NS3a	GNCR

To obtain orthogonal CIP pairs, molecular glues from plant systems were employed. Abscisic acid (ABA) induces Pyl and ABI1 dimerization[13] and gibberellic acid (GA3) induces GID1 and GAI dimerization[14] (**Table 1-1**).

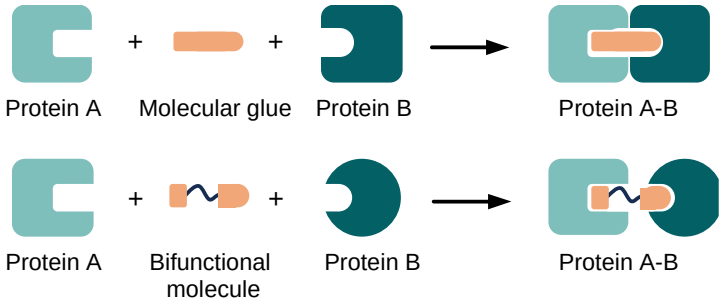


Figure 1-2 CIP inducers can be classified to two types, molecular glues (top) and bifunctional molecules (bottom). Molecular glues bind to protein A first, forming a new binding surface for protein B to interact with. On the other hand, bifunctional molecules compose of two ligands targeting protein A and protein B, respectively, with a linker for connection.

Besides molecular glues, another class of CIP inducer, often referred to as bifunctional molecules, has provided many more opportunities for CIP designs. By connecting two ligands together with a linker, such bifunctional molecules can induce dimerization of two user defined protein domains, providing more CIP pairs (**Figure 1-2**, **Table 1-2**). The very first example was demonstrated by inducing FKBP homodimerization by an FK506 dimer, FK1012[1]. Similarly, a methotrexate (MTX) dimer was synthesized to homodimerize dihydrofolate reductase (DHFR)[15]. Such homodimerizing systems work well when there is only one POI in the system. When applied to dimerize two different POIs (A and B), only part of the dimerization will be heterodimers (A-B), while homodimers (A-A and B-B) will also form. To precisely induce heterodimerization, two distinct ligands could be connected to induce proximity of two different fusion proteins. A heterodimerizing FKCsA generated by connecting FK506 and CsA was used to dimerize FKBP and CyP[16]. Similarly, SLF-TMP dimerizes FKBP and DHFR[17-19]. This approach could also be applied to induce covalent dimerization. SNAP-tag, engineered from human O^6-alkylguanine-DNA alkyltransferase, forms a covalent bond with the O^6-benzylguanine moiety[20,21] (**Table 1-2**). Similarly, HaloTag, which is engineered from a bacterial haloalkane dehalogenase DhaA, forms a covalent bond with the chloroalkane moiety[22] (**Table 1-2**). O^6-benzylguanine connected to MTX[23] dimerizes SNAP-tag and DHFR, and the HaloTag ligand connected to TMP[24] dimerizes HaloTag

and DHFR. SNAP-tag and HaloTag can also be connected to covalently dimerize two fusion proteins[25].

Table 1-2 Summary of common ligand-protein pair for designing bifunctional molecules

Protein	Chemical ligands
FKBP	FK506, SLF
FKBP(F36V)	SLF*, Shield-1
DHFR	MTX, TMP
SNAP-tag	O^6-benzylguanine derivatives
HaloTag	chloroalkane ligand

Unlike with molecular glues, the chemical dosage affects the dimerization efficiency of bifunctional molecules. Excess bifunctional molecules bind to the protein domains separately rather than dimerizing them, whereas molecular-glue-based systems always favor the formation of ternary complexes. Additionally, bifunctional molecules are usually large, perturbing their druggability and performance in animal models. However, bifunctional molecules are easier to design than molecular glues and there are only a few molecular glues available for CIP.

While the first few CIPs were discovered from nature (e.g., FKBP-FRB, Pyl-ABI1, DHFR, HaloTag, SNAP-tag, etc.), new screening and computational design platforms have significantly facilitated the design of new CIPs.

Using advanced library screening and computational design, new molecular glue-based CIP could be developed by designing binders against a known chemical-protein complex. For example, a phage-displaying library encoding the human antibody fragment antigen-binding domain (Fab) was screened against BCL-xL complexed with the small molecule, ABT-737[26]. The resulted Fab, AZ1, binds to the BCL-xL-ABT-737 complex with a K_D of 3 nM and showed no detectable binding in the absence of ABT-737 (**Table 1-1**). In another example, a nano-CLostridial

Antibody Mimetic Proteins (nanoCLAMP) phage display library was screened against the VHH-MTX complex[27]. The resulted nanoCLAMP8 binds to the apo VHH and VHH-MTX complex with a K_D of 3.7 μM and 8.2 nM, respectively (**Table 1-1**).

These two examples rely on a known chemical-protein interaction pair. To de novo engineer a CIP pair, a platform named COMBINES-CID was developed[28]. A nanobody library was first screened against a small molecule of interest, cannabidiol (CBD). Then a second nanobody library was screened against the nanobody-CBD complex. The resulted first nanobody CA-14 binds to CBD with a K_D of 6 μM and the second nanobody DB-21 binds to the CA-14-CBD complex with a K_D of 56 nM (**Table 1-1**).

Computational design has also been employed to aid CIP pair engineering. A small molecule binding site for farnesyl pyrophosphate (FPP) was computationally designed into the binding surfaces of a known protein interaction pair of maltose-binding protein (MBP) and an ankyrin repeat (AR) protein[29] (**Table 1-1**). Using a split murine DHFR reporter system in *Escherichia coli*, FPP induced >100-fold dimerization (K_D changes from >200μM to 2.1 μM in the chemical-bound state). However, this CIP pair has not been tested in mammalian cells. A platform named PROCISiR was also reported[30] to computationally design binders for NS3a protein complexed with different ligands. The resulted CIP, danoprevir/NS3a complex reader (DNCR) and grazoprevir/NS3a complex reader (GNCR), yielded affinities of 36 pM and 140 nM towards the NS3a ligand binding state, respectively, with no detectable affinity towards the apo state (**Table 1-1**). These NS3a-based CIP pairs can be used in mammalian cells, allowing unprecedented multiplexed control with the same ligand binding protein.

CDP is the inversed process of CIP, where small molecule binding dissociates the protein interaction pair, providing complementary strategies for chemogenetic tool design. Several CDPs

have been designed (**Figure 1-1**, **Table 1-3**). For example, self-dimerizing FKBP(F36M) can be disrupted by FKBP ligands[31,32]. The interaction between streptavidin and streptavidin-binding peptide can be disrupted by biotin[33]. The interaction between NS3a and its peptide inhibitor ANR can be disrupted by the NS3a small molecule inhibitors danoprevir, asunaprevir, and grazoprevir[34]. The interaction between BCL-xL and its interaction peptide LD3 could be disrupted by its small molecule ligands A1331852 and A1155463[35].

Table 1-3 Summary of common CDP systems.

Protein A - Protein B	Chemical ligand
FKBP(F36M) - FKBP(F36M)	AP22542, AP21998, SLF
Streptavidin - Streptavidin binding peptide	Biotin
NS3a - ANR	Danoprevir, Asunaprevir, Grazoprevir
BCL-xL - LD3	A1331852, A1155463
NES - FKBP(L106P)	Shield-1

1.1.1.2 Improving spatial temporal resolution of CIPs

Even though the addition of a small molecule provides temporal control, the diffusion of the small molecule to the protein domains is relatively slow. Further, it is challenging to reach subcellular activation. To achieve high spatiotemporal resolution activation in CIP, photocaged small molecules were designed to provide light-dependent gating over the chemogenetic system, thereby improving its spatiotemporal resolution. A typical strategy is to add a bulky, light-degradable chemical moiety to cage the ligand's functionality and prevent it from binding to the protein domain until light illumination removes the bulky group (**Figure 1-3**).

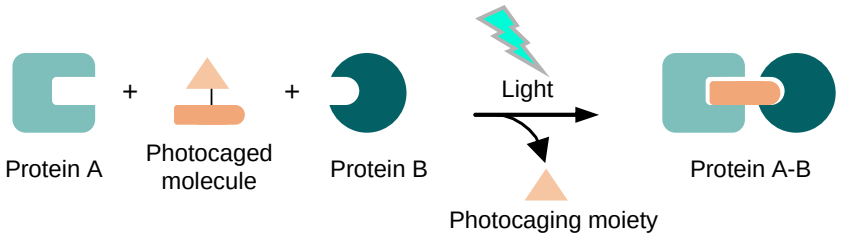


Figure 1-3 A photocaged molecule is composed of a chemical inducer modified by a photocaging moiety. Upon light illumination, the photocaging moiety is cleaved off from the molecule, which then induces the protein A-B dimerization.

A photocaged rapamycin analogue (pRap) was developed by installing a R-methyl-6-nitropiperonyloxycarbonyl group on the C-40 position[36] (**Figure 1-4**). Upon irradiation with 365 nm UV light, pRap is uncaged and induces FKBP-FRB dimerization. pRap has some basal dimerizing activity, which can be abolished through the use of an engineered version of FKBP, iFKBP[37]. Similarly, the rapamycin-biotin conjugate (cRb) was photocaged by connecting the rapamycin C-40 position with biotin using a UV light degradable 4,5-dimethoxy R-methyl nitrobenzyl linker[38] (**Figure 1-4**). Similar to pRap, cRb retains basal dimerizing activity on FKBP and FRB. Such background is eliminated by supplying avidin in cell culture media, thereby retaining cRb in the extracellular media solution by its binding to biotin. Although no modification of the CIP pair is needed, the use of avidin is tedious and the spatial resolution is reduced due to free rapamycin diffusing from the extracellular environment into the cytosol. Another strategy to design a photocaged rapamycin was to connect two rapamycin molecules with a photocleavable linker (dRap, activated by 365 nm light)[39] (**Figure 1-4**). The heterodimerization of FKBP with FRB is blocked until light irradiation cleaves the linker and releases rapamycin.

Photocaged ABA[40] and GA3[41] were synthesized by installing a 4,5-dimethoxy-2-nitrobenzyl group on ABA (ABA-DMNB) and a 2-(o-nitro-phenyl)propyl group on GA3 (pcGA3-3) (**Figure 1-4**). ABA-DMNB can be activated by 365 nm UV light. pcGA3-3 can be activated by

412 nm light. Additionally, two-photon activation at 800 nm can be used to activate pcGA3-3 to avoid phototoxicity and improve tissue penetration of light.

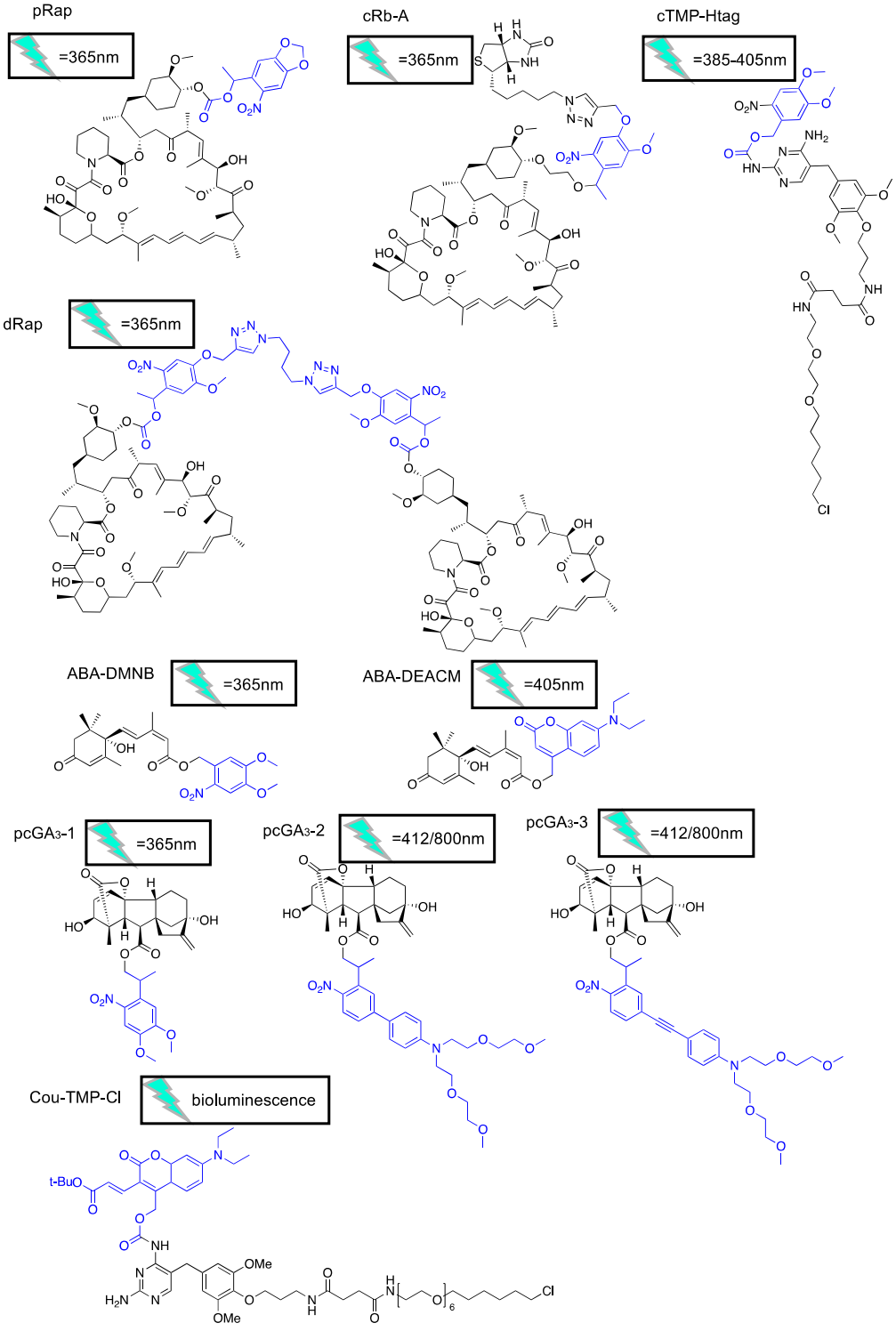


Figure 1-4 Chemical structures of photocaged molecules. The photocaging moieties are highlighted in blue.

The examples discussed above all generate the free form of a small molecule dimerizer after light irradiation, and the diffusion of the small molecules reduces spatial resolution. To address this problem, a photocaged bifunctional dimerizer cTMP-Htag was designed[24] by connecting a photocaged TMP molecule to the HaloTag ligand (**Figure 1-4**). The HaloTag ligand enables anchoring of the cTMP-Htag to selective compartments where HaloTag is localized, such as centromeres, kinetochores, mitochondria, and centrosomes. To avoid phototoxicity from UV light, bioluminescence can be used instead of light irradiation. An engineered HaloTag inserted into a circularly permuted NanoLuc (H-Luc)[42,43] can covalently bind to the HaloTag ligand conjugated with a coumarin-caged TMP (Cou-TMP-Cl) (**Figure 1-4**). The addition of the furimazine substrate uncages the ligand through bioluminescence resonance energy transfer (BRET) and induces dimerization of HaloTag with DHFR.

As discussed above, using light as an extra layer of control could improve the spatiotemporal resolution of CIP-based chemogenetic systems. However, UV light has low tissue penetration and can cause phototoxicity in animal models. The application of a two-photon laser for photo-uncaging could potentially address these issues but has not been tested in vivo.

1.1.1.3 Reversible CIP engineering

Reversible CIP is desirable for investigating dynamic biological processes. However, most CIPs suffer from slow off-kinetics and are usually considered “irreversible one-way on-switches” for the chemogenetic systems. This is due to the strong binding affinity of the small molecule to the protein domain and the high stability of the ternary protein-small-molecule-protein complex. For example, rapamycin binds to FKBP with a K_D of 0.2 nM, and it is predicted that FRB binds to the FKBP-rapamycin with a K_D of 12 nM[44], which would not be considered reversible for

applications in biological processes. To achieve reversibility for a chemogenetic system, several strategies have been employed.

First, on the system design level, POI activity can be switched off by anchoring the POI away from the correct subcellular location, since protein function and activity are highly dependent on location. The POI could be first activated by being localized to its active location (e.g. GTPase on membrane, transcription factor in nucleus), and then deactivated by being relocalized to a different subcellular compartment using an orthogonal CIP pair (**Figure 1-5**). For example, to deactivate functional POIs in the nucleus, they were fused to FRB and an abundant cytosolic protein PMA1 was fused to FKBP to anchor away the POIs to the cytoplasm[45]. The anchor-away strategy can also be applied for cytosolic proteins[46]. POIs were first recruited to the plasma membrane and activated by the FKBP/FRB CIP pair; the complex was then recruited to the mitochondria and deactivated by the GAI/GID1 CIP pair. In this work, one activation-deactivation cycle was successfully achieved.

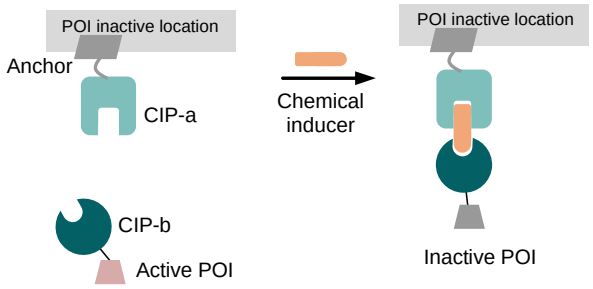


Figure 1-5 Anchor away strategy. The POI becomes inactive when a CIP system brings it to inactive locations.

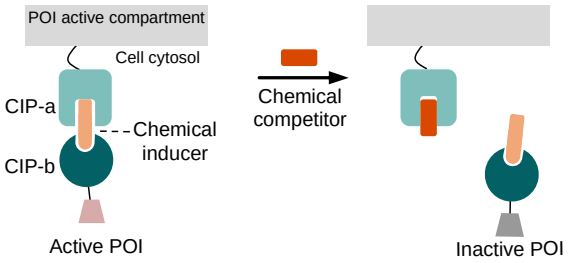


Figure 1-6 Competitive ligand strategy. The addition of a chemical competitor disrupts the CIP system, removing the POI from its active location.

Second, on the small molecule level, the dimerization can be disrupted by competitive binding of a non-dimerizing ligand or photocleavage of the bifunctional molecule (**Figure 1-6**). The competitive binding strategy was applied to FKBP and SNAP-tag-based CIP[47]. A bifunctional molecule, rCD1, was developed by connecting the SNAP-tag ligand to SLF. rCD1 induces membrane localization of the POI. The addition of FK506 disrupts the binding of rCD1 to FKBP and dissociates the POI from the membrane. In a similar strategy, the bifunctional molecule SLF-TMP locates the FKBP-POI fusion protein to the membrane[18]. Later addition of the competitor TMP disrupts the complex and releases the POI to the cytosol. The dissociated POI could be further relocated to the mitochondria by using another CIP pair. To use light as the "reverse signal", a photocleavable bifunctional molecule, MeNV-HaXS, was developed[25] by connecting HaloTag ligand to SNAP-tag ligand with a photocleavable methyl-6-nitroveratryl group (**Figure 1-7**). The addition of MeNV-HaXS covalently links HaloTag and SNAP-tag. UV light (360 nm) irradiation cleaves the linker and dissociates the CIP. Similarly, zapalog was developed[19] by connecting SLF and TMP with a photocleavable dialkoxynitrobenzyl linker. The zapalog-induced dimerization of FKBP and DHFR can be readily reversed by 405 nm light irradiation. The light cleavage step here can also be replaced by BRET. H-Luc was used to cleave the coumarin linker between HaloTag ligand and TMP to disrupt dimerization[42].

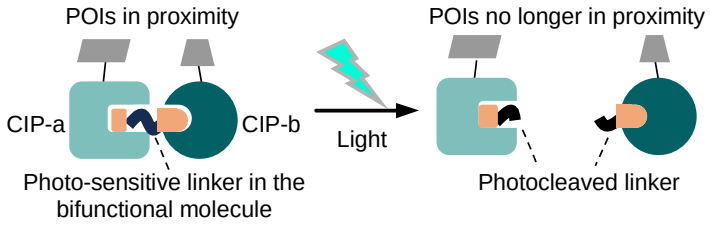


Figure 1-7 Photocleavage strategy. Upon light illumination, the photo-sensitive linker in the bifunctional molecule gets cleaved, disrupting the CIP.

Lastly, on the CIP engineering level, a low affinity CIP pair allows the complex dissociation by washing away the dimerizing molecule. Recently, a CIP system, CATCHFIRE[48] (**Figure 1-8**), was reported that can be readily reversed by washout. The small molecule inducer was named *match*, and the two proteins were named ^{FIRE}tag and ^{FIRE}mate. The system is based on a chemogenetic fluorescent reporter pFAST. ^{FIRE}tag and ^{FIRE}mate are the two split halves of pFAST and *match* is the fluorogenic ligand. The half-time of dimerization upon *match* addition and dissociation upon washout are ~25 s, and the cycle can be repeated multiple times. It is also worth noting that CATCHFIRE allows real-time fluorescence monitoring. The *match* ligand is only fluorescent when the ternary complex is formed, alleviating the need of an extra reporting system to monitor the dimerization process.

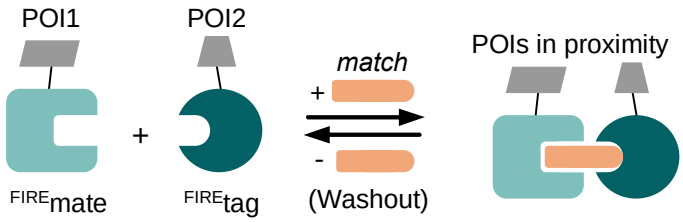


Figure 1-8 CATCHFIRE, a reversible CIP system involves the ^{FIRE}mate and ^{FIRE}tag, inducible by the ligand *match*. Simple washout of the match can reverse the dimerization.

The three strategies described above all provide some degree of "reversibility". The anchor-away and competitive binding strategies do not require extra engineering of the CIP pair but have a limited number of reversible cycles. Additionally, the anchor-away strategy requires multiple components and therefore its efficiency depends on the expression level of each component. Compared to the first two strategies, the "washable" CATCHFIRE system can be reversed multiple times, which is ideal for cell culture experiments. However, the system has low binding affinity to the *match* molecule, and CATCHFIRE is yet to be tested animal models.

1.1.1.4 Applications

CIP and CDP provide spatiotemporal control over protein functions and various biological processes. In this section, we will discuss the design strategies for controlling different cellular processes, including protein transportation, cell signaling, protein degradation, and gene expression.

Protein transportation

CIP and CDP are frequently used to recruit the POI to or release the POI from specific organelles, including the inner plasma membrane, outer mitochondrial membrane, Golgi apparatus, nucleus, etc. Furthermore, organelles could also be transported by motor proteins.

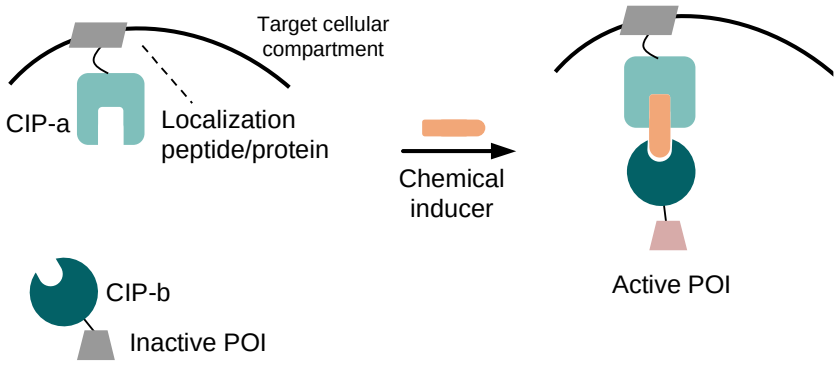


Figure 1-9 POI translocation induced by a CIP system. Localization signals or proteins help localize the CIP-a in the target cellular compartment. Examples of the target cellular compartment include inner plasma membrane, outer mitochondrial membrane, Golgi apparatus, and nucleus.

To control POI localization, the common design is to add a subcellular localization signal or protein to one CIP component as an anchor, and therefore, the POI fused with the other CIP component will be recruited upon CIP induction (**Figure 1-9**). For example, POIs could be recruited to the mitochondria by using TOM20 or MoA[30,41,46,49], to the Golgi using Giantin[49], to the ER using Cb5[49], to the lysosome using LAMP[49], or to the microtubule lumen using β-tubulin[50]. This process can also be reversed by using a light degradable dimerizer[25], the anchor-away strategy[46], or the reversible CIP system, CATCHFIRE[48]. When a CDP pair is used instead of CIP,

POIs can also be delocalized from the organelle's outer membrane[34]. Besides directly manipulating the location of POIs to manipulate their activity, POI inhibitors can also be used to modulate the POI's activity. The peptide inhibitor of AMP-activated protein kinase (AMPK) was localized to different organelle membranes, and it was discovered that inhibition of AMPK at the mitochondria is sufficient to induce cytosolic ATP increase[51].

Nuclear-cytoplasmic distribution is another way to regulate protein functions. With a similar design as above, by anchoring one CIP component in the nucleus or cytoplasm, the POI fused to the other CIP component could be transported in between the nucleus and cytoplasm (**Figure 1-9**). As mentioned above in the anchor-away strategy, by fusing one CIP component to the cytosolic protein PMA1, POIs in the nucleus could be transported into the cytosol[45]. A CIP component tagged with the nuclear localization signal (NLS) can be the anchor in the nucleus, and the addition of the small molecule dimerizer will recruit POIs tagged with the other CIP component into the nucleus[16]. Similarly, cytoplasmic localization can also be achieved by using a nuclear export signal (NES)[52]. When CATCHFIRE is used as the CIP pair, reversible nuclear or cytoplasmic localization is achieved[48]. In another strategy, SNAP-tag and HaloTag were used to covalently anchor the NLS-tagged POI on the Golgi apparatus by a light degradable bivalent dimerizer[25]. Subsequent light irradiation releases the POI to translocate to the nucleus.

Protein secretion is an important protein transportation process for cell communication. This process can be controlled by the dimerization of the POI with an anchor protein in the ER or Golgi apparatus. A series of works controlling glycosyltransferase[17,53,54] and sulfotransferase[55] activities by manipulating their transportation have been reported (**Figure 1-10**). The enzyme and the Golgi anchoring domain are each fused to one of the CIP components. When dimerizer is absent, the enzyme will be secreted directly and give minimal protein glycosylation. When induced

with the dimerizer, the enzyme will be retained in the Golgi apparatus to facilitate protein glycosylation in the secretary pathway.

A general system named RUSH[33] for controlling protein secretion was designed based on the CDP pair of streptavidin and streptavidin-binding peptide (SBP), which could be disrupted by biotin (**Figure 1-10**). By fusing streptavidin to the anchoring protein in the ER and Golgi, SBP-tagged POI is retained intracellularly until the addition of biotin. When CATCHFIRE is used in this system instead of streptavidin and SBP, a retention-secretion-retrieval cycle can be realized by dimerizer's addition-washout-addition process[48].

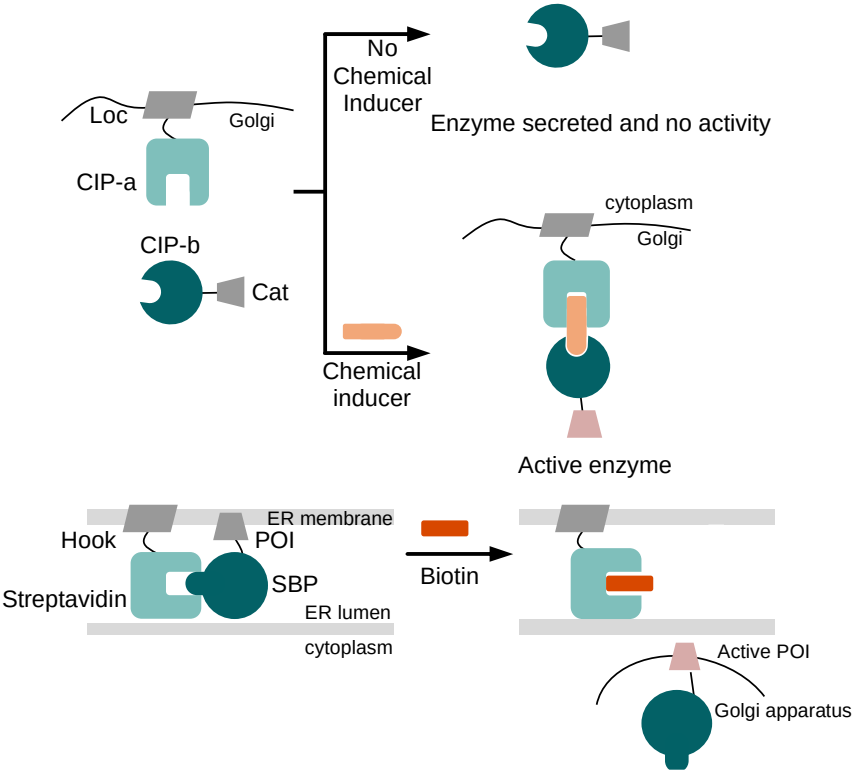


Figure 1-10 Top: Enzyme activation in Golgi apparatus regulated by a CIP system. The catalytic domain of the enzyme (Cat) is secreted outside of the cell, until an induced CIP system retains it in the Golgi apparatus with the Loc (localization domain) and reconstitute the enzyme activity. Bottom: RUSH. Biotin outcompetes the SBP to bind with streptavidin, resulting in the secretion of a POI.

Beyond proteins, organelle transportation can also be controlled using CIP. By fusing both the organelle of interest and a motor protein with the CIP component, the organelle could be transported along microtubules. Some reported work includes: 1) mitochondria transportation with

kinesin[19,56] and dynein[56]; 2) peroxisome transportation with kinesin[56,57], dynein[56,57], and myosin[57]; and 3) lysosome transportation with kinesin[48]. In such systems, the use of a light-degradable dimerizer[19] or CATCHFIRE[48] allows reversible transportation. To inhibit organelle transportation, a CIP pair was fused to synaptic vesicles and presynaptic proteins, inhibiting neurotransmitter release upon dimerizer addition[58]. Further, CIP can also be used to investigate the biology of organelles. By fusing FKBP to a 5-phosphatase INP54P and the two FRB split halves to the plasma membrane and ER, the chemically induced trimerization system is able to selectively deplete the secondary lipid messenger $PI(4,5)P_2$ at the junction sites of the plasma membrane and ER[12].

CIP and CDP have also been used to control artificial subcellular compartment formation to manipulate POI localization (**Figure 1-11**). To generate artificial protein aggregates that mimic stress granules, five FKBP repeats were fused on one component and five FRB repeats on the other[59]. Upon addition of rapamycin, the multivalent components randomly dimerize with others and form protein aggregates. When fused to an RNA recognition motif that binds to polyadenine containing RNA, the artificial aggregates could sequester polyadenine containing mRNAs like endogenous stress granules. On the other hand, CDP can be used to dissociate the aggregate formed with self-aggregating domains to control cellular processes. Based on homodimerizing FKBP(F36M) that could be disrupted by FKBP ligands[31], four repeats of FKBP(F36M) were fused together to form aggregates. Insulin was fused to this aggregating domain with a furin cleavage site so that insulin will be trapped in the ER and cannot be secreted. The addition of the FKBP ligand dissolves the aggregates, exposing the furin cleavage site for cleavage, and releases insulin for secretion. This system was initially developed for mammalian cell applications and further optimized by using an FKBP(F36L, I90V) mutant for yeast applications[32]. Further, a 10-amino-acid NES sequence (LALKLAGLDI) fused to FKBP(L106P) can act as an aggregation domain[60].

The addition of the ligand shield-1 could rapidly dissolve the aggregates, and the removal of shield-1 can induce aggregation again. This aggregation domain alleviates the need of multivalent FKBP repeats, thus minimizing the size of the fusion proteins.

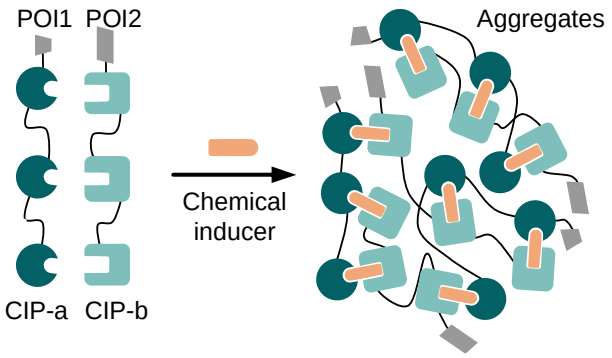


Figure 1-11 Artificial subcellular compartment formation via CIP systems.

Cell signaling

Common strategies that use CIP to control cell signaling include: 1) activating POIs via dimerization or reconstitution of the split POI; and 2) recruiting POIs to their active locations. The landmark work used FK1012 to homodimerize the ζ-chain of TCR to initiate T-cell signaling[1] (**Figure 1-12**). A similar approach has been applied to homodimerize the intracellular signaling domain of the Fas receptor to induce apoptosis[7,8,61] and to homodimerize STIM1 to induce calcium entry[62]. This strategy is not limited to POIs located at the plasma membrane. Induced homodimerization of the cytosolic protein, c-Raf-1 induces Raf signaling[63], and homodimerization of caspase-1 or caspase-3 induces apoptosis[64] (**Figure 1-12**). CDP, on the other hand, can be used to disrupt cell signaling. By dissociating the heterodimeric chimeric antigen receptor (CAR), the T-cell activity can be suppressed by the small molecule disruptor[35].

Splitting protein is a general protein engineering approach to control protein activity by controlling the split protein fragments' reconstitution (**Figure 1-12**). For example, split proteins

were designed for a series of tyrosine kinases, including Lyn, FAK, Src, and PKA[65]. CIP is used to control the reconstitution of these split proteins.

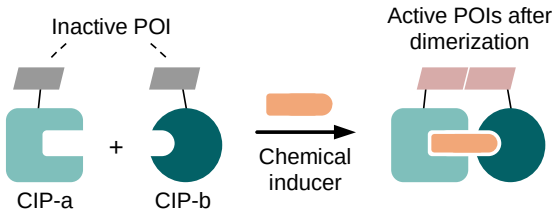


Figure 1-12 Dimerization of POIs induced by a CIP system, applicable to either constitution of functional signaling protein complex or reconstitution of split proteins.

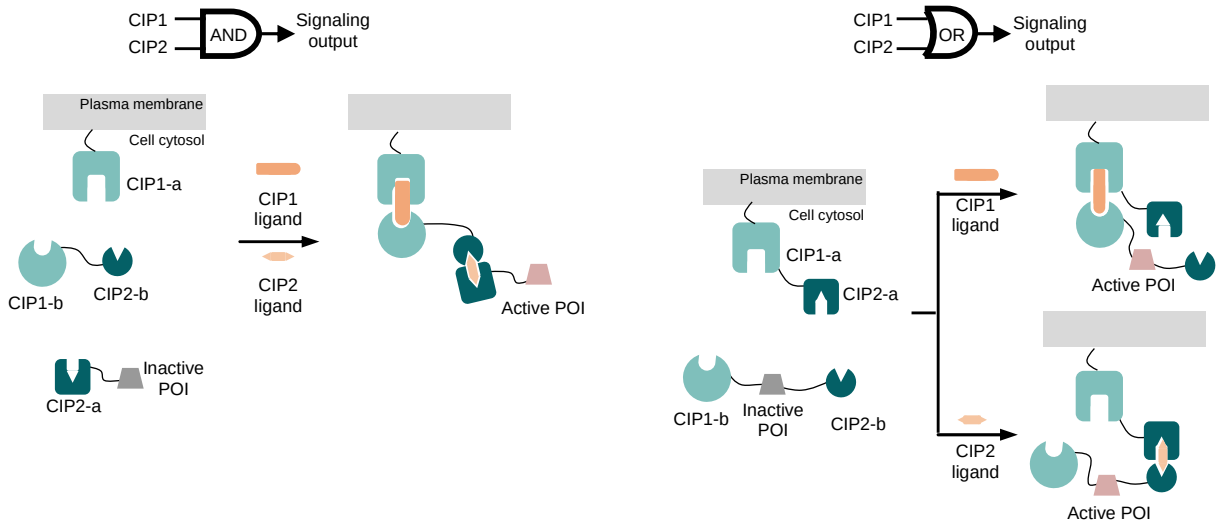


Figure 1-13 CIP-dependent logic gates. In the AND gate (left column), both CIP1 and CIP2 are required to activate the POI. In the OR gate (right column), either CIP1 or CIP2 can activate the POI.

Signaling cascades can also be activated by recruiting POIs to specific locations using CIPs. Recruiting the Rac1 guanine nucleotide exchanger factor (GEF) Tiam1 to the plasma membrane activates the Rac pathway[14,38,40]. The same strategy has been applied to other signaling proteins, such as the Fas signaling domain[16], Rac1[18], Src-like tyrosine kinase[66], SOS[67], Zap70[68], Inp54p[69] and GEF for Ras[49]. Orthogonal CIP pairs also enable the construction of signaling pathways with logic gates. GAI-GID1 and FKBP-FRB can be incorporated to the system to gate translocation with an "AND" gate (GA3 and rapamycin are both supplied) or an "OR" gate (either one is supplied)[14] (**Figure 1-13**). The PROCISiR system discussed in section 2.1, in which the

DNCR and GNCR recognize NS3a bound with danoprevir and grazoprevir respectively, allows similar orthogonal control. This enables the control of two POIs, Tiam and LARG, with one less anchoring component[30].

Because cell signaling is dynamic, reversible systems are beneficial for probing cell signaling events. Using the competing strategy discussed in section 2.3, PI3K can be activated by translocation to the plasma membrane, and deactivated by the addition of a competing binder[47]. The anchor-away strategy was also applied to an inter-SH2 domain from a regulatory PI3K subunit to reversibly modulate PI3K's activity, and to Tiam1 to reversibly induce Rac signaling pathway[46].

Protein degradation

Targeted protein degradation is widely applied to perturb protein functions to study their roles in biological processes and can be used as therapeutics. CIP provides temporal gating to bring POIs into proximity with the cellular protein degradation machineries. Several CIP-based chemogenetic tools have been developed as generalizable platforms to control the stability of POIs. Furthermore, proteolysis targeting chimeras (PROTACs), an emerging class of therapeutic molecules, have been developed to bring a POI into the proximity of E3 ubiquitin ligases for protein degradation[70-72]. Recently, new mechanisms targeting lysosomal pathways have been developed as well[73-75]. Here, I focus on the chemogenetic tools for generally regulating POI degradation. PROTACs and related technologies were comprehensively discussed in these another reviews[70-72].

Platforms that tag POIs with CIP domains could be easily generalized to modulate the degradation of various POIs. The auxin-inducible degron was transferred from plants to yeast and mammalian cells[76] (**Figure 1-14**). The TIR1 from *Oryza sativa* (OsTIR1), an F-box ortholog, was

introduced to cells to form the new E3 ligase complex. Auxins, indole-3-acetic acid and 1-naphthaleneacetic acid, act as molecular glues to induce the interaction between OsTIR1 and IAA17 (called 'AID'), ubiquitinating AID-tagged POIs, and then lead to proteasomal degradation of POIs. The AID tag is compatible at both of the N- and C-terminus of the POI and has been applied for both nuclear and cytosolic proteins. The initial AID tag is relatively large (229 aa). The AID tag was shortened to AID* (residues 71–114 aa) and AID47 (residues 63–109 aa) [77]. To avoid the introduction of the extra OsTIR1 component, three other platforms were developed with similar designs. The dTag system[78], the HaloPROTAC system[79], and TMP PROTAC system[80] all utilize endogenous E3 ligases and tag the POI with a CIP domain. HaloPROTAC utilizes a bifunctional small molecule to dimerize the HaloTag fusion protein with VHL. HaloPROTAC was further developed to target cIAP1 as well[81]. In the dTag system, the bifunctional dimerizer dTag-13 is generated by connecting shield-1 with thalidomide for targeting CRBN. dTag-13 can degrade FKBP(F36V)-tagged POI both in cell culture and a mouse model. dTagV-1 is a further developed version of dTag that targets VHL[82]. The bifunctional molecule TMP-pomalidomide can degrade eDHFR-tagged POIs by inducing their proximity with CRBN[80]. All the three systems are compatible with both N- and C-terminus tagging.

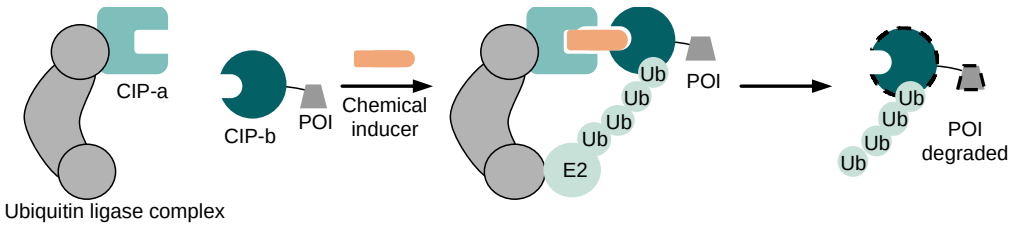


Figure 1-14 CIP-tagging degradation. After a CIP system brings the POI in proximity to a ubiquitin ligase complex, the POI will be degraded. This approach has been used in auxin-induced degradation, dTag and HaloPROTAC.

Some other CIP platforms used for modulating protein existence hijack the protein's posttranslational modifications. The conditional protein splicing (CPS) system[83,84] mimics the

protein splicing process (**Figure 1-15**). In CPS, two POIs are fused to the split VMA intein halves and the FKBP-FRB-induced VMA reconstitution results in splicing, yielding ligated POI. The system was further optimized to control split enzymes in cell culture and *Drosophila melanogaster*[85]. To reduce the basal activity when the split POI halves have high intrinsic affinities, a split ubiquitin for the rescue of function (SURF) system was developed[86] (**Figure 1-16**). SURF utilizes a similar design as CPS but has split ubiquitin(I13A) halves instead of split VMA intein halves. On one of the components, there are three FRB repeats that serve as both a CIP component and a degradation tag. Rapamycin induces FKBP-FRB dimerization and stabilizes the protein from being degraded. The consequent split ubiquitin reconstitution leads to cleavage of the split POI halves, resulting in full length POIs, including caspase-3, v-Src kinase, and the Smad3 transcription factor. In this case, the degradation of one of the components minimizes the basal activity. Compared to the degradation methods using the E3 ligase mechanism, CPS and SURF benefit from producing POIs in native forms.

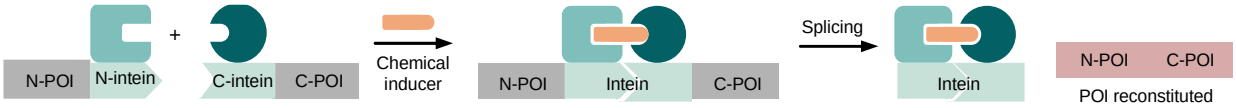


Figure 1-15 CPS. Upon the addition of the chemical inducer, split intein halves reconstitute, resulting in the splicing and the POI reconstitution.

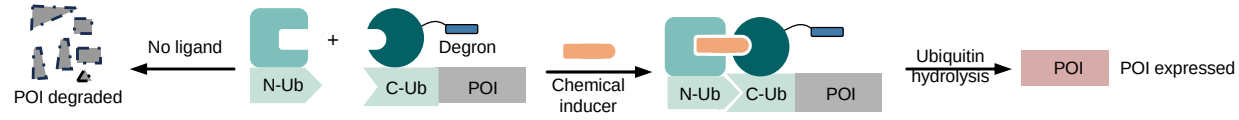


Figure 1-16 SURF. Without the chemical inducer, the POI will be degraded with the degron. CIP-induced ubiquitin reconstitution allows the POI to be expressed and keep stable.

Gene expression

CIP and CDP have been applied to control gene expression via the two-hybrid system and dCas9 system (**Figure 1-17**). The two-hybrid system is a robust platform that was originally used

to detect protein-protein interactions. In the two-hybrid system, the DNA binding domain (DBD) and the activation domain (AD) of a transcription factor are each fused to the CIP or CDP components. There are several commonly used DBD-promoter pairs: 1) Gal4-DBD with UAS; 2) TetR with TRE; 3) LexA-DBD with its binding site; and 4) zinc finger domain, transcription activator-like effector (TALE) and dCAS9-CRISPR protein with their DNA binding sites. Some commonly used AD include Gal4-AD, VP16, VP64 (four repeats of VP16), and VPR (VP64, p65, and Rta fusion protein). The two-hybrid system can be quantified by different assays depending on the genes encoded in the reporter gene. The reporter gene can be secreted embryonic alkaline phosphatase (SEAP) and luciferase for plate reader assays or fluorescent proteins for microscope imaging and flow cytometry.

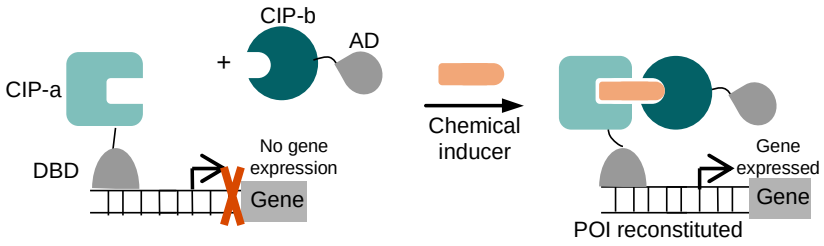


Figure 1-17 CIP-dependent two-hybrid system. The CIP system brings AD in proximity to DBD for activating gene expression.

Controlling protein expression levels in cells is an easy and powerful way to study their functions and the two-hybrid system has been a classic application for CIPs[7,16,40]. A simple CIP-controlled two-hybrid system is a one-way "on" system. To turn the system "off", previously described reversible CIP strategies can be applied to dissociate the DBD-AD complex. For example, a non-dimerizing ligand has been used to outcompete the dimerizer molecule[87]; another competing small molecule could dimerize with a different CIP component and recruit a repressor domain (RD)[88]; and the anchor-away strategy has been applied to transport the transcription factor

out of the nucleus[52]. Being able to initiate and terminate gene transcription with a chemical control enables fine tuning of protein expression levels.

The two-hybrid system amplifies the dimerization signal via a transcriptional readout, and therefore provides a sensitive platform to characterize and screen CIPs. The two-hybrid system has been used to test the orthogonality between the ABA induced Pyl-ABI1 dimerization and the rapamycin induced FKBP-FRB dimerization systems[13]. It has also been used as a platform to screen for FRB mutants that selectively bind to rapamycin analogues[10] and to characterize the binding affinities of different rapamycin analogues with different FRB mutants[11]. These examples discussed above all rely on the proximity induction of the genetically encoded DBD and AD. A chemical-dependent transcriptional system that recruits endogenous transcriptional systems was developed[89,90]. Gal4-DBD was fused to the minimal ligand binding domain of the glucocorticoid receptor. A conjugate molecule composed of Isoxazolidine (recruits endogenous transcriptional machinery) and OxDex (glucocorticoid receptor ligand) could recruit endogenous transcriptional machineries to initiate transcription. This system can upregulate reporter gene expression by 80-fold in Hela cells[90].

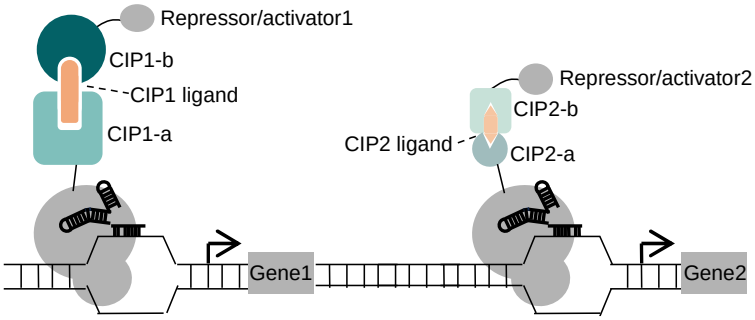


Figure 1-18 Multiplexed control of dCas9 by orthogonal CIP pairs. The chemical ligand activates a specific CIP system and the corresponding gene expression.

A dCas9-CRISPR-based two-hybrid system has been developed, enabling unprecedented modulation of endogenous gene expression. dCas9 activity can be controlled by rapamycin-

induced reconstitution of split dCas9[91]. CIP can also be applied to recruit AD or RD to dCas9 to up- or down-regulate gene transcription (**Figure 1-18**). Multiplexed control of two genes was achieved by using two orthogonal dCas9 systems and CIP systems[92,93]. For example, two dCas9 orthologs from *Streptococcus pyogenes* (Sp-dCas9) and *Neisseria meningitidis* (Nm-dCas9) were fused to two orthogonal CIP pairs, GAI-GID1 and FKBP-FRB[92]. Sp-dCas9 and *Staphylococcus aureus* dCas9 (Sa-dCas9) were also demonstrated with GAI-GID1 and ABI-Pyl CIP pairs[92]. Furthermore, an RD, Krüppel-associated box (KRAB) can be used in parallel with an AD to achieve up- or down-regulation of different genes at the same time (**Figure 1-18**).

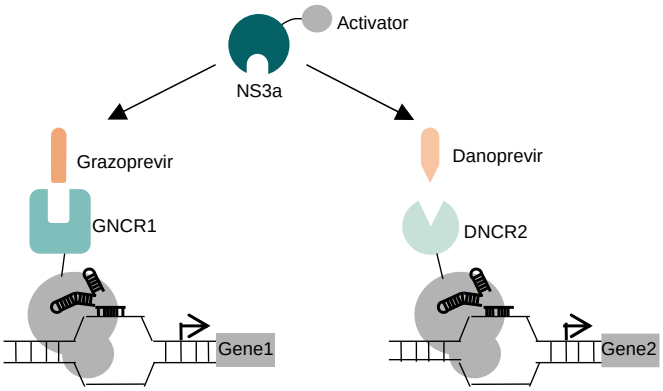


Figure 1-19 Orthogonal control of dCas9-based gene expression by using PROCISiR.

With orthogonal dCas9 proteins, CIP pairs, and CDP pairs, complicated gene expression networks can be built. Multiplexed logic gates for gene regulation can be constructed[92] (**Figure 1-13**, **Figure 1-18**). Orthogonal control of the same DNA locus was also constructed by using one CIP (ABI-Pyl pair) to recruit KRAB for gene down-regulation and another CIP (GAI-GID1 pair) to recruit VPR for gene up-regulation. Similar control could be achieved using the PROCISiR system[30] (**Figure 1-19**), in which NS3a is dimerized to DNCR or GNCR depending on the ligand applied, and thus regulate two different genes. The NS3a-ANR CDP pair has been applied to control VPR dissociation from dCas9[34].

dCas9 could also be used to regulate chromatin modifications. Chromatin regulators, such as Hp1/Suv39h1 heterochromatin complex and mSWI/SNF (BAF) chromatin-remodeling complex, can be recruited via CIP to dCas9 at specific genomic loci[94]. Recruitment of HP1/Suv39h1 results in H3K9me3 deposition and gene silencing, and recruitment of BAF results in H3K27me3 depletion and gene activation. Another strategy is to induce chromatin looping. CIP was used to dimerize two dCas9 proteins targeted to the promoter of the β-globin locus and the HS2 region of the locus control region[95]. Dimerization causes chromatin looping and up-regulates gene expression.

Lastly, Cre recombinase is also frequently used for genomic engineering in animal models. Chemogenetic control over Cre provides another way to control gene expression. The split Cre fragments' reconstitution can be regulated by CIP[96] (**Figure 1-12**).

1.1.2 Conformational and allosteric control

Protein activity can also be regulated by chemically induced conformational change. Many proteins contain a ligand binding site that modulates their activities (e.g., membrane receptors, transcription factors, etc.). For proteins without chemically-inducible properties, small molecule binding domains can be engineered into the protein domain to allosterically modulate the protein's activity and stability.

1.1.2.1 Membrane receptors

Cell membrane receptors can often relay an extracellular ligand binding signal to intracellular signaling activation and provide important engineering elements for chemogenetic

tool design. To avoid cross talk with endogenous receptors and ligands, orthogonal ligands and receptors could be designed to create chemogenetic tools for regulating cellular activities. Alternatively, small molecule ligands could be tethered to the receptors for activity regulation.

Designer G protein-coupled receptors (GPCRs) that are activated by orthogonal ligands have been engineered (**Figure 1-20**) and widely applied in neuroscience to modulate neuronal activity in selective neuronal circuits for studying causal effects on behaviors. The first designer GPCR developed is based on a β2-adrenergic receptor mutant, β2AR(D113S) [97]. β2AR(D113S) can only be activated by synthetic ligands (catechol-containing esters and ketones) but not the endogenous β2AR agonist, adrenaline. However, the orthogonal ligands have low affinity for β2AR(D113S) and do not work well in animal models. A later design of a receptor activated solely by the synthetic ligand, RASSL was developed based on the κ-opioid receptor (κOR)[98]. The binding of the designer κOR with its endogenous peptide ligands was abolished by replacing its second extracellular loop with that of δ-opioid receptor. The resulted RASSL named Ro1 could be activated by κOR's synthetic ligand spiradoline. However, spiradoline also activates endogenous κOR, and therefore, is not completely orthogonal to the endogenous biological system.

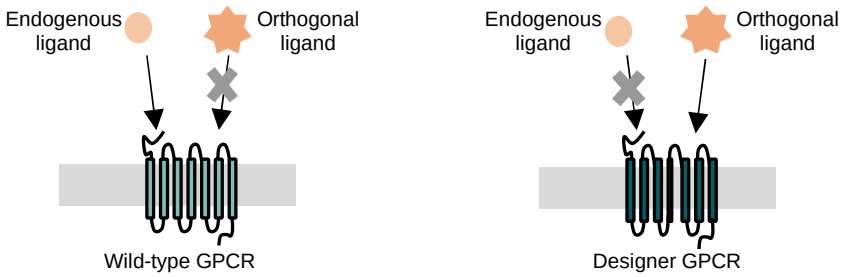


Figure 1-20 Designer GPCRs. Orthogonal ligand can activate the designer GPCR and its downstream signaling, while endogenous ligand can only activate the wild-type GPCR.

Designer receptors exclusively activated by designer drugs (DREADDs) were later designed and are the most widely used designer GPCRs for regulating neuronal activity. The

applications of DREADDs have been extensively reviewed[99-102]. Here, I briefly introduce the existing DREADDs. The M3 muscarinic acetylcholine receptor was engineered to be activated by a biologically inert drug metabolite[103], clozapine-N-oxide (CNO), through a directed evolution platform where the GPCR activation is linked to yeast growth[104]. The resulted DREADD, hM3Dq is activated by CNO but not its endogenous ligand acetylcholine, and CNO does not activate the endogenous M3 receptor. hM3Dq couples to Gq and can be used to activate neurons[105]. The same mutations found in hM3Dq were introduced to the M4 muscarinic acetylcholine receptor, a Gi coupled receptor, to engineer hM4Di[103], which is also activated by CNO and inert towards acetylcholine. hM4Di can be used for neuron silencing[106].

Besides inducing Gq and Gi signaling, DREADD for activating Gs signaling is designed by engineering a chimeric GPCR GsD that incorporates the extracellular ligand binding site of hM3Dq and the intracellular Gs binding site of the turkey erythrocyte β adrenergic receptor[107]. Activation of GsD leads to upregulation of cAMP concentrations. By mutating the DRY motif in hM3Dq, which is known to be important for G protein coupling activity, Rq(R165L) lost its Gq signaling activity and provides a DREADD only with β-arrestin signaling[108]. In addition to a CNO-based DREADDs, a salvinorin-B-based κOR DREADD (KORD) has been engineered for multiplexed control over neuronal activities[109]. Through rational design, a D138N mutation was introduced to KOR to abolish its affinity towards its endogenous ligand, dynorphin A and increase its binding towards salvinorin B, which does not activate endogenous κOR. KORD is coupled with Gi and silences neuron on activation[109].

A similar engineering strategy has been applied to ligand-gated ion channels (LGICs). The glutamate-gated chloride channel (GluCl) from *Caenorhabditis elegans* was engineered to have reduced glutamate activity and retained affinity towards its synthetic ligand, ivermectin[110,111]

(IVM). GluCl does not exist in mammals, so IVM only activates the engineered GluCl. However, the application of GluCl is limited by its poor expression and folding due to the required co-expression of two subunits and its non-human origin. To overcome this limitation, a human α1 glycine receptor was engineered to be activated by IVM but have no activity for glycine[112].

A lot of engineering effort is required for modifying the receptor binding pockets to recognize new ligands. Alternatively, selective tethering of ligands to the receptor can be used to control receptor activity (**Figure 1-21**). This strategy relies on bifunctional small molecules consisting of a reactive warhead and a receptor ligand. The reactive warhead can covalently tether the small molecule to the modified receptors. In addition, using an azobenzene moiety as the linker in the small molecule creates photoswitchable tethered ligands (PTLs), allowing photoswitchable activation of the receptors. A PTL for Shaker K^+ channel was constructed by maleimide (for tagging cysteine residues), azobenzene, and tetraethylammonium (for blocking the K^+ channel) [113]. By introducing a cysteine residue near the ligand binding site, PTLs are covalently tethered to the engineered ion channel. With the same strategy, the NMDA receptor[114], potassium channel TREK1[115], metabotropic glutamate receptor mGluR2[116], and $GABA_A$ receptor[117] have been reported to be controlled in the same fashion.

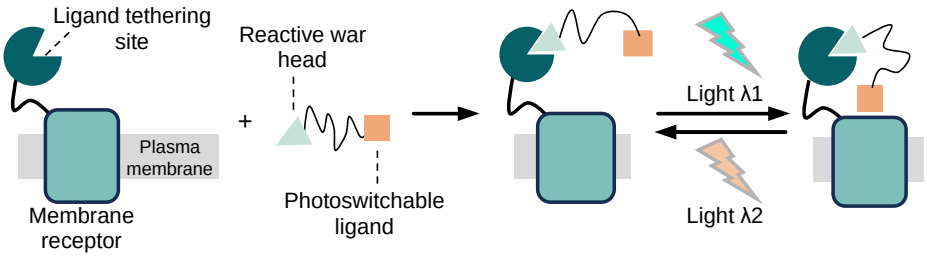


Figure 1-21 POI ligand tethering strategy. For the ligand tethering, PTL utilizes covalent bonding via cysteine residues on the POI, while PORTL fuses SNAP-tag to the N-terminus of the POI for interacting with the SNAP-tag ligand. Light illumination switches the ligand between two conformations.

The use of maleimide-cysteine chemistry has several limitations[118] for in vivo applications of PTLs: 1) the native cysteine residues on cell surfaces could be affected; 2) maleimide could

react with other nucleophiles in the environment such as water and glutathione; 3) the reaction efficiency is not high enough. To overcome such limitations, a new class of tethered ligands, photoswitchable orthogonal remotely tethered ligand (PORTL) (**Figure 1-21**), was developed[118]. Instead of using cysteine as the tethering site, PORTL uses SNAP-tag to achieve specific labeling. SNAP-tag is fused to the N-terminus of the metabotropic glutamate receptor 2 (mGluR2), which also avoids introducing cysteine mutation in the receptor. By using orthogonal tethering domains, multiplexed control over two types of receptors could be achieved by fusing a SNAP-tag variant, CLIP-tag, to mGluR2, and SNAP-tag to mGluR7[119]. Importantly, the wavelength of the activation and deactivation light were different for the two PORTLs to allow multiplexed control.

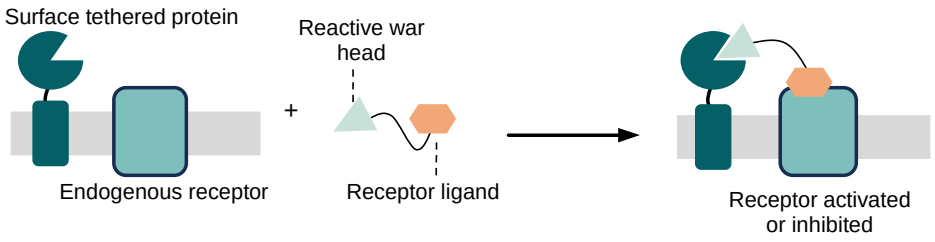


Figure 1-22 DART. A protein tether on the plasma membrane through a transmembrane domain can activate or inhibit a receptor, with the addition of a bifunctional molecule.

All of the PTL or PORTL examples discussed above utilize engineered receptors. To probe endogenous receptors, the ligand could be tethered to a separate transmembrane domain (**Figure 1-22**). A technology named drugs acutely restricted by tethering (DART) has been developed[120]. By fusing HaloTag to a transmembrane domain, a chloroalkane-conjugated ligand YM90K could be tethered to the membranes of selected neuronal populations to antagonize α-amino-3-hydroxy-5-methylisoxazole-4-propionicacid receptor (AMPAR). The ligand in DART could be replaced with PORTLs to achieve photoswitching. SNAP-tag was fused to a single-pass transmembrane segment derived from low density lipoprotein receptor, so that endogenous mGluR2 could be

activated by the tethered ligand[121]. The same strategy has also been applied to the dopamine D1 receptor[122].

1.1.2.2 Transcription factors

Transcription factors that are regulated by small molecules can serve as another class of chemogenetic tools for regulating gene transcription. The commonly used chemically induced gene expression systems in mammalian cells and animal models are the Tet-Off[123] and Tet-On[124] systems (**Figure 1-23**) derived from the tetracycline responsive transcriptional system found in *Escherichia coli*. The two systems are controlled by tetracycline or doxycycline (Dox, a stable analogue of tetracycline), where Dox turns off gene expression in Tet-Off and turns on in Tet-On. The detailed development and applications of the two systems are available in these reviews[125,126].

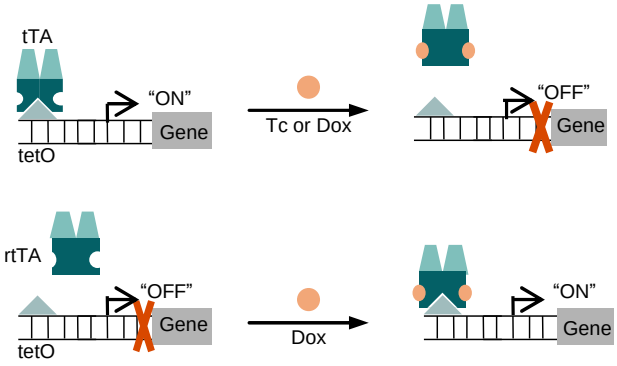


Figure 1-23 Tet-Off (top) and Tet-on system (bottom). In the Tet-Off system, either Tetracycline (Tc) or Dox can Turn off the gene expression. In the Tet-on system, the addition of Dox can turn on the gene expression.

Tet-Off[123] is based on a tetracycline-controlled transactivator (tTA) and its promotor, tetracycline response element (TRE). tTA is composed of the DNA binding domain, tetracycline repressor (TetR), fused with a strong AD VP16. The continuous supplement of Dox keeps the gene silent, and the removal of Dox initiates gene expression. This process works well for cell culture experiments by replacing the culture media but is more challenging for animal experiments

because the removal of DOX may take days in mouse models[125]. Because the administration of Dox is faster than removal, Tet-On was developed[124] by engineering a reverse-TetR, which binds to the promoter tetracycline response element (TRE) upon Dox binding. The resulted reverse-tTA (rtTA) works in the opposite way as tTA and provides a Dox-induced gene activation system. Tet-On was further optimized[127] to be more sensitive to Dox to improve the performance in a mouse brain and reduce the basal activity under the Dox free condition.

Similarly, transcription factors TtgR from *Pseudomonas putida*[128] and PadR from *Bacillus subtilis*[129] were transferred to mammalian cell systems for gene regulation. TtgR and PadR dissociate from the promoters upon induction of resveratrol and ferulic acid, respectively. By fusing to an AD, TtgR-VPR and VP64-PadR function in the same way as tTA. To engineer chemically induced transcription activation "on" systems, a RD was fused to TtgR and PadR so that upon chemical induction, the transcription repressor dissociates, and gene transcription is activated[128,129].

Another approach to obtain an orthogonal induced gene expression system is to engineer the transcription factors to respond to new ligands. Computational modeling and library screening were employed to engineer the *Escherichia coli* derived LacI transcription factor, which is induced by isopropyl β-D-1-thiogalactopyranoside (IPTG). A LacI mutant (Q291H, T276L, S279G) was identified, which gained high activity towards gentiobiose and completely lose activity for IPTG[128].

1.1.2.3 Protein stability modulation

To modulate protein stability, a destabilizing domain (DD) that is degraded in the absence of a small molecule ligand and is stabilized upon ligand binding can be added to the POI (**Figure**

1-24). DDs are fused to POIs as tags, so that the stability of POI-DD fusion protein is chemically controlled. Compared to the E3 ligase dependent techniques discussed in section 2.4, such designs benefits from containing only one component.

It was first discovered that protein half-life is dependent on its N-terminal residue, which is also known as the N-end rule[130,131]. An unstable DHFR mutant with an N-terminal Arg residue (Arg-DHFR) could be stabilized by the ligand MTX[132]. Further, the stability of the POI-DD fusion protein could also be modulated. The FRB*(K2095P, T2098L, W2101F) mutant could destabilize the fusion protein with glycogen-synthase-kinase-3β and be stabilized upon binding of the rapamycin analog, MaRap and the consequent dimerization with FKBP[133,134].

DDs that regulate POI-DD fusion protein degradation solely by ligand binding were then developed (**Figure 1-24**). Several DDs were developed based on FKBP with shield-1[9], DHFR with TMP[135], estrogen receptor ligand binding domain (ERLBD) with 4OHT and CMP8[136], and the UnaG fluorescent protein with bilirubin[137]. The four DDs are orthogonal to each other and could be applied for multiplexed control of protein stability of multiple POIs. The UnaG-DD also provides a fluorescent readout of POI expression level, alleviating the need for an extra expression marker. It is worth noting that the FKBP-DD is more efficient when fused to the N-terminus of the POI, while the others are not affected by fusion geometry. These DDs have been successfully demonstrated in animal models as well. FKBP-DD has been used to control the CAR protein level in CAR-T cells in a mouse model[138]; DHFR-DD was used to control Cre protein level to control POI abundance in a mouse brain[139].

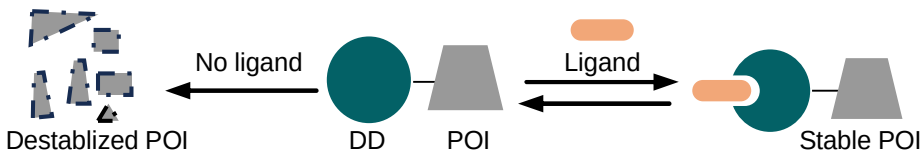


Figure 1-24 Ligand-induced protein stabilization. The POI is destabilized by DD, unless a ligand is introduced.

For POIs that require unmodified terminal regions to be functional, liberation-prone degron (LIBRON) was developed[140] (**Figure 1-25**). LIBRON is composed of DD (FKBP- and DHFR-DD were used) and a ubiquitin mutant (Ub*) whose polyubiquitination ability is abolished. When the fusion protein is stabilized by ligand binding, DD-Ub* will be cleaved from the POI, and the POI will be released in its native form. Another advantage of LIBRON is that the released POI is still stable even after the stabilizing ligand is removed from the system. This could be beneficial in animal applications, because a single dose of ligand supplementation can release the protein from the DD, alleviating the need of constant drug administration. Release from the DD can also be achieved by protease cleavage. SMASh tag was developed for time-gated production of soluble protein in its native form[141] (**Figure 1-25**). SMASh tag is composed of a NS3 protease cleavage site followed by the NS3 protease and a degron. SMASh tag can be connected to the C-terminus of the POI. When the NS3 activity is inhibited by its inhibitor, asunaprevir, the degron stays connected and leads to whole protein degradation. When inhibitor is removed, the cleavage site will be cut by NS3 to separate the degron from the POI, so that the native form of the POI is released and stabilized. SMASh tag has been applied to control Zap70 kinase to modulate CAR-T cell activity[142].

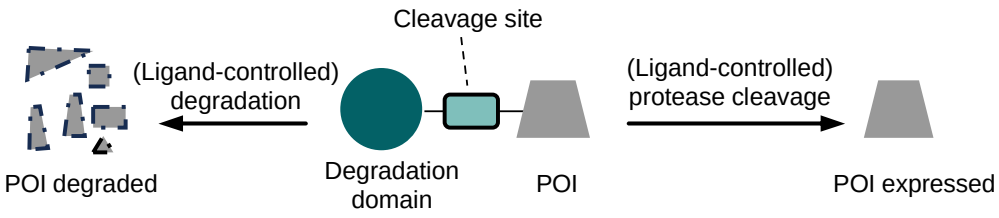


Figure 1-25 Ligand-controlled destabilization. In LIBRON, the ubiquitin is employed to release the POI from DD. In SMASh tag, the protease NS3 is utilized to cleave the POI off from a degron.

All of the DDs described above are stabilized by ligand binding. Chemical-dependent destabilization was achieved by tagging a hydrophobic adamantyl moiety to the POI-HaloTag fusion protein[143,144]. Upon HaloTag labeling, the adamantyl group mimics the partially denatured

state of the protein and causes degradation of the HaloTag-POI fusion protein by the proteosome. This system was successfully applied in mammalian cell culture, zebrafish embryos, and mice. However, the HaloTag-based strategy is not as efficient as the DDs, with only ~65% degradation observed when HaloTag7 is used.

1.1.2.4 Allosteric switch insertion

Chemical-dependent allosteric control of protein activity can also be achieved by inserting a chemogenetic switch into the allosteric sites of POIs (**Figure 1-26**). The conformational change of the inserted chemical-dependent protein switch domain could be propagated to the POI, modulating the POI's activity. Tools and models for designing allosterically controlled proteins have been comprehensively discussed[145,146]. Here in this section, I discuss the chemogenetic allosteric switches developed.

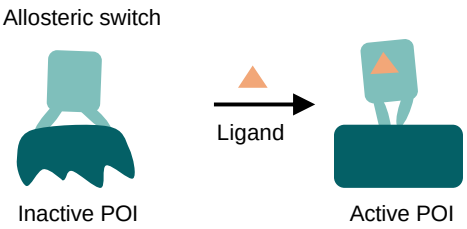


Figure 1-26 Manipulation of protein functions through allosteric switches. . Examples of allosteric switch includes iFKBP, uniRapR, cpDFHR. The allosteric switch is usually inserted into a loop in the POI. After ligand binding, the allosteric switch changes conformation and thus affect the POI's function.

A common mechanism of the chemogenetic allosteric switches is that the apo structure of the chemical-sensing domain is partially unfolded, and ligand binding stabilizes and rigidifies the structure. Additionally, to not completely disrupt the structure of the POI when inserting the chemical-sensing domain, the two termini of the chemical-sensing domain should be spatially close to each other. This can be achieved by truncation of the chemical-sensing domain from one

of the termini until the two new termini are close in space (e.g., iFKBP and uniRapR) or by circular permutation (e.g., cpDHFR).

A series of chemically-controlled kinases were engineered using insertable FKBP (iFKBP) (**Figure 1-26**). iFKBP was designed by truncating the first β-sheet from the N-terminus to make the N- and C-termini closer in distance[37]. Molecular dynamic studies suggested that the iFKBP conformation is rigidified by its interaction with rapamycin and its consequent dimerization with FRB. A rapamycin controlled focal adhesion kinase (FAK) was designed by screening and optimization of the insertion site and the linkers connecting iFKBP with FAK[37]. The resulted rapamycin-regulated FAK could be activated as efficiently as the wild type with 50 nM rapamycin when FRB is present, while the required rapamycin concentration increases to 4 μM in the absence of FRB. Additionally, the photocaged rapamycin analogue, pRap, was used to further improve the spatiotemporal resolution[36]. This design strategy was also applied to design rapamycin-activatable Src and p38[37]. By fusing FRB to the downstream targets of Src, FAK and p130Cas, the mechanism of allosteric control and proximity control were utilized simultaneously, which is useful for dissecting the kinase activity on different targets[147]. Other kinases engineered with the same strategy include Fyn[148], Src[148], Lyn[148], Yes[148], c-Src[149], and p38α[149].

When using the iFKBP domain, a second component of FRB needs to be co-expressed. To engineer single-component allosterically controlled POI, uniRapR domain (**Figure 1-26**) was designed[150] by inserting the circular permuted FRB between the β5 and β6 barrel of iFKBP. The uniRapR was successfully inserted into Src kinase[150], Vav2[150], ITSN[150], and Pak1[150] to achieve rapamycin-dependence.

Another example of a chemogenetic allosteric switch is circularly permuted DHFR (cpDHFR)[151]. cpDHFR is partially unfolded in the absence of its cofactor NADPH or ligand TMP.

cpDHFR can be inserted into nanobodies, generating the ligand-modulated antibody fragment (LAMA)[152] (**Figure 1-26**). ^{GFP}LAMA was generated by the insertion of cpDHFR into the complementary-determining region 3 (CDR3) of a GFP nanobody. The apo ^{GFP}LAMA has low nanomolar affinity towards GFP, and the presence of NADPH and TMP abolishes the binding. Notably, the binding of ^{GFP}LAMA to GFP is reversible upon media washout on the order of minutes. The crystal structure of ^{GFP}LAMA suggests that the abolished binding of ^{GFP}LAMA towards GFP is from both allosteric effects as well as the steric hinderance of cpDHFR. Similarly, cpDHFP inserted into the CDR3 of a nanobody against p24 HIV capsid protein (^{p24}LAMA) and into the CDR2 of a nanobody against ALFA-tag ($^{ALFA\text{-}tag}$LAMA) both showed reduced binding affinity towards their targets upon NADPH and TMP binding. The LAMA for a nanobody against lamina-associated polypeptide 1 showed increased affinity upon ligand binding.

1.1.3 Steric hinderance control

To directly control the activity of a protein, chemical-dependent control of the accessibility of the active site can be applied. This offers an alternative approach to control protein activity directly in addition to controlling the protein conformation as discussed in Chapter 1.1.2.

1.1.3.1 Steric unblocking of protein active site

By fusing a CDP pair to the N- and C-termini of the POI, the POI active site can be blocked via steric hinderance. Addition of the small molecule disruptor dissociates the two CDP components and liberates the POI (**Figure 1-27**). The first example using this design was demonstrated by controlling a Rho family GEF, intersectin[153]. The two CDP components, BCL-xL and BH3 were fused to the termini of the catalytic domain of intersectin. Over a five-fold

chemical-dependent activation could be achieved in an in vitro assay. The same design was applied to control the catalytic domain of a RAS activator, Son of Sevenless (SOScat), called chemically inducible activator of RAS (CIAR)[154]. A computational model, RosettaRemodel[155], was used to guide the optimization of the linker lengths. The small molecule disruptor, A-385358 could induce over 10-fold activation of CIAR in a luciferase reporter assay. To show the generalizability of this design, the computational optimization approach was further applied to intersectin and another Rho family GEF VAV, resulting in 4- and 8-fold activation by A-385358 in an in vitro nucleotide exchange assay. Another CIAR with the CDP pair NS3a-ANR has also been developed[34], further showing that this approach is generalizable to other orthogonal CDP pairs.

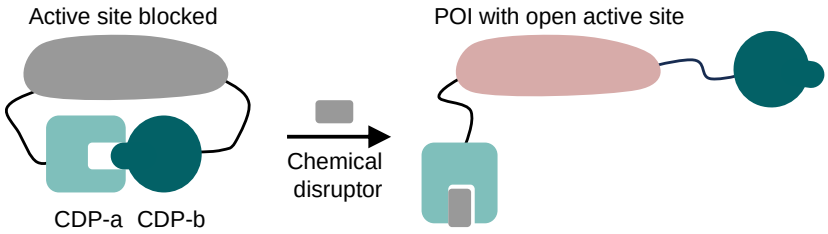


Figure 1-27 Manipulation of protein functions through steric hindrance. CDP for unblocking protein active sites. The POI is sterically blocked by an interactive pair of proteins, until a chemical disruptor is introduced and unlock the protein pair.

Besides the GEFs, the design mentioned above was used to engineer chemically inducible Cas9 (ciCas9)[156]. ciCas9 was engineered by replacing the nonessential REC2 domain of Cas9 with BCL-xL, and fusing BH3 to the C-terminus (**Figure 1-27**). ciCas9 is self-inhibited at the basal state and is activated by A-385358 in a dose-dependent manner. It is not clear whether ciCas9 is controlled by a steric hinderance mechanism similar to GEFs or by the allosteric effects caused by the BCL-xL-BH3 interaction. The basal activity of ciCas9 could be reduced by using BH3 variants with higher affinity, but the activation is reduced simultaneously. A more potent small molecule disruptor, A-1155463 was used to dissociate the strong CDP pair to achieve high activation[157]. The

same design can be applied to other Cas9 effectors to chemically control base editing, prime editing, and gene transcription[158].

1.1.3.2 Steric blocking of peptides

Although many chemogenetic domains and systems have been developed to control protein functions, chemogenetic control of short peptides is less explored. Peptides could regulate various biological functions and chemically dependent protein domain for regulating peptide activity will expand the design of chemogenetic tools. Protein activity is highly dependent on tertiary structure and conformations. However, short peptide functions (especially for those shorter than 10 amino acids) mostly depend on their sequences and the recognition by their targets. Therefore, peptides could not be controlled using the mechanisms for regulating proteins, such as splitting and conformational change. The existing approach to modulate peptide functions include sterically blocking peptides by another protein domain and using CIP/CDP to control their localizations.

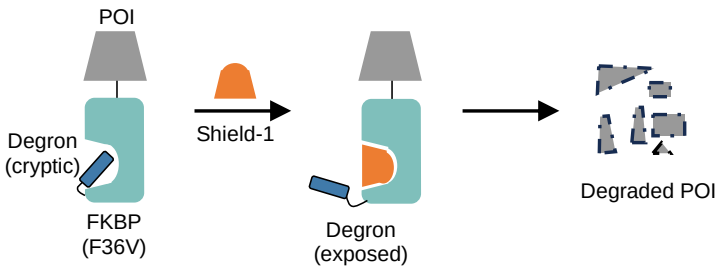


Figure 1-28 LID. Shield outcompetes the degron from the FKBP ligand binding site, leading to the exposure of the degron and the degradation of the POI.

Ligand-induced degradation (LID) domain was designed by chemically regulating the accessibility of a degron peptide[159]. LID domain is an FKBP protein with a C-terminal fusion of a 19-amino-acid degron which binds to the FKBP binding pocket (**Figure 1-28**). The presence of the FKBP ligand shield-1 competes binding and unblocks the degron, leading to protein degradation. It is also determined that the last five amino acids of the degron are responsible for

degradation. The LID system has been successfully applied to control CAR degradation for CAR-T cells in mice[160]. However, the LID system has not been applied to generally modulate peptide functions other than the degron.

1.2 Optogenetics tool design

Optogenetic and chemogenetic tools share the genetically encoding feature but the signals used to control the system are different. By introducing light-sensing domains into genetic systems, protein functions can be optically modulated. To enable control over different types of proteins, a wide range of optogenetic domains and design strategies have been developed. These optogenetic tools also share same mechanisms as chemogenetic ones, which are proximity regulation, conformational and allosteric control, and steric hinderance. Due to this similarity, the most commonly used optogenetic tools are briefly introduced with the same structure as Chapter 1.1, focusing on the light-sensing domains and the designs using them. For more details about optogenetic tools, OptoBase is a good resource summarizing the available optogenetic domains and their applications[161], as well as the reviews here[162-166].

1.2.1 Proximity control

Light induced proximity (LIP) utilizes light to induce a conformational change in a light-sensing domain, and further cause homodimerization or heterodimerization (**Figure 1-29**). The conformational change is usually realized through the cofactor or the chromophore. Similar to CIP, LIP mechanism is also mainly used to 1) relocate POIs to their functional locations or away from them; 2) reconstitute split proteins or inducing interactions of POIs. The inverse process of LIP,

light disrupted proximity (LDP), has also been found in nature. With this mechanism, complementary designs have also been well explored.

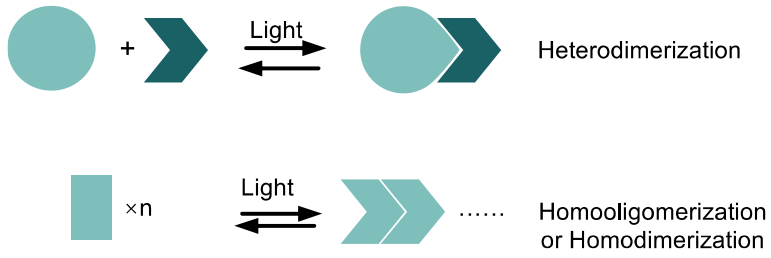


Figure 1-29 Light induced heterodimerization, homodimerization and homooligomerization. The interaction can be reversed by resting in dark or another wavelength of light.

One major difference between the light-controlled and the chemically controlled systems is the reversibility. The light-sensing domains can be switched between two conformation states by light irradiation and dark resting, or by two different wavelengths of light irradiation. This reversibility allows rapid reversible control over dynamic processes. Here in this section, the LIP and LDP domains and the tool designs incorporating them will be discussed.

1.2.1.1 Available optogenetic protein domains for proximity control

A wide range of optogenetic domains have been discovered from microbiomes and plants, and further transferred to mammalian systems. Here in this section, some commonly used LIP and LDP domains will be introduced and categorized by their activation wavelengths.

Light induced proximity pairs

VVD (or Vivid) is a light, oxygen, voltage sensing (LOV) domain derived from *Neurospora crassa*. Within the LOV domain, the Per-Arnt-Sim (PAS) core bearing the flavin adenine dinucleotide (FAD) can be activated by blue light, and results in a cysteine-flavin adduct[167,168]. The conformational change propagates to the N-terminus of the protein and forms a rapidly exchanging homodimer[169] (**Table 1-4**).

To achieve precise control over POIs, heterodimerization is more advantageous. VVD is further engineered to a heterodimerization system Magnets[170] by introducing complementary charged residues to the binding surface of VVD homodimer (**Figure 1-29**). nMag and pMag dimerize with each other upon blue light irradiation, and their homodimerization property is abolished due to the charge (**Table 1-4**). Magnets have low dimerizing efficiency and also require a preincubation at low temperature (28 °C) to allow expression and folding. Magnets were then optimized to avoid these constraints and with higher efficiency[171]. The dimerization of VVD and Magnets can be reversed when rested in dark.

Cryptochrome 2 (CRY2) from *Arabidopsis thaliana* oligomerizes upon blue light irradiation[172,173] (**Figure 1-29**, **Table 1-4**). It was identified that the photolyase homology region (PHR) of CRY2 is sufficient for the oligomerization[172]. Further, the blue light activated CRY2 also dimerizes with cryptochrome-interacting basic-helix-loop-helix (CIB1) protein[174] (**Figure 1-29**, **Table 1-4**). The truncated version of CIB1 without the basic-helix-loop-helix domain (CIBN) is sufficient for dimerization with CRY2[175]. Also, the CRY2-CIBN dimerization can be induced by two-photon stimulation at 860 nm[175]. CRY2-CIBN was further optimized to have reduced dark dimerization activity, and can be tuned to have prolonged or shortened dimerization in response to a pulse of light[176].

Light scattering is a major concern when applying optogenetic tools in non-transparent tissues. Red and near-infrared light sensitive domains were developed to overcome the low tissue penetration problem when using blue light[177]. Phytochromes are light-sensing signaling proteins mediating processes in plants. Red and near-infrared light induces photoisomerization of their tetrapyrrole chromophores, switching the protein between two conformation states of Pr (absorbing red light) and Pfr (absorbing far-red light). Phytochrome B (PhyB) from *Arabidopsis*

thaliana dimerizes with phytochrome interaction factor 3 (PIF3)[178,179] and phytochrome interaction factor 6 (PIF6)[180,181] on its Pfr state (**Figure 1-29**, **Table 1-4**). Phytochrome A (PhyA) from *Arabidopsis thaliana* dimerizes with far-red elongated hypocotyl 1 (FHY1)[182,183] and FHY1 like protein (FHL)[183] on its Pfr state (**Table 1-4**). Both PhyA and PhyB employ phycocyanobilin (PCB) as their cofactor. One common limitation of these two systems is that the large size of phytochromes (> 4.2 kilobases) limits their use in animal models when using adeno-associated viruses (AAVs) as the delivery method[184]. Recently, a truncated version of PhyA (ΔPhyA, 3.2 kilobases) has been developed so that it is compatible with AAV delivery[185].

When applied to mammalian systems, PCB needs to be exogenously supplied or synthesized from heme by introducing extra enzymes[186,187]. To alleviate the need of this extra requirement, bacterial phytochrome P1 (BphP1) and its natural partner PpsR2 from *Rhodopseudomonas palustris* could be used[188] (**Figure 1-29**, **Table 1-4**). BphP1 utilizes endogenously existing biliverdin as its cofactor. PpsR2 tend to oligomerize and also suffers from its large size. Q-PAS1 was further engineered from PpsR2 to be both three-times smaller and non-oligomerizing[189,190].

Plants need to respond to ultraviolet light in order to survive. The *Arabidopsis thaliana* protein UVR8 was identified to respond to ultraviolet light[191]. UVR8 converts from homodimer to monomer upon ultraviolet-B irradiation, and the monomeric UVR8 further dimerizes with constitutively photomorphogenic 1 (COP1)[191,192] (**Table 1-4**). Unlike the light-sensing proteins mentioned above, UVR8 does not contain external cofactors as a chromophore. Instead, it was proposed to utilize the cation-π interactions of two tryptophan residues (Trp285 and Trp233) with the two arginine residues (Arg286 and Arg338) on side chain to induce conformational change[193].

Table 1-4 Summary of LIP pairs introduced in this section, adapted from OptoBase[161]. Note that CRY2 undergo oligomerization rather than dimerization upon activation.

Photoreceptor	Binding partner	Cofactor	Excitation wavelength	Reversion wavelength	Excitation time	Reversion time
VVD	VVD	FAD	450 nm	Dark	seconds	hours
pMag/nMag	nMag/pMag	FAD	450 nm	Dark	seconds	seconds to hours
*CRY2	CRY2	FAD	450 nm	Dark	seconds	minutes
CRY2	CIB1, CIBN	FAD	450 nm	Dark	seconds	minutes
PhyB	PIF3, PIF6	PCB	660 nm	740 nm	milliseconds	milliseconds
PhyA	FHY1, FHL	PCB	660 nm	740 nm	NA	NA
BphP1	PpsR2, Q-PAS1	Biliverdin	760 nm	640 nm or Dark	seconds	seconds
UVR8	COP1	/	300 nm	Dark	milliseconds	hours

Light disrupted proximity pairs

The second LOV domain from *Avena sativa* phototropin 1 (AsLOV2) is one of the most widely used optogenetic domain. Similar to VVD, upon blue light irradiation, the cysteine residue (Cys450) in the PAS core forms a covalent bond with the cofactor flavin mononucleotide (FMN), resulting in conformational change and unwinds the N-terminal A'α helix[194] and C-terminal Jα helix[195,196]. A LDP pair named LOVTRAP was developed utilizing this light induced conformational change in AsLOV2 (**Figure 1-30**, **Table 1-5**). A small protein domain Zdk was developed to selectively bind the dark state AsLOV2[197]. Upon blue light irradiation, Zdk dissociates from AsLOV2. Another LOV domain from *Rhodobacter sphaeroides* (RsLOV) was also employed for LDP (**Table 1-5**). RsLOV homodimerizes in its dark state and blue light irradiation dissociates the homodimer[198].

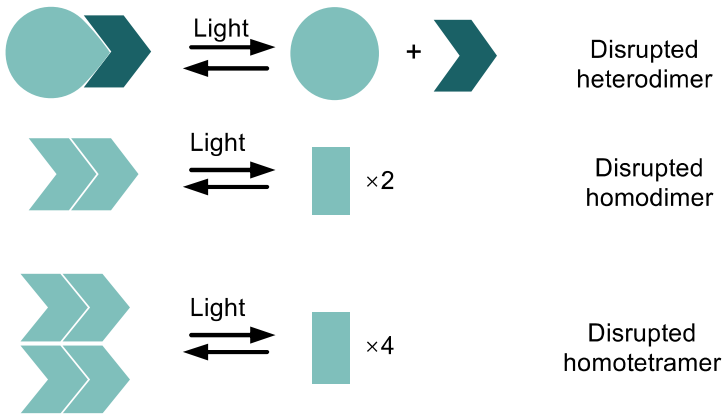


Figure 1-30 Light disrupted proximity. The interaction can be reversed by resting in dark or another wavelength of light.

Green light sensing domains fill in the gap in the wavelength spectrum. The cobalamin-binding domain of the CarH protein from *Thermus thermophilus* (TtCBD) transitions from a homotetramer to monomer when activated with green light (**Figure 1-30**, **Table 1-5**). A fluorescent protein, Dronpa145N, tetramerizes in its off-state, and green light irradiation dissociates the tetramer[199] (**Figure 1-30**, **Table 1-5**). The conformational change is initiated by a cis-trans isomerization of the chromophore[200]. Dronpa145N was further engineered to pdDronpa1, which is dimeric and with brighter fluorescence as well as less aggregation[201] (**Figure 1-30**, **Table 1-5**). Engineered from Dronpa145N, the Dronpa145K dimerizes with Dronpa145N. The Dronpa145K-Dronpa145N fusion protein performs as a homodimer at the off state, and green light irradiation can then dissociate this dimer[199] (**Figure 1-31**, **Table 1-5**).

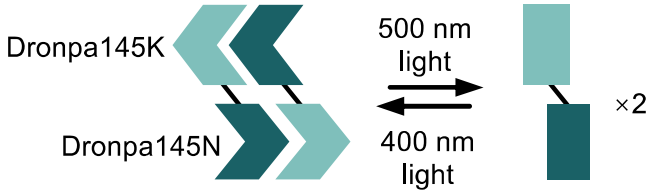


Figure 1-31 Dronpa145K-Dronpa145N as LDP switch.

In addition, as described above, the UVR8 homodimer dissociates when irradiated by ultraviolet light, and can be applied for LDP as well[193] (**Figure 1-30**, **Table 1-5**).

Table 1-5 Summary of LDP pairs introduced in this section, adapted from OptoBase[161]. Note that the ones labeled with * forms tetramer.

Photoreceptor	Binding partner	Cofactor	Excitation	Reversion	Excitation time	Reversion time
AsLOV2	Zdk	FMN	450 nm	Dark	seconds	seconds to minutes
RsLOV	RsLOV	FMN	450 nm	Dark	NA	NA
*TtCBD	*TtCBD	Adenosylcobalamin	545 nm	Dark	NA	NA
*Dronpa145N	*Dronpa145N	/	500 nm	400 nm	seconds	seconds
pdDronpa1	pdDronpa1	/	500 nm	400 nm	seconds	seconds
Dronpa145N/K	Dronpa145K/N	/	500 nm	400 nm	seconds	seconds
UVR8	UVR8	/	300 nm	Dark	milliseconds	hours

In summary, LIP and LDP pairs controlled by different wavelengths of light have been developed. The major advantage of such optogenetic domains over chemogenetic ones is that they are usually reversible (either by resting in dark or by another wavelength of light), so that no extra engineering is required to achieve reversible activation in optogenetic systems. When using these domains, the LIP or LDP pair can be selected based on required kinetics and desired stimulation light wavelength. Further, the kinetics can also be tuned by protein engineering[176].

1.2.1.2 Applications

LIP and LDP could be utilized in different applications with similar designs as CIP and CDP. Two of the main advantages of optogenetic systems is the fast kinetics and the reversibility. Here in this section, the typical designs using LIP and LDP are introduced.

Protein transportation

With similar designs as CIP and CDP, protein transportation can be achieved by light-sensing domains as well.

POI membrane localization could be achieved by anchoring one LIP domain on cell membrane and fusing POI with the other LIP domain. When using CRY2/CIBN[175], the POI diffuses away on order of minutes, and membrane translocation can be induced by blue light again. When PhyB/PIF3[179] and BphP1/PpsR2[188] are used, the depletion of POI from the membrane can be induced by infrared light.

The same strategy can be applied for controlling nuclear-cytoplasmic distribution as well. By using NLS as the nuclear localization tag, one of a LIP domain can be anchored in nucleus. Upon stimulation with the corresponding light, POIs fused with the other LIP domain can be transported into nucleus. This has been achieved by using BphP1/PpsR2[188] and UVR8/COP1[202].

Membrane depletion can also be achieved by using LDP. By anchoring one of a LDP domain on membrane, the POI fused to the other LDP domain could be located to the membrane as well. Upon stimulation with the corresponding light, the POI will be dissociated from the anchored membrane. AsLOV2/Zdk has been used to deplete POI from outer mitochondria membrane[197]. The homotetramerizing Dronpa145N and homodimerizing UVR8 has been used to deplete POI from plasma membrane[199,203].

Beyond proteins, LIP can be further applied to transport organelles as well. Some examples include transporting mitochondria[204], lysosome[204], endosome[205] and peroxisome[204,205] with dynein and kinesin. When PhyB/PIF6 is used, the transportation could be rapidly stopped by infrared light irradiation[205].

LIP can be further used to develop light induced artificial subcellular compartment formation. CRY2 oligomerize upon blue light excitation, providing unique advantages for designing such systems. When fusing intrinsically disordered region (IDR) to CRY2, protein aggregates can be efficiently induced[206]. However, this system induces random aggregation. To

precisely control the clustering, CRY2/CIBN can be used instead. By fusing CIBN to an ordered multimer forming core, CRY2 can then oligomerize and recruit the cores together (**Figure 1-32**)[207,208]. Such system could efficiently trap proteins into these artificial compartments and perturb POI functions[207,208]. Similar to the example discussed in Chapter 1.1.1.4 using FKBP(F36M) clustering to control protein secretion, LDP can also achieve similar outputs. By fusing two UVR8 repeats to POI, the fusion protein forms aggregates and are stuck in the ER. Ultraviolet light irradiation dissociates the aggregates and allows the POI to be secreted[203].

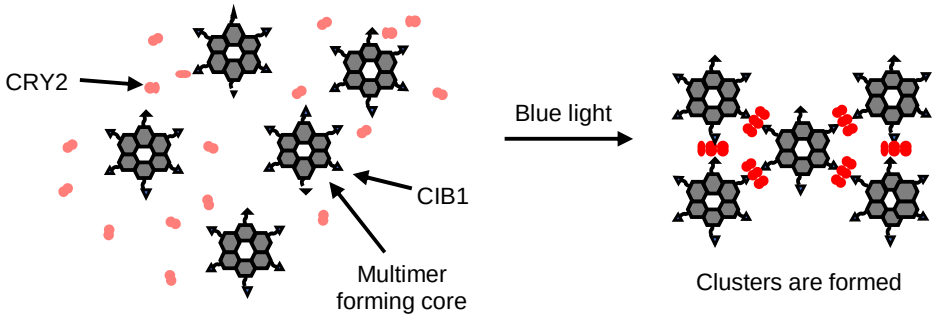


Figure 1-32 Formation of artificial subcellular compartment. CIB1 is fused to a protein that forms multimer. Blue light oligomerizes CRY2 and bring the cores together.

Cell signaling

LIP and LDP can be applied to control dynamic cell signaling with similar designs as the chemogenetic ones: 1) activating POIs via dimerization or reconstitution of the split POI; 2) recruiting POIs to the correct locations for function.

CRY2 has been applied to oligomerize many different POIs to induce cell signaling, including: LRP6 C-terminal domain for β-catenin signaling[172]; RhoA for Rho GTPase signaling[172,209]; N-terminal src-homology 2 (SH2) domain from PLC-ɣ for receptor tyrosine kinase (RTK) signaling[209]; a guidance receptor named Deleted in Colorectal Cancer[210]; STIM1 for Ca^{2+} modulation[211]; an RTK tropomyosin- related kinase[212]. VVD has also been applied to dimerize an

RNA binding protein, transcriptional antiterminator protein termed LicT, for regulating RNA metabolism[213]. The inverse process can be achieved as well by using LDP. CBD has been applied to dissociate a homodimer RTK, murine fibroblast growth factor receptor1[214]. LOVTRAP has been applied to dissociate split protein tyrosine phosphatase 1B (PTP1B)[215].

LIP can be applied to translocate POIs to their active locations to induce cell signaling. PhyA/FHY1 has been applied to translocate SOScat to membrane for MAPK signaling[185]; BphP1/PpsR2 can induce recruitment of a DHPH domain of intersectin 1 to the plasma membrane for activation of the small GTPase Cdc42[188]. In another design, LDP is applied to deplete POIs from their inactive locations and further induce signaling. By this strategy, LOVTRAP has been applied to control Vav2, Rac1, and RhoA activity[197].

Gene expression

Light induced gene expression can be achieved with similar designs as described in Chapter 1.1.1.4: 1) controlling two-hybrid systems; 2) controlling dCas9 systems; 3) controlling recombinases.

Two-hybrid systems could be controlled through manipulating the DBDs as well as the recruitment of AD. The DBDs can be modulated by controlling their location and dimerization.

A functional TF TtCBD-TetR-VPR was located away from nucleus by another TtCBD anchored on plasma membrane and is therefore not functional[216]. Green light stimulation disrupts the TtCBD homotetramer and allows the TF to enter the nucleus and initiate gene transcription. A similar design has been applied by using BphP1/PpsR2, where the AD VP16 was located in nucleus by tagging with NLS, and the TetR was located in cytoplasm. Red light induces translocation of TetR into nucleus and further initiates gene transcription[188].

Another way to control the DBD is to modulate their homodimerization. Some DBDs, including Gal4, LexA and TetR, require being in their dimeric form to bind to their target DNA sequence (**Figure 1-33**). RsLOV was fused to LexA[217] and TetR[218] to control this process. In the dark state, RsLOV homodimerizes and retain LexA and TetR in their functional dimeric form. Blue light irradiation dissociates RsLOV, and the conformational change in RsLOV perturbs the homodimer, inactivating the DBD (**Figure 1-33**). An inversed strategy has been applied to induce the formation of the homodimeric Gal4-DBD[219]. By truncating the homodimerizing domain from Gal4-DBD, the DNA-binding property is then solely controlled by the LIP domains fused to it. The resulted truncated-Gal4-VVD fusion protein only binds to the DNA upon blue light irradiation (**Figure 1-33**).

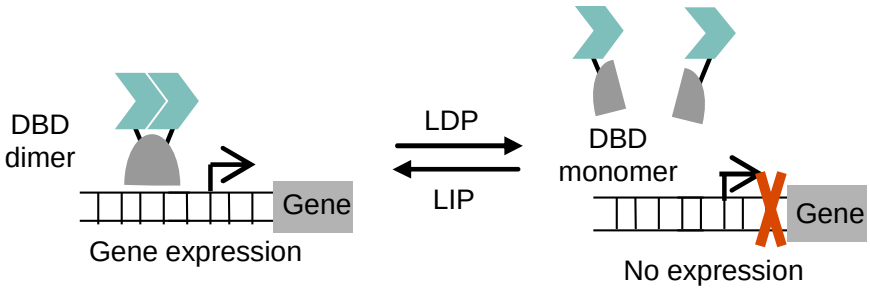


Figure 1-33 The DBD needs to be in dimeric form to bind to DNA. The constitution of DBD can be controlled by LDP or LIP.

In a typical two-hybrid system design (**Figure 1-17**), AD is dimerized with DBD to trigger gene transcription. In addition to the chemogenetic versions introduced in Chapter 1.1.1.4, LIP can also be applied to achieve same results. Gal4-DBD and TetR are the two typical DBDs to use. Since it requires a certain length of time for gene transcription, the LIPs with longer reversion time are used. The dimerization of DBD with AD can be induced by blue light with CRY2/CIBN[175,176,220-222]; by red light with PhyA/FHY1[185], PhyB/PIF3[178,223], PhyB/PIF6[181,223], and BphP1/Q-PAS1[190]; by ultraviolet light with UVR8/COP1[202]. It is worth to note that when using TetR as the DBD, the gene transcription can be regulated by both Dox and light[221].

Cas-protein-based gene expression can also be regulated with a similar fashion. LIP can be used to reconstitute split Cas proteins. nMag/pMag has been applied to control split Cas9[224] and Cas12a[225] reconstitution by blue light. In dCas9-based two-hybrid systems, CRY2/CIB1 has been applied to recruit AD by blue light[226-228].

Controlling split recombinases offers another way for gene expression and genome engineering. Split Cre recombinase could be controlled by CRY2/CIBN[175,176,229], nMag/pMag[230], and VVD[231]; split Flp has been controlled by nMag/pMag[232] and VVD[231]; split Dre has been controlled by VVD[231].

1.2.2 Conformational and allosteric control

As discussed above, the proximity controlling systems are rooted from the conformational change of the optogenetic domains induced by light. This conformational change is usually more significant than those of the chemogenetic domains (e.g., the AsLOV2 domain unwinds it's A'α and Jα helix upon blue light irradiation), which makes them suitable for modulating POIs allosterically. Further, a wide range of light-sensing membrane receptors exist in microbial systems and have been transferred as important tools for neuron activity modulation. Here in this section, optically controlled membrane receptors and allosteric switches are discussed.

1.2.2.1 Membrane receptors

Neuroscience is arguably the field where optogenetic tools are mostly applied due to the precise and rapid control over cellular events provided by these tools. Photoreceptors have opened doors to a wide array of applications, from probing neural circuits and studying cell signaling to dissecting the complicated animal behaviors. The detailed applications of these receptors have

been extensively reviewed[233-238], a brief overview of these tools will be provided here in this section. The currently used photoreceptors can be roughly divided into two categories[233]: 1) type I rhodopsins directly regulating membrane voltage by ion transportation; 2) type II rhodopsins regulating GPCR signaling.

Photoreceptors regulating ion transportation

Type I rhodopsins are often also referred to as microbial rhodopsins and are transferred from archaea, bacteria, algae, and fungi[233]. Due to their function of generating ion gradient, type I rhodopsins were transferred to neurons to modulate membrane voltage. These photoreceptors function through the light-induced isomerization of the retinal chromophore from 11-cis to all-trans conformation, which leads to conformational change in rhodopsin[239]. Based on the mechanisms, type I rhodopsins can be further divided into ion pumps and ion channels.

For ion pumps, the photoactivation of the receptors induces the transportation of H^+ or Na^+ out of the cells or the transportation of Cl^- into the cells, resulting in hyperpolarization and suppression of neurotransmitter release.

Archaerhodopsin-3 (Arch) from *Halobacterium salinarum* is a yellow-light-sensing H^+ pump[240]. Arch has been optimized to be a more sensitive version ArchT[241] as wells as eArchT to trigger greater photocurrent[242]. Prolonged activation of H^+ pump perturbs H^+ concentrations in neurons and results in intracellular pH values out of the typical physiological ranges, leading to suppression of firing and rebound firing.

To avoid such drawbacks, Na^+ pumps and Cl^- pumps were applied to hyperpolarize neurons. KR2 is a green-light-sensing Na^+ pump from *Krokinobacter eikastus*[243]. Halo[244] and NpHR[245] are both yellow-light-sensing Cl^- pumps from *Natronomonas pharaonic*. NpHR was further optimized to eNpHR2.0 with better membrane trafficking and less aggregation[246].

eNpHR3.0 was further developed to have 20-fold stronger inhibition than the original version and can be activated by far-red light as well[247].

Ion pumps require high irradiation intensity and high protein expression level to achieve high efficiency because only one ion is transported upon one photon absorption. In contrast, channelrhodopsins (ChRs) are light-sensing ion channels that transport ions passively and have higher efficiency. Although ChRs are discussed here after ion pumps, they are considered as the birth of using optogenetic approach in neuroscience, and are the most well-known optogenetic tools.

Channelrhodopsin-1 (ChR1)[248] and channelrhodopsin-2 (ChR2)[249] from *Chlamydomonas reinhardtii* represent the first ChRs. Blue light activated ChR2 conducts H^+ and Na^+ ions, and the influx of these ions further leads to depolarization of neurons[250-252]. Many ChR variants have been engineered for broader applications (reviewed here with more details[238]). Here, the several engineering directions are briefly introduced.

ChR2 was optimized to have increased photocurrent. Some examples include ChR2-H134R[253], ChR2-T159C[254], and ChR2-XXL[255]. An alternative approach to reach higher ion conducting efficiency is to reduce the reversion kinetics. Bistable step-function-rhodopsins (SFRs) were engineered to have a prolonged time staying in the open state after being activated. Mutants on ChR2 C128 and D156 residues at the retinal binding pocket significantly reduce the reversion kinetic, and keeps the open state for a longer time[255-258].

As complimentary tools for SFRs, ChRs with faster kinetics were also developed. The wild type ChR2 cannot successfully trigger spikes at frequencies greater than 40 Hz. Some mutants with faster kinetics include ChR2-E123T/T159C[254] (up to 60 Hz) and ChETA[259] (up to 200 Hz). Chrono, a ChR from *Stigeoclonium helveticum*, also provides faster kinetics[260] (up to 60 Hz).

To achieve multiplexed control over multiple neuron populations, ChRs sensing different wavelengths of light are desired. Chrono can be activated by blue and green light[260]. Chrimson, derived from *Chlamydomonas noctigama*, responds to red light[260]. PsChR from *Platymonas subcordiformis* is blue shifted and allows activation below 400 nm light[261]. VChR1 from *Volvox carteri* is red shifted and can be activated by yellow light[262]. VChR1 is further engineered to ReaChR, which can be activated by orange to red light[263]. C1V1 is a chimeric ChR between ChR1 and VChR1 that could be activated by green to orange light[258,264].

ChRs can be engineered to conduct currents with other ions. CatCh was engineered from ChR2[265] and has increased conductivity towards Ca^{2+}. The ChRs mentioned above all conduct cations and depolarize neurons. ChRs can also be engineered to conduct Cl^- anion to hyperpolarize neurons. With the starting point of a ChR1/ChR2 chimera, C1C2[266], the conductance of cations can be reversed to Cl^- anion by replacing the negatively charged Glu90 residue with positively charged residues Arg and Lys, yielding a chloride channel ChloC[267]. The systematic replacement of Glu residues in C1C2 conduction pathway also yielded another chloride channel iC1C2[268]. ChloC was further improved to iChloC[269]; iC1C2 was further improved to iC++ and SwiChR++[270]. Although successfully engineered to conduct anions, these ChRs still retain some cation conductance and were engineered to be more sensitive at the expense of slower kinetics. Natural anion channel rhodopsins (ACRs) were then identified from other species. ACR from *Guillardia theta*[271] and *Proteomonas sulcate*[272,273] can be activated by blue and green light, respectively.

To overcome the low tissue penetration of blue light, BRET offers an alternative way to activate ChRs. A class of luminescent ChRs, luminopsins, was developed. By fusing *Gaussia princeps* luciferase (GLuc) to ChR2's extracellular N-terminus, the supplement of coelenterazine substrate activates ChR2 and lead to neuron depolarization[274]. The efficiency of this luminopsin

can be further optimized by improving the GLuc's brightness[275]. Similarly, an inhibitory luminopsin was engineered by fusing *Renilla reniformis* luciferase (Rluc) to NpHR's intracellular C-terminus[276]. Coelenterazine substrate supplement leads to hyperpolarization of neurons.

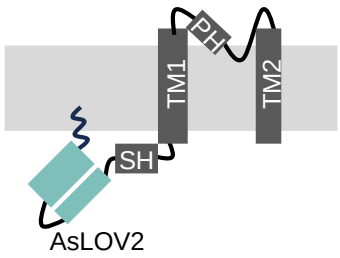


Figure 1-34 Design of BLINK. Kcv comprises slide helix (SH), pore-helix (PH), and transmembrane domains (TM1 and TM2). Wiggle line is the myristoylation/palmitoylation sequence anchoring AsLOV2 to the membrane.

Besides using rhodopsins, an alternative approach to optically modulate membrane voltage is to use optogenetic module AsLOV2 to induce conformational change in ion channels. Blue-light–induced K^+ channel (BLINK)[277] was designed based on a small viral K^+ channel Kcv[278] (**Figure 1-34**). By fusing the N-terminal slide helix to AsLOV2 that is anchored on plasma membrane by myristoylation/palmitoylation sequence, blue light irradiation induces the conformational change in AsLOV2 and activates the Kcv and hyperpolarize neurons. BLINK2 was further optimized to show better expression level in neurons[279]. In addition, BLINK2 stays in the activated form over tens of minutes in dark condition and can thus be applied to freely moving rats.

Photoreceptors regulating GPCR signaling

Type II rhodopsins are light-sensing GPCRs that mediate visual process in animals. Vertebrate rhodopsin 4 (RO4), a rod opsin, has been applied to activate postsynaptic K^+ currents and inhibits presynaptic Ca^{2+} currents by $G_{i/o}$ signaling[252]. Since RO4 suffers from bleaching and reduced activity after repeated stimulation, opsins from visual cones were applied to sense other

wavelengths of light and enable repetitive activation[280]. An invertebrate opsin from box jellyfish *Carybdea rastonii*, JellyOP, is bleach resistant and induces G_s signaling[281]. Melanopsins, Opn4 from human (hOpn4L) and mouse (mOpn4L), can be activated and deactivated by blue and yellow light. hOpn4L and mOpn4L trigger $G_{i/o}$ and G_q signaling, respectively[282].

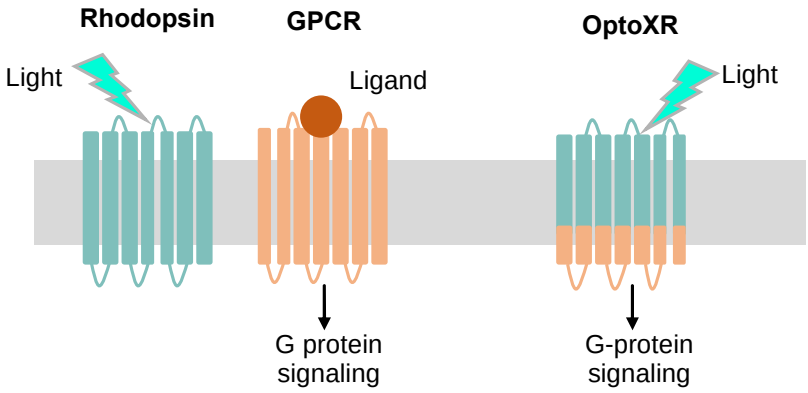


Figure 1-35 Design of OptoXR. The light sensing section is the transmembrane domains and the extracellular loops of rhodopsin. The G protein binding section is the intracellular parts of the GPCR with desired signaling.

To expand the signaling pathways that type II rhodopsins can induce, chimeric GPCR strategy (also known as OptoXR) has been applied (**Figure 1-35**). As introduced in Chapter 1.1.2.1, the GsD DREADD was developed by combining the ligand binding section of a GPCR with the intracellular G protein binding section of another GPCR. With a similar strategy, when combining the light-sensing section of rhodopsins with G protein binding section of other GPCRs (e.g., β2AR, μOR, etc.), more light-sensing GPCRs triggering different downstream signaling pathways could be developed. The first OptoXR, Opto-β2AR was designed by building the chimera of bovin rhodopsin and hamster β2AR[283,284], resulting in a G_s-coupled OptoXR. Same strategy was applied to the G_q-coupled human α1a-adrenergic receptor (α1AR) to yield Opto-α1AR[284].

The same strategy has been applied to generate many other OptoXRs: $5HT_{1A}$ chimera with RO4 for $G_{i/o}$ signaling[285]; μOR chimera with RO4 for $G_{i/o}$ signaling[286,287]; dopamine receptor D1 (DRD1) chimera with rhodopsin for G_s signaling[288,289]; dopamine receptor D2 (DRD2) chimera

with rhodopsin for $G_{i/o}$ signaling[289]; chemokine C-X-C motif receptor 4 (CXCR4) chimera with rhodopsin for G_i signaling[290]; adenosine A_{2A} receptor ($A_{2A}R$) with rhodopsin for G_s signaling[291]; muscarinic acetylcholine receptor M1, 2, and 3 (mAChR1, 2, 3) with rhodopsin for G_q, G_i, and G_q signaling, respectively[289]; metabotropic glutamate receptor 6 (mGluR6) chimera with melanopsin for G_i signaling[292]; 5-HT_{2A} receptor chimera with melanopsin for G_q signaling[293]. More importantly, the OptoXR strategy could be applied to orphan GPCRs and would have the potential to study their biological functions[289].

By introducing GPCR α-helical residues that form contacts with G_α protein, Opto-β2AR was further optimized to Opto-β2AR-2.0 with 20-fold signaling increase[294]. With the same strategy, the Opto-$A_{2A}R$ was also optimized to Opto-$A_{2A}R$-2.0[295].

Genetically encoded membrane-tethered ligand

Genetically encoded ligand can be displayed on cell membrane and can be controlled optogenetically (**Figure 1-36**). The function of K^+-channel-specific peptide toxin, α-dendrotoxin (αDTX), can be optically controlled by being fusing to AsLOV2, so that voltage-gated K^+ (Kv) channels can be modulated[296]. Under the dark condition, αDTX is restricted to cell membrane surface and inhibits Kv. The light irradiation causes AsLOV2 Jα helix unfolding, and leads to a more flexible αDTX displayed on membrane, reducing the local concentration of αDTX and unblock Kv.

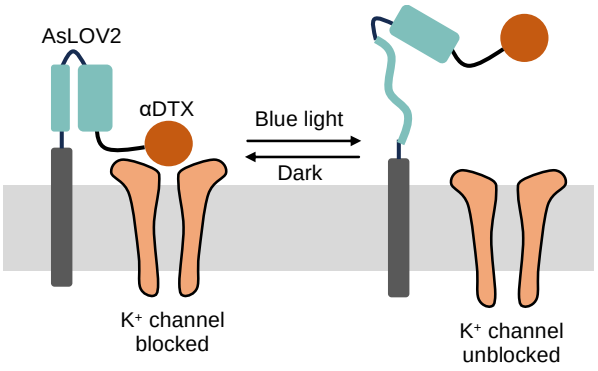


Figure 1-36 The inhibitor αDTX and AsLOV2 are displayed on cell membrane. The αDTX blocks the K^+ channel due to high local concentration. Blue light irradiation unfolds the AsLOV2 Jα helix, reduces the local concentration, and unblocks the K^+ channel.

1.2.2.2 Allosteric switch

As discussed above, optogenetic control over protein conformations can be achieved by transferring light-sensing functional proteins from other biological systems as well as engineering the light-sensing sections into other proteins based on homology. For proteins lacking such properties, allosteric switch insertion is a more generalizable approach. Here in this section, the developed optogenetic allosteric switches are introduced.

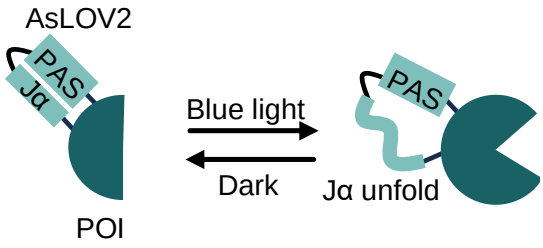


Figure 1-37 Scheme of AsLOV2 insertion into POI. Whether the POI gets activated or deactivated upon blue light irradiation varies from case to case.

AsLOV2 is the most widely used allosteric switch (**Figure 1-37**) for two reasons: 1) the blue light activation of AsLOV2 leads to a big conformational change, in which the N-terminal A'α helix[194] and the C-terminal Jα helix[195,196] are unwound; 2) the N-terminus and C-terminus are

close to each other, making it suitable for inserting into loops of POI without completely disrupting POI's structure. However, the output of such control cannot be well predicted. Whether the protein inserted with AsLOV2 gets activated or deactivated upon light irradiation varies from case to case.

AsLOV2 can be inserted into loops of protein binders to modulate their binding activities (**Figure 1-37**). By inserting AsLOV2 into AcrIIC3, an inhibitor protein of NmCas9, the activity of NmCas9 can be inhibited upon light irradiation[297]. Intrabodies are antibody-like proteins that work within live cell that bind to intracellular proteins. Optogenetic control over intrabodies provides temporal control on target proteins. Two classes of intrabodies, nanobody and monobody, have been recently reported to be optogenetically controlled by AsLOV2 domain insertion. The two light-controlled nanobodies, the mCherry binding LaM8-AK74 and the actin binding OptoNB, can reversibly bind to their targets with a light-switchable fashion[298]. The light-controlled monobody binds the SH2 domain of Abl kinase upon light irradiation[299].

Beyond protein binders, enzymes can also be regulated with the same fashion (**Figure 1-37**). AsLOV2 has been inserted into several proteins to optogenetically modulate cell signaling, including Src[300], Rac1[300], Vav2[300], GEF-H1[300], intersectin1[300], protein tyrosine phosphatase 1B (PTP1B)[215], T cell protein tyrosine phosphatase (TCPTP)[215], and pyruvate kinase M2[301]. For all the examples above, light irradiation deactivates the POIs. This approach can also offer optogenetic control over tool enzymes. AsLOV2 inserted DHFR[302] and a proximity labeling enzyme, LOV-Turbo[303], have been developed to have precise spatiotemporal control of enzyme activities. These two enzymes are activated upon light irradiation.

VVD can serve as allosteric switches as well. By fusing two VVD proteins together, the light irradiation will cause dimerization and lead to a change in the distance between the N- and

C-terminus. This switch has been applied to control several kinases including Src, ABl, b-Raf, and the Cre recombinase by the same insertion design[304].

Besides the insertion strategy, the rigid dark state conformation of AsLOV2 (fully folded) provides opportunities for allosteric control by direct fusion. The allosteric effect can be caused by 1) the interface between the POI and AsLOV2 (**Figure 1-38**); 2) the unwinding of the A'α helix and Jα helix of AsLOV2 can propagate to the fusion α helix of POI (**Figure 1-39**).

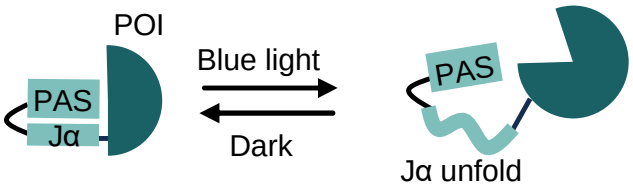


Figure 1-38 Scheme of AsLOV2 fusion strategy. The allosteric effect is caused by the interface between AsLOV2 and POI.

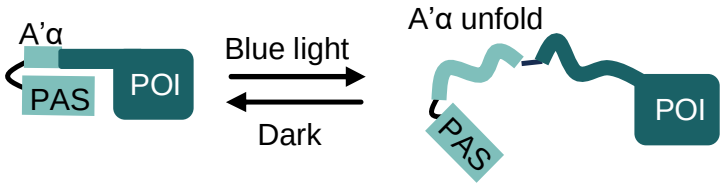


Figure 1-39 Scheme of AsLOV2 fusion strategy. The allosteric effect is caused by the propagation to the POI α helix.

By fusing AsLOV2 to the N-terminus of Rac1, the resulted photoactivatable-Rac1 (PA-Rac1) activity is blocked under dark condition[305] (**Figure 1-38**). Blue light irradiation unwinds the Jα helix and relieves Rac1 from being blocked. The mechanism had been initially proposed to be steric unblocking by the removal of the AsLOV2 PAS domain when first being developed, but was later identified to be the allosteric effect from the AsLOV2-Rac1 contact interface[306]. The PA-Rac1 is a successful optically controlled kinase and has been applied to multiple in vivo studies including drosophila[307], zebrafish[308], and mouse brain[309,310]. The same strategy has been applied to the bacterial transcription factor trp repressor (TrpR) and the catalytic domain of caspase 7.

TrpR gets activated upon light irradiation and binds to its operator DNA, protecting DNA from nuclease digestion[311]. When activated with blue light, the caspase 7 becomes functional and induces apoptosis[312].

The conformational change of AsLOV2 A'α helix and Jα helix can be used to manipulate the connected α helix of POI (**Figure 1-39**). When fused to the C-terminus of PTP1B, the light-induced unwinding of AsLOV2 A'α helix destabilizes the continuous α7 helix in PTP1B, so that PTP1B gets partially unfolded and deacticated[313]. TCPTP is a homologue of PTP1B and their catalytic domains share 70% sequence identity. This strategy was then applied to TCPTP by fusing AsLOV2 to the α7 helix of TCPTP. The resulted TCPTP gets deactivated upon light irradiation[215,314].

While the strategy of allosteric switch insertion seems generally applicable for many proteins, the light dependence is hard to predict. As shown above, light irradiation could lead to either activation or deactivation for different POIs. Same results have been observed even in the process of screening for insertion sites[298,303], where some insertion sites provide positive dependence and some lead to negative dependence.

1.2.3 Steric hinderance control

Steric hinderance blocking can be achieved by optogenetic domains as well. The strategy discussed in Chapter 1.1.3.1, where CDP pairs block protein active sites, can be replaced by LDP pairs and achieve similar results. In addition, AsLOV2 domain can be directly fused to POIs to introduce steric hinderance, and therefore block the POIs from functional forms. As a comparison to Chapter 1.1.3.2, where limited chemogenetic approaches have been reported to modulate

peptide functions, AsLOV2 has been generally applied to cage a wide range of peptides. Here in this section, the optogenetic designs for modulating steric hinderance will be introduced.

1.2.3.1 Steric unblocking of protein

Steric unblocking of proteins can be achieved by two design strategies: 1) AsLOV2 direct fusion to POI (**Figure 1-40**); 2) a LDP pair fused to both termini of POI (**Figure 1-41**).

The AsLOV2 C-terminal Jα helix is unfolded in its light state. When the Jα helix is fused to the POI, the steric hinderance from the PAS domain is removed upon light irradiation. The change in steric hinderance can therefore be used to modulate POI activities.

When fused to the N-terminus of the Cre recombinase, AsLOV2 creates the steric hinderance to block the formation of the Cre tetramer functional form (LiCre)[315] (**Figure 1-40**). Blue light irradiation removes the steric hinderance and activates the Cre recombinase. Compared to the split recombinase strategy, LiCre showed faster and stronger activation by light as well as a lower residual activity in the dark. In addition, LiCre is a single component system, so it is easier to be introduced into animal models as well.

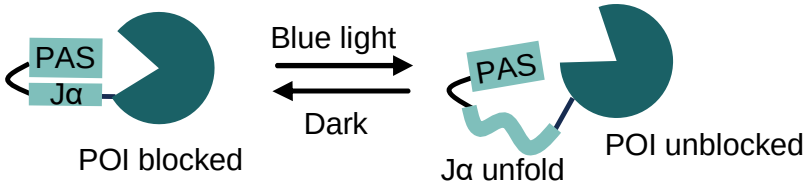


Figure 1-40 Scheme of AsLOV2 controlling steric hinderance. Blue light irradiation unfolds Jα helix and removes the hinderance.

Beyond perturbing self-multimerization, the same strategy can be used to block POI's interaction with its target to control POI activity. Light irradiation unblocks the fusion POI and activates it (**Figure 1-40**). With this design, AsLOV2 has been fused to STIM-ORAI activating

domain (SOAR) to control ORAI channel signaling[316], as well as the repressor element 1 silencing transcription factor (REST) proteins to regulate neural differentiation[317].

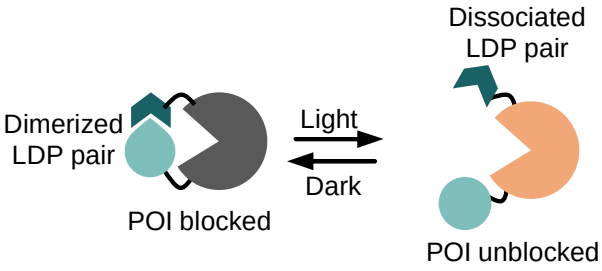


Figure 1-41 Scheme of using LDP to block protein active site. Light irradiation dissociates the LDP pair and removes the hinderance.

LDP pair can be fused to the two termini of a POI to block its active site. Light irradiation dissociates the LDP pair and activates the POI (**Figure 1-41**). This design using LOV/Zdk as the LDP was named Z-lock. Z-lock has been shown to reversibly control the activity of cofilin and a tubulin acetylase αTAT[318]. In addition, the dimeric Dronpa145N-Dronpa145K fusion protein and dimeric pdDronpa have been incorporated into this design, controlling Cdc42 GEF intersectin[199], hepatitis C virus NS3-4A protease[199], MEK1[201], RhoA GEF[319], Cas9[320], and IRSp53-MIM homology domain[321]. When using the Dronpa145N-Dronpa145K fusion protein as LDP pair, there will be in total four Dronpa proteins fused to POIs, while the use of pdDronpa results in two Dronpa proteins fused and is more efficient.

1.2.3.2 Steric blocking of peptides

As discussed in Chapter 1.1.3.2, peptide functions are less dependent on their tertiary structure, and could not be modulated through typical protein engineering approaches such as splitting proteins and allosteric switching. The photoswitching mechanism of AsLOV2 provides unique advantage to generate steric hinderance, which has made it to be the most widely used

optogenetic domain for designing photoswitchable peptides. The functional peptides could be fused to Jα helix so that the peptide will be sterically block under the dark state and unblocked under the light state (**Figure 1-42**). Depending on the peptide it cages, AsLOV2 has been applied to control dimerization, enzyme inhibition, protein degradation, protein transportation, and protein splicing. Here in this section, the LOV domain based photoswitchable peptides will be introduced.

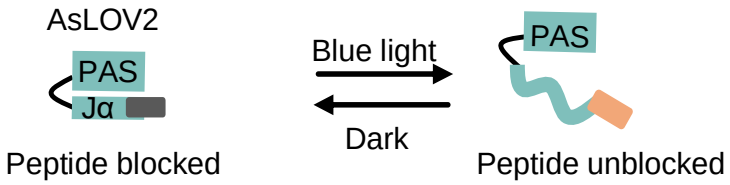


Figure 1-42 Scheme of using AsLOV2 to block peptides. Light irradiation unfolds Jα helix and activates the peptide.

Two AsLOV2-based dimerization pairs have been developed by caging binder peptide. TULIP was developed by appending a peptide epitope (SSADTWV) to the Jα helix[322]. Blue light irradiation unblocks the peptide epitope and further induces dimerization with its binder, an engineered PDZ domain (ePDZ)[322]. However, PDZ domains also exists in endogenous systems, and could lead to cross-talk with endogenous signaling pathways[323]. With the same design, AsLOV2 was applied to cage peptides that do not exist in mammalian systems. ipaA is a vinculin binding peptide from the invasin protein, and SsrA peptide is from *Escherichia coli*, which binds the protease delivery protein SspB. By appending ipaA or SsrA peptide to Jα helix, light-induced dimerization can be achieved[324]. The initial version of AsLOV2 caged SsrA showed 8-fold light dependence on dimerization, and was further improved to iLID showing over 50-fold dependence[325]. Both TULIP and iLID has been applied to some typically used proximity-based designs. For example, TULIP for peroxisome transportation[326], iLID for dimerizing SOAR and STIM1 to induce ORAI signaling[316], iLID for reconstituting a split nanobody[327], iLID for inducing protein aggregation[328], iLID for reconstituting split SNAP-tag for induced protein labeling[329].

A head-to-head comparison has been conducted on iLID and TULIP[330]. Depending on the SspB mutant, iLID has an affinity of 0.13 μM (light) and 4.7μM (dark) when using high affinity SspB, and an affinity of 0.8 μM (light) and 47 μM (dark) when using low affinity SspB. TULIP has an affinity of 18 μM (light) and 150 μM (dark) towards ePDZ.

AsLOV2 can also induce covalent dimerization by caging SpyTag. When induced with blue light, the SpyTag is unblocked and then covalently dimerizes with SpyCatcher[331].

Beyond dimerizing peptides, AsLOV2 has been applied to cage more functional peptides for diverse applications. AsLOV2 has been applied to cage enzyme modulating peptides, including cyclic-AMP dependent kinase inhibitory peptide (PKI)[332], myosin light chain kinase inhibitor peptide 18 (MKI)[332], G-protein activating peptide α-binding-and-activating (GBA) motif[333], and tobacco etch virus protease cleavage site (TEVcs)[334-338].

As described in Chapter 1.1.1.4 and Chapter 1.2.1.2, protein translocation requires two components when using proximity controlling mechanism, with one fused to POI and the other anchored at a subcellular location. By caging NLS or NES, AsLOV2 can induce protein nucleus-plasma transportation with only one component. Examples include LINuS (AsLOV2 caged NLS)[339], LANS (AsLOV2 caged NLS)[340,341], and LEXY (AsLOV2 caged NES)[342].

LOV domain can also regulate protein existence by controlling stability and splicing. Protein splicing can be controlled by AsLOV2 through caging a split intein peptide[343]. Light activation unblocks the split intein peptide and allows the reconstitution of intein, further leading to protein splicing and generation of full length POI. The peptide degron developed in the ligand-induced degradation system LID (discussed in Chapter 1.1.3.2) has been transferred to be caged by AsLOV2[344]. Light exposure unblocks the degron and induces protein degradation. AtLOV2 from *Arabidopsis thaliana* was used to cage murine ornithine decarboxylase-like degradation

sequence cODC1[345]. Another special example utilized the N-end rule. Ubiquitin and a subsequent Arg residue were fused to the N-terminus of AsLOV2[346]. The cleavage of ubiquitin after protein translation leaves an N-terminal Arg residue on the AsLOV2-POI. The blue light irradiation unwinds the A'α helix and exposes the Arg residue, which leads to degradation of the whole protein due to the N-end rule.

In summary, AsLOV2 domain is arguably the most versatile optogenetic domain partially due to its general ability to cage peptides.

1.3 Comparison of chemogenetic and optogenetic tools

Here in this chapter, I discussed three mechanisms of chemogenetic and optogenetic tool design based on proximity, conformational and allosteric control, and steric hinderance. The available chemical- and light-sensing protein domains offer various design strategies for manipulating biological processes and have provided valuable methods to study biological questions.

Chemogenetic and optogenetic tools share the same "genetic" feature and differ in the signal used to control the systems. The "genetic" feature enables the precise delivery of the systems to specific cell types, neuronal circuits, and tissues. Some methods for selective gene delivery and expression include: 1) cell-type specific promoters[347]; 2) Cre-*loxP* system[348]; and 3) target-specific virus[349,350]. These methods could control the genetic tool expression in desired cell populations and thus minimize the influence and side effect on unrelated cells. Further, as mentioned in Chapter 1.1.1.4 and Chapter 1.2.1.2, the proteins could be directed to desired subcellular organelles and locations, further increasing the spatial resolution of the tools.

The difference between these two classes of tools comes from the use of either small molecules or light, which determines their suitable applications for different purposes. When choosing from the two classes of tools, there are several factors that could be taken into concern: 1) spatial resolution; 2) temporal resolution and reversibility; 3) delivery of chemical and light; and 4) side effects.

Spatial resolution. Spatial resolution is usually considered when applying to animal models, since cell culture applications usually only require the spatial resolution at the "well" level. Small molecules are usually globally administrated, so that the expressed chemogenetic system could be controlled in multiple locations and organs simultaneously. In addition, even when injected locally, the "diffusion" nature of small molecules could reduce the spatial resolution. In contrast, light could be precisely delivered to restricted and specific locations, and the use of two-photon activation could further improve the spatial resolution.

Temporal resolution and reversibility. Temporal resolution is a major consideration because the activation by small molecules or light gives very different time scale, while spatial resolution could still be restricted by the selective tool expression as mentioned above.

In general, chemogenetic tools are considered to have slower kinetics and lower reversibility compared to optogenetic tools. When applied in cell culture, small molecules need to diffuse in media and sometimes cross the cell membrane, which takes around tens of seconds to activate the chemogenetic system. In addition, usually the small molecules have high affinity towards the protein, the reversibility is low and need to be overcome by methods mentioned in Chapter 1.1.1.3. When applied in animal models, since the administrated small molecule needs to circulate in body until reaching the desired location, chemogenetic tools provide a relatively longer time scale, on order of minutes. For example, the CNO molecule used for DREADDs could reach

to mouse brain tissue within 15 minutes[351]. In addition, the reversibility of the chemogenetic systems depends on the metabolism and clearance of the small molecule. CNO concentration in mouse brain significantly decreases at 60 minutes[351]. On the other hand, optogenetic tools could provide ultrafast temporal control (as quick as milliseconds[177]) and reversible control. Since no diffusion is required and light travels at light speed, this temporal resolution is determined by how fast the light-sensing protein domain can be activated and reversed to basal state.

Delivery of chemical and light. The experimental challenges need to be taken into consideration when making the choice of using small molecule or light. In general, small molecules are experimentally easier to use. For cell culture experiments, small molecules are delivered by adding solutions directly to cells, which is not much different than conventional cell culture protocols. When using light as a control, plates need to be wrapped with aluminum foils, and experiments need to be performed in a dark room. When applied in animal models, small molecules can be delivered globally, allowing the use of chemogenetic tools in deep tissues. Also, for long term experiments, drugs can be delivered to animal models by relatively easy ways such as oral dosing and injection. On the other hand, many of the currently used optogenetic proteins are stimulated by blue light, which suffers from weak tissue penetration (< 1 millimeter for brain[177]). Light delivery setup for animal experiments is invasive and is not ideal for experiments involving free-moving mice. Another important consideration would be the efficiency of delivery. Light can always be efficiently delivered to the desired location. But small molecules have different pharmacology properties. Not all small molecules can cross blood-brain barrier efficiently, making those systems not suitable for neuroscience applications.

Side effects. Exogenous small molecules can crosstalk with endogenous biological systems. Some drugs have side effects and need to be considered when studying biological

questions. For example, CNO was typically considered as orthogonal to endogenous systems, however, it was later discovered that CNO could be metabolized into clozapine in vivo, which also contributes to the activation of DREADD and endogenous receptors[352,353]. Light is generally considered as orthogonal, but the phototoxicity needs to be taken into concern when prolonged stimulation is required.

In summary, chemogenetic tools and optogenetic tools are complementary to each other: optogenetics are more suitable for studying rapid reversible processes and chemogenetics are suitable for long term chronic studies.

1.4 Knowledge gap to fill and dissertation overview

Besides large proteins whose function can be genetically modulated, peptides also play irreplaceable roles in biological events but genetic tools controlling peptides have been less explored. In addition to some biological events that has been extensively achieved by different chemogenetic and optogenetic designs, such as dimerization, protein localization, enzyme modulation, protein degradation, etc., peptides can serve some unique functions. For example, neuropeptides are important elements for neuromodulation[354], but no example prior to this thesis has shown to genetically control the function of a specific neuropeptide.

As discussed in Chapter 1.2.3.2, the AsLOV2 domain was the only optogenetic domain that has shown general ability to control peptide functions, which has made it to be a versatile and useful switch in optogenetic tool design. However, prior to this thesis, there has not been a chemogenetic domain to generally modulate peptide functions. Such a domain could greatly expand the chemogenetic toolbox and achieve new applications such as neuropeptide modulation. Further, the AsLOV2 domain is fused to the N-terminus of functional peptides, which prohibits its

application for peptides that requires a fusion-free N-terminus. A new optogenetic domain for modulating peptides in their C-terminal fusion geometry can further expand the utility of AsLOV2.

This dissertation aims to bridge the gap described above. Together with other collaborators, I engineered a pair of chemogenetic switches and an optogenetic switch to generally control a range of peptides. Further, we developed a yeast surface display platform for the directed evolution of protein switches that enables the selection of two criteria properties, in which enhanced selection efficiency is obtained.

Chapter 2 describes the engineering of a pair of chemically activated protein domains (CAPs) for controlling the accessibility of both the N- and C-terminal portion of peptides. The switches use FKBP(F36V) protein to sterically block functional peptides. Upon the FKBP(F36V) ligand binding, the functional peptide is unblocked and becomes functional. I describe the design and the directed evolution to develop CAPs. I further show the application of using CAPs to control a range of short peptides, including a protease cleavage site, a dimerization-inducing heptapeptide, a nuclear localization signal peptide, and an opioid peptide, with a chemical dependence up to 156-fold. We show that the CAPs system can be utilized in cell cultures and multiple organs in living animals.

Chapter 3 describes the development of a new yeast surface display directed evolution platform with higher selection efficiency. To evolve a switch with both good activation efficiency and low background activity, the conventional selection method requires alternative positive and negative selections to obtain the clones that meet with both of the two criteria. In this new platform, activation and leakage signals of a clone are shown on the same cell, so that the protein switches with improved dynamic ranges are selected in each round. This new selection platform reduces the

library diversity more efficiently and yielded improved CAPs with reduced background activity. The evolved CAPs were further applied to cage three neuropeptides.

Chapter 4 describes the engineering of a new optogenetic switch, circularly permuted AsLOV2 (cpLOV). To optogenetically control functional peptides, they are fused to the C-terminus of AsLOV2 domain, which is not applicable for peptides that require free N-terminus for functionality. We designed cpLOV, in which the functional peptides are fused via their C-terminus, so that new geometry could be adapted for optogenetic tool design. We show that cpLOV could be applied to cage TEVcs in a transcription reporter assay and could be used with AsLOV2 tandemly to tune the dynamic range of the transcription reporter assay.

Chapter 5 provides a summary of the key outcomes and highlights the novelty and impact of this dissertation. Additionally, it outlines perspective directions based on the protein switches and the directed evolution platform described in this thesis.

Note that the two acronyms LIP and LDP are not commonly used and are just for the ease of writing in this thesis.

Chapter 2 Development of Chemical-Activated Protein Switches for Modulating Peptide Functions

Chapter 2 is largely adapted from: Shen, J., Geng, L., Li, X., Emery, C., Kroning, K., Shingles, G., Lee, K., Heyden, M., Li, P. and Wang, W. A general method for chemogenetic control of peptide function. *Nature Methods* (2023), 20, 112-122.

Natural or engineered peptides serve important biological functions by acting as partners in a dimerization pair[355], inhibitors of enzyme activity[332,356], protein localization[357,358], and protein degradation[159,345]. Although many chemogenetic domains and systems have been developed to control protein functions, chemogenetic control of short peptides is less explored. The existing approach to modulate peptide functions include chemically modified synthetic peptides, sterically blocking peptides by another protein domain, and using CIP/CDP to control their localizations.

Chemically-modified synthetic peptides incorporate removable protecting groups[359] or an azobenzene-based chromophore[360]. While useful in vitro, this method has limited use in endogenous biological environments due to challenges in cell delivery and subsequent degradation of external peptides, as well as the lack of the ability to target specific cell populations.

Another way to control peptide function is by using genetically-encoded protein domains[159,324]. By fusing a peptide to a protein that changes its conformation in the presence of light, caging and uncaging of the peptide can be achieved. This approach allows genetic introduction of peptides into living organisms and cell type specific protein expression; therefore,

it is well suited for biological studies. The second light, oxygen, voltage sensing domain from *Avena sativa* phototropin 1 (AsLOV2) has been used as a generally applicable protein domain to control peptide function[324,325,332,339,340,342,344,345]. As discussed in Chapter 1.2.3.2, the light-dependent conformational change of AsLOV2's C-terminal α helix can affect the accessibility of the peptide fused to it. In the dark, the peptide is blocked by the core domain of AsLOV2. In the light, a conformational change takes place in AsLOV2; consequently, the peptide is unblocked and becomes accessible[195,196]. Although light provides fast temporal control, its use in non-transparent organisms is limited by poor tissue penetration[177].

As discussed in Chapter 1.1, many chemical-dependent protein domains have been previously developed and used to obtain chemical control of protein proximity[361,362] and conformation[145,146,363], however, much less examples on chemogenetic control of peptides were reported. Protein activity is highly dependent on tertiary structure and conformation, providing many different opportunities to modulate on it. However, peptides are less dependent on these properties. Especially for short peptides (less then 10 amino acids), their functions are mostly dependent on their sequences and the recognition by their targets. Therefore, it remains challenging to control peptide functions by chemogenetic means, and no chemical-dependent protein domain has been shown to have general applicability in directly controlling peptide functions prior to this dissertation.

The closest example of a chemogenetic tool to control peptide functions is the LID system[159] (discussed in Chapter 1.1.3.2 with more details), where a five-amino-acid peptide that can induce protein degradation is blocked by an FK506 binding protein until a small molecule, shield-1, displaces the peptide from the ligand binding site. However, LID has not been shown to be able to control other peptides.

Additionally, some peptides could not be regulated through induced proximity mechanism (e.g., neuropeptides). A chemically activated protein domain that could generally control peptide functions is valuable for four reasons: 1) it can be easily delivered to animal models by genetic means; 2) it can complement the disadvantage of low tissue penetration when using light; 3) it can be used for peptides that cannot be modulated through proximity; 4) it will open up many more new opportunities for chemogenetic tool design.

To address the absence of chemogenetic protein switches for modulating peptide functions, I together with Lequn Geng engineered a pair of chemically activated protein domains (CAPs) that can be generally applied to control a fused peptide. CAPs were developed via directed evolution and include two protein domains, CapN and CapC, for sterically blocking the N- and C-terminal portion of a peptide, respectively (**Figure 2-1**). Addition of shield-1 relieves steric blocking and activates peptide functions. We demonstrate the use of CAPs to cage tobacco etch virus protease cleavage site (TEVcs), the SsrA peptide, a nuclear localization signal peptide (NLS), and an opioid peptide, [Met^5]-enkephalin. We also show that CAPs can be used in mammalian cell culture, mouse brains, and mouse livers, demonstrating both in vitro and in vivo utilities.

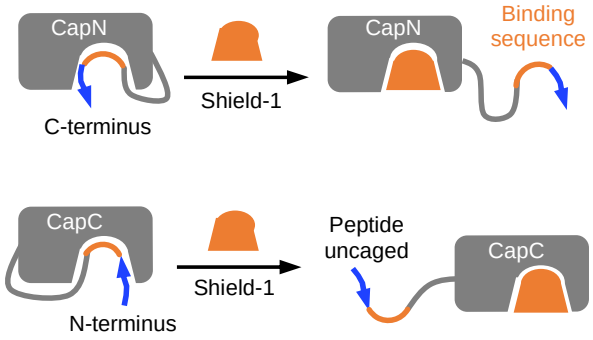


Figure 2-1 Design of CAPs. CapN and CapC block the N- and C- terminal portion of a peptide, respectively. Addition of shield-1 releases the binding sequence from the ligand binding site, unblocking the peptide.

2.1 Design and directed evolution of CapN

As discussed in Chapter 1.1.3.2, LID was designed by caging a short degron peptide sequence by steric hinderance. Although the degron is the only peptide shown to be caged, the mechanism using steric hinderance shall be generally applicable for caging other peptides. We then used LID as the starting point and engineered CapN by directed evolution. CapN was further characterized on yeast surface display platform.

2.1.1 Design of CapN

We started with designing a CAP for caging the N-terminal portion of a peptide (**Figure 2-1**). The design of CapN was inspired by LID[159], which is based on FKBP12(F36V), hereafter referred to as FKBP in this chapter. In LID, a five-amino-acid peptide degradation sequence is blocked by FKBP via fusion of the degradation sequence to a binding sequence that interacts with FKBP. The degradation sequence can be activated when an FKBP ligand, shield-1, is added to displace the binding sequence. LID has several desirable features as our starting point. First, the functionality of LID has been proven on a five-amino-acid peptide, and we hypothesized that a similar caging mechanism may apply to other peptides of similar sizes. Second, the ligand for LID, shield-1, has low nanomolar affinity for FKBP, which is over 1000-fold higher than that for the wild-type FKBP[364]. Therefore, shield-1 has minimal interference with the physiological functions of endogenous FKBP. Lastly, shield-1 is membrane permeable and has been previously used in live animals[365].

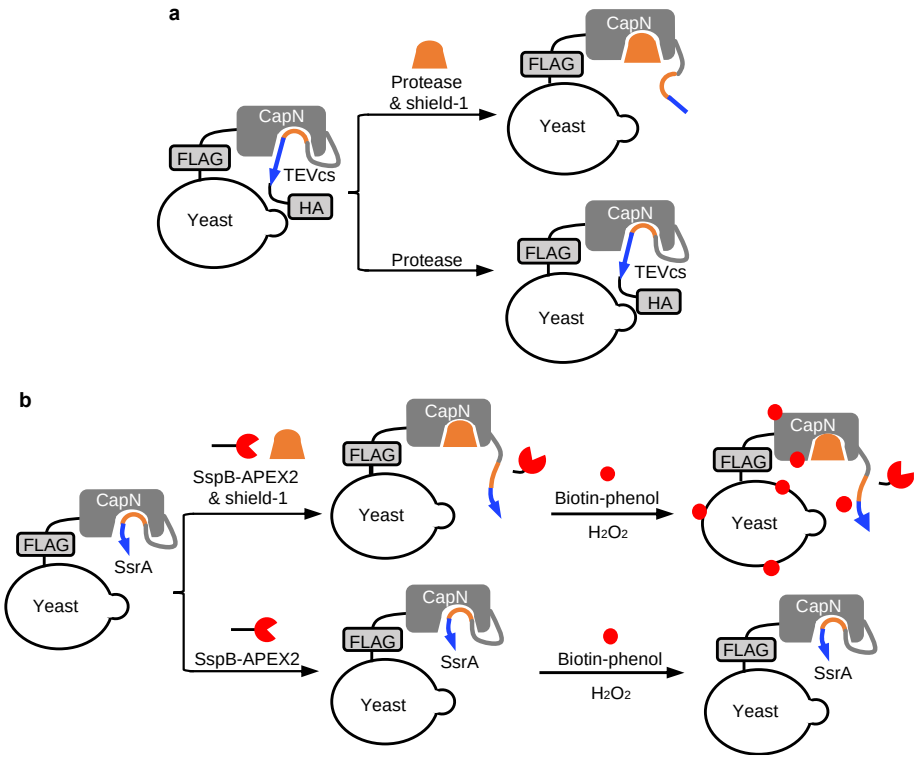


Figure 2-2 a, Labeling scheme for CapN-caged TEV protease cleavage site. TEVcs, TEV protease cleavage site (ENLYFQG). FLAG and HA are epitope tags. **b,** Labeling scheme of CapN-caged SsrA. CapN-SsrA is displayed on yeast surface by fusing to the yeast Aga2p protein. APEX2 labels protein within close proximity with biotin-phenol.

We first tested the ability of LID to cage two short peptides that have been previously controlled by the AsLOV2 domain[324,325,334,335]. A heptapeptide, TEVcs, and an octapeptide, modified SsrA, were individually displayed on the yeast surface, and their accessibility was evaluated using a protease cleavage assay and a binding assay (**Figure 2-2**), respectively. The degradation sequence (RRRGN) in LID was replaced with TEVcs or SsrA, and the rest of the binding sequence and the linker were kept the same. In this initial test, we did not observe shield-1 dependence for either TEVcs or SsrA (**Figure 2-3**), suggesting that LID is ineffective in introducing steric blocking to peptides beyond the original degradation sequence. To achieve generally applicable peptide switch, further engineering is required.

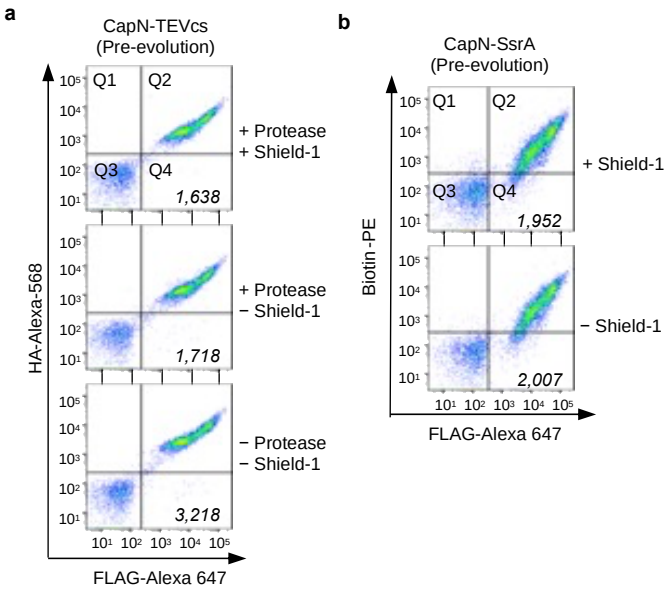


Figure 2-3 a, FACS analysis of CapN-caged TEVcs before and after directed evolution. Values are median HA intensity of FLAG-positive cells (Q2 + Q4). **b,** FACS analysis of CapN-caged SsrA before and after directed evolution, using labeling scheme as shown in c. Values are median biotin intensity of FLAG-positive cells (Q2 + Q4).

2.1.2 Directed evolution of CapN

We hypothesized that the binding sequence is the key for peptide caging. The amino acids near the C-terminus of the binding sequence are critical for interacting with the hydrophobic ligand binding site of FKBP (**Figure 2-4**), so we focused on tuning these residues.

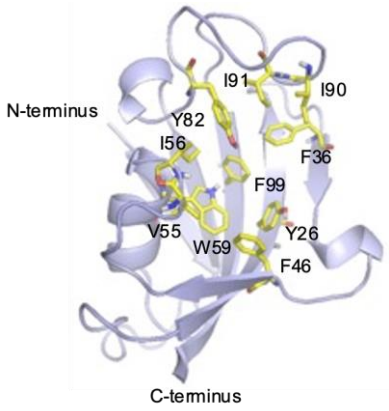


Figure 2-4 Crystal structure of FKBP12 (PDB:1FAP). The hydrophobic residues around the ligand binding site are shown in yellow and stick representation.

We applied yeast surface-based directed evolution to improve the shield-1 dependence of CapN, using TEVcs as the caged peptide. The libraries were generated by introducing site-saturated mutagenesis to the last six amino acids of the binding sequence right before TEVcs (**Figure 2-5**).

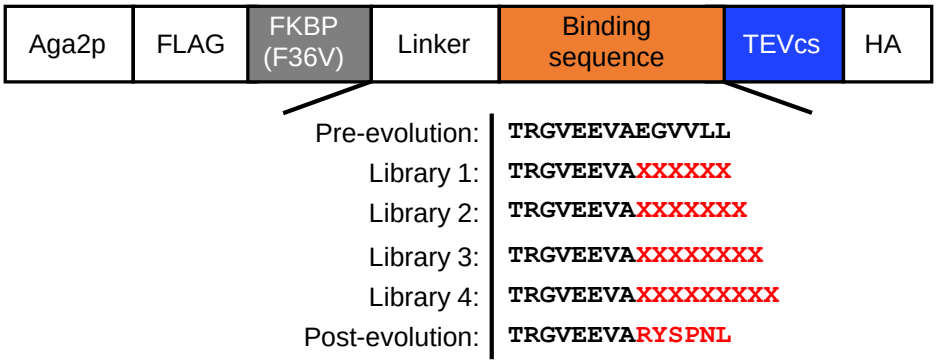


Figure 2-5 CapN sequence before, during, and after directed evolution. "X" indicates any of the twenty amino acids. Amino acids that are different from the original LID sequence are highlighted in red.

Four libraries were selected according to the scheme shown in **Figure 2-6**. First, we performed one round of negative selection to retain CapN variants with tight caging in the absence of shield-1 (high expression and low cleavage). This also ensured all clones entering the second round of selection expressed the DNA construct without early stop codons. A round of positive selection was then performed to enrich clones that can efficiently uncage TEVcs upon addition of shield-1. After these two rounds of selections, the libraries of CapN variants started to show shield-1 dependence. After four rounds of selections, a clear shield-1 dependence and tight cell population was observed, suggesting that the directed evolution was successful and sufficient.

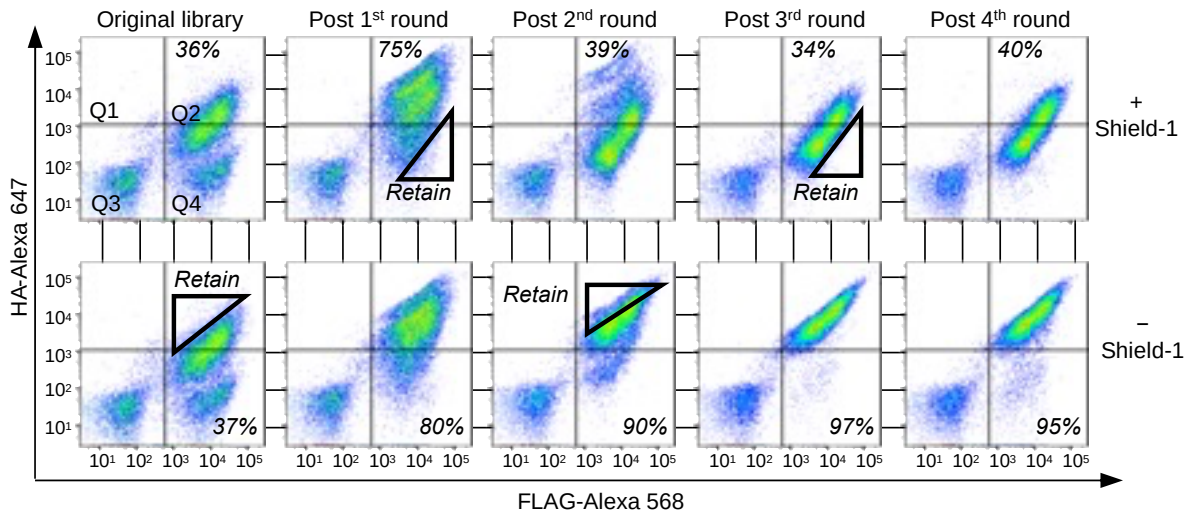


Figure 2-6 FACS selection of CapN libraries to improve shield-1 dependence. Libraries were combined after the 1st round of selection. Values are percentage of cells in Q2 over (Q2 + Q4).

2.1.3 Characterization of evolved CapN

After four rounds of directed evolution and sequencing of forty individual clones, we found twenty-three distinct sequences that were rich in hydrophobic residues (**Table 2-1**), which agrees with our initial hypothesis that the interaction between binding sequence and FKBP binding pocket plays an important role in caging peptides.

We characterized eight clones that appeared more than once and found a similar level of improvement (**Appendix Figure A-1**). We selected the most enriched clone as the post-evolution CapN (**Figure 2-5**, **Table 2-1**) for the rest of this study. Compared with the pre-evolution CapN, the post-evolution one showed both tighter caging and a larger shield-1 dependence (38-fold, difference between "+ shield-1" and "– shield-1" conditions) (**Figure 2-7**). Importantly, when we applied the same post-evolution CapN to SsrA, enhanced caging and a larger shield-1 dependence (59-fold) were also observed (**Figure 2-7**). This demonstrated the general applicability of CapN in controlling short peptides.

Table 2-1 Sequences of forty clones from the post 4th round CapN library. Twenty-three distinct sequences were identified. Clone #1 is the final CapN used for the rest of this study.

Label	Frequency	Sequence
#1	6	RYSPNL
#2	5	RREVHVPM
#3	3	LQSWYPPL
#4	3	AKKYSPNL
#5	2	PTNWRLPY
#6	2	MRRDWHPP
#7	2	NFHIPM
#8	2	RHWNPPM
#9	1	KRTGQLIP
#10	1	YRWHPPY
#11	1	LLWHPPSS
#12	1	PRGWRMP
#13	1	GACLNVPC
#14	1	SRSHIPY
#15	1	DISQCKSC
#16	1	SCRFRPPC
#17	1	QRPKFVPP
#18	1	QTNWYPPL
#19	1	PTCWHVPL
#20	1	TTGGWHV
#21	1	STPRHIPM
#22	1	TSNWHPPM
#23	1	PKRSWHLP

We next characterized the dose response of evolved CapN system by the SsrA binding assay since it showed a good shield-1 dependence. The assay showed half maximum response at 53 nM, and 100nM is sufficient to reach maximum response (**Appendix Figure A-2**).

We also investigated the reversibility of CapN. Yeasts expressing the CapN-SsrA were first incubated with 10 μM shield-1 for 10 minutes. Shield-1 was then washed away and the yeasts were allowed to incubate in media for 0 hour, 1 hour, and 12 hours before performing the binding assay. CapN remained open 12 h after washing away shield-1 (**Appendix Figure A-3**). This study suggested that shield-1 stably binds to CapN, making CapN an irreversible "on" switch.

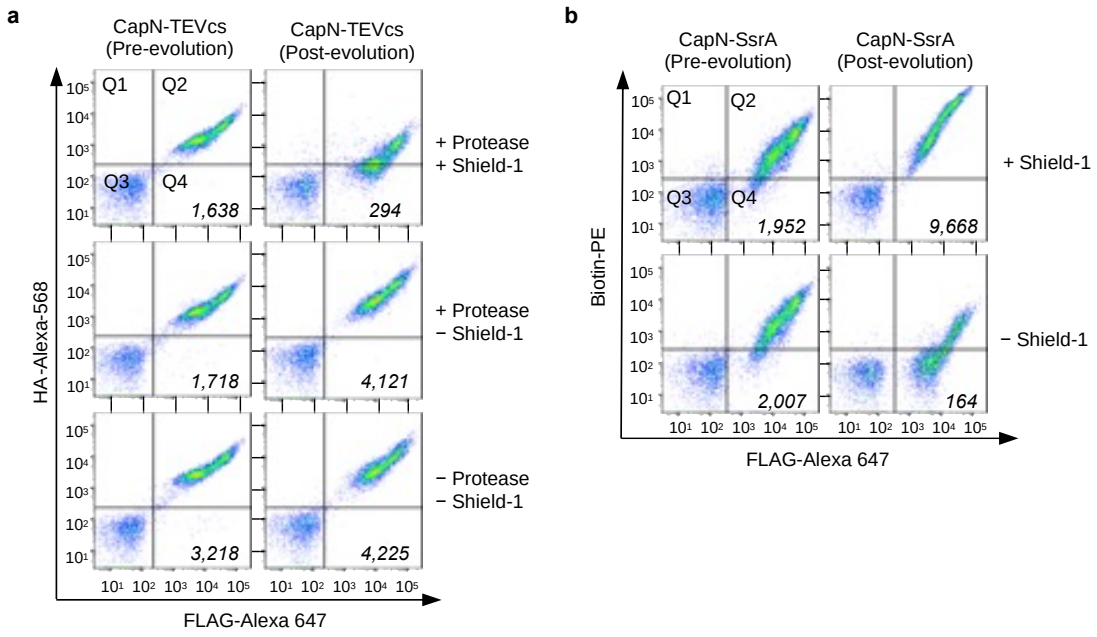


Figure 2-7 a, FACS analysis of CapN-caged TEVcs before and after directed evolution. Values are median HA intensity of FLAG-positive cells (Q2 + Q4). Protease cleavage ("+ shield-1" or "− shield-1") is defined as the difference of the median HA signals between the + protease and – protease conditions. The shield-1 dependence is calculated by the ratio of protease cleavage of the "+ shield-1" and "− shield-1" conditions. **b,** FACS analysis of CapN-caged SsrA before and after directed evolution. Values are median biotin intensity of FLAG-positive cells (Q2 + Q4). The shield-1 dependence is calculated by the ratio of the median PE signal of the "+ shield-1" and "− shield-1" conditions.

2.2 Design and directed evolution of CapC

Next, we sought to engineer an analogous CAP that can cage the C-terminal portion of a peptide (CapC). The motivation to engineer this version is three-fold. First, peptide caging is often the most effective when the key residues on a peptide are closely bound to the CAP proteins. Peptides with a crucial C-terminal portion could be better caged by CapC than CapN. Second, CapN is not feasible for peptides that require a free N-terminus to be functional. Third, CapN and CapC can be used in tandem to further reduce "leaking" in the absence of shield-1. This is especially critical for applications in live animals where the protein is expressed often for weeks to months, before the experiment is performed.

Based on CapN, we next engineered CapC with the similar design and further optimized it by directed evolution. The evolved CapC is further characterized on yeast surface display platform.

2.2.1 Design of CapC

To design CapC, we simply fused the binding sequence of CapN (RYSPNL) to the N-terminus of FKBP with a flexible linker (SGAGSGGSGTGSGSGGS) and tested its efficacy in caging the SsrA peptide (**Figure 2-8**). SsrA was chosen because we found CapN showed a larger shield-1 dependence in caging SsrA (59-fold) than caging TEVcs (38-fold), which facilitates easier selection during directed evolution. The initial CapC construct showed promising shield-1 dependence (1.4-fold) (**Figure 2-9**, first column) but the shield-1 dependence was much smaller than CapN (59-fold). We then performed directed evolution to further improve the shield-1 dependence of CapC.

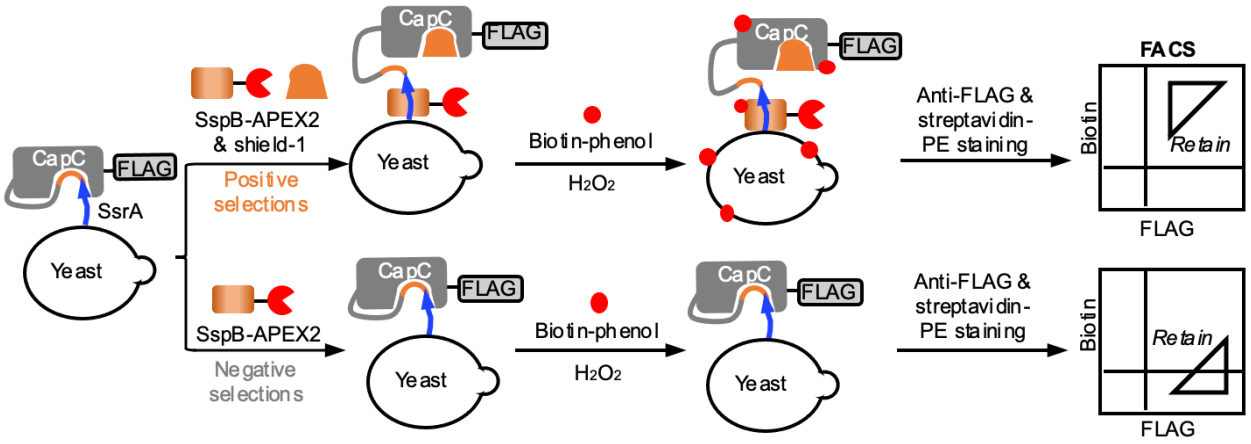


Figure 2-8 Labeling and library selection scheme for CapC-caged SsrA. For library selection, retained populations are shown in triangles on the FACS plots.

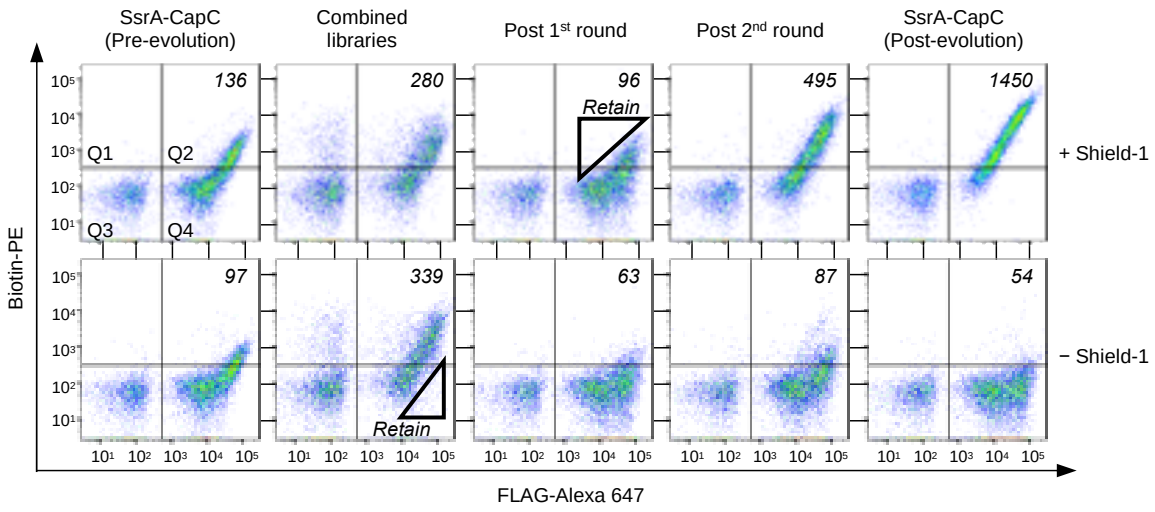


Figure 2-9 FACS selection of CapC libraries to improve shield-1 dependence. Libraries were combined for selection. Values shown in plots represent median biotin intensity of FLAG-positive cells (Q2 + Q4). Experiment was conducted by collaboration with Dr. Lequn Geng.

2.2.2 Directed evolution of CapC

We then used directed evolution to enhance the shield-1 dependence of CapC. We kept the last three amino acids (Pro-Asn-Leu) of the binding sequence because they provide a hydrophobic surface which might be important for binding. This is confirmed by the computational simulation work from Dr. Matthias Heyden. The importance of the hydrophobic leucine residue was supported by an all-atom molecular dynamics simulation that showed leucine has a high tendency to interact with the FKBP hydrophobic ligand binding site (**Appendix Figure A-4**).

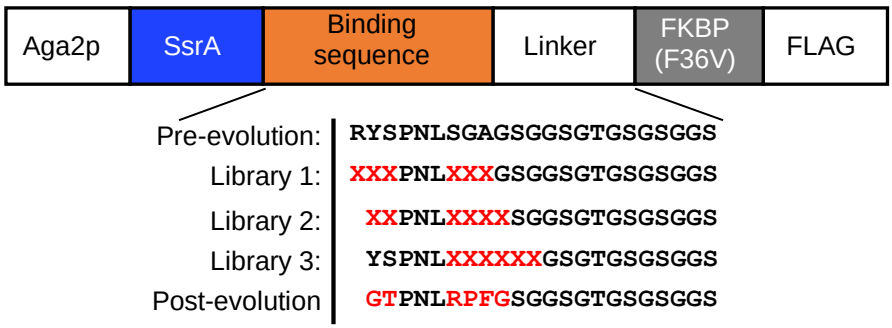


Figure 2-10 CapC sequence before, during, and after directed evolution. "X" indicates any of the twenty amino acids.

Table 2-2 Sequences of twenty clones from the post 2nd round CapC library. Eighteen distinct sequences were identified. Clone #18 is the final CapC used for the rest of this study. Experiment was conducted by Dr. Lequn Geng.

Library	Label	Frequency	Sequence
Library 1	#1	1	DWPPNLMTS
	#2	1	LFRPNLGSC
	#3	1	PWYPNLDST
	#4	1	WRKPNLPYS
	#5	1	VGSPNLFYM
	#6	1	ICRPNLCAT
	#7	1	KWRPNLSSC
	#8	1	VGLPNLGGY
	#9	1	IGSPNLNLY
Library 2	#10	1	TSPNLRWRW
	#11	1	PFPNLDYGL
	#12	2	PFPNLGEPP
	#13	1	YSPNLCGAG
	#14	1	PFPNLRSYL
	#15	1	GFPNLGYRY
	#16	1	CKPNLNRHS
	#17	1	MWPNLDNQW
	#18	1	GTPNLRPFG

With this information, we applied site-saturated mutagenesis to the regions before and after the Pro-Asn-Leu sequence (**Figure 2-10**). The libraries were combined and selected according to the scheme shown in **Figure 2-8** and **Figure 2-9**. After two rounds of sorting, a 5.7-fold shield-1 dependence of the CapC libraries was obtained. We characterized twenty individual clones (**Table 2-2**) and identified one with tight caging and a large shield-1 dependence (27-fold) as our post-evolution CapC (**Figure 2-8**, **Figure 2-9**, and **Appendix Figure A-5**).

2.2.3 Tandem use of CAPs

For most of the chemogenetic tools, further optimization is usually required to reduce the background leakage. However, CAPs provide a much easier and modular design to overcome this problem since CapN and CapC could be fused to both of the termini of a peptide.

We next tested if using CapN and CapC together could provide tighter caging and reduce background. Since CapN already showed satisfying caging efficiency, a more efficient protease (truncated TEV protease was used for directed evolution, full length TEV protease was used for this experiment) was used to fully show the caging efficiency. As shown in **Appendix Figure A-6**, both CapN and CapC caged TEVcs showed significant leakage in the absence of shield-1. The tandem use of both CAPs in caging TEVcs significantly reduced the leakage compared to using either CapN or CapC alone. Consequently, the shield-1 dependence of tandem CAPs (2.4-fold) was higher than individual CAPs (both 1.1-fold). We, therefore, used tandem CAPs for the following experiments in cell culture and animals whenever possible.

2.3 CAPs for protein translocation

As discussed in Chapter 1, protein translocation is a widely used method to regulate protein functions. This could be realized through inducing or disrupting proximity, protein cleavage from an anchor, and caging signaling peptide. Here we employ the two peptides that CAPs were developed with, SsrA and TEVcs, to manipulate protein translocation in mammalian cell culture.

2.3.1 Shield-1 induced plasma membrane translocation by CAPs-caged SsrA

To demonstrate the use of CAPs for protein translocation, we designed a shield-1 induced membrane localization system (**Figure 2-11**). In this design, CAPs-caged SsrA peptide is localized on the plasma membrane of HEK 293T cells by being fused to CAAX, a membrane-anchoring domain. The interacting partner of SsrA, SspB protein, is fused to the POI. Here, we used enhanced green fluorescent protein (EGFP) as the proof-of-principle POI. Without shield-1, no binding

between SsrA and SspB should occur, and EGFP should be found throughout the cell. Addition of shield-1 should uncage SsrA from CAPs and allow it to bind with SspB-EGFP, translocating EGFP to the membrane.

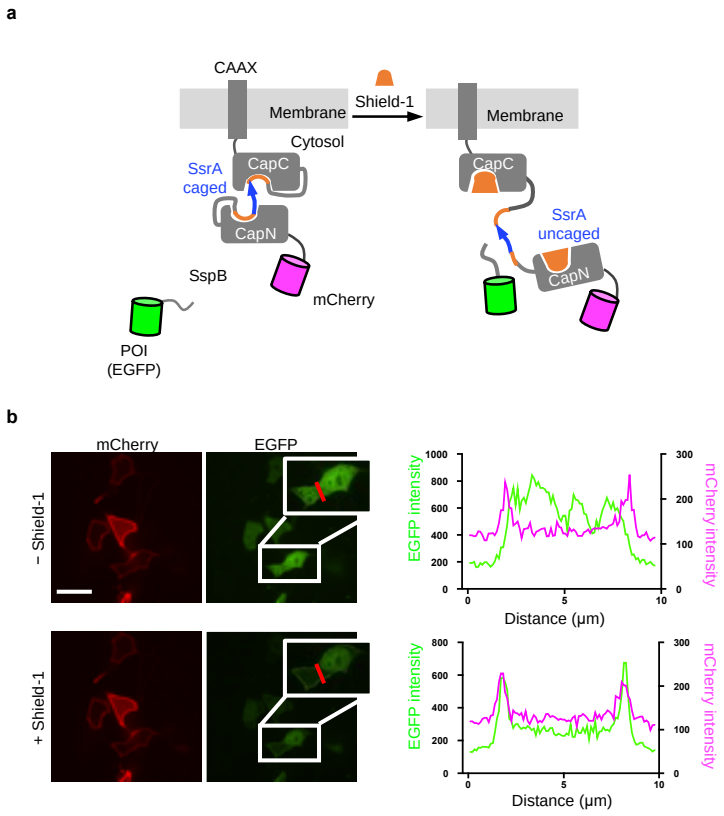


Figure 2-11 a, Scheme of shield-1 induced protein translocation to plasma membrane. Membrane-anchoring domain is CAAX. POI, protein of interest. **b,** Left: Representative fluorescence microscopy images of HEK 293T cells expressing the constructs shown in **a**. Right: Intensity profiles of mCherry and EGFP along the red line in images. Scale bar, 20 μm. Data was acquired by collaboration with Dr. Kayla Kroning.

As shown by **Figure 2-11**, under the no shield-1 condition, EGFP was localized in the cytosol and nucleus with no apparent membrane pattern. Upon adding shield-1, we observed significant EGFP translocation to the plasma membrane within 30 seconds (**Figure 2-11**).

This showed that shield-1 can easily permeate through the mammalian cell membrane and uncage CAPs on the order of seconds, making the CAPs-shield-1 system useful for experiments that require a fast temporal control. By directing CAPs-SsrA to different subcellular

compartments, a POI could be recruited to various locations of interest in a shield-1-dependent manner.

2.3.2 Shield-1 induced plasma membrane delocalization by CAPs-caged TEVcs

In this application, we aimed to remove the POI from the plasma membrane. This could be useful for temporally-controlled perturbation of proteins that only function when localized on the plasma membrane. We anchored CAPs-caged TEVcs to the plasma membrane of HEK 293T cells by fusing CAPs to CAAX, followed by the POI, using EGFP as an example (**Figure 2-12**). Under the no shield-1 condition, the TEVcs should be inaccessible to the co-expressed TEV protease, and EGFP should remain localized on the plasma membrane. With shield-1, the TEVcs should be uncaged and cleaved by the TEV protease, releasing EGFP from the cell membrane. **Figure 2-12** shows that prior to shield-1 addition, EGFP was almost exclusively bound to the plasma membrane. The addition of shield-1 depleted EGFP from the plasma membrane and significantly increased its presence in the cytoplasm.

Therefore, this system can delocalize proteins from the membrane to the cytoplasm in a shield-1-dependent manner. Similar to CAPs-caged SsrA, when directly transferred from the yeast surface to mammalian cell culture, CAPs were still capable of caging TEVcs efficiently. The only modification needed was to change the P1' position (the last amino acid) on TEVcs from the canonical glycine to a less active methionine to reduce unwanted cleavage, as TEV protease is present for a much longer time in this assay (2-3 days) than in the yeast surface assay (3 hours).

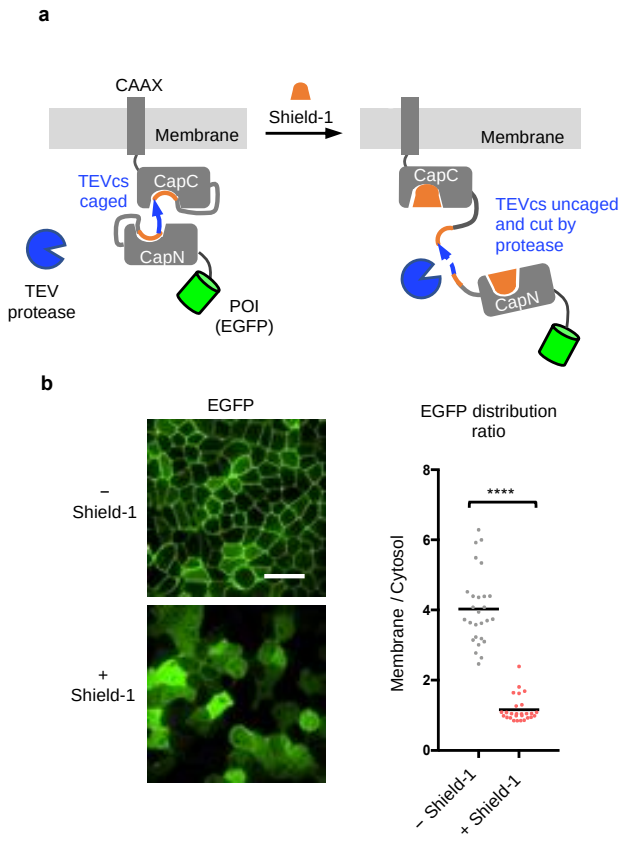


Figure 2-12 a, Scheme of shield-1 induced protein translocation from plasma membrane. TEVcs(ENLYFQM). **b,** Left: Representative fluorescence microscopy images of HEK 293T cells expressing the constructs shown in **c**. Scale bar, 20 μm. Right: Quantification of EGFP total intensity distribution. The ratio is calculated by the EGFP mean intensity on membrane to that in cytosol. n = 27 cells from one replicate for both conditions. The center lines indicate mean values of the ratio. *P* value is determined by unpaired two-tailed *t*-tests. ****$P < 0.0001$. Data was acquired by collaboration with Dr. Lequn Geng.

2.4 Shield-1 induced gene expression by CAPs caged SsrA

Induced gene expression could serve as a useful method to study POI functions. Two-hybrid platform is a robust design using induced protein-protein interactions for controlling transgene expression in cell culture and animal models. To achieve this, we adapted the two-hybrid platform and applied CAPs caged SsrA peptide and SspB as the interaction pair (**Figure 2-13**). In this design, CAPs-caged SsrA is fused to a DBD, and SspB to an AD, VP16. In the presence of shield-1, SspB-VP16 will be recruited to DBD-SsrA and initiate transcription of the reporter gene.

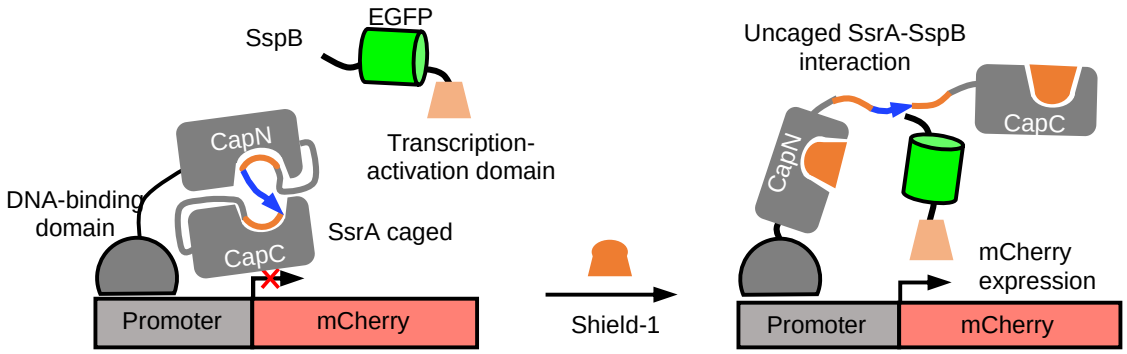


Figure 2-13 Scheme of shield-1-induced gene transcription. Transcription-activation domain is VP16 for all following experiments. DNA-binding domain is specified under each experiment.

2.4.1 Optimization of the gene expression system

We first tested this system in HEK 293T cells using the Gal4 DBD and the *UAS-mCherry* reporter gene. With the eight-amino-acid SsrA peptide (**Figure 2-14**) used in previous experiments, a ~4-fold and ~25-fold shield-1-induced increase in mCherry intensity was observed for CapN and CAPs caged SsrA, respectively (**Figure 2-14** and **Appendix Figure A-7**). As expected, the tandem use of CAPs showed both lower background and higher shield-1 dependence than using CapN alone. This was consistent with our prior observation of the CAPs-caged TEVcs on yeast surface (**Appendix Figure A-6**).

We then sought to further lower the background because transcriptional systems are highly sensitive and are often only practically useful when undesired transcription is reduced to a minimum. It has been previously shown that the amino acid sequence of SsrA is tunable[325,355], so we truncated the SsrA sequence from both the N- and C-termini (**Figure 2-14** and **Appendix Figure A-8**) to achieve more efficient caging. The truncation of the N-terminal Ala residue abolished the interaction between SsrA and SspB, and no transgene expression was observed. The best result was from a seven-amino-acid SsrA sequence (AANDENY) (**Figure 2-14**, **Figure 2-15**,

and **Appendix Figure A-8**), which showed 156-fold shield-1 dependent reporter gene expression change, with two-fold lower background and six-fold higher shield-1 dependence (**Figure 2-14**) than the original sequence tested.

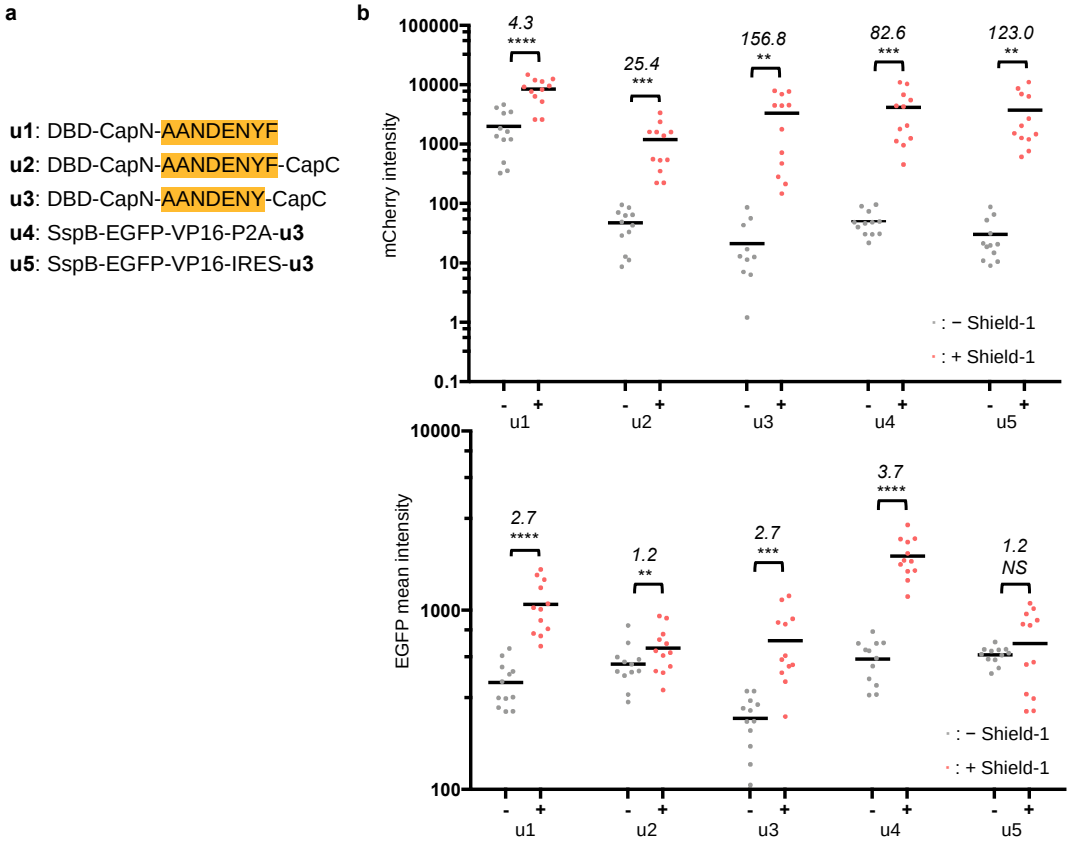


Figure 2-14 a, Summary of main constructs tested. Amino acid sequences of SsrA are highlighted. **b,** Quantification of mCherry expression level for constructs shown in **a**. Values on the plot are the ratio of mean mCherry intensity of + shield-1 to that of − shield-1 conditions for each construct. The center lines indicate mean values of mCherry intensity. Images are shown in **Appendix Figure A-7, 8, 9.** For this experiment, Gal4 was used as DBD, and *UAS-mCherry* was used as reporter gene. n = 12 fields of view from one replicate for all conditions. *P* values are determined by unpaired two-tailed *t*-tests. ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001; *NS*, not significant.

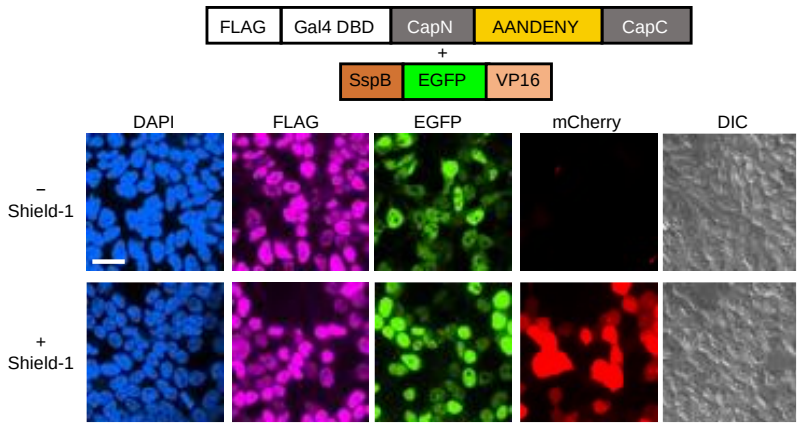


Figure 2-15 Representative fluorescence microscopy image of u3. Gal4 was used as DBD, and *UAS-mCherry* was used as reporter gene. Scale bar, 20 µm.

2.4.2 Characterization of the gene expression system

In the set of experiments screening for a better SsrA sequence, we also observed that the expression level of SspB-VP16, shown by EGFP signal, was positively correlated to the addition of shield-1 and reporter gene expression (**Figure 2-14**). Also, in the conditions where robust reporter gene expression was observed, a nuclear localization pattern was found for EGFP (**Appendix Figure A-7**, **8**, **9**). To investigate whether the reporter gene expression is attributed to the increased AD expression or the dimerization with SsrA, we further designed a control experiment where the SspB-EGFP-VP16 is supplied with shield-1 or media. No significant SspB-EGFP-VP16 expression level change was observed, and no reporter gene expression was observed under either conditions (**Figure 2-16**). We hypothesized that the SsrA-SspB interaction stabilized the SspB protein, thereby increasing the EGFP signal. In addition, shield-1 induced a much greater increase in reporter gene expression than in EGFP protein level (**Figure 2-14**), so the reporter gene activation is primarily due to the dimerization of the two components rather than the increased SspB-VP16 protein expression level.

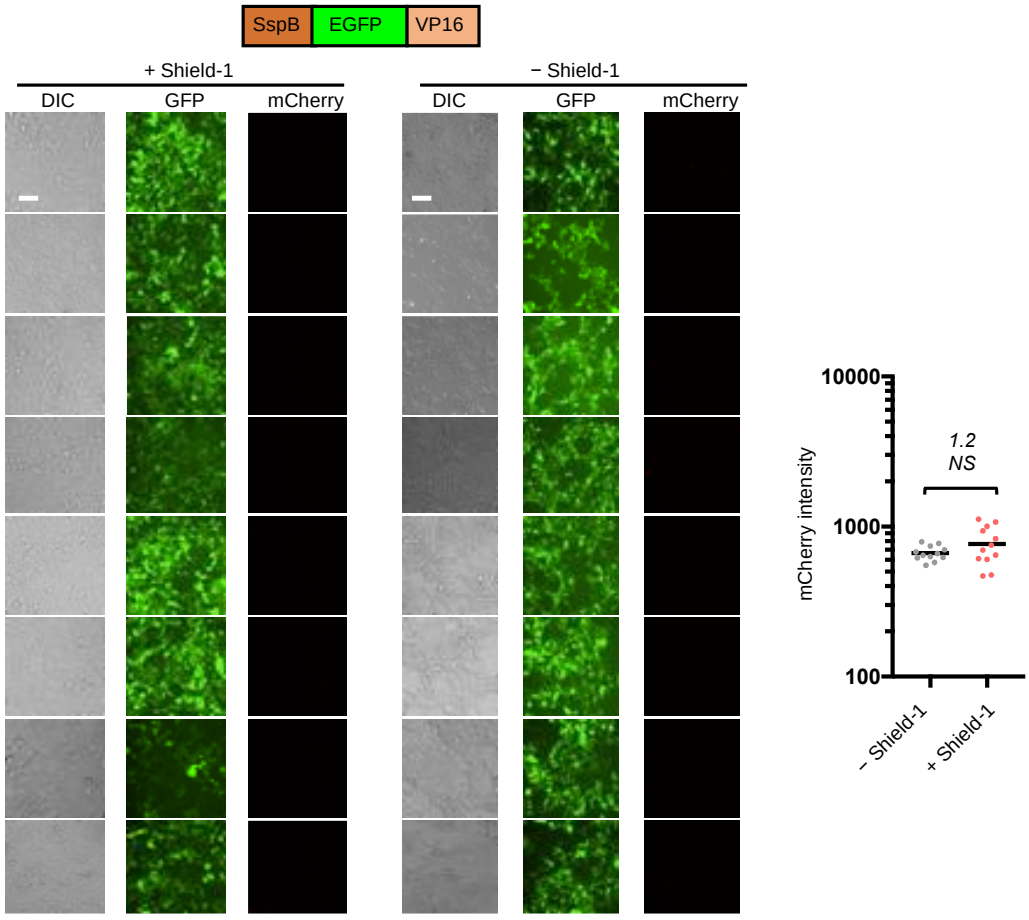


Figure 2-16 Effect of shield-1 on transcription-activation domain expression. No difference in reporter expression was found between the + shield-1 and – shield-1 conditions in the control study with SspB-EGFP-VP16 only. All scale bars, 50 µm. *P* values are determined by unpaired two-tailed *t*-tests. *NS*, not significant.

To investigate the sensitivity of this system, we characterized the shield-1 concentration and shield-1 incubation time required to induce robust gene expression. We first characterized shield-1 sensitivity for the CAPs double-caged SsrA system by fluorescence imaging. With overnight incubation, a shield-1 concentration below 10 nM did not induce gene expression above background, while 50 nM or above induced > 30-fold gene expression (**Appendix Figure A-10**). Consistent with our observations above, although all conditions were transduced with the same amount of virus, those with higher shield-1 concentrations also showed higher EGFP signal, possibly due to the stabilization effect of the SspB-SsrA heterodimerization.

For shield-1 incubation time characterization, an incubation period as short as 30 min was sufficient to induce gene expression at 1 μM shield-1 concentration, but a longer incubation time, on the order of hours, was required for more robust gene expression (**Appendix Figure A-11**).

We further performed a shield-1 titration of this transcriptional system with both CAPs double-caged SsrA and CapC single-caged SsrA using FACS for analysis. As shown in **Figure 2-17** and **Appendix Figure A-12**, the CAPs double-caged and CapC single-caged SsrA each showed an EC50 of 55 nM and 21 nM, respectively.

We also applied this two-hybrid system for plate reader quantification, which is with higher throughput and easier experiment setup. With a luciferase reporter gene, luciferase is expressed upon stimulation instead of fluorescent proteins, and could be further detected by luminescence. The luminescence assay provided a similar result as the fluorescence assay using FACS (**Figure 2-18**).

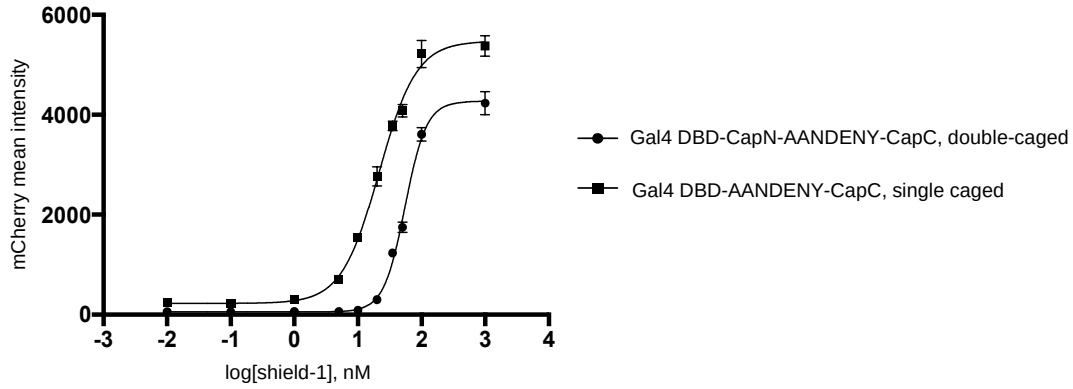


Figure 2-17 Dose response curve of CapC and CAPs caged SsrA using FACS analysis. The mean mCherry intensities of EGFP positive cell population are plotted against shield-1 concentration. Half maximum response was observed at 21 nM for CapC single-caged SsrA (95% confidence interval = 19 nM ~ 24 nM) and 55 nM for CAPs double-caged SsrA (95% confidence interval = 51 nM ~ 59 nM). Errors, s.e.m.

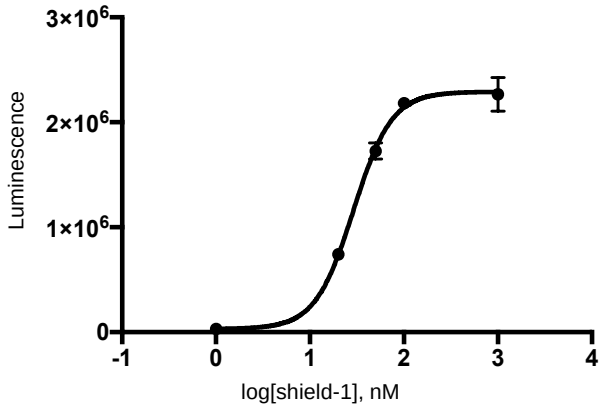


Figure 2-18 Dose response curve of CAPs caged SsrA using luciferase assay. The mean luminescence signal is plotted against shield-1 concentration. Half maximum response was observed at 29 nM for CAPs double-caged SsrA (95% confidence interval = 24 nM ~ 35 nM). Errors, s.e.m.

2.4.3 Gene expression in animal models

Before we applied this chemical-dependent transcriptional system in neuroscience studies, we tested it in cultured neurons. To enable homogeneous expression of DNA in these stringent experiments, we made single viral constructs that expressed both the DNA-binding and the transcription-activation domains via the self-cleaving peptide P2A or the internal ribosome entry site (IRES) (**Figure 2-14**).

In HEK 293T cells, upon adding shield-1, the P2A and IRES constructs showed 83- and 123-fold increase in mCherry expression, respectively, which are comparable to the two-component system (**Figure 2-14** and **Appendix Figure A-9**). The P2A construct was then used to introduce the TetR DBD and VP16 transcription-activation domain into cultured rat cortical neurons through adeno-associated viruses (AAV), together with another AAV encoding the TRE-mCherry reporter gene. Shield-1 induced a 44-fold increase in mCherry reporter gene expression compared to the no shield-1 condition (**Figure 2-19** and **Appendix Figure A-13**), showing that this system could work robustly in cultured neurons.

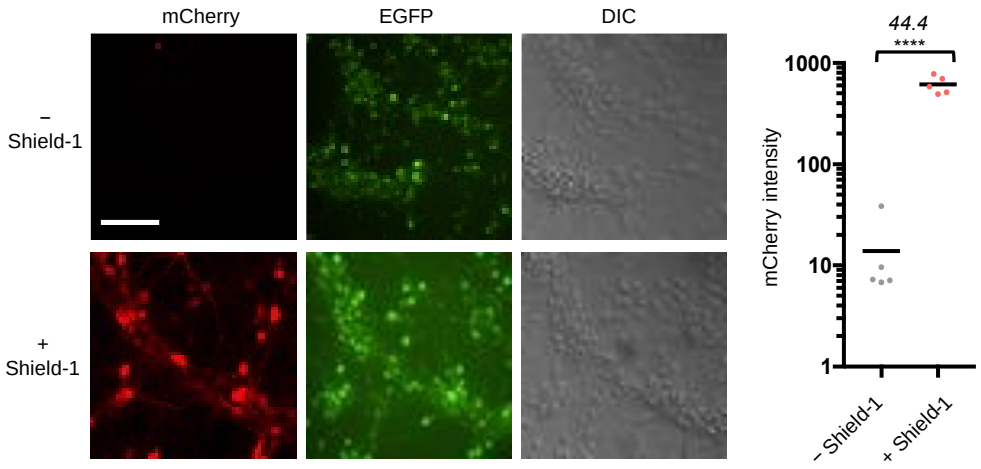


Figure 2-19 Shield-1 induced gene expression in rat cortical neurons using u4 construct. Left: Representative fluorescence microscopy images. For this experiment, TetR was used as DBD, and *TRE-mCherry* was used as reporter gene. Right: Quantification of mCherry expression level. The number on the plot is the ratio of mean mCherry intensity of + shield-1 to that of – shield-1 conditions. The center lines indicate mean values of mCherry intensity. n = 5 fields of view from one replicate for both conditions. Scale bar, 100 μm. *P* values are determined by unpaired two-tailed *t*-tests. ****$P < 0.0001$.

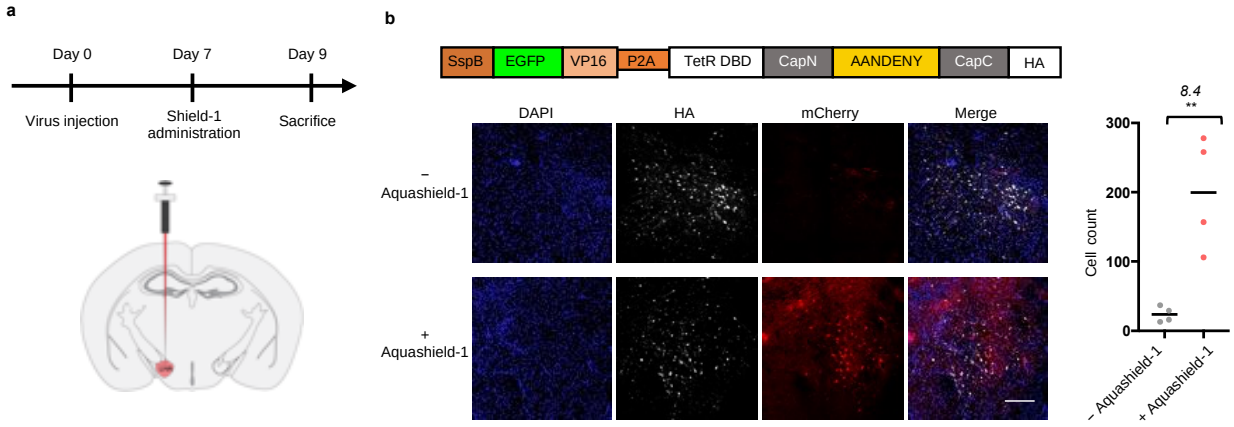


Figure 2-20 AquaShield-1 induced transgene expression in mouse brain. **a,** Timeline for the aquashield-1-induced transgene expression in mouse brain. AAV is locally injected to LHA. Aquashield-1 is locally administered to mice (1 μL, 1 mM). **b,** Left: Representative fluorescence microscopy images of brain sections of the lateral hypothalamic area. Scale bar, 200 μm. Right: Quantification of total number of cells expressing mCherry. Numbers on the plot are the ratio of mean cell count of + aquashield-1 to that of − aquashield-1 conditions. The center lines indicate mean values of cell count. n = 4 injection sites for both conditions. *P* values are determined by unpaired two-tailed *t*-tests. **$P < 0.01$. Experiment was conducted by collaboration with Dr. Xingyu Li and Catherine Emory.

To enable chemical-dependent peptide regulation in animal models, we next tested the CAPs system in mouse brains and livers. For the application in the brain, we performed stereotactic

injection of AAV1/2 mixed serotype viruses encoding shield-1-dependent gene regulation constructs into the lateral hypothalamic area (LHA). Seven days after viral delivery, aquashield-1 (a water-soluble analogue of the shield-1 molecule[160]) was locally administered into LHA (**Figure 2-20**). HA immunostaining of brain tissues confirmed AAV viral delivery and targeting in the LHA (**Figure 2-20**). In the saline treated control brains (− aquashield-1), there were only a few sparse neurons with mCherry expression throughout the entire LHA region. In contrast, mCherry was observed in a large cluster of LHA neurons in the aquashield-1 treated brains (+ aquashield-1). mCherry expression in the presence of aquashield-1 was 8.4-fold higher than the control, suggesting a shield-1 dependent gene expression.

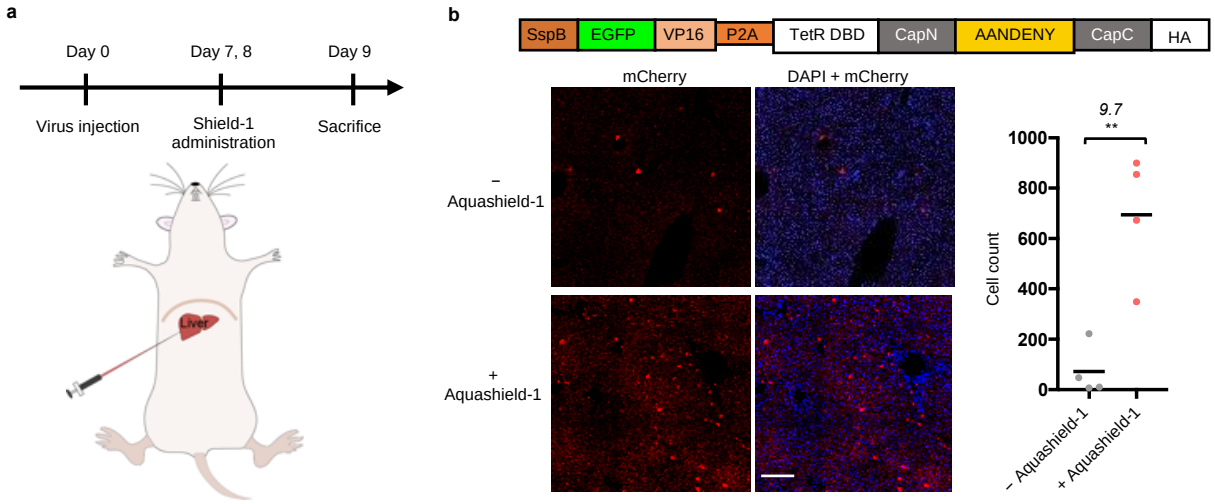


Figure 2-21 AquaShield-1 induced transgene expression in mouse liver. **a,** Timeline for the aquashield-1-induced transgene expression in mouse liver. AAV is locally injected to liver. Aquashield-1 is administered to mice via two intraperitoneal (IP) injections (40 mg/kg) with 24 hours apart. **b,** Left: Representative fluorescence microscopy images of liver sections from injection site. Scale bar, 200 µm. Right: Quantification of total number of cells expressing mCherry. Numbers on the plot are the ratio of mean cell count of + aquashield-1 to that of − aquashield-1 conditions. The center lines indicate mean values of cell count. n = 4 injection sites for both conditions. *P* values are determined by unpaired two-tailed *t*-tests. $^{**}P < 0.01$. Experiment was conducted by collaboration with Dr. Xingyu Li.

We next tested the system in the mouse liver. On the seventh and eighth day after viral delivery, aquashield-1 or saline was administered via intraperitoneal (IP) injection twice (**Figure 2-21**). In the control group, there were limited cells expressing mCherry; whereas the livers of the animals treated with aquashield-1 had a significantly greater number of cells expressing mCherry

(**Figure 2-21**), demonstrating that systematic injection of aquashield-1 can control CAPs in mouse liver tissues.

Taken together, these results demonstrate that the aquashield-1-induced gene regulation works in multiple organs in mice, including the brain and the liver. In addition, aquashield-1 can be readily administered through local or IP injection to activate gene transcription in animal tissue of interest. This is advantageous over light-induced gene transcription systems, as chemicals provide a global control and introduce less disturbance to the animals.

2.5 CAPs are generalizable for other peptides

Modular chemogenetic and optogenetic domains are highly valuable for genetic tool design, because they can be easily incorporated into new systems for new applications. For example, as discussed in Chapter 1, CIP and CDP pairs could be readily applied for various purposes; AsLOV2 domain can be engineered to cage different peptides as well as serving as an allosteric switch. For chemogenetic domains, there has not been a comparable domain as AsLOV2 that can be generally applied to cage peptides yet.

Presumably, CAPs cage peptides by steric hinderance, and this mechanism shall be generally applicable beyond the SsrA and TEVcs that was employed in the directed evolution process. Here we are aiming to show that CAPs could be generally applied to cage other peptides, including NLS and [Met^5]-enkephalin (hereafter referred to as enkephalin).

2.5.1 Shield-1 controlled protein nuclear cytosolic distribution by CAPs-caged NLS

The nucleocytoplasmic distribution of many eukaryotic proteins is a common determinant of their functions35. Previously, NLS has been controlled by light through customized engineering of the AsLOV2 domain[339,340]. To test if NLS can be controlled by CAPs in a shield-1-dependent manner, we fused EGFP to CAPs-caged NLS (**Figure 2-22**).

In the no shield-1 condition, NLS should be sterically blocked, and EGFP should be found throughout the cell. Addition of shield-1 should uncage CAPs and expose NLS to endogenous importins, bringing more EGFP to the nucleus. Due to the strength of the NLS, we added a weak nuclear export signal, PKIt NES[339], to the construct to reduce the nuclear localization of the protein without shield-1. **Figure 2-22** show that EGFP was found both in the cytosol and the nucleus when there was no shield-1. Upon addition of shield-1, EGFP was depleted from the cytosol and preferentially localized in the nucleus with a statistically different cytosol-to-whole-cell ratio than in the basal state.

However, there was some nuclear localization pattern in the absence of shield-1, suggesting that the NLS is not completely blocked by CAPs. This is possibly because the residues in the middle of the NLS sequence are also important for importin recognition and nuclear localization, while CAPs are best at caging residues at the N- and C-terminal portion of a peptide. To further eliminate nuclear pattern in the absence of shield-1, further optimization of CAPs for caging NLS could be performed. Alternatively, NLS sequences with a weaker strength could be used[339]. We also noticed that under the "+ shield-1" condition, there was a weak cytosolic localization pattern in cells with a high level of EGFP expression. This was possibly due to the equilibrium of protein distribution between the nucleus and the cytosol. The current system is therefore not suited for

protein activation by translocating proteins to the nucleus. Instead, it can be used to remove the function of cytosolic proteins by depleting them from the cytosol.

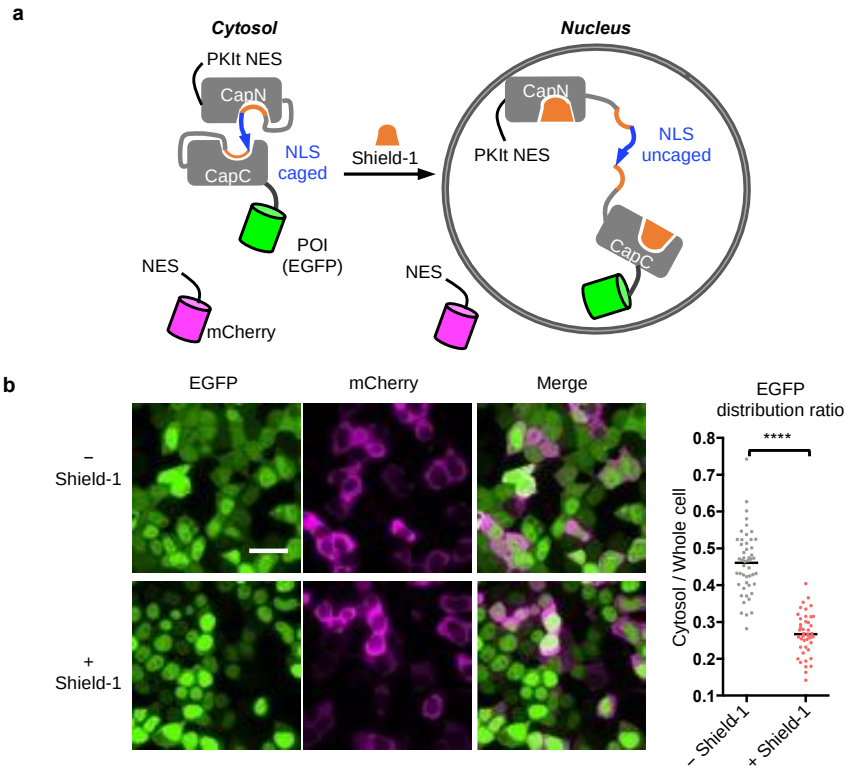


Figure 2-22 Applying CAPs to cage nuclear localization signal peptide. **a,** Scheme of CAPs controlling nuclear localization signal peptide. NES-mCherry is used to indicate the cytosol. NLS (PKKKRKV). NES (LQLPPLERLTLD). PKIt NES, (LALKLAGLDI). **b,** Left: Representative fluorescence microscopy image of HEK 293T cells expressing the constructs shown in **a**. Scale bar, 20 μm. Right: Quantification of EGFP total intensity distribution. The ratio is calculated by the EGFP total intensity in cytosol to that in whole cell. The center lines indicate mean values of the ratio. n = 43 cells (+ shield-1) and n = 49 cells (– shield-1) from one replicate. This experiment was performed three times with similar results. *P* value is determined by unpaired two-tailed *t*-tests. ****$P < 0.0001$.

2.5.2 Shield-1 induced opioid signaling by CapC-caged enkephalin

We sought to apply CapC to cage peptide agonists for G-protein-coupled receptors (GPCRs), which are an important class of receptors regulating various physiological processes[366].

Genetically encoded chemical-switchable peptide agonists can be targeted to specific cellular compartment or cell types.

To test this application, we used CapC to cage enkephalin, an endogenous opioid peptide agonist for the μ-opioid receptor (μOR)[367]. DAMGO, a synthetic analog of enkephalin (**Figure 2-23**) binds to μOR with the N-terminal tyrosine residue bound deep in the binding pocket[368] (**Figure 2-23**). This indicates that the N-terminus of enkephalin must be free of fusions in order to enter the binding pocket and interact with μOR; consequently, only the C-terminus can be tethered to a protein domain.

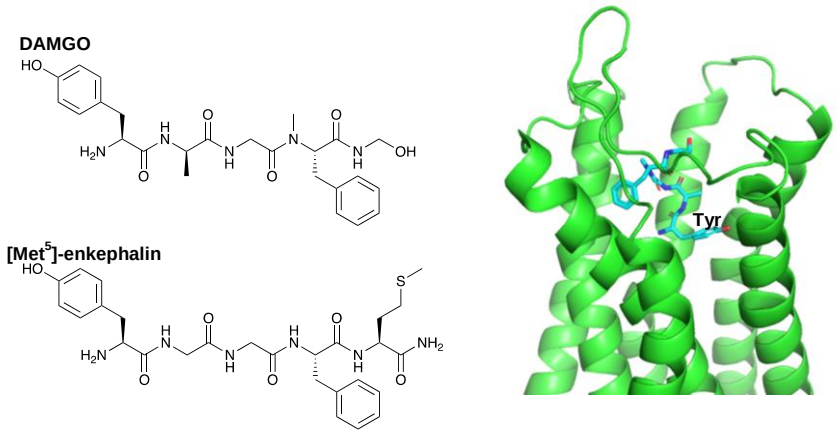


Figure 2-23 Structures of μOR peptide ligands and binding pocket. **a,** Left: Structure of [Met^5]-enkephalin and its analog DAMGO. Right: Crystal structure of μOR binding pocket. PDB: 6DDF.

We tethered CapC-caged enkephalin to the extracellular N-terminus of μOR (**Figure 2-24**). A cleavable signal sequence (KTIIALSYIFCLVFA)[369] is further fused to the N-terminus of enkephalin to direct membrane trafficking without perturbing its function. Enkephalin will not be able to activate μOR under the no shield-1 condition. When shield-1 is added, we expected enkephalin to be unblocked and then it can activate μOR (**Figure 2-24**). We designed two assays to evaluate the ability of CapC to control enkephalin's function. The first one is to detect the μOR conformation change upon activation. In the second assay, we detected the downstream signaling events.

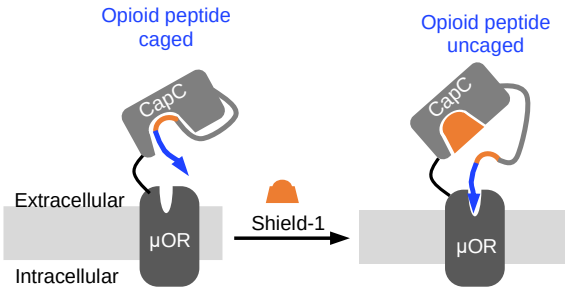


Figure 2-24 Scheme of CapC-controlled opioid peptide.

For the first assay, we utilized a conformation specific binder Nb44, which specifically binds to the ligand-activated state μOR. In this design, a split luciferase assay, NanoBiT[370], was employed to quantify the interaction between Nb44 and μOR (**Figure 2-25**). The binding of Nb44 and μOR results in the reconstitution of the two split luciferase halves, SmBiT and LgBiT, and therefore provides a luminescence readout. Shield-1 was able to activate μOR in this assay, with a slightly lower level compared to loperamide (**Figure 2-25**). However, addition of naloxone showed a lower luminescence signal compared to the no drug condition, indicating μOR partial activation under no shield-1 condition.

To further investigate whether CapC caged enkephalin could be applied to modulate opioid signaling, we next investigated the downstream cAMP level, which could be detected by the GloSensor[371]. We tested this system in HEK 293T cells pre-treated with forskolin to increase the cAMP level, providing a better readout of μOR's inhibitory effect on cAMP level. As shown in **Figure 2-26**, cAMP level increased upon forskolin treatment at 15 minutes. Later, addition of shield-1 induced a 62% decrease in GloSensor signal compared to cells with no drug treatment at 80 minutes (**Figure 2-26**). DAMGO addition at 45 minutes exhibited a similar trend to the shield-1 condition, causing a 76% decrease in GloSensor signal by activating μOR. Together with the results in **Figure 2-25**, this suggests that enkephalin can be controlled by CapC in a shield-1-dependent manner, and successfully induce opioid signaling event.

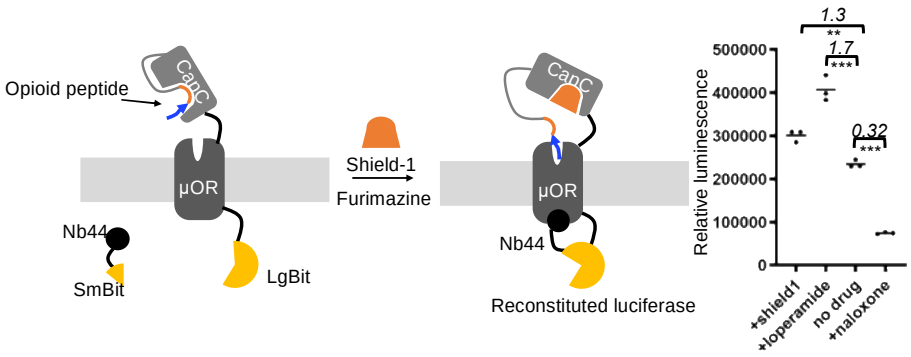


Figure 2-25 Luciferase assay for CapC-controlled opioid peptide. Cells were stimulated with 10 μM of drugs. Values on the plot are the luminescence signal ratios of two conditions. *P* values are determined by unpaired two-tailed *t*-tests. ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001; *NS*, not significant.

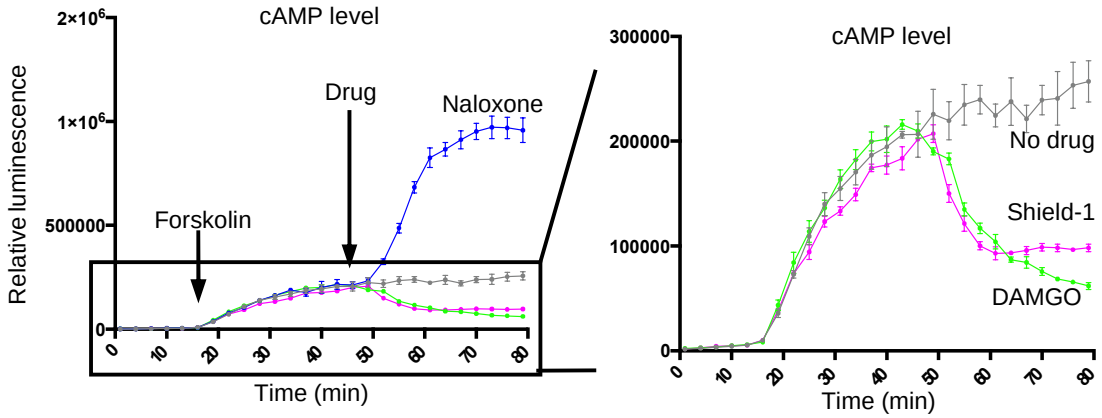


Figure 2-26 The cAMP assay for CapC-controlled opioid peptide. A cAMP biosensor, GloSensor, was co-transfected to indicate cAMP level. Cells were first stimulated with forskolin (1 μM) at 15 min, then stimulated with different drugs (10 μM) at 45 min. Left: Full graph. Right: Zoomed-in graph of the boxed region on the left. n = 3 wells from one replicate for all conditions. This experiment was performed three times with similar results. Errors, s.e.m. Experiment was first performed by Dr. Lequn Geng and Kerry Lee.

To further evaluate the caging efficiency in the basal state, we added the antagonist naloxone at 45 minutes. A 273% GloSensor signal increase was observed, indicating that the enkephalin was not completely caged by CapC, which was consistent with the observations in **Figure 2-25**.

We next investigated if the inverse agonism effect of naloxone attributes to the GloSensor signal increase. In a control study with HEK 293T cells expressing μOR only (with the signal sequence and without the CapC caged enkephalin), DAMGO could still decrease cAMP level by activating μOR, suggesting the μOR is functional (**Figure 2-27**). Neither shield-1 or naloxone

altered the cAMP level at around 45 minutes after the initial forskolin treatment (**Figure 2-27**). This result suggests that naloxone did not function as an inverse agonist under this experiment condition, and the GloSensor signal increase in **Figure 2-26** is due to the leakage of CapC.

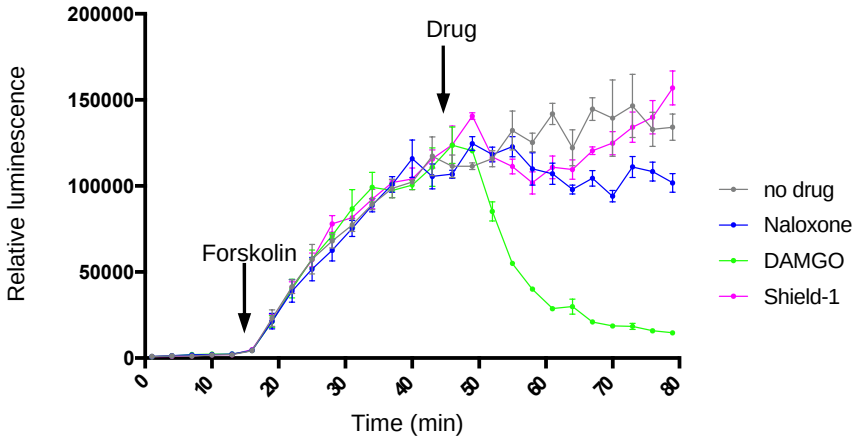


Figure 2-27 The cAMP assay for μOR construct. Cells were stimulated with forskolin (1 μM) at 15 min, and then different drugs (10 μM) at 45 min.

This is the first example of manipulating a peptide agonist for a GPCR by a chemically activated protein domain. Despite the background leakage, shield-1 could uncage the CapC-caged enkephalin, providing a chemical-gating for regulating the activity of μOR. The background is possibly due to the intramolecular binding competition between enkephalin to μOR and the binding sequence to CapC. To reduce such leakage, the binding sequence of CapC could be further optimized. An alternative approach is to fuse the CapC-caged enkephalin to a separate transmembrane domain, so that the binding of enkephalin to μOR becomes intermolecular, which could potentially provide a valuable approach to regulate endogenous μOR activities. Endogenous opioid peptide release induces different downstream signaling responses than synthetic opioids used for clinical treatment[372]. Compared to the other strategy of using DREADD, where KORD is activated by an bioorthogonal agonist SALB[109], regulating opioid receptor activities with opioid peptides is more biologically relevant.

The experiment methods used to produce data in Chapter 2 are provided in **Appendix Methods A-1**.

I acknowledge the following individuals for their contribution on this project. Design and directed evolution of CapC, as well as some of the cell culture applications were from the collaborative work with Dr. Lequn Geng, a previous Ph.D. student from the Wenjing Wang lab, University of Michigan. Application in animal models were from the collaborative work with Dr. Xingyu Li and Catherine Emery from the Peng Li lab, University of Michigan. Computational simulation was from Dr. Matthias Heyden at Arizona State University. Previous Ph.D. student Dr. Kayla Kroning, my undergraduate mentee Gwendolyn Shingles, and previous undergraduate researcher Kerry Lee contributed to some of the cell culture applications.

I sincerely thank Dr. Kayla Kroning, my lab mate and desk mate for five years, for her countless discussions, insightful suggestions, as well as sarcastic comments on this project. I also appreciate her substantial help on the manuscript preparation from which this chapter is largely adapted.

Chapter 3 Development of High Efficiency Yeast Display Directed Evolution Platform with Double Criteria Selection to Optimize Protein Switches

Chapter 3 is prepared into a manuscript: Vazquez-Rivera, L., Shen, J., Shingles, G., Zhou, G., Li, X., Babar, W., Chen, Y., Li, P., Wang, W. A high efficiency yeast surface display platform for double criteria selection. (Manuscript in preparation)

Optogenetic and chemogenetic tool design heavily relies on efficient protein switches. Protein switches are modular protein domains that can be toggled between active and inactive states by external signals such as light and small molecules. Although these protein switches could be readily incorporated for modulating various POIs, optimization on the switches is usually required to achieve high dynamic range.

To develop a protein switch, rational design and computational simulation could serve as initial steps to provide a foundational starting point, while directed evolution is often employed as a critical subsequent step to achieve optimal performance. Therefore, an efficient directed evolution platform would be beneficial for protein switch development.

To evolve protein switches, yeast display[303,334,373], phage display[26-28,325], and ribosomal display[197] have been employed. Phage display[374] and ribosome display[375] benefit from their large library size and high throughput. Phage display offers a library size[376,377] at 10^{10} and up to 10^{11}. Ribosome display provides a library size[376,377] of 10^{12}-10^{13} and can reach up to 10^{14}. Typically screened by affinity selection method, phage and ribosome display are suitable for improving

binding affinities. However, in these two platforms, the POIs are expressed in bacterial context or in vitro, and therefore protein misfolding and immaturity (e.g. antibodies) could be a concern. In addition, the dynamic range of protein switches cannot be showed during the affinity selection processes. On the other hand, yeast display[378] offers protein expression in a eukaryotic environment and allows protein maturation in yeast secretion pathway. When using yeast display, FACS is used as the selection method, which allows monitoring of affinity changes during selection process. The yeast display library size[379] is limited to 10^7, which is lower than that of phage and ribosome display, and the FACS screen throughput[376,380] is limited to 10^8-10^9 (~10^8 per hour[381]). Thus, within the library size and throughput limit, yeast display provides unique advantages for protein switch directed evolution.

To obtain the response to stimulation signal for protein switches, both positive and negative selections are required. Phage display method provides large library size and throughput, so the negative selection and positive selection could be performed within the same round. For example, to develop chemogenetic proximity switches[26-28] (details discussed in Chapter 1.1.1.1), the phage displaying library was first subjected to negative selection on the apo state protein binder, and then the positive selection on the protein-small-molecule complex. After this whole round of selection, the phage can then be collected and amplified. The same workflow has been adapted for the development of optogenetic proximity switch iLID[325] except that the negative and positive selection was based on dark and light conditions.

When using yeast display, only one type of selection could be done in each round, which limits the selection efficiency. Due to the throughput, the collected yeast from FACS in a negative or positive selection need to be amplified before the next round of selection. For example, to evolve AsLOV2 inserted enzyme[303] and AsLOV2 caged TEVcs[334], high activation clones and low leakage

clones were collected in separate positive (light) and negative (dark) selections. Similarly, as described in Chapter 2.1.2 and Chapter 2.2.2, to evolve CapN caged TEVcs and CapC caged SsrA, the positive selections (with shield-1) and negative selections (no shield-1) were separately performed. For yeast display platforms, alternative positive and negative selections have to be performed to obtain the mutants with improved dynamic range.

Although the directed evolution of CAPs was successful and they showed good dynamic range in caging TEVcs and SsrA (Chapter 2.3 and Chapter 2.4) with limited leakage, CAPs caging NLS and enkephalin still showed significant background (Chapter 2.5) and requires further improvement. In addition, while evolving CapC, the binding of CapC caged SsrA to SspB increased in both shield-1 and no shield-1 conditions after positive selection; the binding decreased in both conditions after negative selection (Chapter 2.2.2, **Figure 2-8** and **Figure 2-9**). A directed evolution platform with higher selection efficiency is desired to further improve CAPs and is also beneficial for future development of new protein switches.

Here in this chapter, I discuss the development of a yeast-display based directed evolution platform named DuoSelect, where both the activation and leakage signals are showed on same yeast cell. DuoSelect platform enables selection of clones with both high activation and low leakage within the same round. I demonstrate that this DuoSelect platform provides higher selection efficiency compared to the conventional one, and its application in further improving CAPs. The improved version of CAPs were then applied to cage three neuropeptides: enkephalin, pituitary adenylate cyclase-activating polypeptide (PACAP), and α-Melanocyte-stimulating hormone (α-MSH).

3.1 Design and development of the double criteria platform DuoSelect

It was noticed that CapC library remained very diverse after two rounds of selection (**Table 2-2**), where 18 distinct clones were identified from 20 sequences. The final clone with the best performance (binding sequence: GTPNLRPFG) only showed up once and provided much better shield-1 dependence than most of the other clones. In addition, during the directed evolution process, each round of positive and negative selection shifts the SspB binding for both "with shield-1" and "no shield-1" conditions. Therefore, the conventional yeast display selection method (**Figure 3-1**) is not efficiently selecting the clone with best dynamic range.

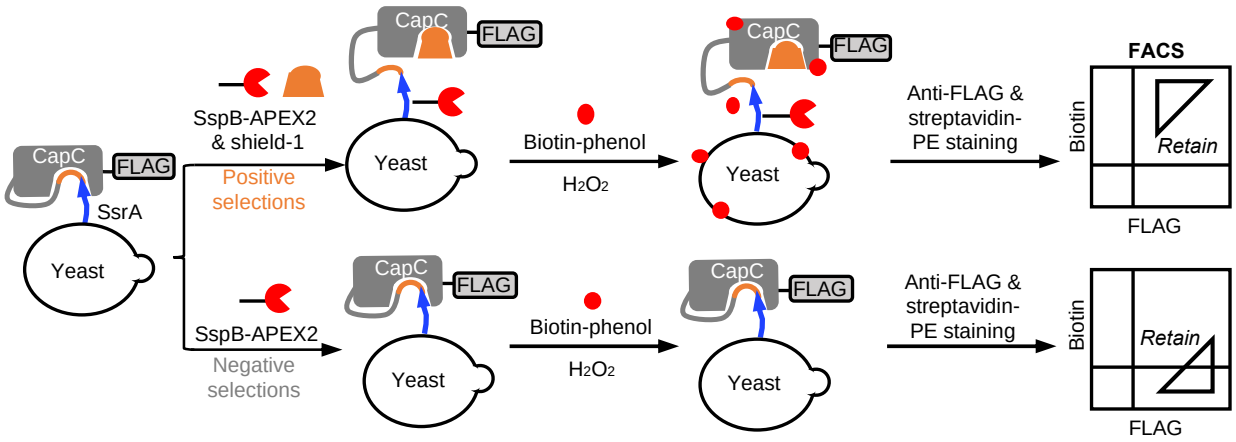


Figure 3-1 Scheme of conventional yeast display platform selection method. CapC caged SsrA is displayed on yeast surface by fusing to the yeast Aga2p protein. APEX2 labels protein within close proximity with biotin-phenol. Alternative positive and negative selections are performed round by round.

To improve the selection efficiency, I then aimed to design a yeast display platform, where "negative selection" and "positive selection" could be done on the same yeast cell. To achieve this, DuoSelect was designed to show both leakage and activation signals on the same yeast cell. A triple gating strategy was applied in DuoSelect to eliminate false positive readouts.

3.1.1 Design of DuoSelect

To show both leakage and activation signals on the same yeast cell, the binding assay with "no shield-1" and "+ shield-1" conditions are sequentially performed on the same cells (**Figure 3-2**). In the leakage labeling step, SspB-APEX2 fusion protein was allowed to bind to yeast surface under "no shield-1" condition. The covalently attached biotin molecules on yeast surface via peroxidase reaction are then stained with streptavidin-PE. The resulted yeasts are then subjected to an activation labeling step, where the yeast cells are treated with SspB-APEX2 fusion protein in the presence of shield-1. The newly attached biotin molecules are then labeled with streptavidin-647. Therefore, the two fluorescence signals, PE and 647, represent the leakage and activation signals, respectively. To ensure that the two fluorescence signals accurately represent leakage and activation, excess amount of streptavidin-PE is used in the leakage labeling round.

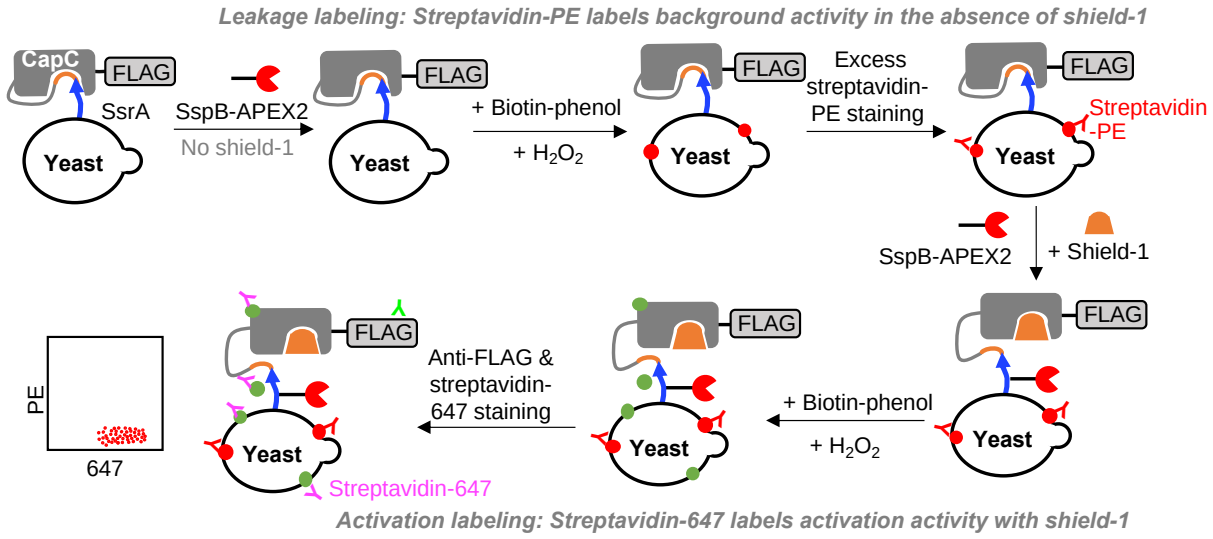


Figure 3-2 Scheme of DuoSelect. CapC caged SsrA is displayed on yeast surface by fusing to the yeast Aga2p protein. APEX2 labels protein within close proximity with biotin-phenol. Two rounds of APEX2 labeling were performed under no shield-1 and with shield-1 conditions. The covalently attached biotin from the two rounds was stained with streptavidin-PE and streptavidin-647, respectively. FLAG, epitope tag.

FACS is used to quantify the activation, leakage, and expression level signals. The desired clones shall present both high activation and low leakage, and therefore should always locate in the bottom right conner on the PE-647 plot (**Figure 3-2**).

3.1.2 Example selection of DuoSelect

To confirm whether the DuoSelect can separate the populations with a higher shield-1 dependence, an example selection was first performed. Yeast cells expressing the post-evolution CapN-caged SsrA (positive cells) are mixed with yeast cells expressing the pre-evolution FKBP-caged SsrA (negative cells). CapN presents a shield-1 dependence, so that positive cell population in DuoSelect shall differ from the negative cell population.

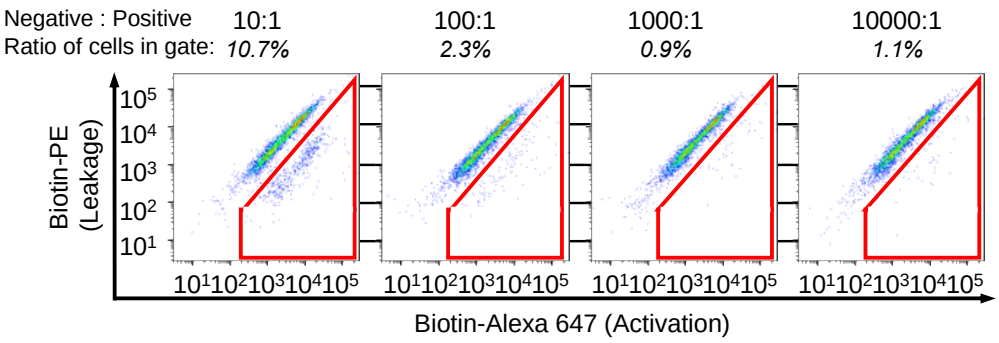


Figure 3-3 FACS plots of example sorting. Positive cells are yeasts expressing the post-evolution CapN caged SsrA. Negative cells are yeasts expressing the pre-evolution FKBP caged SsrA. Only cells expressing FLAG are shown on plots. For each sample, positive cells and negative cells are mixed according to the ratio indicated in the first row. The percentage is the ratio of the cells dropped in the gate. 10 μM shield-1 was used for activation labeling. Experiment was conducted by collaboration with Luis Vazquez-Rivera.

Positive cells and negative cells were mixed with different ratios, and were subjected to the DuoSelect labeling (**Figure 3-2** and **Figure 3-3**). When mixing the negative cells and positive cells with 10:1 ratio, the two populations were successfully separated (**Figure 3-3**), indicating that DuoSelect method can separate CapN variants with different drug-dependence.

The ratio of cells within the gate corresponds well to the mixing ratio when the positive cells are abundant in the sample. However, when the mixing ratio of positive cells reduces to

1000:1, the ratios of cells within the gate do not correspond to the mixing ratios anymore. When diluting the positive cells to 10000:1, there was still 1.1% of the cells within the gate. This result suggests that DuoSelect labeling strategy has high false positive rates. In a real directed evolution experiment, true positive cells also exist with a low ratio, so the false positives need to be taken into concern when performing sorting experiment.

We noticed that the false positive cells in 10000:1 sample all exhibited very high activation signals (Alexa 647), above the main population of cells. This is possibly due to the nonspecific labeling after two rounds of APEX2 labeling. Therefore, to avoid such false positives, a consecutive triple gating strategy is performed (**Figure 3-4**). In addition to the first gate that collects the cells with low leakage activities, we added a gate to select the cells that fall into the main population in the activation graph (**Figure 3-4**, middle). The second gate avoids the super high Alexa 647 signal cells, which could result in false positives in the third gate where we collect the clones exhibiting both high activation and low leakage.

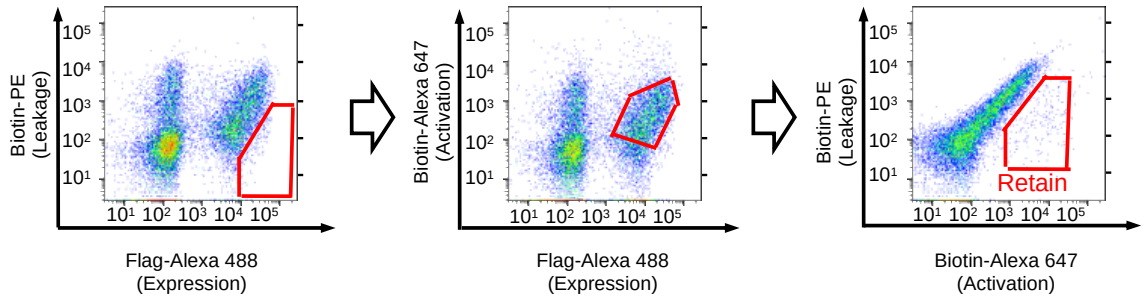


Figure 3-4 Triple gating strategy. Only the cells fall in all the three gates are collected in sorting experiments.

3.2 DuoSelect for CapC improvement

As discussed above in Chapter 3.1, CapC showed high leakage in caging enkephalin and the directed evolution of CapC was not efficient enough. As a proof of principle study, DuoSelect

was first applied for the directed evolution of CapC using the same library built in Chapter 2.2.2. A side-by-side comparison of DuoSelect with conventional selection was performed to investigate the selection efficiency difference. The newly evolved CapC was further characterized on yeast display platform and mammalian two-hybrid system.

3.2.1 Comparison of DuoSelect and conventional selection for the CapC directed evolution

To demonstrate DuoSelect's utility in directed evolution, it was first applied for the CapC library 2, from which CapC was obtained (**Figure 3-5**). Herein, the first version of CapC is referred to as CapC1.0.

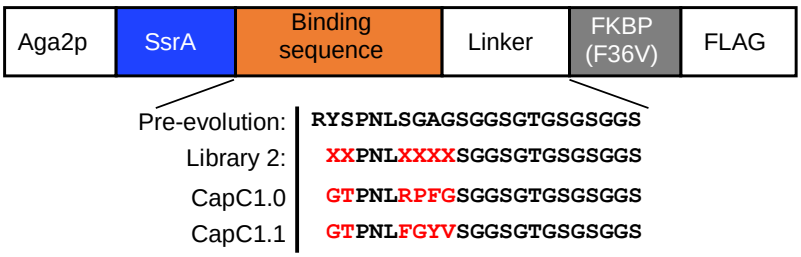


Figure 3-5 CapC sequence before, during, and after directed evolution. CapC1.0 is the clone obtained in Chapter 2.2.2[373]. CapC1.1 is the clone obtained by DuoSelect. "X" indicates any of the twenty amino acids.

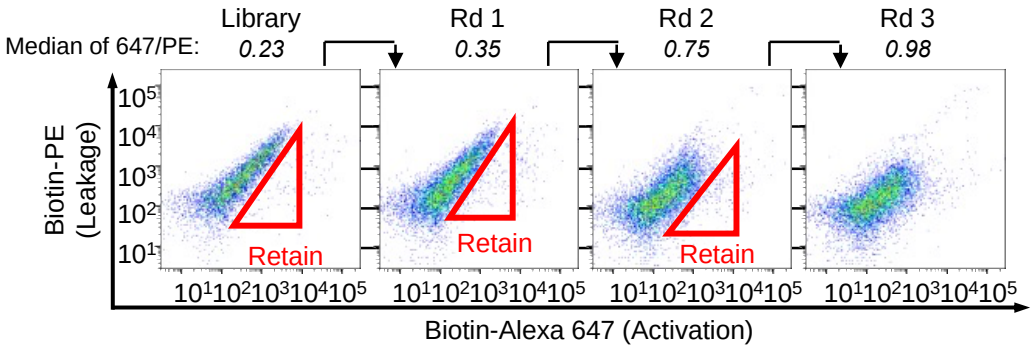


Figure 3-6 FACS selection of CapC library by DuoSelect. Values above the plots are the median value of the Alexa 647/PE signal ratios. Only Flag positive cells are shown in plots. Shield-1, 10 μM. Experiment was conducted by collaboration with Luis Vazquez-Rivera.

The CapC library was subjected to both DuoSelect (**Figure 3-2**) and the conventional selection method (**Figure 3-1**) for a side-by-side comparison. For DuoSelect, the consecutive triple gating strategy was employed (**Figure 3-4**), and three rounds of selection were performed according to the scheme shown in **Figure 3-5 and Figure 3-6** to enrich the low leakage and high activation clones. After three rounds of selection, the population shifted towards the bottom right corner. The Alexa 647/PE signal ratios, meaning activation/leakage, was used to represent and quantify the dynamic range. The median value of the Alexa 647/PE signal ratios was calculated by R programming (for more analysis details, see **Appendix Method B-1**). The median value of Alexa 647/PE also continuously increased as the selection went. This result suggests that the DuoSelect successfully achieved the directed evolution of CapC.

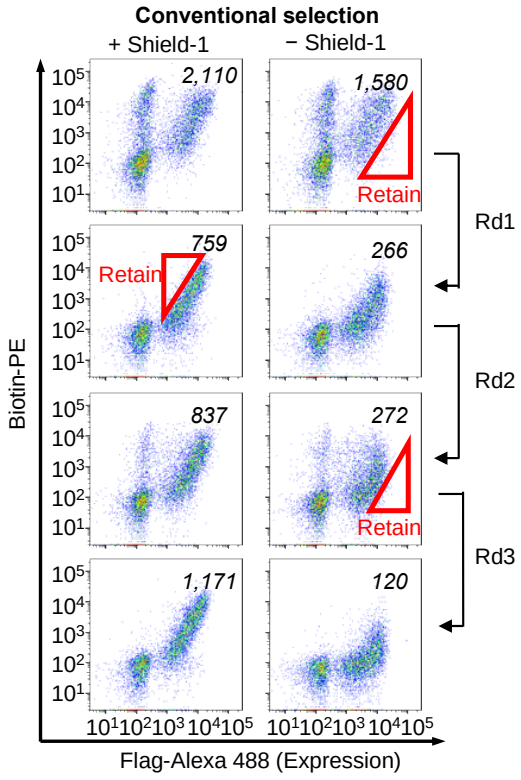


Figure 3-7 Selection scheme of conventional selection method. The values in the plots are the median PE signal of the Flag positive cells. Shield-1, 10 μM. Experiment was conducted by collaboration with Luis Vazquez-Rivera.

In the conventional selection process, a negative selection, a positive selection, and a negative selection were sequentially performed (**Figure 3-7**). After three rounds of sorting, an obvious shield-1 dependence was reproduced.

We next investigated the selection efficiency from two aspects: 1) library diversity reduction; and 2) shield-1 dependence improvement in each round. A more efficient selection method should reduce the library diversity and improve the shield-1 dependence in less rounds of selection.

Twenty-four clones from each of the post 3rd round library were sequenced (**Table 3-1**, **Table 3-2**). DuoSelect resulted in only nine distinct sequences with one clone highly enriched (seven times). In comparison, the conventional method resulted in nineteen distinct sequences, and the most enriched one only showed up three times. This result suggests that DuoSelect reduces library diversity and enriches the desired clones faster than the conventional method. The most enriched clone from DuoSelect was named as CapC1.1, and the one developed in Chapter 2.2.2 was named as CapC1.0 for differentiation.

Table 3-1 Sequences of twenty-four clones from the post 3rd round CapC library by DuoSelect. Nine distinct sequences were identified. CapC1.1 is the improved version of CapC used in the rest of this chapter.

Label	Frequency	Sequence
CapC1.1	7	GTPNLFGYV
CapC1.2	4	GSPNLSHRR
CapC1.3	4	GSPNLRASY
CapC1.4	3	GSPNLVPHI
CapC1.5	2	GSPNLRIPD
CapC1.6	1	GFPNLGSWN
CapC1.7	1	GSPNLGVRA
CapC1.8	1	TFPNLNQTD
CapC1.9	1	GFPNLRNVE

Table 3-2 Sequences of twenty-four clones from the post 3rd round CapC library by conventional selection method. Nineteen distinct sequences were identified.

Label	Frequency	Sequence	Label	Frequency	Sequence
#1	3	GSPNLSSYI	#10	1	GSPNLYGES
#2	2	GTPNLRHYA	#11	1	GTPNLSSTS
#3	2	GSPNLQGRS	#12	1	AWPNLYPLP
#4	2	GSPNLRKYL	#13	1	AWPNLYRME
#5	1	GSPNLGKPD	#14	1	IPPNLSRLL
#6	1	GSPNLHSRV	#15	1	KLPNLKESI
#7	1	GSPNLNRVW	#16	1	KSPNLHWRV
#8	1	GSPNLRKGW	#17	1	NFPNLSGKI
#9	1	GSPNLRVSR	#18	1	SPPNLFGDR
			#19	1	TWPNLRGIE

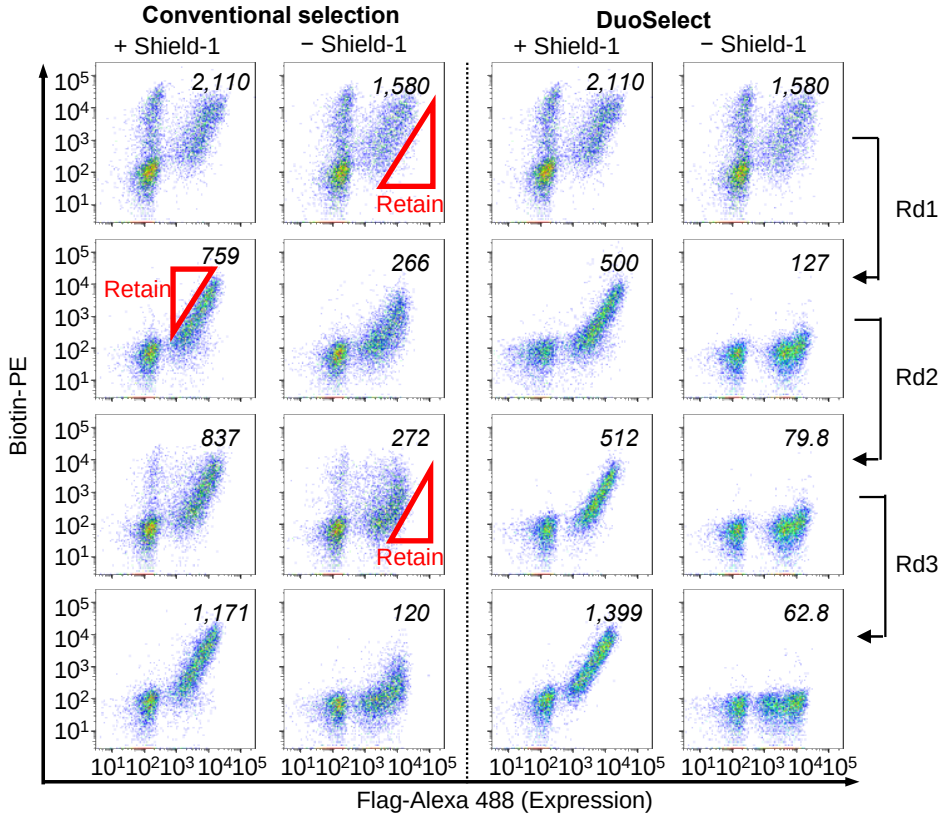


Figure 3-8 One-color characterization of libraries sorted by conventional selection method and DuoSelect. Values in the plots are the median PE signal of the Flag positive cells. Shield-1, 10 μM. Experiment was conducted by collaboration with Luis Vazquez-Rivera.

To compare the shield-1 dependence improvement, the libraries sorted from each round were first characterized by the one-color platform (**Figure 3-1**), where the cells were labeled under with or without shield-1 conditions and stained with streptavidin-PE. As shown in **Figure 3-8**, after one round of DuoSelect, the library already showed significant shield-1 dependence as well as very low leakage signal. On the other hand, after one round of negative selection by the conventional method, only subtle shield-1 dependence was observed, and the leakage signal is still significant. In addition, the median PE signal of both "no shield-1" and "with shield-1" conditions decreased after negative selection and increased after positive selection. After three rounds of sorting, the library from DuoSelect showed both higher activation and lower leakage signal than that from conventional method (**Figure 3-8**).

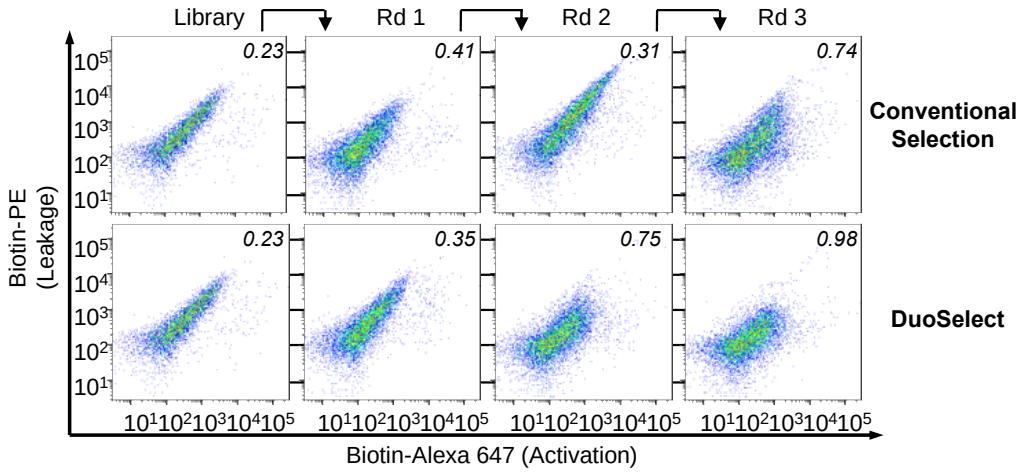


Figure 3-9 Two-color characterization of libraries sorted by conventional selection method and DuoSelect. Values in the plots are the median Alexa 647/PE signal ratio of the Flag positive cells. Only Flag positive cells are shown in plots. Shield-1, 10 μM. Experiment was conducted by collaboration with Luis Vazquez-Rivera.

We then investigated whether the DuoSelect can be used as a characterization platform for protein switches. The libraries sorted from each round were labeled according to the DuoSelect scheme (**Figure 3-2**). The Alexa 647/PE signal ratios were used to quantify the shield-1 dependence. Parallel to the observation from the one-color characterization, the shield-1

dependence of conventionally sorted library shifted back and forth due to the negative and positive selection, while the DuoSelect platform kept increasing the shield-1 dependence (**Figure 3-9**).

Therefore, DuoSelect reduces library diversity and improves shield-1 dependence more efficiently than the conventional method. In addition, DuoSelect could also serve as a platform to characterize the drug dependence.

3.2.2 Characterization of CapC on yeast display platform

Nine distinct clones were obtained after three rounds of DuoSelect (**Table 3-1**). The ones appeared more than once (CapC1.1~1.5) were then characterized (**Figure 3-10**, **Appendix Figure B-1**) in parallel with CapC1.0.

All the five clones showed shield-1 dependence (**Appendix Figure B-1**). The most enriched CapC1.1 showed lower leakage than CapC1.0 and still maintained good activation. Previous data in Chapter 2.5.2 suggested that the main concern using CapC to cage enkephalin is the high basal activity but not the activation. Therefore, CapC1.1 could be a better candidate for such applications where leakage need to be minimized.

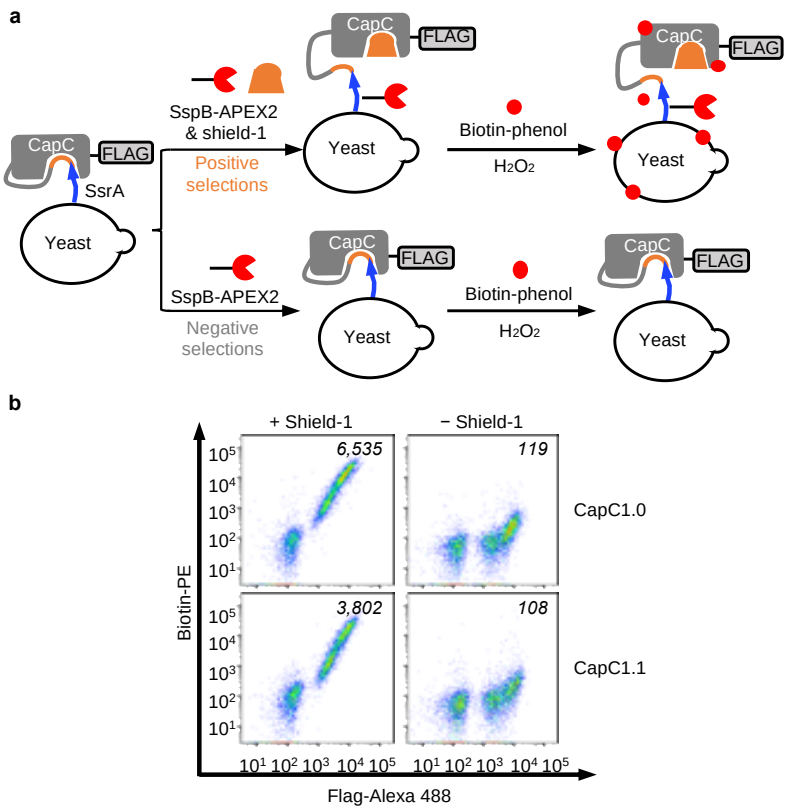


Figure 3-10 a, Labeling scheme of CapC-caged SsrA. CapC-SsrA is displayed on yeast surface by fusing to the yeast Aga2p protein. APEX2 labels protein within close proximity with biotin-phenol. **b,** Comparison of CapC1.0 with CapC1.1 by the assay shown in **a**. The values in the plots are the median PE signals of the Flag positive populations. Shield-1, 10 μM. Experiment was conducted by collaboration with Luis Vazquez-Rivera.

3.2.3 Characterization of CapC in shield-1 induced mammalian gene transcription system

To further examine the utility of newly evolved CapC variants in mammalian systems, the shield-1 induced gene transcription system was employed to evaluate their performance (**Figure 3-11**).

In this assay, CapC-caged SsrA is fused to a Gal4-DBD, and SspB to VP16. In the presence of shield-1, SspB-VP16 will be recruited to the Gal4-DBD and initiate transcription of the reporter gene. A *Photinus pyralis* luciferase was used as the reporter gene and the expression level was

quantified by Bright-Glo luciferase assay. To fully investigate the leakage of these variants, only CapC was used to cage SsrA.

As shown in **Figure 3-12**, all the CapC variants except CapC1.4 showed significant shield-1 dependence in this assay. CapC1.1 showed similar dynamic range and basal activity as CapC1.0. Nonetheless, CapC1.1 could still be transferred to mammalian system for efficient caging of SsrA.

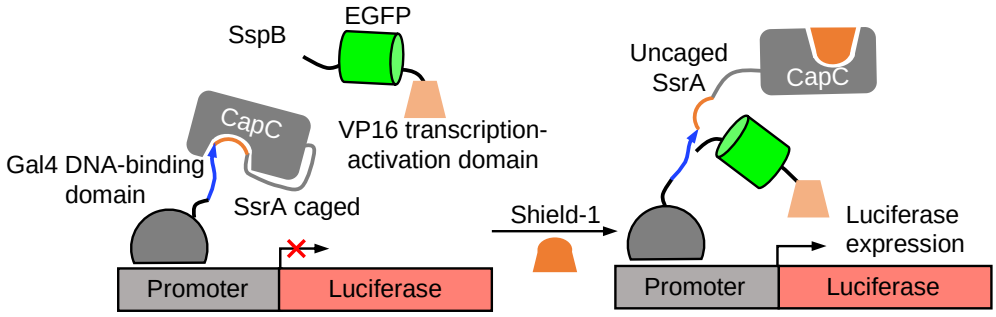


Figure 3-11 Scheme of shield-1-induced gene transcription assay. Addition of shield-1 uncages SsrA and recruits VP16 into proximity with the reporter gene to activate gene transcription. The reporter gene encodes a *Photinus pyralis* luciferase. The expression was quantified by Bright-Glo assay.

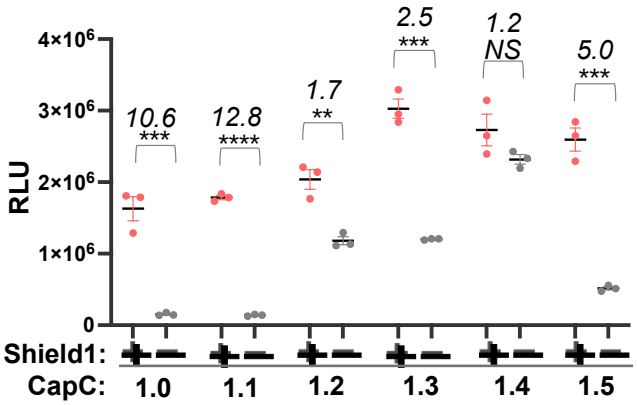


Figure 3-12 Quantification of luciferase expression level. Values on the plot are the ratio of mean luminescence signal of + shield-1 to that of − shield-1 conditions for each construct. The center lines indicate mean values of luminescence intensity. Shield-1, 10 μM. *P* values are determined by unpaired two-tailed *t*-tests. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; *NS*, not significant. Errors, s.e.m. Experiment was conducted by Gwendolyn Shingles.

3.3 DuoSelect for CapN directed evolution

Given that DuoSelect successfully evolved CapC, we then sought to employ DuoSelect to further improve CapN. A new library was designed based on the CapN developed in Chapter 2.1 (hereafter referred to as CapN1.0). DuoSelect was applied to perform the CapN directed evolution. The improved version of CapN was then characterized on yeast display platform.

3.3.1 Directed evolution of CapN by DuoSelect

To further improve CapN1.0, we sought to employ DuoSelect for a new yeast display library based on CapN1.0. We hypothesized that the affinity of binding sequence to FKBP could be improved by optimizing the preceding residues within the binding sequence. Therefore, the library was generated by introducing site-saturated mutagenesis to the four amino acids before the critical PNL sequence (**Figure 3-13**).

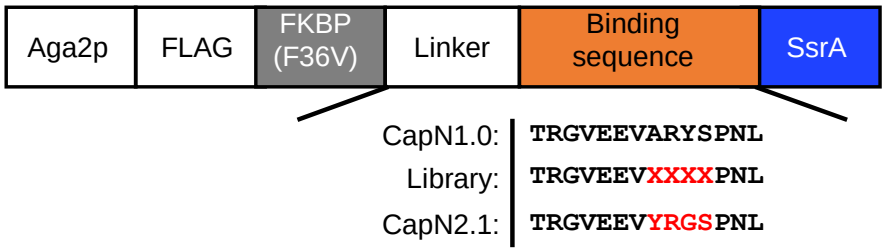


Figure 3-13 CapN sequence before, during, and after directed evolution. "X" indicates any of the twenty amino acids. Amino acids that are different from the CapN1.0 sequence are highlighted in red.

The CapN library was then subjected to DuoSelect sorting according to the scheme shown in **Figure 3-14**. By four rounds of selection on the population exhibiting both high activation and low leakage signals, a very significant improvement was observed. The Alexa 647/PE signal ratio was improved from 4.9 for CapN1.0 to 35.9 for the third-round library. Therefore, we further proved that DuoSelect is an efficient platform for directed evolution.

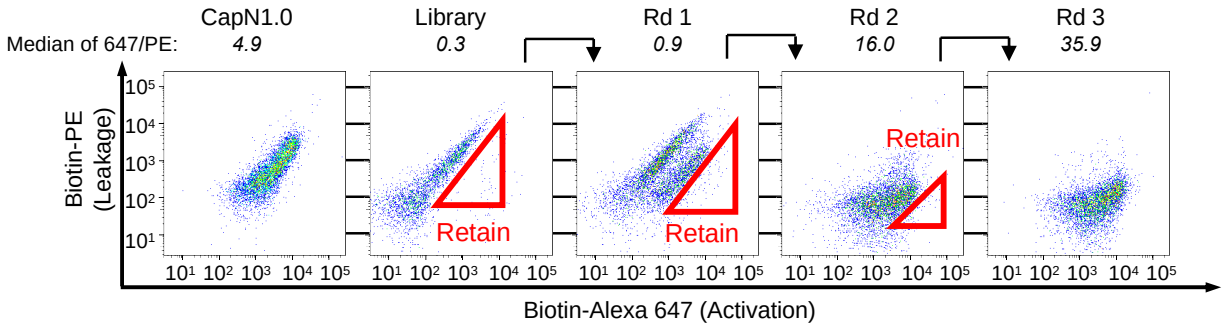


Figure 3-14 FACS selection of CapN library by DuoSelect. Values above the plots are the median value of the Alexa 647/PE signal ratios. Only Flag positive cells are shown in plots. Shield-1, 10 μM. Experiment was conducted by Luis Vazquez-Rivera.

3.3.2 Characterization of CapN on yeast display platform

After three rounds of DuoSelect sorting and sequencing twenty-five individual clones, three distinct clones were identified (**Table 3-3**). When evolving CapN1.0, the last six residues in the binding sequence were mutated (generated RYSPNL). Here in this library, the Arg-Tyr-Ser sequence was further randomly mutated, but the CapN2.1~CapN2.3 sequences showed similar results on these three positions. The small hydrophobic Ala residue was mutated to negatively charged Glu and bulky hydrophobic Tyr residues.

The three newly evolved clones were then characterized by the one-color platform (**Figure 3-1**, **Figure 3-15**). All the three clones showed significantly lower leakage than CapN1.0. The most enriched CapN2.1 was employed for the rest of this study.

Table 3-3 Sequences of twenty-five clones from the post 3rd round CapN library by DuoSelect. Three distinct sequences were identified. CapN2.1 is the improved version of CapN used in the rest of this chapter.

Label	Frequency	Sequence
CapN1.0	/	TRGVEEVARYSPNL
CapN2.1	19	TRGVEEVERYMPNL
CapN2.2	5	TRGVEEVERYSPNL
CapN2.3	1	TRGVEEVYRGSPNL

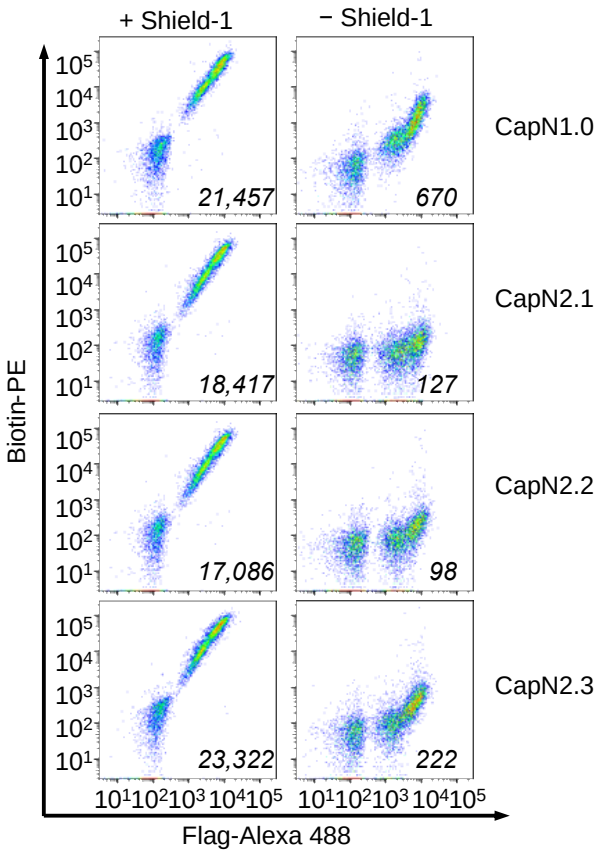


Figure 3-15 FACS analysis of the three clones from DuoSelect, corresponding to CapN2.1 ~ CapN2.3 shown in **Table 3-3**. Values are median PE intensity of FLAG-positive cells. All three clones showed shield-1 dependence. Experiment was conducted by Luis Vazquez-Rivera.

3.4 Shield-1 induced GPCR signaling by CAPs caged peptide ligands

CAPs caged membrane receptor ligand peptide serves as a unique and valuable approach to investigate the receptors' biology, where the endogenous receptors are chemogenetically stimulated by their endogenous peptide ligand. With improved versions of CAPs, we sought to apply them to modulate more peptide ligands. , I discuss the application of CAPs by me and by other lab members to cage three neuropeptides including enkephalin (me, Lequn, Luis and Guanwei), PACAP (Gwendolyn Shingles), and α-MSH (Ryan Singer). CapC caged enkephalin was tethered on the extracellular side of μOR (**Figure 3-16**). For PACAP and α-MSH, a DART-

like strategy[120-122] (more details are discussed in Chapter 1.1.2.1) was employed (**Figure 3-16**) to allow the targeting of endogenous receptors. CAPs caged PACAP and α-MSH are displayed on cell surface in a separate component than their receptors, pituitary adenylate cyclase 1 receptor (PAC1R) and melanocortin-4 receptor (MC4R).

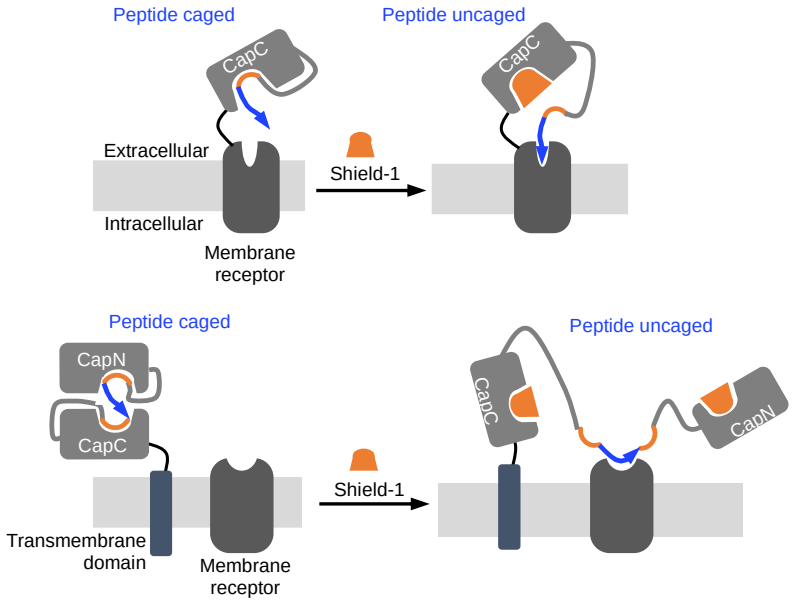


Figure 3-16 Scheme of CAPs-controlled receptor ligand peptides. Top: The CapC caged peptide is tethered to the extracellular part of the receptor. Bottom: DART-like strategy. CAPs caged peptide is displayed on cell surface by a transmembrane domain. Only CapC is used for PACAP. Both CapN and CapC are used for α-MSH.

3.4.1 Shield-1 induced μOR signaling by CapC1.1 caged enkephalin

To demonstrate the application of CapC1.1, we used it to control enkephalin by the design shown in Chapter 2.5.2 (**Figure 3-17**). By detecting the amplified downstream signaling of μOR activation, this system could also better differentiate the caging efficiency difference between CapC1.0 and CapC1.1. The downstream signaling is evaluated by the second messenger cAMP, which is detected by GloSensor[371]. A cleavable signal sequence (KTIIALSYIFCLVFA)[369] is fused to the N-terminus of enkephalin to direct membrane trafficking without perturbing its function.

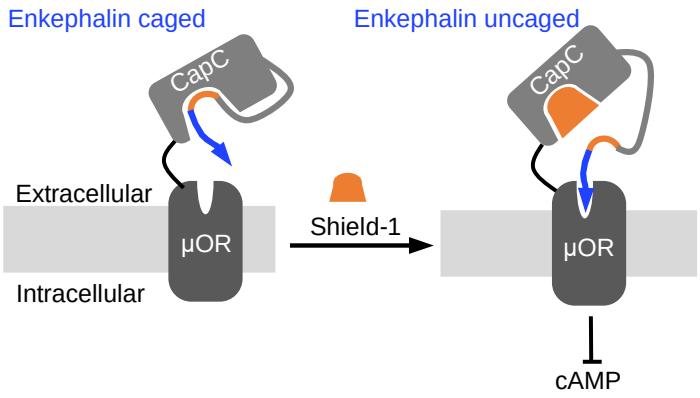


Figure 3-17 Scheme of CapC-controlled enkephalin. Addition of shield-1 uncages enkephalin and therefore activates the μOR, leading to inhibition of cAMP production. cAMP is detected by GloSensor[371].

We tested this system in HEK 293T cells pre-treated with forskolin to increase the cAMP level, and then the cells were treated with shield-1, loperamide (agonist), naloxone (antagonist) and media. As shown in **Figure 3-18**, addition of shield-1 successfully reduced the cAMP level, indicating that both CapC1.0 and CapC1.1 can cage enkephalin to some extent. The addition of the full agonist loperamide[382] also induced a decrease in cAMP level. The addition of the antagonist naloxone increased cAMP level, indicating that both CapC1.0 and CapC1.1 are not fully caging enkephalin. These results are consistent with the observations in Chapter 2.5.2.

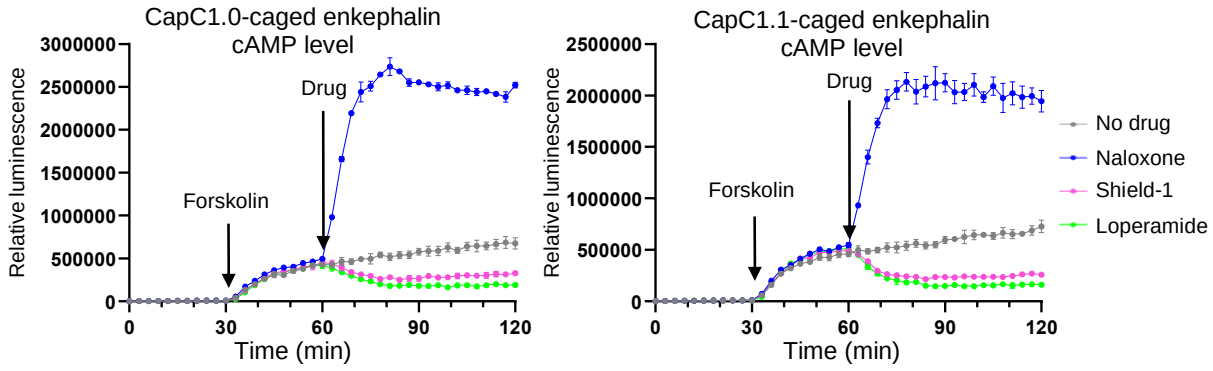


Figure 3-18 The cAMP assay for CapC1.0- and CapC1.1-controlled opioid peptide. A cAMP biosensor, GloSensor, was co-transfected to indicate cAMP level. Cells were first stimulated with forskolin (1 μM) at 30 min, then stimulated with different drugs (10 μM) at 60 min. Left: CapC1.0-caged enkephalin. Right: CapC1.1-caged enkephalin. n = 3 wells from one replicate for all conditions. Errors, s.e.m. Experiment was conducted by collaboration with Guanwei Zhou.

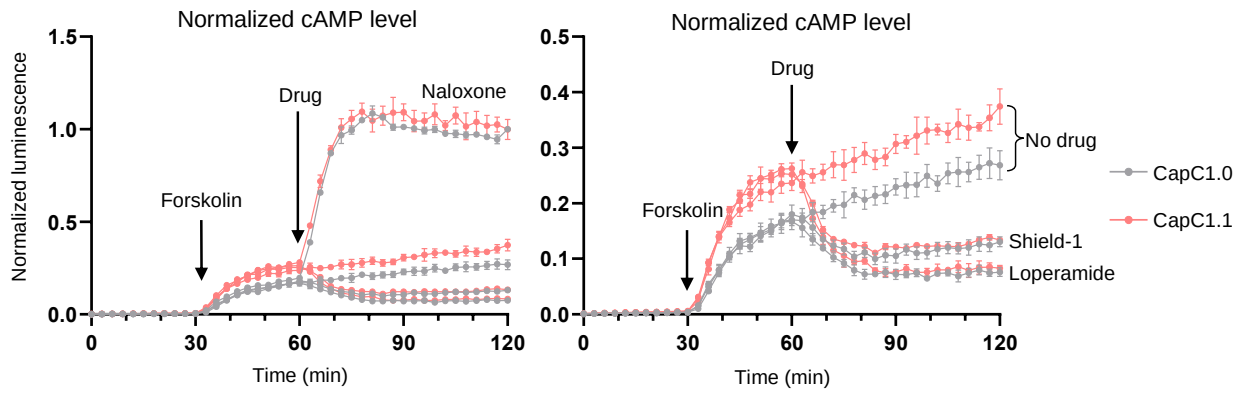


Figure 3-19 Normalized luminescence graph of **Figure 3-18** comparing CapC1.0 and CapC1.1. Left: Full graph. Right: Zoomed in graph with no drug, shield-1, and loperamide conditions. n = 3 wells from one replicate for all conditions. Errors, s.e.m. Experiment was first performed by collaboration with Guanwei Zhou.

The absolute luminescence signals between the CapC1.0 and CapC1.1 are different (**Figure 3-18**). To compare the caging efficiency between the two constructs, all the signals are normalized to the mean luminescence of naloxone condition at the last time point (**Figure 3-19**). This will normalize the responses to the basal activity of μOR, so that the differences across the two transfection conditions are eliminated. When all the luminescence signals with each construct are normalized to the naloxone condition, the luminescence signals of shield-1 and loperamide with each construct aligns, suggesting that this data processing method is reasonable.

For the shield-1 condition, the CapC1.1-caged enkephalin showed higher GloSensor signal than CapC1.0. To quantify this difference, the leakage is defined as the normalized luminescence signal difference between the naloxone and no drug conditions; the activation is defined as the difference between the shield-1 and no drug condition. Compared to CapC1.0, the leakage of CapC1.1 decreased by 14.5%, and the shield-1 induced activation increased by 74.8%. Therefore, CapC1.1 cages enkephalin better than CapC1.0 does, without the sacrifice of activation efficiency.

3.4.2 Shield-1 induced PAC1R signaling by CapC1.1 caged PACAP

PAC1R plays a crucial role in diverse physiological processes by mediating the signaling with its ligand, the PACAP neuropeptide. PAC1R is primarily expressed in central and peripheral nervous systems. The PACAP induced PAC1R signaling has been shown to be associated with stress related behaviors, including depression and posttraumatic stress disorder[383,384]. Chemogenetically modulated PACAP will serve as a useful tool to study these important biological processes.

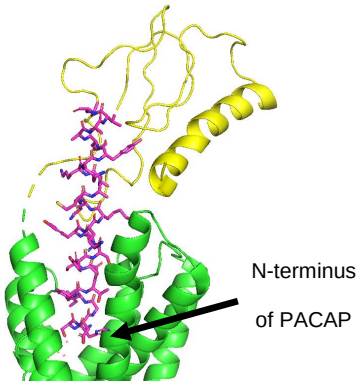


Figure 3-20 Crystal structure of the PAC1R/PACAP(1-27) complex. Green: PAC1R. Magenta: PACAP(1-27). Yellow: Extracellular domain of PAC1R. PDB: 8E3X.

PAC1R is a member of class B GPCR family, which consists of a large N-terminal extracellular domain (ECD). The N-terminus of PACAP is reaching into the deep binding pocket, so only CapC could be used to modulate its function (**Figure 3-20**, **Appendix Figure B-2**). It seemed to be challenging to fully cage PACAP by CapC because PACAP is much longer (> 20 amino acids) than the peptides CAPs were used to cage (7 amino acids for NLS and SsrA). However, the recognition of PACAP C-terminal region by ECD is critical for the receptor activation[383,385](**Figure 3-20**, **Appendix Figure B-2**), which provides the opportunity of modulating the peptide recognition.

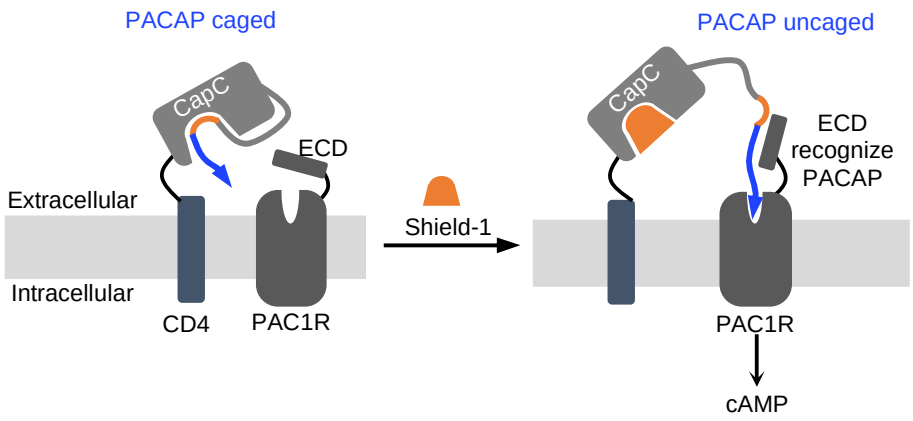


Figure 3-21 Scheme of CapC-controlled PACAP. Addition of shield-1 uncages PACAP and therefore activates the PAC1R, leading to increase of cAMP production. cAMP is detected by GloSensor[371]. CD4 is a transmembrane domain. ECD is the N-terminal extracellular domain of PAC1R.

A separate chain design was employed to enable the modulation of endogenous PAC1R and to minimize the basal activity (**Figure 3-21**). The CapC1.1-caged PACAP is displayed on cell membrane by fusing to a transmembrane domain CD4. A cleavable signal sequence (KTIIALSYIFCLVFA)[369] is fused to the N-terminus of PACAP to aid membrane trafficking. PACAP will be caged until the addition of shield-1. Activation of PAC1R will lead to the increase of intracellular cAMP level, which can be detected by GloSensor[371].

Table 3-4 EC_{50} and efficacy of different PACAP variants towards PAC1R. Data is adapted from previous report[386].

PACAP variants	Binding IC_{50}	EC_{50} ± s.e.m. (nM)	Efficacy (%)
PACAP(1-27)	6.7 ± 1.2	5.2 ± 1.0	100 ± 2
PACAP(1-23)	32 ± 8	51 ± 7	96 ± 2
PACAP(1-22)	740 ± 135	960 ± 154	89 ± 4
PACAP(1-21)	> 1000	> 1000	NA
PACAP(1-20)	> 1000	> 1000	NA
PACAP(1-19)	> 1000	Inactive up to 1 μM	Inactive up to 1 μM

To achieve best caging efficiency, we first tested different C-terminal truncated PACAP variants. It has been previously reported that truncating the C-terminal residues in PACAP can alter the binding affinity and EC_{50} of them towards PAC1R[386] (**Table 3-4**). The truncated PACAP

variants showed lower affinity and truncation beyond 6 residues abolished the activity. Therefore, only three variants were tested for this assay: PACAP(1-27), PACAP(1-23), and PACAP(1-22).

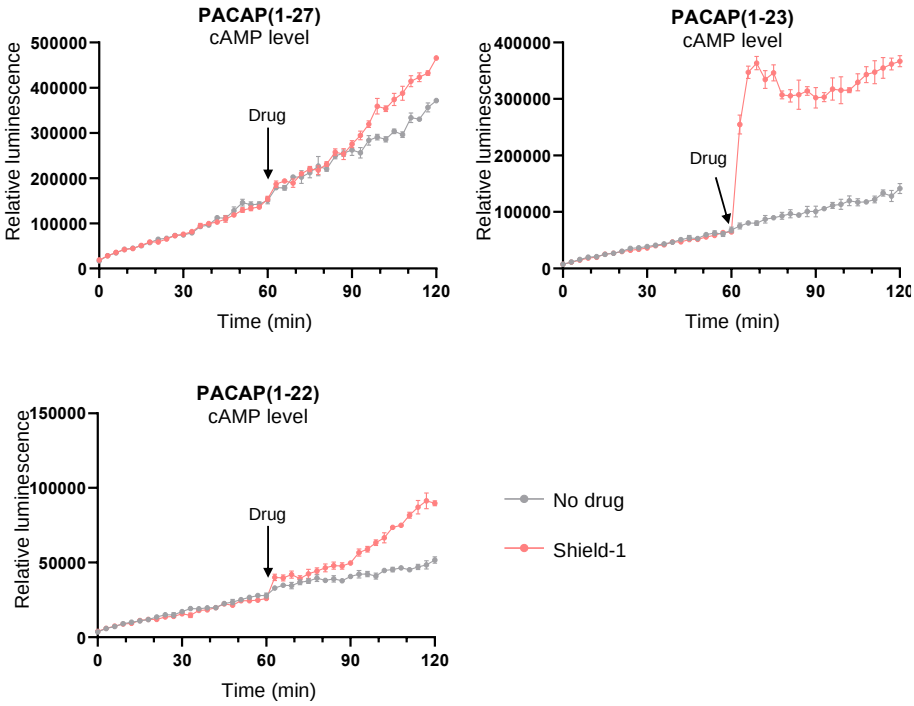


Figure 3-22 The cAMP assay for CapC1.1-controlled PACAP variants. GloSensor and PAC1R were co-transfected. Cells were stimulated with shield-1 (10 μM) or media at 60 min. n = 3 wells from one replicate for all conditions. Errors, s.e.m. Experiment was conducted by Gwendolyn Shingles.

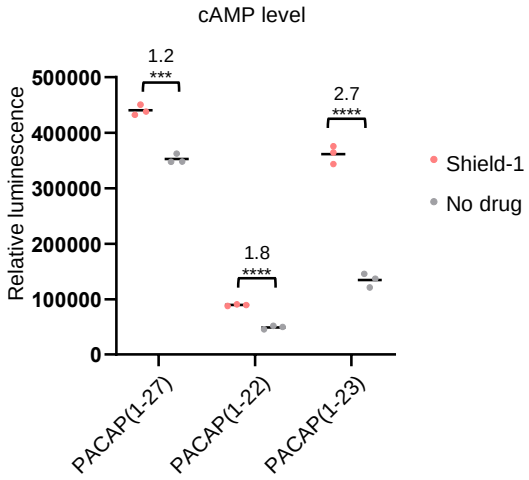


Figure 3-23 Shield-1 dependence for CapC1.1-controlled PACAP variants. Each dot stands for the mean luminescence signal of one well at the last three time points (114 min, 117 min, 120min). The values in the plot are the ratio of the mean luminescence of shield-1 to no drug conditions. n = 3 wells from one replicate for all conditions. *P* values are determined by unpaired two-tailed *t*-tests. ****P* < 0.001; *****P* < 0.0001.

All the three variants showed response to shield-1 stimulation. To quantify this response, the mean luminescence signals of the last three time points were used to determine the shield-1 dependence (**Figure 3-23**). CapC1.1-caged PACAP(1-23) showed the highest shield-1 dependence, and provided 2.7-fold shield-1-dependent PAC1R activation.

We reasoned that shield-1 dependence could be further improved by adjusting the protein expression level. Reducing the PACAP concentration on the membrane will lower the background activation of PAC1R (**Appendix Figure B-3**). When the CapC1.1-caged PACAP plasmid used for transfection is reduced to 10 ng/well and 5 ng/well (50 ng/well was used for previous experiments), the shield-1 dependence increased to over 6-fold (**Appendix Figure B-3**). Therefore, 5 ng/well transfection condition is used for latter experiments.

To confirm that the cAMP increase is due to the PAC1R activation, a negative control was performed in parallel (**Figure 3-24**), where the cells expressed CapC1.1-CD4 without PACAP. The addition of shield-1 induced minimal effect for the negative control. The direct addition of the PACAP(1-27) agonist significantly increased the cAMP level, suggesting that PAC1R is functional. In comparison, when CapC1.1 caged PACAP(1-23) is expressed, shield-1 induced a significant increase in cAMP level (3.4-fold higher than no drug condition). PACAP(1-27) also activated the PAC1R and increased cAMP level (70-fold higher than no drug condition).

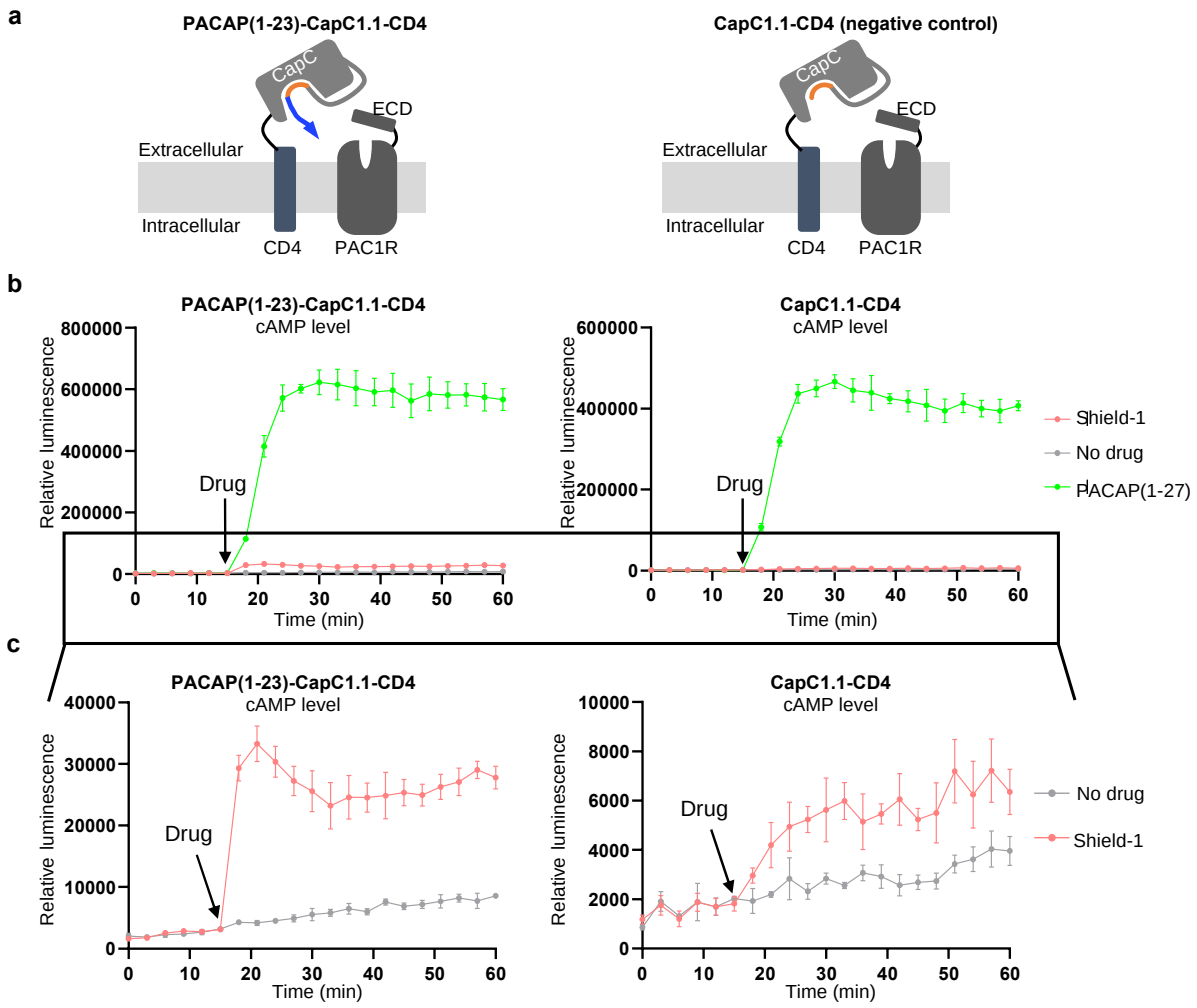


Figure 3-24 a, Scheme of the constructs tested. The CapC1.1-CD4 construct is the negative control and does not contain the PACAP peptide. **b,** The cAMP assay for CapC1.1-controlled PACAP. Cells were stimulated with shield-1 (10 μM), PACAP(1-27) (10 μM) or media at 15 min. **c,** Zoomed-in graph of the boxed region in **b**. n = 3 wells from one replicate for all conditions. Errors, s.e.m. Experiment was conducted by collaboration with Gwendolyn Shingles.

The absolute luminescence signals between the CapC1.1-caged PACAP and the negative control are different. To evaluate the basal activity of CapC1.1-caged PACAP, all the signals are normalized to the mean luminescence of the "+ PACAP(1-27)" condition at the last time point (**Figure 3-25**). For the CapC1.1-caged PACAP construct, the "no drug" condition overlaped with the negative controls, indicating minimal basal activity was present.

These results suggests that PACAP(1-23) can be successfully caged by CapC1.1 and control PAC1R activity in a shield-1-dependent manner. The direct addition of agonist induced a

much higher activation under this condition. This is possibly due to the low expression level of the CapC1.1-caged PACAP construct.

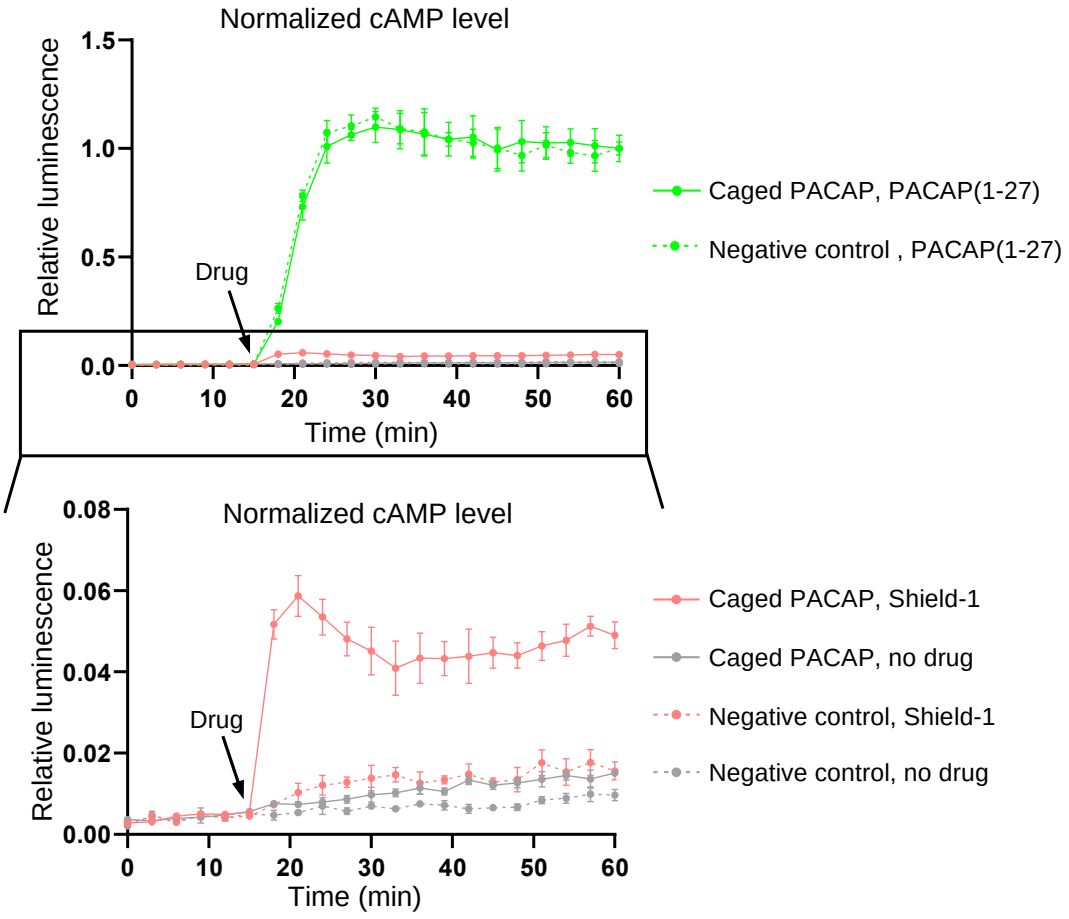


Figure 3-25 Normalized luminescence graph of **Figure 3-24** comparing CapC1.1-caged PACAP with negative control. Top: Full graph. Bottom: Zoomed in graph with no drug, and shield-1 conditions. Solid lines are the traces of CapC1.1-caged PACAP construct. Dashed lines are the traces of the negative control. n = 3 wells from one replicate for all conditions. Errors, s.e.m. Experiment was first performed by collaboration with Gwendolyn Shingles.

3.4.3 Shield-1 induced MC4R signaling by CapN2.1 and CapC1.1 double caged α-MSH

α-MSH is a neuropeptide of the melanocortin family receptors. One of its receptor, MC4R, is known to be associated with food intake and energy homeostasis[387]. Tools to dissect the role of α-MSH in central nervous system could further advance the understanding of these biological

processes. To further show CAPs' general ability to control peptide functions, I discuss the utility of CAPs to control α-MSH's activity towards MC4R in this section (**Figure 3-26**).

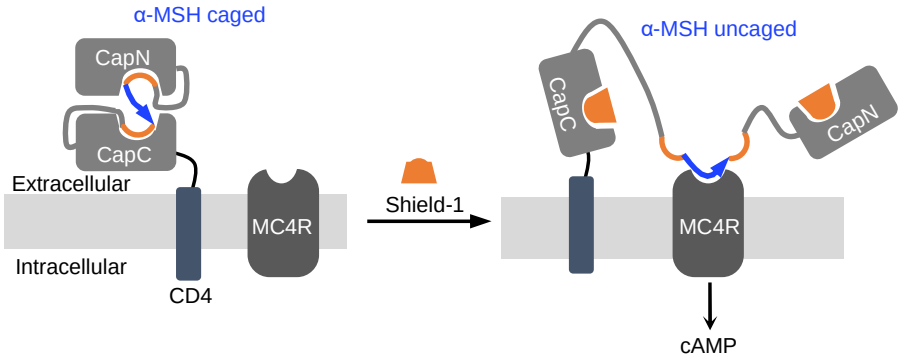


Figure 3-26 Scheme of CAPs-controlled α-MSH. Addition of shield-1 uncages α-MSH and activates the MC4R, leading to increase of cAMP production. cAMP is detected by GloSensor. CD4 is a transmembrane domain.

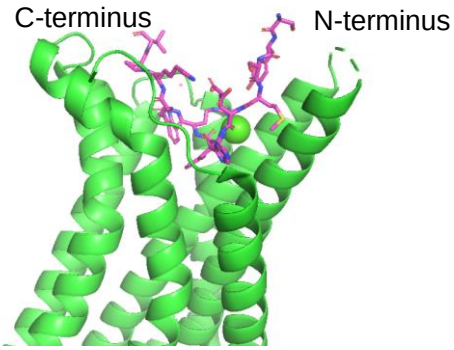


Figure 3-27 Crystal structure of MC4R/α-MSH complex. Green: MC4R. Magenta: α-MSH. PDB: 7F53.

Both the N- and C-terminus of α-MSH are exposed to extracellular environment when binding to MC4R (**Figure 3-27**). Therefore, we employed both CapN2.1 and CapC1.1 to modulate α-MSH accessibility (**Figure 3-26**) and displayed the protein on cell surface by fusing to CD4. A cleavable signal sequence (KTIIALSYIFCLVFA)[369] is fused to the N-terminus of the protein to direct membrane trafficking. Addition of shield-1 uncages α-MSH and activates MC4R, leading to increase of cAMP level.

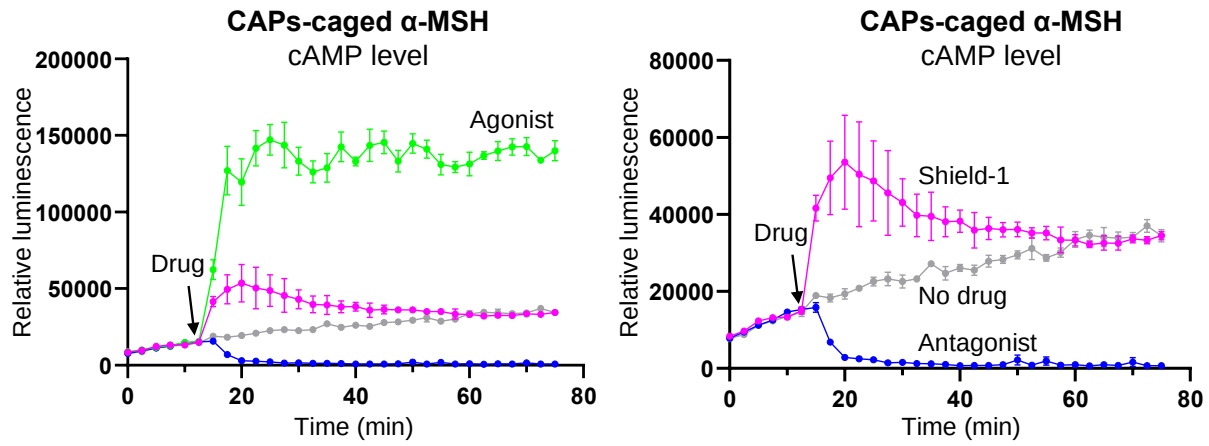


Figure 3-28 The cAMP assay for CAPs-caged α-MSH. GloSensor and MC4R were co-transfected. Cells were stimulated with 10 μM agonist (Bio-Ahx-Melanotan I), 10 μM shield-1, 100 μM antagonist (SNT-207707) or media at 15 min. Left: Full graph. Right: Zoomed in graph with no drug, shield-1, and antagonist conditions. n = 3 wells from one replicate for all conditions. Errors, s.e.m. Experiment was conducted by Ryan Singer.

As shown in **Figure 3-28**, addition of shield-1 and the agonist Bio-Ahx-Melanotan I both increased the cAMP level (2.8-fold and 6.2-fold, respectively), suggesting that CAPs can modulate α-MSH function with a shield-1 dependent manner. However, addition of the antagonist SNT-207707 reduces cAMP level by 85%, suggesting that α-MSH was not fully caged by CAPs.

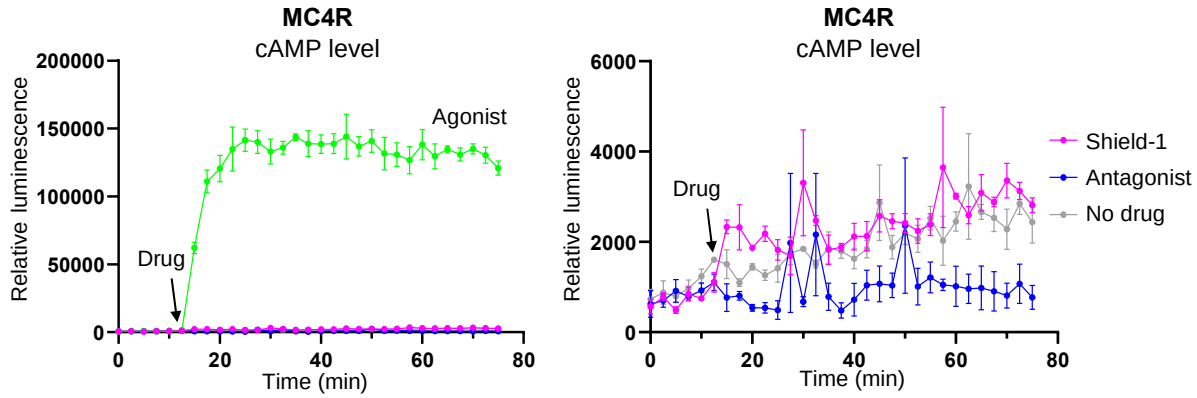


Figure 3-29 The cAMP assay for the control study. Only GloSensor and MC4R were cotransfected. Cells were stimulated with 10 μM agonist (Bio-Ahx-Melanotan I), 10 μM shield-1, 100 μM antagonist (SNT-207707) or media at 15 min. Left: Full graph. Right: Zoomed in graph with no drug, shield-1, and antagonist conditions. n = 3 wells from one replicate for all conditions. Errors, s.e.m. Experiment was conducted by Ryan Singer.

To further confirm that shield-1 functions by uncaging α-MSH, a control study was conducted where no CAPs-caged α-MSH was expressed in cells. The agonist Bio-Ahx-Melanotan

I drastically increased the cAMP level, suggesting that MC4R is functional. No obvious cAMP level was detected under the shield-1 and antagonist conditions. Therefore, shield-1 could not alter cAMP level by itself alone, and MC4R has minimal basal activity.

To investigate the biology of α-MSH, previous studies typically aimed to modulate neurons that express pro-opiomelanocortin (POMC), from which α-MSH is generated. DREADDs and rhodopsins have been applied to modulate the POMC neurons. The CAPs controlled α-MSH could potentially offer a new way to study this biology by directly modulating the endogenous receptors of α-MSH.

The experiment methods used to produce data in Chapter 3 are provided in **Appendix Methods B-2**.

I acknowledge the following individuals for their contribution on this project. The characterization data of DuoSelect and CapC1.1 was reproduced by Luis Vazquez-Rivera, a Ph.D. student from the Wenjing Wang lab, University of Michigan. Directed evolution of CapN was performed by Luis Vazquez-Rivera. CapC1.1-controlled enkephalin was from the collaborative work with Guanwei Zhou, a Ph.D. student from the Wenjing Wang lab, University of Michigan. CapC1.1-controlled PACAP was from the collaborative work with my undergraduate trainee Gwendolyn Shingles from Department of Chemistry, University of Michigan. CAPs-controlled α-MSH was performed by Ryan Singer, a Ph.D. student from the Wenjing Wang lab, University of Michigan.

Chapter 4 Development of a Light-Activated Protein Switch for Modulating Peptide Functions

Chapter 4 is largely adapted from: Geng, L., Shen, J. and Wang, W. Circularly permuted AsLOV2 as an optogenetic module for engineering photoswitchable peptides. *Chemical Communications* (2021), 57(65), 8051-8054.

Optogenetic tools have been transformative by enabling manipulation of specific cellular processes using light[388-390]. Their genetic encodability and light-dependence allow fast and reversible control of cellular events in specific cell types. Light-sensing proteins are crucial building blocks for engineering optogenetic tools. Among them, the AsLOV2 domain has been most well-studied[196,391-393] and most widely applied to modulate the activity of various proteins and peptides[277,311,312,322,324,325,339,344,394-396].

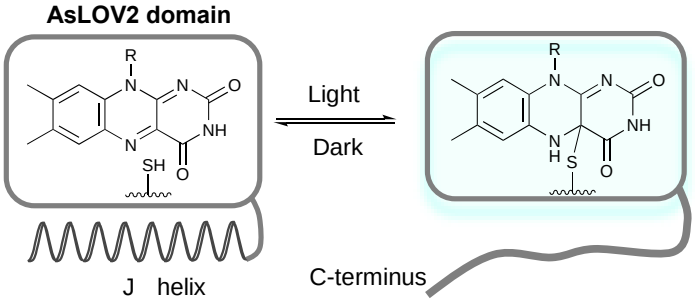


Figure 4-1 Reversible light-dependent conversion of the AsLOV2 domain.

As described in Chapter 1.2, the broad applicability of AsLOV2 is due to its unique mechanism for light-dependence. In the dark state of AsLOV2, its C-terminal Jα-helix is packed

against the PAS domain through hydrophobic packing and hydrogen bonding interactions[195,196,391] (**Figure 4-1**). With light irradiation, a cysteine residue (C450) in the protein core forms a covalent bond with the cofactor flavin mononucleotide, and the conformational change propagates to the Jα-helix, causing it to unwind from the rest of the protein[195,196,391]. The typical approach to render a peptide or protein photoswitchable is therefore to fuse it to the C-terminus of the Jα-helix directly. This introduces steric hindrance ("blocking") in the dark state which is "unblocked" in the light state.

While this light-dependent conformational change of AsLOV2 has been useful for protein engineering, it limits the use of AsLOV2 to only cage the N-terminal portion of a peptide and precludes the possibilities to cage peptides where their function requires a free N-terminus (e.g., enkephalin and PACAP as described in Chapter 3.4). In addition, for some peptides, it will be more effective to cage the C-terminus that contains critical residues. Therefore, a similar light-sensing domain that can cage the C-terminal portion of a peptide complements AsLOV2 and expands the designs of optogenetic tools.

To bridge this gap, Lequn and I engineered a circularly permuted AsLOV2 domain, cpLOV. cpLOV retains light-sensing capability while allowing modulation of peptides' activities by controlling their C-terminal portion. A conceptually similar design[397] has been reported to cage the C-terminal portion of SsrA peptide for inducing dimerization, and also embedded NLS and NES in the Jα helix to control protein transportation. An earlier work[398] fused a 4E-BP[399] peptide between the Jα helix and PAS core, which enables photoswitchable inhibition of human eukaryotic initiation factor 4E (eIF4E). Our design focuses on caging of short peptides (< 10 amino acids) with a free N-terminus. In addition, similar to the strategy described in Chapter 2, where CapN and CapC could be tandemly used, simultaneous caging of both the N- and C-termini of a peptide

using cpLOV and AsLOV2 together provides enhanced caging of the peptide target and alters the dynamic range of an existing optogenetic tool. Therefore, cpLOV represents a new light-sensing module useful for engineering optogenetic tools.

4.1 Rational design of cpLOV

To engineer cpLOV, we started from an AsLOV2 variant, hLOV1[337], which contains 15 mutations from the wild-type AsLOV2 and has been shown to have superior caging in the dark state. We first connected the hLOV1 termini with a flexible linker. Based on the crystal structure[391] (**Figure 4-2**), we reasoned that a four-amino-acid (GSGS) linker is sufficient to connect the original N- and C-termini of the LOV domain. We then introduced a new opening at the original "hinge region" connecting Jα-helix to the protein core. We split between amino acids L520 and H521, as H521 is the first helical residue on the Jα-helix.

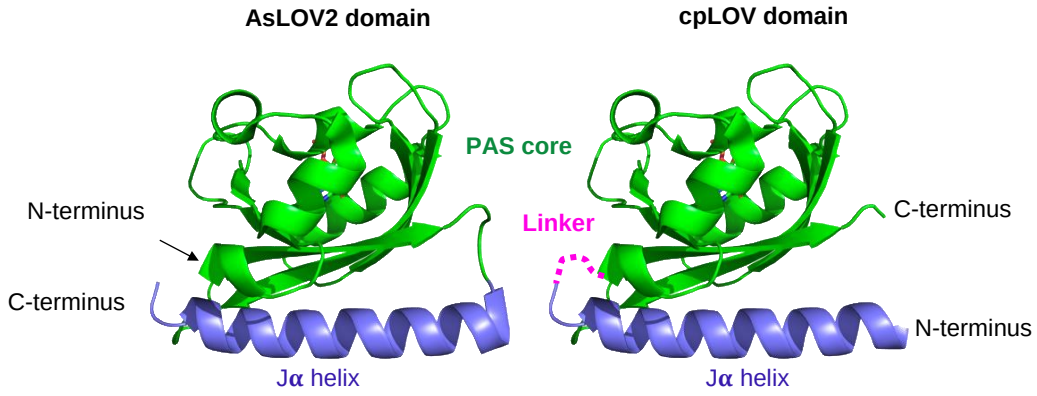


Figure 4-2 Design of the circularly permuted AsLOV2 (cpLOV) based on the AsLOV2 domain (PDB: 2V1A).

To check whether the light-induced conformational change of Jα-helix could still take place in cpLOV, we used cpLOV to cage a heptapeptide, SsrA, and used a yeast surface-based binding assay similar to the one used for CAPs described in Chapter 2.2.2 (**Figure 4-3**) to evaluate its

accessibility in the dark and light states. Yeast surface display was used because this allows future directed evolution to improve cpLOV.

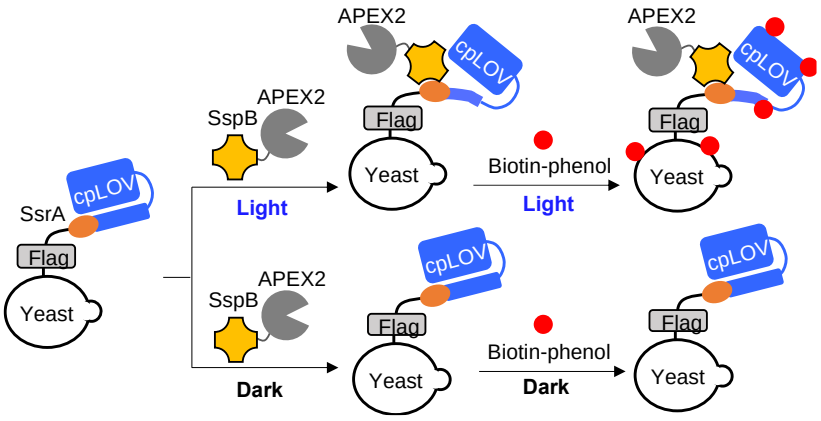


Figure 4-3 Yeast surface display assay for testing cpLOV's caging of SsrA peptide. cpLOV caged SsrA is expressed on the yeast surface. Under light irradiation, SsrA is uncaged and recruits SspB-APEX2 fusion protein. The APEX2 can covalently label proteins in proximity with biotin-phenol molecule. Flag tag indicates protein expression and was measured by anti-Flag antibody labeling. Biotin indicates SsrA-SspB interaction and was measured by streptavidin labeling.

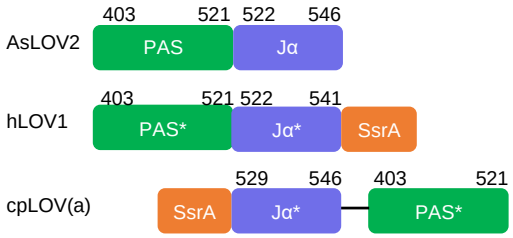


Figure 4-4 The architecture of AsLOV2, hLOV, and cpLOV(a) domain. In cpLOV(a), the original N- and C- termini are linked through a four-amino-acid GSGS linker.

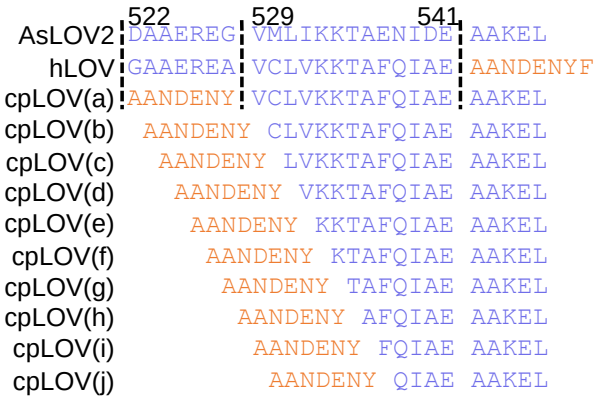


Figure 4-5 The Jα helix sequences and truncation sites for different cpLOVs. SsrA peptide sequence, AANDENY. The SsrA sequence for hLOV is adapted from the previously reported iLID[325].

We screened 10 fusion sites along the Jα-helix (**Figure 4-4**, **Figure 4-5**), from V529 to Q538, as it has been shown that the fusion sites on the Jα-helix affect the light-dependence of each peptide[322,339,400]. Since the absolute difference of the biotin signal is low (**Appendix Figure C-1**), to better quantify the dynamic range, the binding was measured by the biotin/flag signal ratio of each individual cell, so that the sample-to-sample differences (e.g. technical errors, yeast growth difference and difference in antibody labeling) are minimized. The mean of the ratios was used to represent the SspB binding of each clone. The significance test was performed by using R coding (see **Appendix Method C-1**).

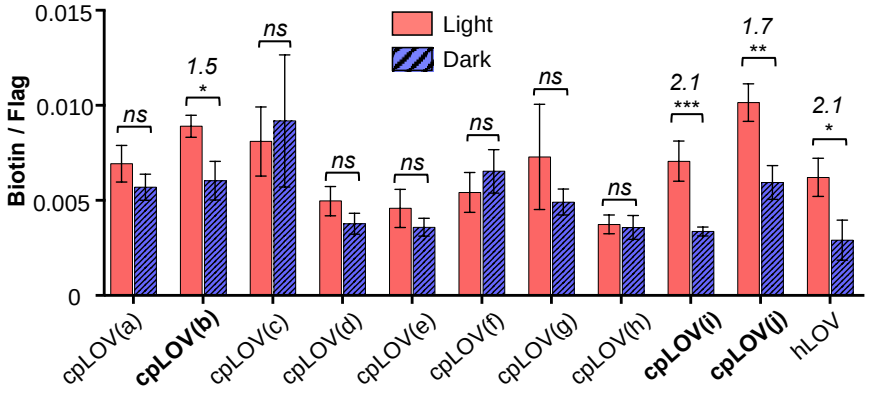


Figure 4-6 Bar plot of biotin/Flag signal ratio for different constructs under light and dark conditions. Biotin and Flag signals were measured by flow cytometry. Biotin/Flag signal ratio of cells expressing Flag tag were calculated. The bars in the plot indicate the mean of the ratio and the error bars indicate the standard error of the mean. The values above bars indicate the ratio of the mean biotin/Flag signal ratio between light and dark conditions (only for the ones with significant difference). *P* values were determined by unpaired two-tailed *t*-test. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *ns*, not significant. Data was acquired by collaboration with Dr. Lequn Geng.

Out of the ten constructs tested, three (cpLOV(b), cpLOV(i), cpLOV(j)) showed significant light-induced signal increase (**Figure 4-6**). Among them, cpLOV(i) showed comparable light-to-dark ratio as hLOV1. This shows that cpLOV retains light-sensing capability and its dynamic range can be similar to hLOV1. Our results also highlight the importance of fusion site screening when using cpLOV. Similar to previous studies using LOV domain, identifying the optimal fusion

position is key to the successful engineering of the photoswitchable peptides, and the optimal fusion site for each peptide should be tested individually.

4.2 cpLOV as the replacement of existing AsLOV2 for caging TEVcs

We then investigated the performance of cpLOV comparing to the well-established domain hLOV1 in a sensitive transcriptional assay SPARK[337], where TEVcs is caged.

Before moving to SPARK assay, we first tested the fusion site using the yeast surface display platform (**Figure 4-7**). We tested the light-dependent protease cleavage of the TEVcs-cpLOV fusion construct (**Figure 4-8**). Similar to the analysis in **Figure 4-6**, the HA/Flag signal ratio is used to minimize the sample differences (**Figure 4-9**). The cleavage is represented by the HA/Flag ratio difference between the "+ protease" and "− protease" conditions (**Figure 4-10**). The construct showed a light-dependent protease cleavage efficiency in between that of the AsLOV2 and hLOV1 domains.

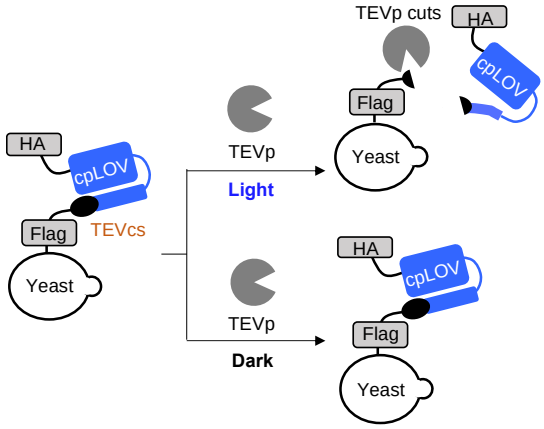


Figure 4-7 Yeast surface display assay for testing the cpLOV caged-TEVcs. Under light irradiation, TEVcs is uncaged and cleaved by TEV protease (TEVp), causing a reduction of HA signal. Flag tag indicating protein expression and HA indicating protease cleavage are measured by Flag and HA antibody labeling.

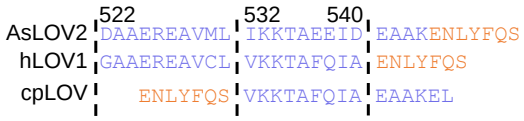


Figure 4-8 The Jα helix sequences and truncation sites for AsLOV2, hLOV, and cpLOV. TEVcs sequence, ENLYFQS.

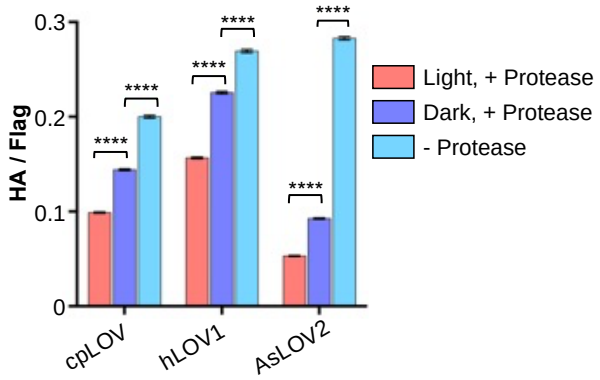


Figure 4-9 Bar plot of HA/Flag signal ratio for different constructs. HA and Flag signals were measured by flow cytometry. HA/Flag signal ratio of cells expressing Flag tag were calculated. The bars in the plot indicate the mean of the ratio and the error bars indicate the standard error of the mean. *P* values are determined by unpaired two-tailed t-test. ****$P < 0.0001$. Data was acquired by collaboration with Dr. Lequn Geng.

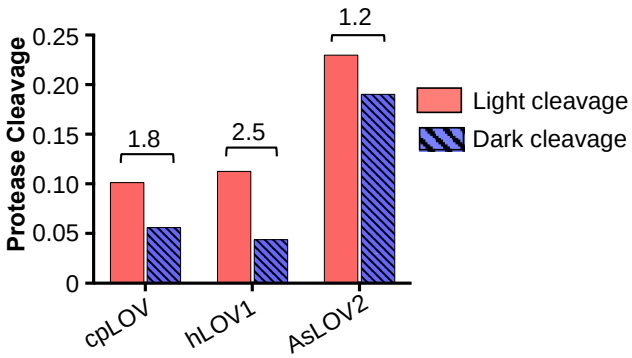


Figure 4-10 Bar plot of protease cleavage level of different constructs under light and dark conditions. Light or dark cleavage efficiency is defined as the difference of the HA/Flag ratio value between + protease, light or dark condition and – protease condition. The values above bars indicate the ratio of light cleavage to dark cleavage.

We then applied this cpLOV for caging TEVcs in a sensitive DRD1-SPARK assay[337] to evaluate the caging efficiency of cpLOV (**Figure 4-11**). Only under light and dopamine condition, can the transcription factor be cleaved from membrane and translocate to nucleus to induce reporter gene expression. Using mCherry as the reporter, the cells treated with four different conditions were imaged and quantified by mCherry sum intensity (**Figure 4-12**). cpLOV provided a good

light dependence of 8.4-fold, and the cpLOV based SPARK showed as high as 15.4-fold signal to noise ratio (**Figure 4-12**).

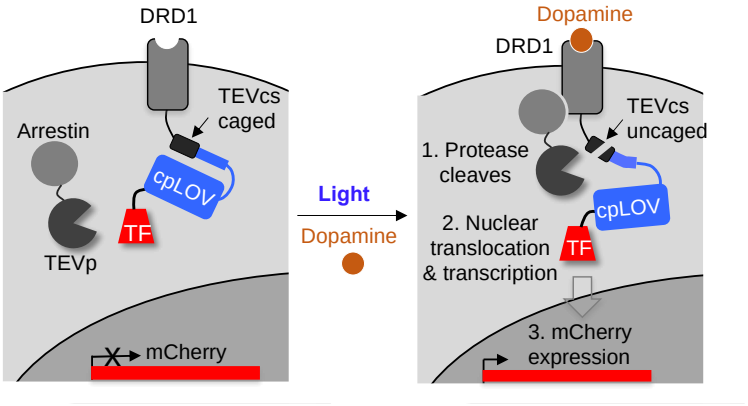


Figure 4-11 Scheme of the DRD1-SPARK assay in HEK293T cells. Under dark and no dopamine condition, the TEVcs is caged and not cleaved. Hence, no mCherry is expressed. Under light and dopamine stimulation, the TEVcs is uncaged and cleaved by the protease brought into its proximity via the DRD1-arrestion interaction. The transcription factor is released from the membrane to initiate mCherry expression. TEVcs, ENLYFQM.

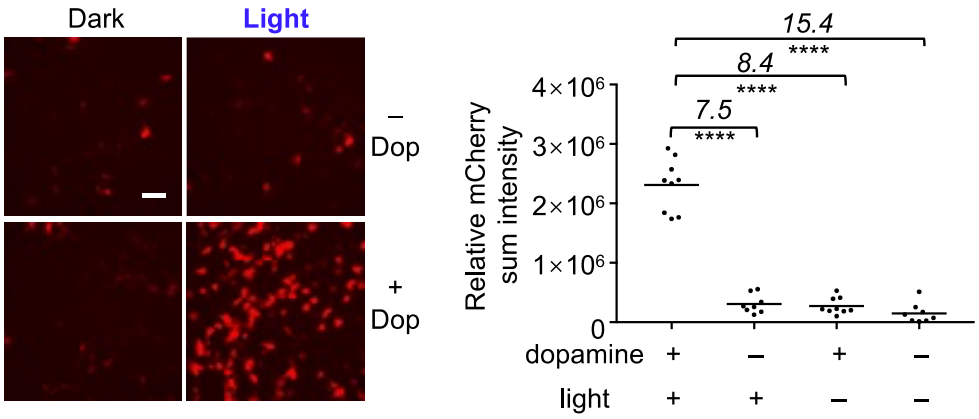


Figure 4-12 Left: Confocal fluorescence images of the transcriptional assay in **Figure 4-11**. The cells were stimulated with light and/or 100 μM dopamine for 10 minutes. Scale bar, 50 μm. Right: Dot plot of relative mCherry sum intensity in each image. Eight to ten images were analyzed for each condition. The values above the dots indicate the ratio of total intensity between two conditions. *P* values were determined by Wilcoxon-Mann-Whitney test. ****$P < 0.0001$. Data was acquired by collaboration with Dr. Lequn Geng.

4.3 Tandem caging strategy with cpLOV for background reduction and dynamic range tuning

We next investigated the effects of tandem caging by both cpLOV and AsLOV2 for improving the peptide caging efficiency in the dark. This is because one of the long-standing challenges associated with using light-sensing proteins is the insufficient caging and the consequent high background activity in the dark state. Although extensive engineering efforts have improved the dark-state caging of AsLOV2 via rational design[392] and directed evolution[325,334], an easy and readily applicable approach to improve the peptide caging efficiency in the dark will be highly advantageous. Tighter caging in the dark is especially important for experiments that are highly sensitive or require expression of the protein for an extended period of time.

We investigated the strategy of tandem caging by AsLOV2 variants and cpLOV comparing to the conventional single caging by the efficient hLOV1 (**Figure 4-13**). The SPARK assay shown in **Figure 4-11** was used to measure caging and uncaging.

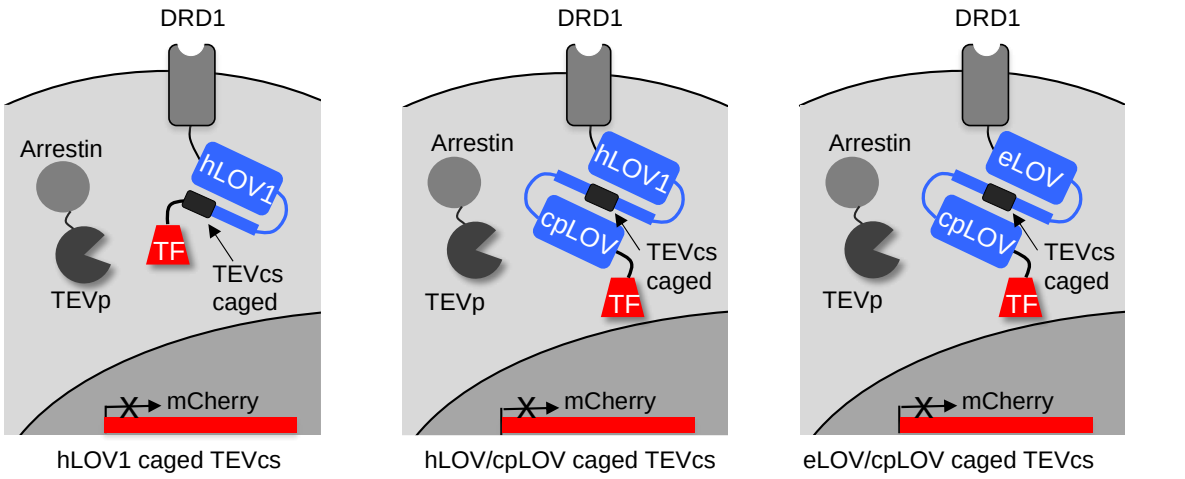


Figure 4-13 Scheme of single and tandemly caged TEVcs in SPARK assay. TEVcs, ENLYFQM.

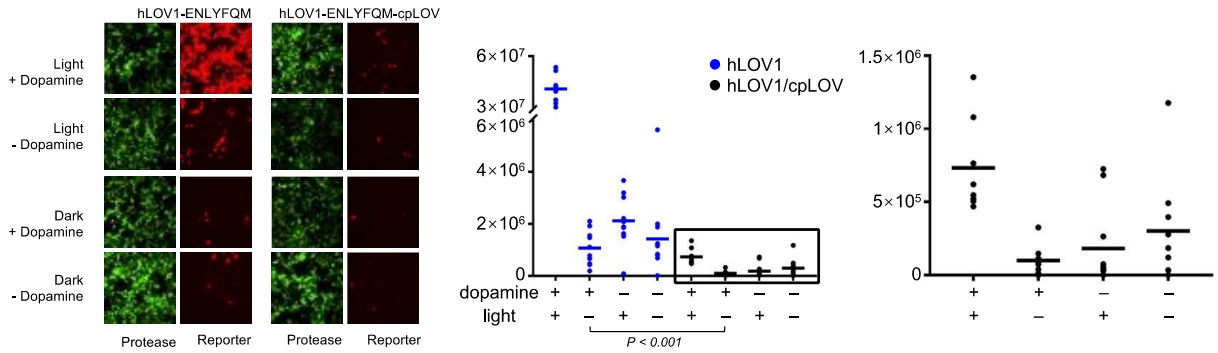


Figure 4-14 Left: Confocal fluorescence images of the dual-caged-TEVcs transcriptional assay in HEK293T cells. The cells were stimulated with light and/or 100 μM dopamine 30 minutes. Scale bar, 50 μm. Right: Quantification of the mCherry reporter gene expression. *P* value is determined by unpaired two-tailed *t*-test. Data was acquired by collaboration with Dr. Lequn Geng.

We first tested dual caging with hLOV1 and cpLOV (**Figure 4-13**, middle), because hLOV1 provides the best caging efficiency among all AsLOV2 variants. During a 72-hour lentiviral transduction period, tandem caging reduced the background by over 90% and effectively shifted the dynamic range to the lower end (**Figure 4-14**). However, this design resulted in significantly reduced activation signal (**Figure 4-14**), presumably because the TEVcs is deeply embedded in hLOV1 and might have lost accessibility to TEV protease in the dual-caging configuration.

Therefore, we replaced hLOV1 with another AsLOV2 variant, eLOV[334], which has a lower caging efficiency. This eLOV/cpLOV construct (**Figure 4-13**, right) reduced the assay background by up to 15-fold (**Figure 4-15**) while still achieving a significant amount of activation. Although the activation was reduced by 7-fold, we still obtained a high light-dependent dynamic range of 42-fold. This demonstrates that dual-caging can effectively shift the light-dependent dynamic range to a lower end. In principle, this strategy can be generally applied to photoswitchable peptides, and we expect optogenetic tools that suffer from high background to particularly benefit from such design.

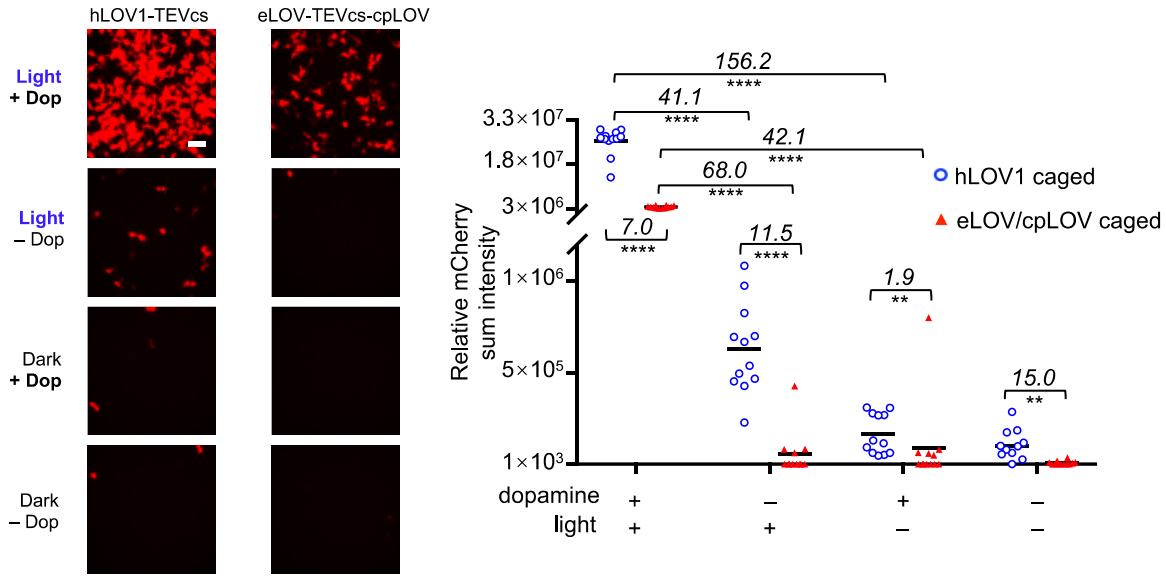


Figure 4-15 Left: Confocal fluorescence images of the single and tandemly caged SPARK assay. The cells were stimulated with light and/or 100 μM dopamine for 10 minutes. Scale bar, 50 μm. Right: Dot plot of relative mCherry sum intensity in each image. Twelve images were analyzed for each condition. The values indicate the ratio of total intensity between two conditions. *P* values were determined by Wilcoxon-Mann-Whitney test. $^{**}P < 0.01$; $^{****}P < 0.0001$. Data was acquired by collaboration with Dr. Lequn Geng.

The experiment methods used to produce data in Chapter 4 are provided in **Appendix Methods C-2**.

The experiments were conducted by the collaboration work with Dr. Lequn Geng, a former Ph.D. student from Wenjing Wang lab, University of Michigan.

Chapter 5 Summary of Results and Future Directions

This thesis demonstrated the development of two genetic tools: a pair of chemogenetic protein domains CAPs and an optogenetic domain cpLOV for modulating the accessibility of peptides. For future optimization of these protein switches, an efficient yeast surface based directed evolution platform was established. In Chapter 5, I summarize the results form Chapters 2-4 and discuss the possible future directions of this research.

5.1 Summary and discussion of research results

In Chapter 2, we engineered a pair of small-molecule-dependent protein domains, CAPs, to control peptide activity. CAPs were shown to have general applicability in caging short peptides, including TEVcs, SsrA, NLS, and enkephalin. We demonstrated the versatility of CAPs by using them to translocate proteins to various cellular locations, induce μOR signaling, and to control gene transcription in a chemical-dependent manner. CapN and CapC provide alternative caging strategies for peptides with geometry restrictions, as shown by the CapC-caged [Met^5]-enkephalin. We also introduced the tandem caging strategy by CAPs to reduce basal activity, where a peptide is fused to both CapN and CapC. Comparing to further optimizing protein switch performance, this strategy provides an easier and modular approach for engineering switchable peptides. We further showed that CAPs are useful in a variety of biological contexts, including HEK293T and neuronal cell culture, as well as in living animals.

In Chapter 3, we established the DuoSelect directed evolution platform employing yeast surface display for efficient selection of mutants with an enhanced dynamic range. Through two rounds of binding assays under stimulation and non-stimulation conditions, DuoSelect allowed the simultaneous display of both leakage and activation signals on a single yeast cell. This approach accelerates the enrichment of desired clones compared to conventional methods, and eliminates the need for alternating positive and negative selections. DuoSelect facilitated the optimization of CapC and CapN, resulting in CapC1.1 and CapN2.1, respectively, both exhibiting lower background activity. The improved CAPs were applied to modulate three neuropeptides, with CapC1.1 demonstrating enhanced caging efficiency for enkephalin and minimal background activity for caged PACAP. Tandem use of CapN2.1 and CapC1.1 illustrated the broader applicability of this approach, particularly in modulating the activity of α-MSH towards MC4R.

In Chapter 4, we re-engineered the AsLOV2 domain through circular permutation, yielding cpLOV. This novel domain allowed for the modulation of peptides requiring a free N-terminus for function. cpLOV was applied to cage TEVcs in mammalian cell culture. When used in tandem with the AsLOV2 domain, background accumulation could be minimized, and the dynamic range was tuned to the lower end.

5.2 Novelty and impact of the developed domains and platform

CAPs feature the first chemogenetic domain that can be generally applied to control peptide functions, as showcased by TEVcs, SsrA, NLS, enkephalin, PACAP, and α-MSH. Compared to the existing methods using the optogenetic AsLOV2 domain, CAPs provide an alternative approach by using a small molecule, shield-1 or aquashield-1, which is easier to use and deliver compared to light. Importantly, shield-1 has good cell permeability and can be administered

systematically in living animals. This enables the use of CAPs in most parts of a living organism, including those that are difficult or otherwise too invasive to reach by light, such as the brain and the liver.

While protein dimerization and transportation can be achieved by many other chemogenetic and optogenetic tools, CAPs-caged neuropeptides feature the first example of direct modulation over receptor ligands. Compared to the commonly used DREADD and rhodopsin approach, this design enables the probing of endogenous receptors. In addition, the signaling events are triggered by endogenous ligands, so this method represents a more biologically relevant approach to study the biological functions of both neuropeptides and receptors. CapC1.1-caged PACAP also demonstrated the potential of CAPs to control longer peptides, whose terminal residues are crucial for their functions.

As other chemogenetic domains, activation of CAPs relies on the binding of the ligand shield-1 to FKBP, which is an irreversible process. Optogenetic domains features the reversibility as an advantage. To enable similar control over peptides with optogenetic control, the cpLOV domain was developed to control peptide activities by fusion to their C-terminus. This complemented the AsLOV2 domain by enabling the caging of C-terminal portions.

These novel protein switches should provide protein engineers with new tool building blocks and strategies for genetic tool design. The DuoSelect platform could facilitate this by offering faster and more efficient directed evolution method to optimize these protein switches. The platform could efficiently improve protein switches with lower leakage. We envision that DuoSelect can be applied to optimize other types of proteins as well, such as induced or disrupted proximity pairs and switchable proximity labeling enzymes[303,401] (**Figure 5-1**).

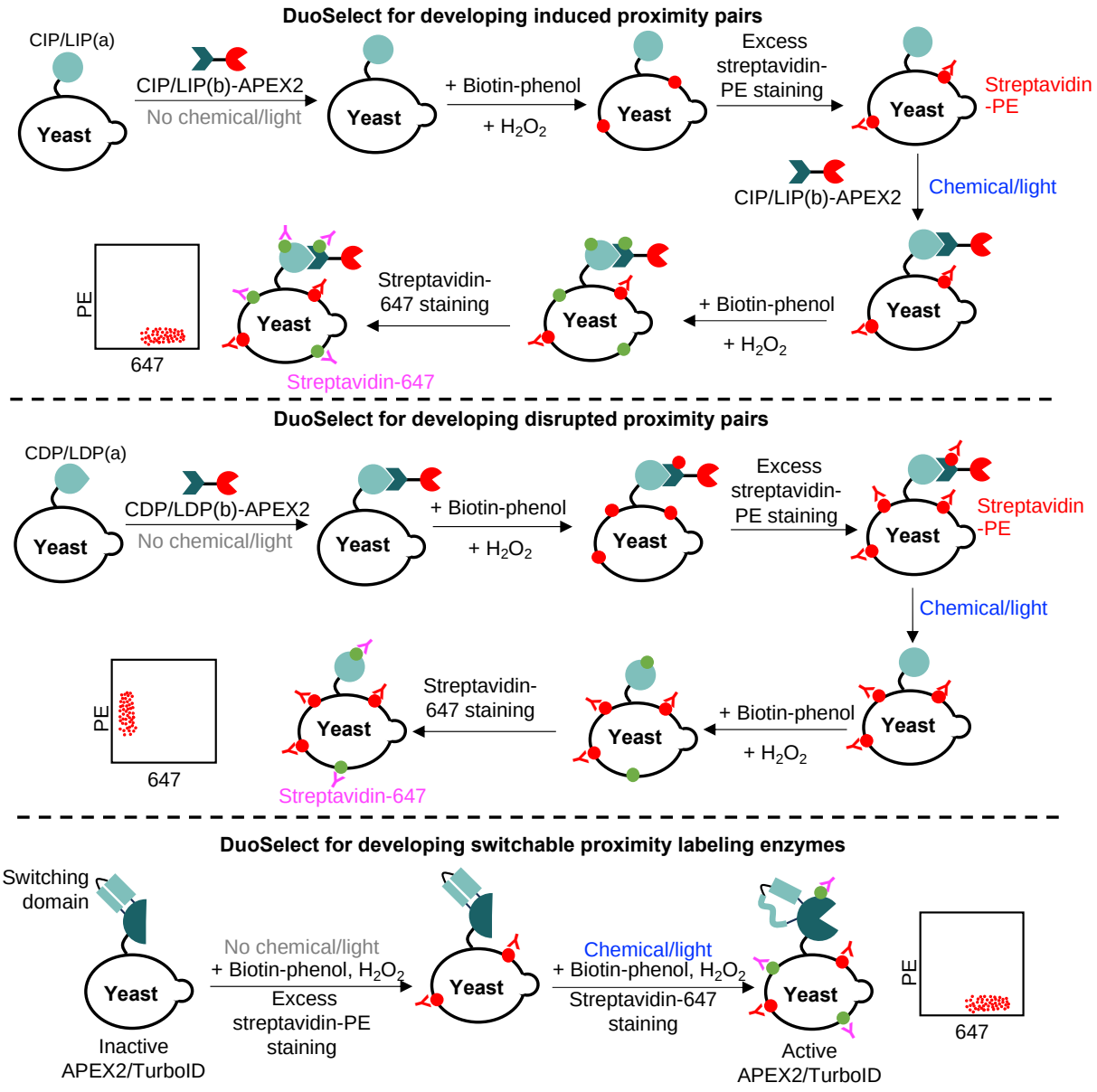


Figure 5-1 Scheme of DuoSelect for developing induced proximity pairs, disrupted proximity pairs, and switchable proximity labeling enzymes (APEX2 and TurboID).

5.3 Future directions

Future directions of this research include optimization of the current protein switches, expansion of current chemogenetic toolbox, and application in animal models. The caging efficiency of CAPs need to be optimized and the dynamic range of cpLOV need to be improved.

CAPs-controlled switchable neuropeptides remain to be tested in animal models. Furthermore, with the CAPs design, more chemogenetic switches could be developed to achieve multiplexed control.

5.3.1 Optimization of CAPs caging efficiency

Although they could be controlled in a shield-1 dependent manner, CapC-caged enkephalin and CAPs-caged α-MSH have significant basal activity. To further improve the caging efficiency of CAPs, using CapC1.1 as an example, several potential approaches could be taken in steps.

First, the linker connecting the binding sequence with FKBP could be optimized to appropriate length. The length of the linker in CapC1.1 (shown in magenta) seems to be excessive according to the AlphaFold predicted model (**Figure 5-2**). The extra length of the linker would bring extra entropy penalty and solvation enthalpy in the caging state. Therefore, a thorough investigation on the linker length could potentially help improve caging efficiency.

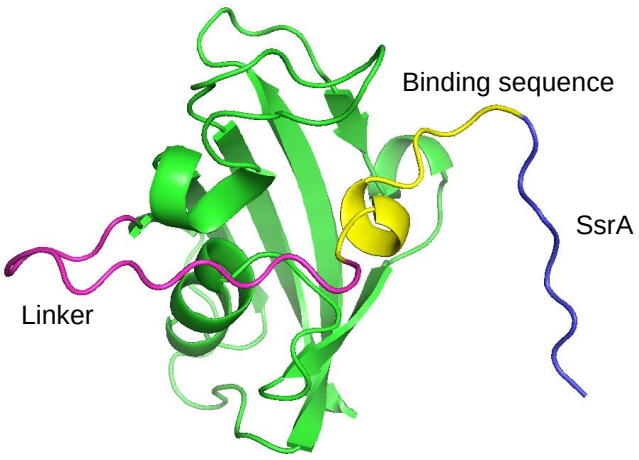


Figure 5-2 AlphaFold predicted model of CapC1.1. Blue: SsrA (AANDENYF). Yellow: binding sequence (GTPNLFGYV). Magenta: linker (SGGSGTGSGSGGS).

Next, optimization of the linker sequence could improve the affinity and therefore improve caging efficiency. Since the flexible Gly-Ser linker would not contribute enough for binding, a followed site-saturated mutagenesis could be performed on the current linker region. The caging efficiency can be then improved by directed evolution.

Finally, to further increase the binding affinity between the binding sequence and FKBP, a whole protein random mutagenesis could be performed. In this case, the residues out of the binding pocket could be utilized to enhance this binding.

Figure 5-3 Residues colored in blue are the SsrA sequence. Residues colored in red are the mutagenesis sites. "X" indicates any of the twenty amino acids. Underlined residues are from enkephalin.

Beyond these classic optimization approaches, "customized" optimization for caging each functional peptide is potentially feasible. CapC was developed based on the SsrA binding assay. The directed evolution results showed a clear pattern preferring a "GTPNL" and "GSPNL" (**Table 3-1**, **Table 3-2**) sequence as the binding sequence. However, Gly-Thr and Gly-Ser sequence would not possibly contribute a lot to the hydrophobic binding to the FKBP ligand binding site. We hypothesized that the C-terminal hydrophobic portion of SsrA (AANDENYF) is contributing to the binding instead. To overcome such problem, the CAPs caging could be potentially customized for the functional peptide by introducing some of its sequence between the CAP binding sequence and SsrA (**Figure 5-3**). Using [Met^5]-enkephalin (YGGFM) as an example, the C-terminal Phe-Met residues could be inserted between SsrA and the binding sequence (**Figure 5-3**). In this case, if the directed evolution works for caging SsrA, then the evolved CapC can at least cage the Phe-Met residues well, providing a chance to cage enkephalin better.

The CapN2.1 sequence does not differ from CapN1.0 a lot. Therefore, the residues before the current mutation region can be optimized first, and then the whole protein random mutagenesis could be performed. DuoSelect could provide a good platform for all the directed evolution experiment mentioned above.

5.3.2 Expanding the scope of chemically switchable neuropeptides and application in animal models

The CapC-controlled enkephalin system suffers from high basal activity and low shield-1 dynamic range. In addition to improving CapC caging efficiency, the separate chain approach as CapC-controlled PACAP could be applied to reduce the basal activity. The leakage most likely results from the competing binding process between the binding sequence with CapC and the enkephalin with μOR. Separating the enkephalin and μOR into two protein chains will turn their interaction into intermolecular and also enable its ability to probe endogenous μOR.

The steric hinderance generated by CAPs is not sufficient to control long peptides (> 10 aa), however, the example of CapC-controlled PACAP suggests that the caging can be achieved by perturbing the neuropeptide ligand recognition. Many neuropeptides contain post-translational modifications in their endogenous forms, which seems incompatible for genetic encoding. For example, α-MSH contains the N-terminal acetylation and C-terminal amidation modifications. However, the example of CAPs caged α-MSH showed that such modifications are not always critical for their activity towards the receptor. Therefore, we envision that CAPs can be potentially applied to modulate many other longer neuropeptides as well[402], such as substance P[403], neurokinin A[404], orexin-B[405], corticotropin-releasing factor[406], galanin[407], and calcitonin gene-related peptide[408].

The application of CAPs-controlled neuropeptides remains to be tested in animal models. These switchable neuropeptides will allow cell-type-specific and neuronal circuit-specific endogenous receptor activation and will be suitable for studying casual effect on animal behaviors. CapC-controlled enkephalin will be with specific interest in studying pain perception, rewarding, and breathing.

5.3.3 Improvement of cpLOV dynamic range and application for photoswitchable neuropeptide

Compared to CAPs, cpLOV showed much smaller dynamic range and need to be improved for controlling neuropeptides. To improve cpLOV, directed evolution is the most promising approach. For directed evolution, a site-saturated mutagenesis on the hinge region and a whole protein random mutagenesis could be performed sequentially.

In the two previous reports on directed evolution of AsLOV2[325,334], some mutants in the hinge region improved the performance, which are hypothesized to contribute to the dark state stability. When designing cpLOV, the original N- and C-termini were connected by a flexible GSGS linker. Therefore, this linker would be an easy starting point for cpLOV improvement. Site-saturated mutagenesis on these four positions could potentially yield cpLOV variants with improved dynamic range. A followed whole protein random mutagenesis could be applied to further seek for mutations improving cpLOV. DuoSelect can serve as an efficient platform for these selections.

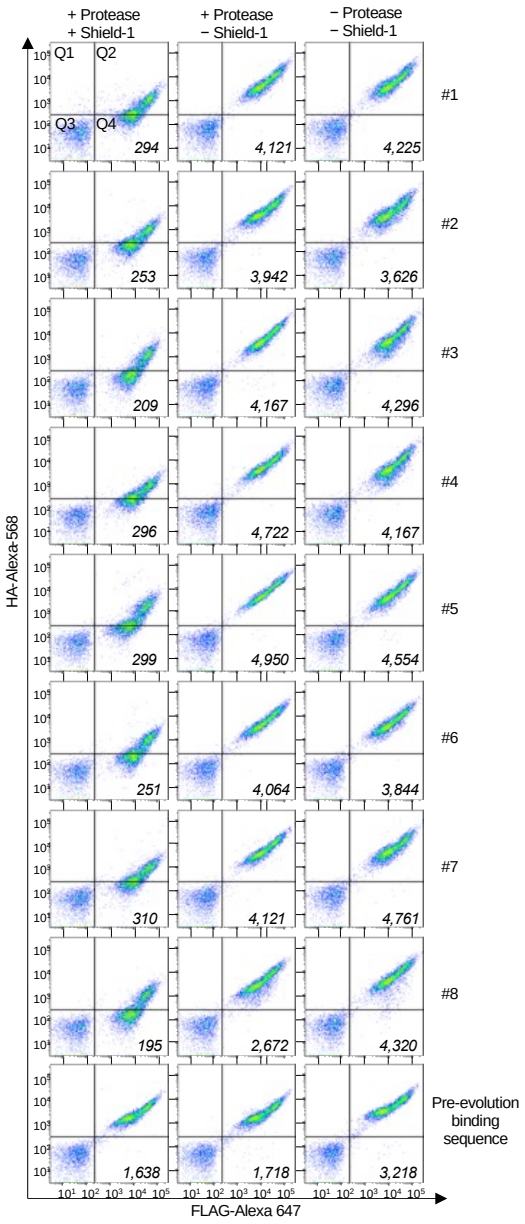
We envision that AsLOV2 and cpLOV could also be applied to control neuropeptides. There is only one report applying AsLOV2 on mammalian cell surface[296] and the use of these two domains for controlling neuropeptides remains to be tested.

5.3.4 Expanding the scope of chemogenetic switches for peptide modulation

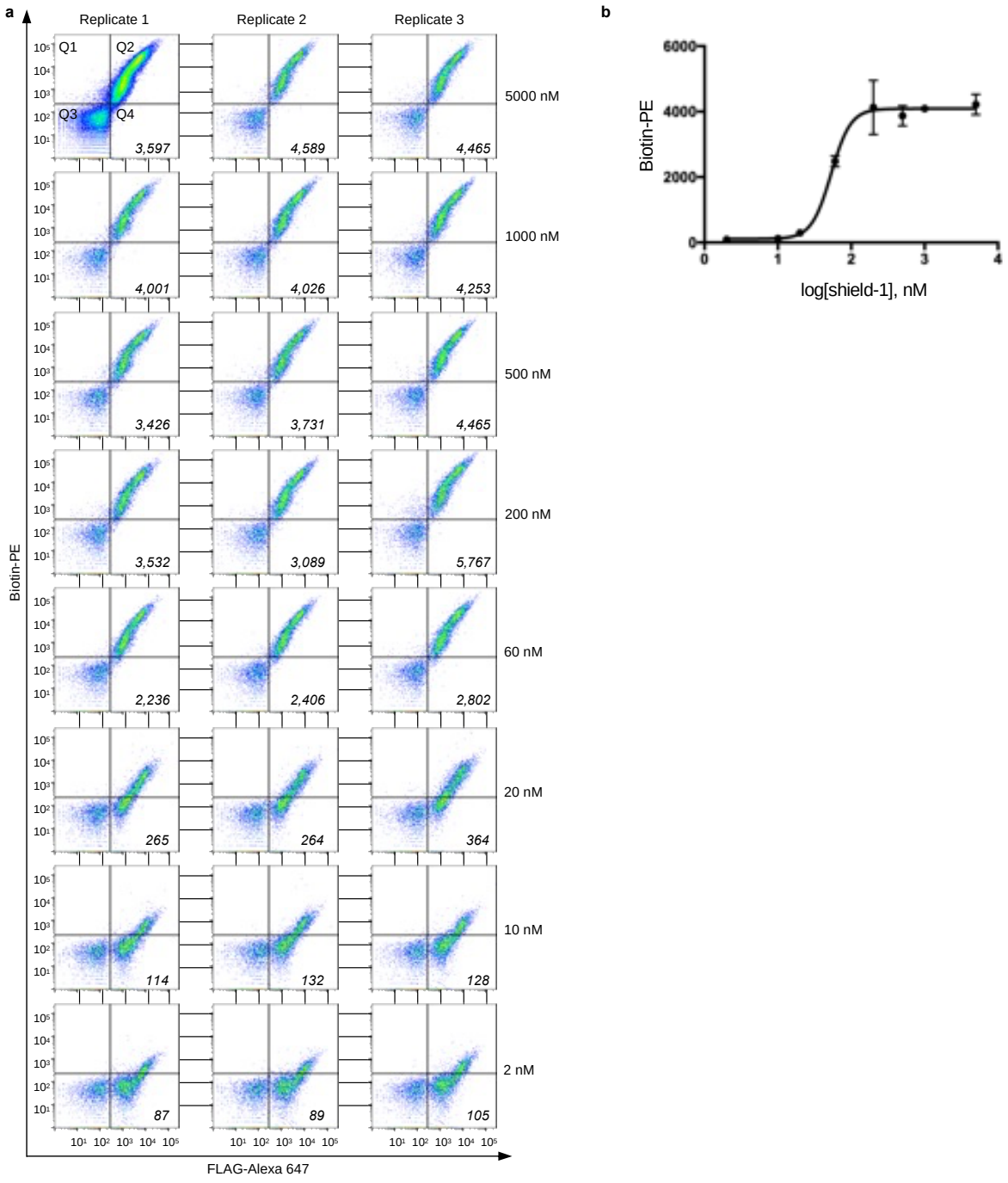
CAPs utilize the steric hinderance from the ligand binding pockets to modulate peptide functions. This mechanism should be not only generalizable for various functional peptides, but also adaptable for other ligand binding protein domains. We expect that the engineering strategy for CAPs is also applicable to other chemical-dependent protein domains such as the hepatitis C virus protease NS3a[409], BCL-xL protein[35], and streptavidin[33], which have been shown to have peptide binders. This approach might be also feasible for those chemogenetic protein domains such as DHFR, Pyl, and GID1, whose peptide binders could be potentially developed. Successful engineering of these domains as general protein cages will enable multiplexed control of cellular processes and expand the toolbox of available chemogenetic tools.

Appendices

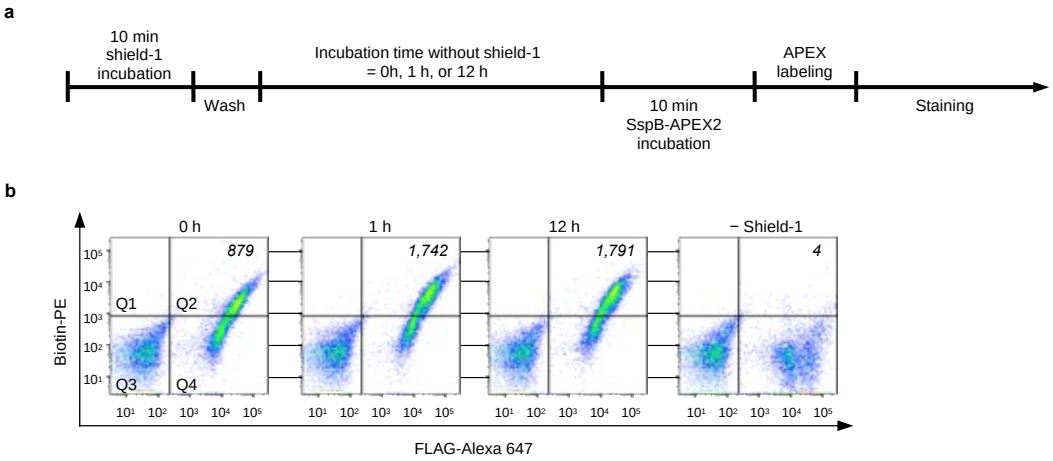
Appendix A: Appendix Figures and Methods Related to Chapter 2



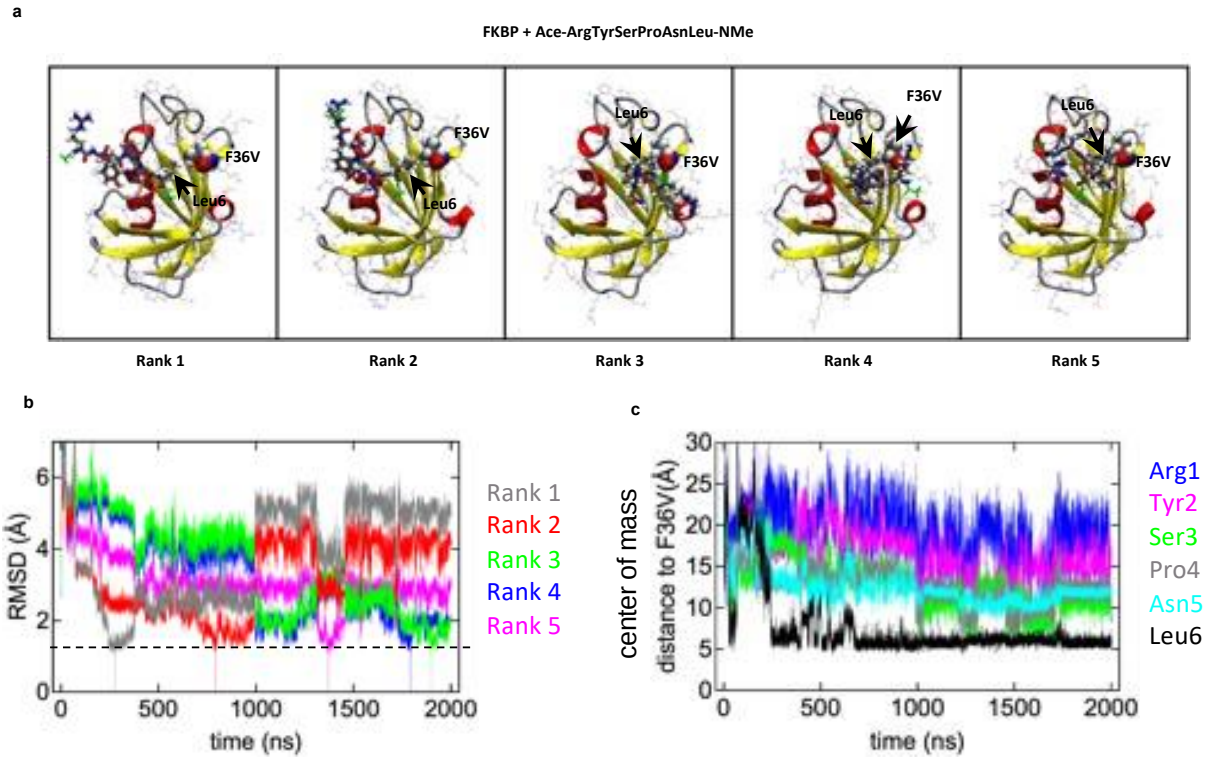
Appendix Figure A-1 FACS analysis of the most enriched eight clones, corresponding to clones #1-#8. Values are median HA intensity of FLAG-positive cells (Q2 + Q4). All eight clones showed similar results. This experiment was performed once.



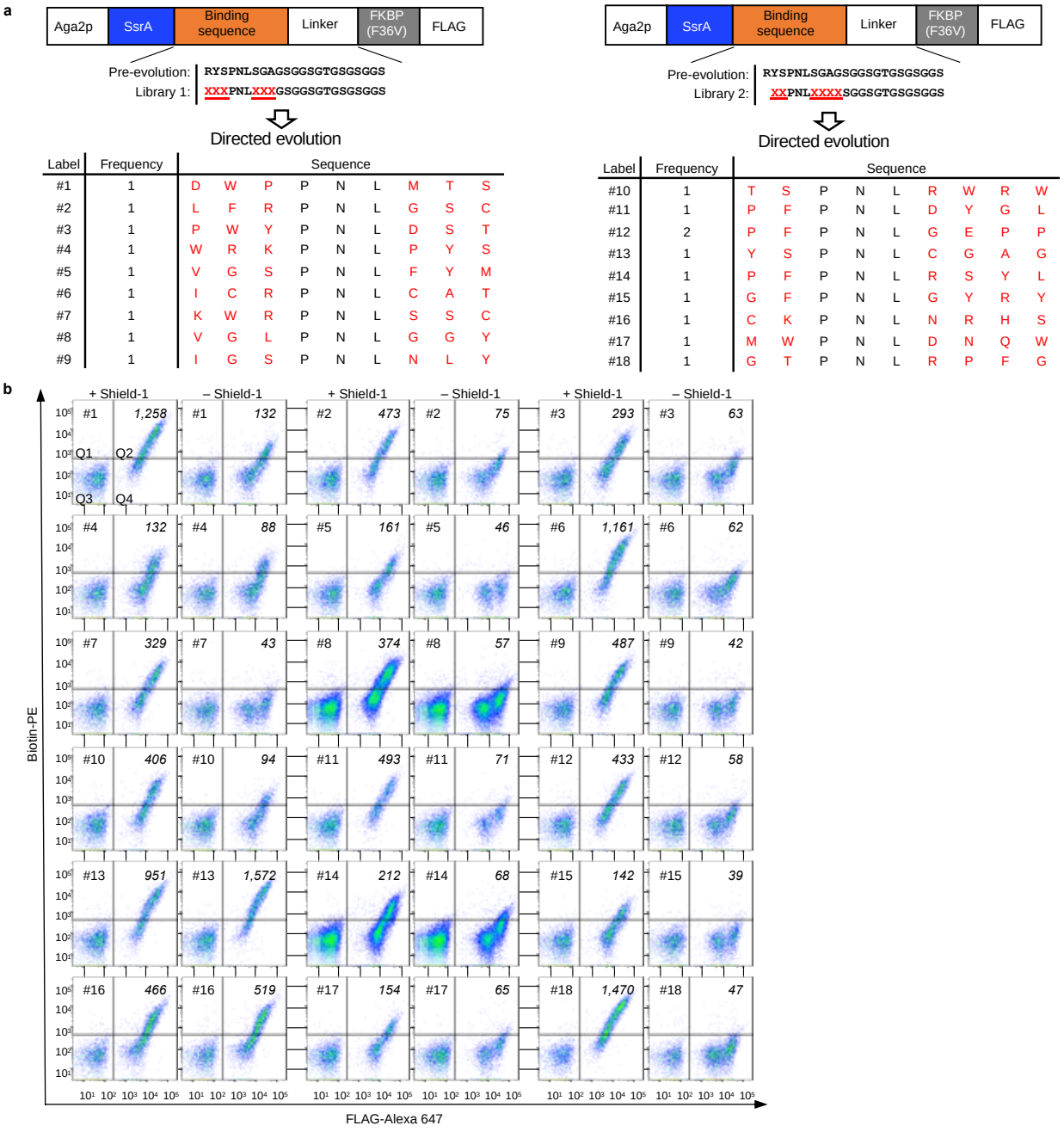
Appendix Figure A-2 Shield-1 dose response characterization with CapN. **a**, FACS analysis of CapN-caged SsrA on yeast surface treated with different concentrations of shield-1. Three technical replicates were performed for each condition. Values are median biotin intensity of FLAG-positive cells (Q2 + Q4). **b**, Dose-response curve using data from a. The median biotin signal is plotted against shield-1 concentration. Half maximum response was observed at 53 nM. 95% confidence interval = 38 nm ~ 67 nM. Errors, s.e.m.



Appendix Figure A-3 Shield-1 reversibility characterization with CapN. **a**, Timeline of shield-1 reversibility characterization. Yeast cells were incubated with shield-1 for 10 min, followed by washing to remove excess shield-1. Yeast cells were then incubated at room temperature for 0-12 h before the accessibility of SsrA was evaluated using SspB-APEX2 and biotin-phenol labeling as shown in **Figure 2-2**. **b**, FACS analysis of the yeast cells from a. Values are median biotin intensity of FLAG-positive cells (Q2 + Q4). No decrease in biotin signal was seen even after cells were incubated for 12 h without shield-1.



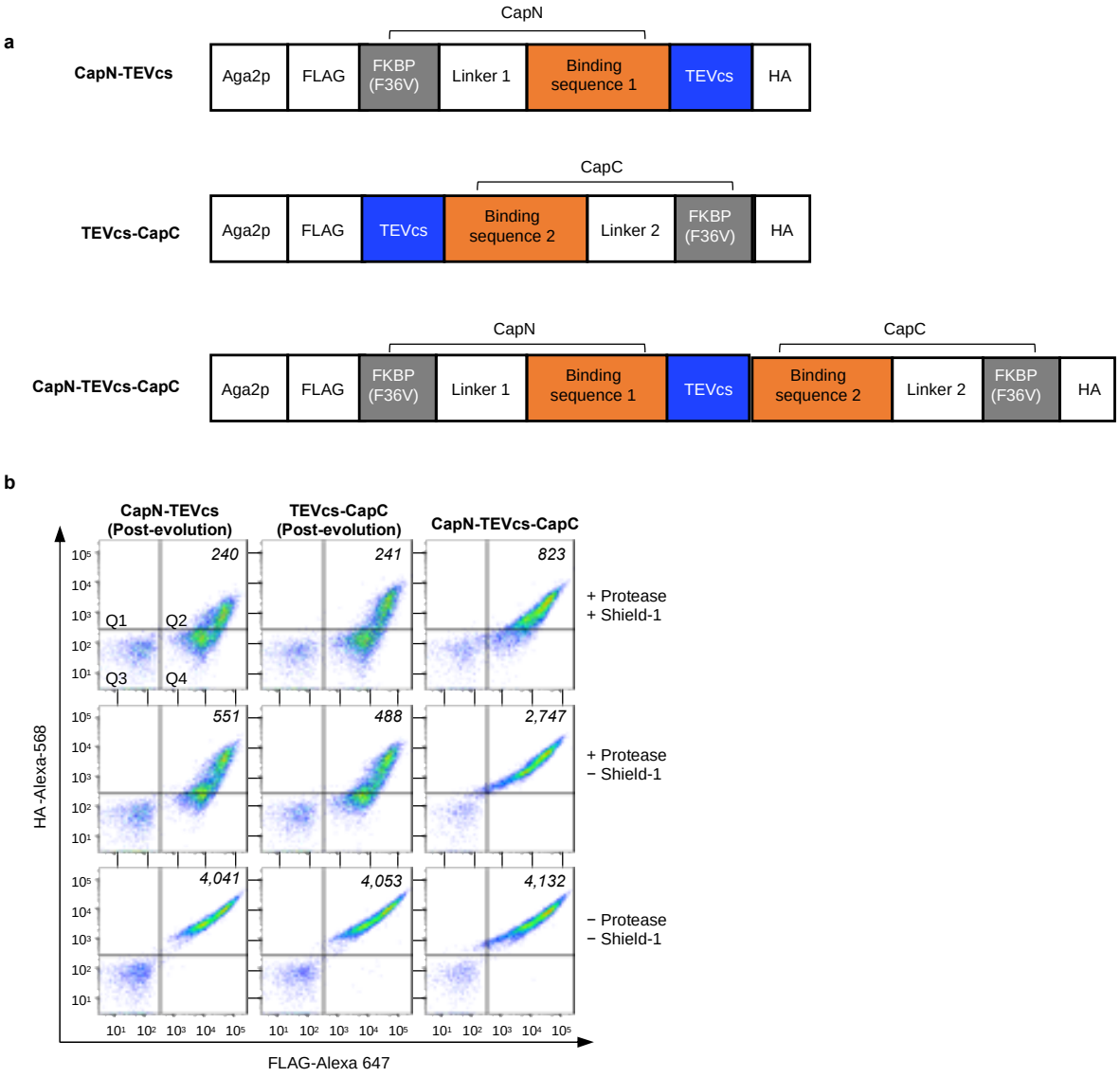
Appendix Figure A-4 Results from a 2 microsecond molecular dynamics simulations of FKBP and a capped ArgTyrSerProAsnLeu peptide in 150mM buffer. **a,** the central configurations for the top 5 clusters (Rank 1-5) obtained from RMSD clustering indicate direct interactions between Leu6 of the peptide (shown in a "licorice" representation; cap residues are shown in green, other atoms in CPK colors with gray carbons) and the F36V binding site of FKBP (shown as van-der-Waals spheres). The secondary structure of the FKBP protein is shown in a cartoon representation with red α-helices and yellow β-sheets. **b,** RMSD time traces with respect to the structures shown in **a** indicate the longevity of the respective conformations within the simulations. RMSD's of 0 indicate the simulation time points corresponding to the structures in **a**. A horizontal dashed line indicates the 1.5 Å cutoff used for clustering. **c,** time traces of the center of mass distances between each individual sidechain of the peptide and the sidechain of the F36V binding site indicate a persistent proximity of Leu6 to the binding site for a large fraction of the simulation trajectory (distances of 5-6 Å). Fractions of the simulation trajectory with close proximity of Leu6 to the F36V binding site include all configurations associated with the top 5 clusters shown in **a**. The simulation experiment was conducted by Dr. Matthias Heyden, Arizona State University.



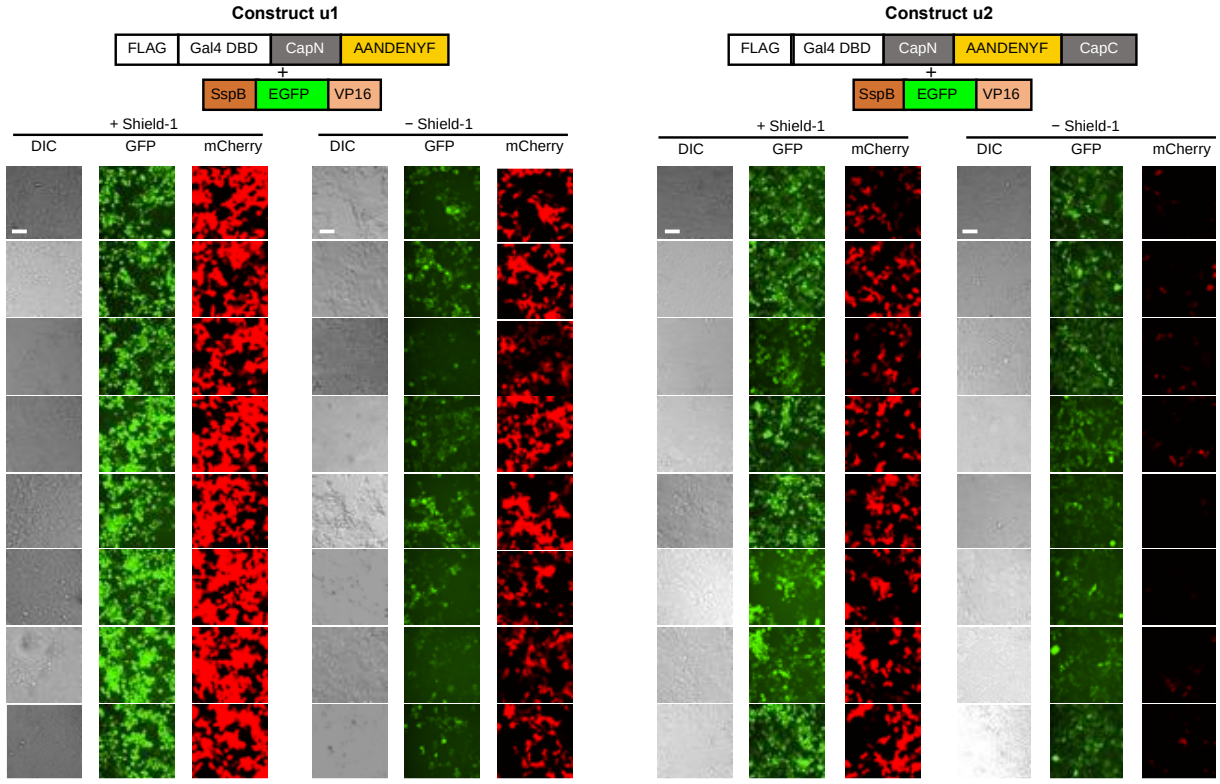
Label	Frequency	Sequence
#1	1	D W P P N L M T S
#2	1	L F R P N L G S C
#3	1	P W Y P N L D S T
#4	1	W R K P N L P Y S
#5	1	V G S P N L F Y M
#6	1	I C R P N L C A T
#7	1	K W R P N L S S C
#8	1	V G L P N L G G Y
#9	1	I G S P N L N L Y

Label	Frequency	Sequence
#10	1	T S P N L R W R W
#11	1	P F P N L D Y G L
#12	2	P F P N L G E P P
#13	1	Y S P N L C G A G
#14	1	P F P N L R S Y L
#15	1	G F P N L G Y R Y
#16	1	C K P N L N R H S
#17	1	M W P N L D N Q W
#18	1	G T P N L R P F G

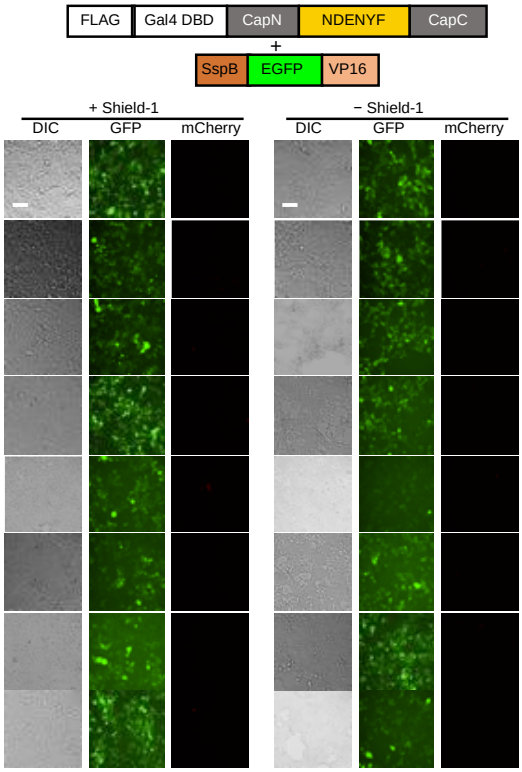
Appendix Figure A-5 a, Sequences of twenty clones from the post 2nd round CapC library as shown in main **Figure 2-9.** Eighteen distinct sequences were identified and characterized. One sequence with early stop codon is not shown. Clone #18 is the final CapC used for the rest of this study. **b,** FACS analysis of the eighteen clones shown in **a**. Values are median HA intensity of FLAG-positive cells (Q2 + Q4). Data was acquired by Dr. Lequn Geng.

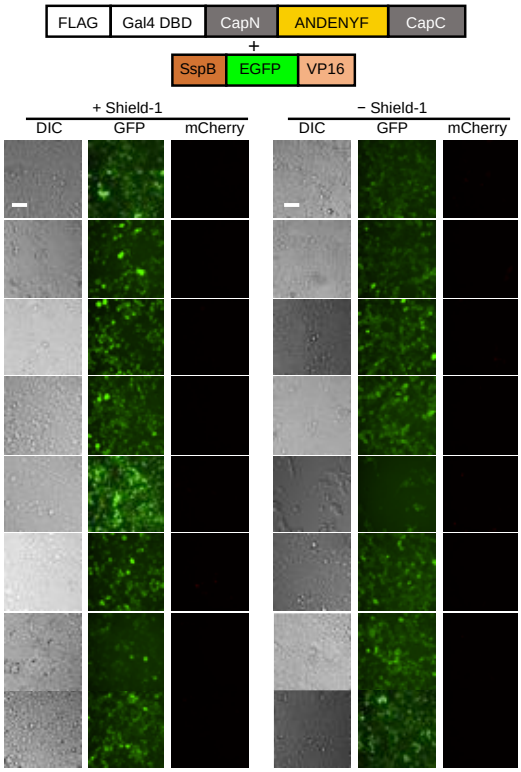


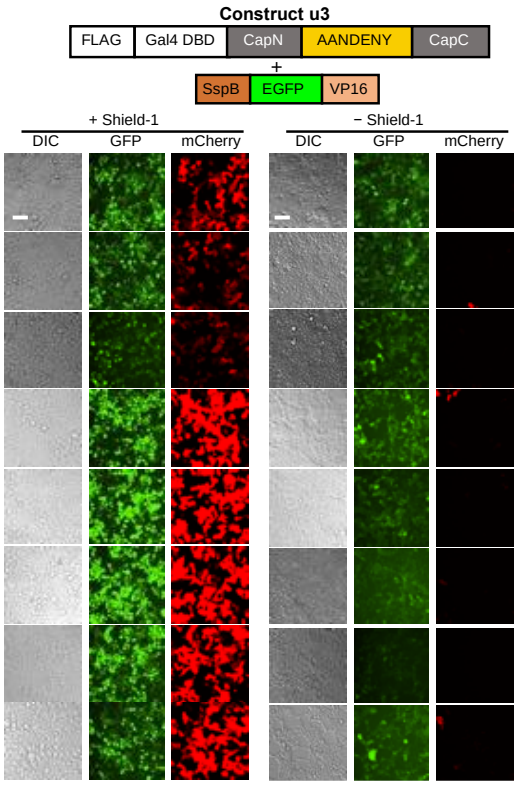
Appendix Figure A-6 a, Scheme of the three constructs tested. CapN-TEVcs-CapC is the combined use of both post-evolution CAPs. FLAG and HA are epitope tags. **b,** FACS plots of the three constructs shown in **a**. Values are median biotin intensity of FLAG-positive cells (Q2 + Q4). This experiment used a stronger TEV protease condition. Protease cleavage ("+ shield-1" or "− shield-1") is defined as the difference of the median HA signal between the + protease and – protease conditions. The dynamic ranges are calculated by the ratio of protease cleavage of the "+ shield-1" and "− shield-1" conditions. Data was acquired by Dr. Lequn Geng.



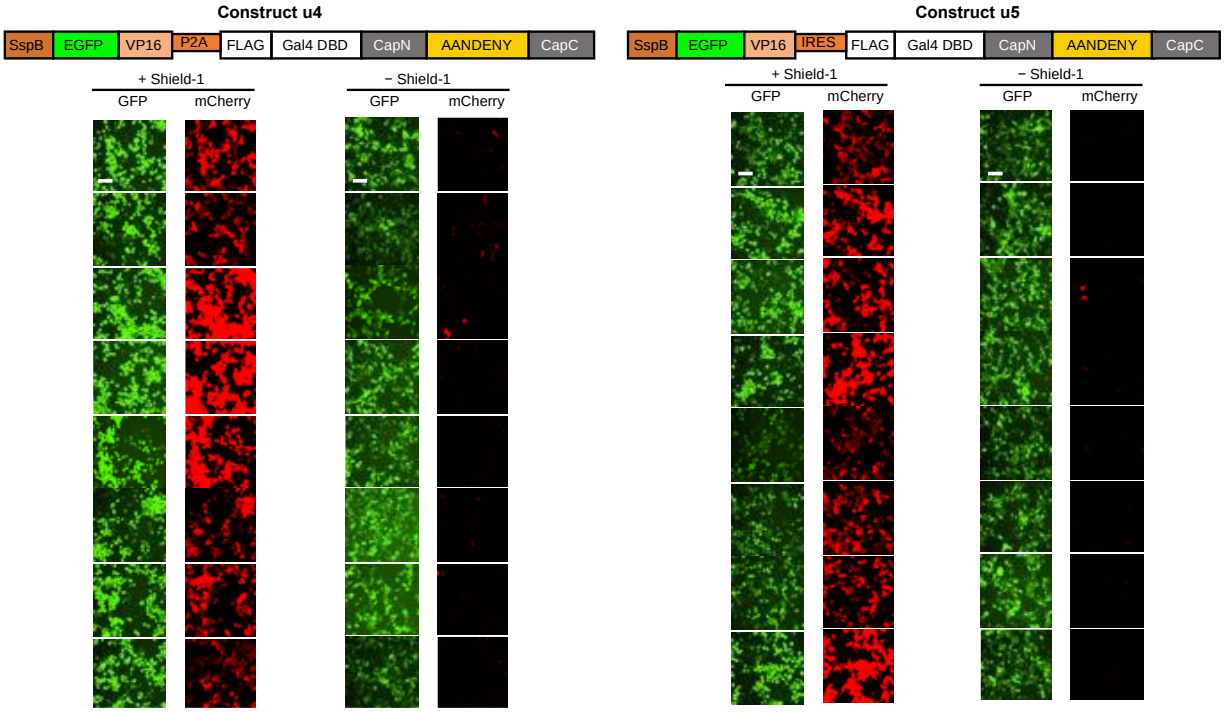
Appendix Figure A-7 Additional images of constructs u1 and u2. Fluorescence microscopy images of HEK 293T cells expressing the constructs shown and UAS-mCherry reporter gene. All scale bars, 50 µm.



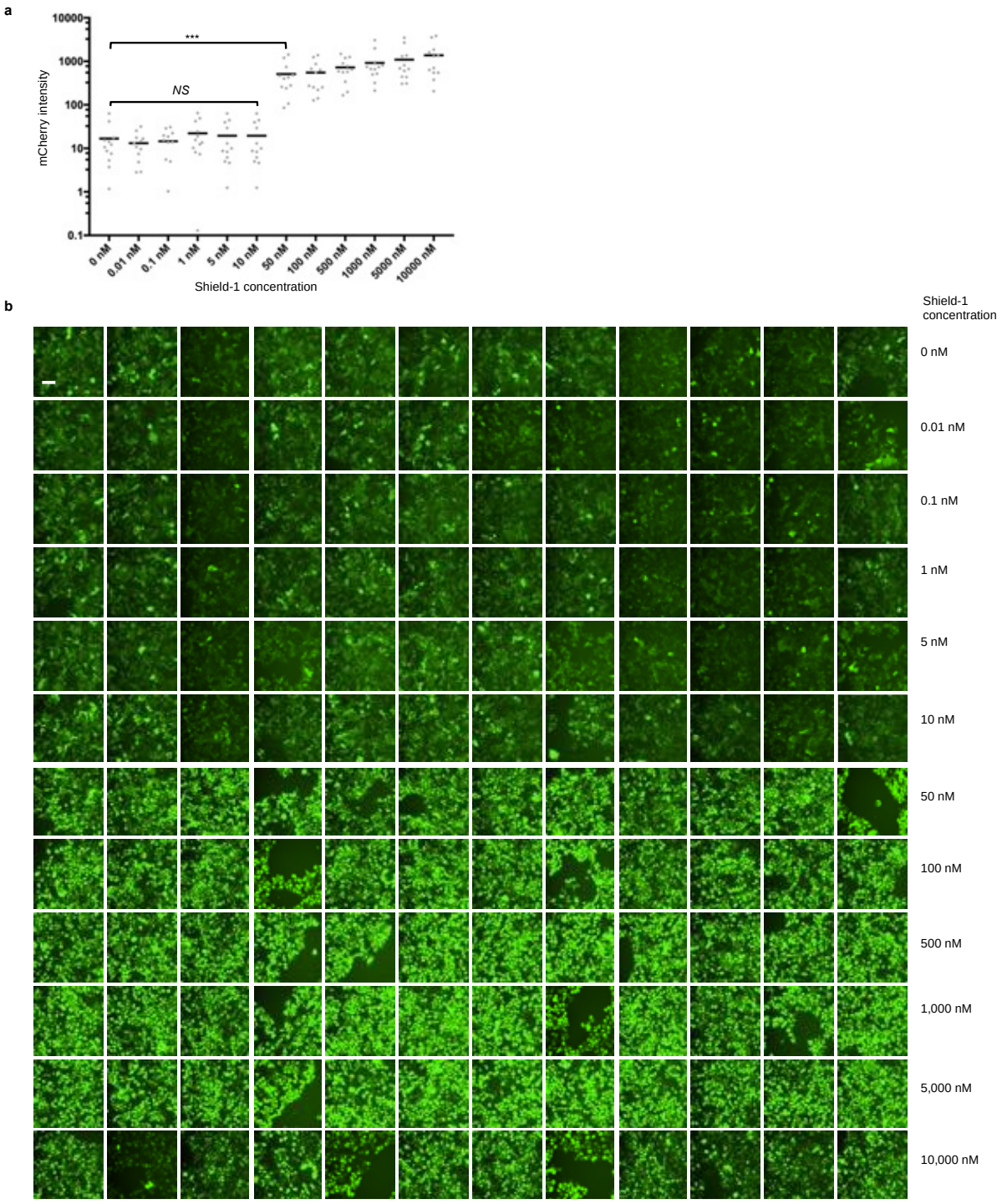




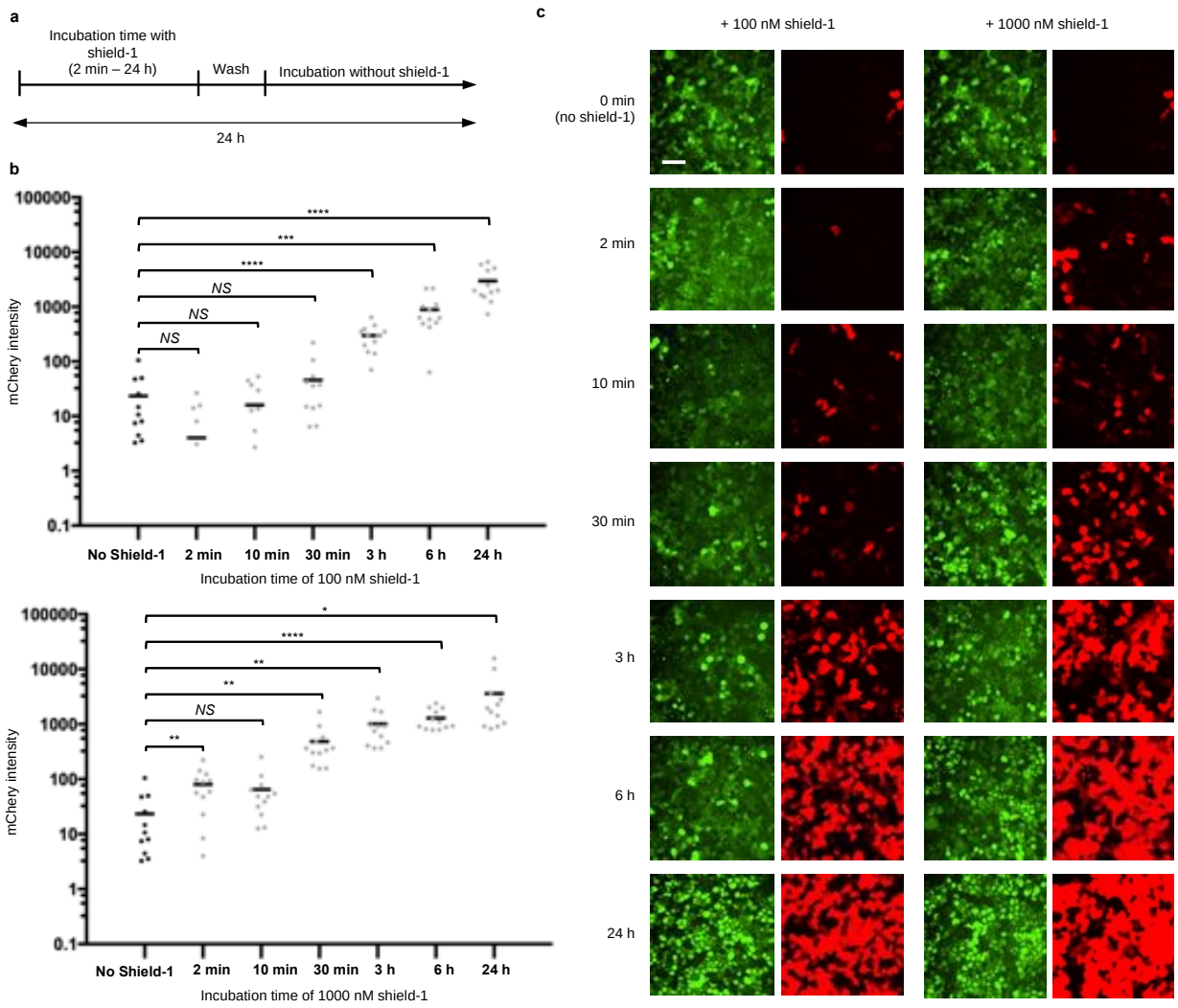
Appendix Figure A-8 Images of construct u3 and two other constructs with different SsrA sequences. Fluorescence microscopy images of HEK 293T cells expressing the constructs shown and UAS-mCherry reporter gene. These images were used for quantification of construct u3. Two other truncations shown here (truncating one or both N-terminal alanine from SsrA) did not show improvement. All scale bars, 50 µm.



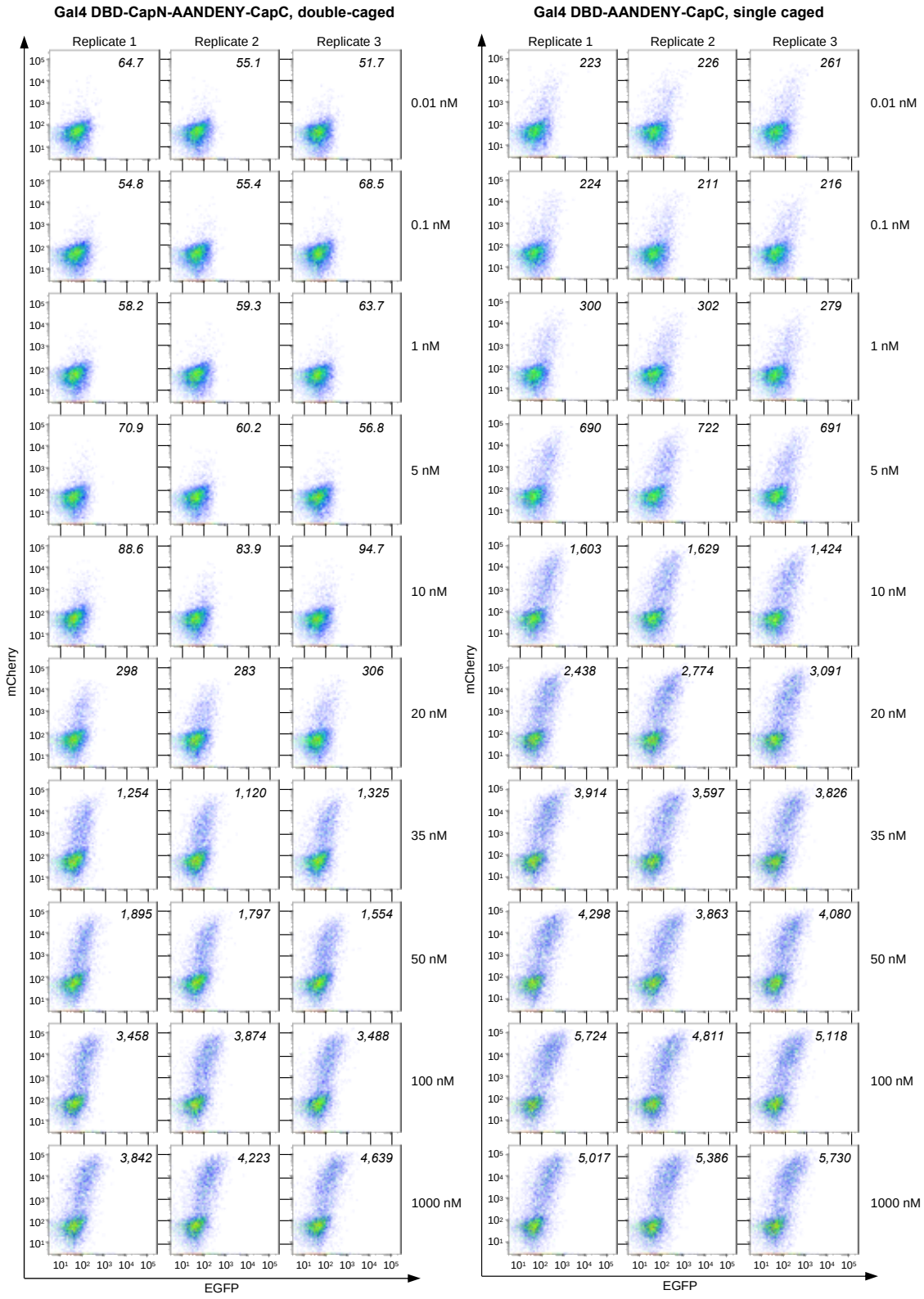
Appendix Figure A-9 Fluorescence microscopy images of HEK 293T cells expressing the constructs shown and UAS-mCherry reporter gene. These images were used for quantification of constructs u4 and u5. All scale bars, 50 µm.



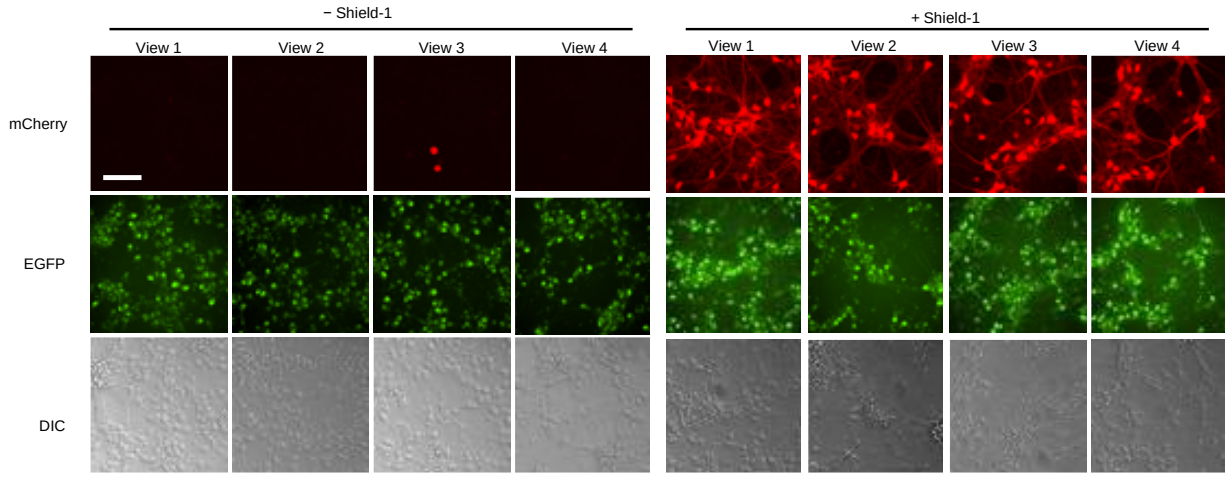
Appendix Figure A-10 Gene expression at different shield-1 concentrations with construct u3**. a,** Quantification of mCherry expression level under different shield-1 concentrations. 50 µM or above induced robust gene expression. *P* values are determined by unpaired two-tailed *t*-tests. ****P* < 0.001; *NS*, not significant. n = 12 for all conditions. **b,** Fluorescence microscopy images of HEK 293T cells, showing EGFP (transcription-activation domain expression) only. For this experiment, Gal4 was used as DBD, and UAS-mCherry was used as reporter gene. Scale bar, 50 µm. Data was acquired by Dr. Lequn Geng.



Appendix Figure A-11 Gene expression after incubating with shield-1 for different amount of time using construct u3. **a,** Scheme of the experiment. HEK 293T cells expressing construct u3 were incubated with 100 nM or 1000 nM shield-1 for different amount of time. Cells were then washed and incubated without shield-1, and imaged 24 h after initial shield-1 addition **b,** Quantification of mCherry expression level with different shield-1 incubation time, at 100 nM or 1000 nM shield-1 concentrations. *P* values are determined by unpaired two-tailed *t*-tests. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001; *NS*, not significant. n = 12 for all conditions. **c,** Representative fluorescence microscopy images of HEK 293T cells used for quantification in **b**. For this experiment, Gal4 was used as DBD, and USA-mCherry was used as reporter gene. Scale bar, 50 µm. Data was acquired by Dr. Lequn Geng.



Appendix Figure A-12 Source FACS data of dose response curve of CapC and CAPs caged SsrA. For this experiment, Gal4 was used as DBD, and *UAS-mCherry* was used as reporter gene. Values are mCherry mean intensities of cells expressing EGFP.



Appendix Figure A-13 Additional images of shield-1 induced gene expression in rat cortical neurons using u4 construct. Four additional views for each shield-1 condition are shown. Scale bar, 100 µm.

Appendix Method A-1 Experiment methods and materials used in Chapter 2.

Cloning.

Constructs for yeast surface display were cloned into the pCTCON2 vector. Constructs for protein expression in HEK 293T cells were cloned into the pAAV viral vector for transfection or the pLX208 lentiviral vector for transduction. Constructs for protein expression in neuronal culture, mouse brain, and mouse liver were cloned into the pAAV vector. FKBP for CapN was amplified from YFP-LID (Addgene plasmid #31767, Thomas Wandless laboratory). Codon optimized FKBP for CapC was synthesized by IDT.

For cloning, PCR fragments were amplified using Q5 or Taq DNA polymerase (New England Biolabs (NEB)). The vectors were double-digested with restriction enzymes (NEB), gel purified, and ligated to gel-purified PCR fragments by T4 ligation, Gibson assembly, or the In-Fusion HD Cloning Plus kit (Takara Bio). Ligated plasmid products were introduced into competent XL1-Blue *Escherichia coli* cells by heat shock transformation, or in the case of In-Fusion cloning, into the Stellar competent *Escherichia coli* cells from the kit following the corresponding protocol. For economical In-Fusion cloning, we used a modified protocol that proportionally decreased the amount of each reagent or competent cell by half or up to three quarters than the recommended amount.

Expression and purification of TEV protease.

Full-length TEV protease (TEVp, S219V) was expressed as a fusion to maltose binding protein (MBP) with a polyhistidine-tag. His-tag-MBP-TEVp(S219V) in a pYFJ16 vector was introduced into homemade competent BL21-CodonPlus (DE3)-RIPL *Escherichia coli* cells by heat shock transformation. Cells were cultured in 5 mL Miller's LB medium (Bio Basic) supplemented with 100 mg/L ampicillin at 37 °C with shaking at 220 r.p.m. for 6 h. Then, this

saturated culture was transferred to 500 mL LB with 100 mg/L ampicillin, which was grown at 37 °C with shaking at 220 r.p.m. for roughly 2-3 h until OD_{600} = 0.4-0.8. IPTG (isopropyl β-D-1-thiogalactopyranoside, EMD Millipore) was added to the culture to a final concentration of 1 mM, and the culture was grown at 16 °C with shaking at 220 r.p.m. overnight. All following procedures were done at 4 °C unless otherwise specified. Cells were harvested by centrifugation at 5,000 r.p.m. for 5 min. The cell pellet was lysed and resuspended with 15 mL ice-cold B-PER bacterial protein extraction reagent (Thermo Fisher Scientific). DTT (Fisher, freshly made) was supplemented to a final concentration of 1 mM. Benzonase nuclease (Millipore-Sigma) was added to a final concentration of ~ 100 units/ml. The mixture was incubated on ice for 5 min, and centrifuged at 10,000 r.p.m. for 15 min. The supernatant was incubated with 3 mL Ni-NTA resin (Thermo Fisher Scientific) for 10 min with rotation and then transferred to a gravity column. The resin was washed with 5 mL washing buffer (30 mM imidazole, 50 mM Tris, 300 mM NaCl, 1 mM DTT, pH = 7.8), then protein was eluted with 3 mL elution buffer (200 mM imidazole, 50 mM Tris, 300 mM NaCl, 1 mM DTT, pH = 7.8). The eluent was concentrated with a 15 mL 10,000 Da cutoff centrifugal unit (Millipore), flash frozen in liquid nitrogen, and stored at -80 °C.

To obtain effective TEVp, we concentrated each batch of TEVp by at least 10-fold. We noticed batch-to-batch variation in TEVp yield and activity.

Expression and purification of SspB-APEX2.

SspB-APEX2 in pYFJ16 vector was expressed with polyhistidine-tag in homemade competent BL21 *Escherichia coli* cells same as the expression of TEV protease described above (under "Expression and purification of TEV protease") except that DTT was not added into the cell lysate, washing buffer, or elution buffer. No protein concentration was performed.

Yeast strain and non-library culture.

Non-library yeast culture was generated by chemical transformation of the yeast surface display plasmid pCTCON2 into *Saccharomyces cerevisiae* strain EBY100 competent cells. Preparation of EBY100 competent cells has been described elsewhere[334]. To transform, 1 µg of the plasmid DNA was mixed with 5 µL competent cells. 200 µL of Frozen-EZ Yeast Solution 3 (Zymo Research) was added and thoroughly mixed. The mixture was incubated at 30 °C for 30 min to 2 h, and then transferred to 5 mL SDCAA (synthetic dextrose plus casein amino acid media, 2% dextrose, 0.67% yeast nitrogen base without amino acids (BD Difco), 0.5% Casamino acids (BD Difco), 0.54% Disodium phosphate, 0.856% Monosodium phosphate) lacking tryptophan for growth at 30 °C with shaking at 220 r.p.m. After the initial saturation (OD_{600} > 10) in 2-3 d, the yeast culture was passaged at least once prior to experiment, or was stored at 4 °C for up to half a year. Passaging was done by adding 500 µL of saturated culture into 5 mL fresh SDCAA media, and growing at 30 °C and 220 r.p.m. overnight. A negative control with no plasmid DNA accompanied each batch of transformation to ensure the media was selective.

Yeast library generation.

CapN and CapC mutant libraries were generated by first producing plasmid libraries using targeted mutagenesis followed by transformation into EBY100 yeast competent cells by electroporation. Targeted mutagenesis was done by regular PCR (polymerase chain reaction) using primers with mixed bases (IDT). The first two bases in each codon corresponding to a mutant amino acid were designed to be an equal mix of A, C, G, T, and the third base was an equal mix

of G and T. The PCR fragment was amplified such that it had forty extra bases beyond the two restriction sites used for linearizing the vector.

500 ng of the template DNA plasmid (Aga2p-FLAG-FKBP-binding sequence-TEVcs-HA) was mixed with 100 µmol forward and reverse primers annealing to the outside of FKBP gene, 1× Q5 High GC Enhancer, 1× Q5 Reaction Buffer, 1 unit of Q5 High-Fidelity Polymerase, 10 nmol dNTP (VWR) in a total volume of 50 µL (two reaction for each library).

CapN libraries:

Forward primer:

All libraries:

GGCTCTGGTGCTAGCGACTACAAGGATGACGACGATAAGACTAGT

Reverse primer:

Library 1:

GGATCCACCCTGGAAGTAGAGATTTTCMNNMNNMNNMNNMNNMNNCGCCA
CTTCCTCCACTCCACGC

Library 2:

GGATCCACCCTGGAAGTAGAGATTTTCMNNMNNMNNMNNMNNMNNMNNC
GCCACTTCCTCCACTCCACGC

Library 3:

GGATCCACCCTGGAAGTAGAGATTTTCMNNMNNMNNMNNMNNMNNMNNM
NNCGCCACTTCCTCCACTCCACGC

Library 4:

GGATCCACCCTGGAAGTAGAGATTTTCMNNMNNNMNNNMNNNMNNNMNNNM

NNMNNCGCCACTTCCTCCACTCCACGC

The PCR was run for 20 cycles with annealing temperature of 60 °C. The product DNA was gel-purified, and amplified by the following primers:

Forward primer:

All libraries:

GGCTCTGGTGCTAGCGACTACAAGGATGACGACGATAAGACTAGT

Reverse primer:

All libraries:

CTCGAGCTATTAAGCGTAATCTGGAACGTCATATGGGTAGGATCCACCCTGG

AAGTAGAGATTTTC

CapC libraries:

Forward primer:

Library 1:

GCTAGCGCAGCGAATGATGAAAATTACTTCNNKNNKNNKCCTAATTTGNNKN

NKNNKGGATCAGGCGGTTCTGGTACTG

Library 2:

GCTAGCGCAGCGAATGATGAAAATTACTTCNNKNNKCCTAATTTGNNKNNKN

NKNNKTCAGGCGGTTCTGGTACTGG

Library 3:

CTGCAGCAAGGTCTGCAGG

Reverse primer:

Library 1:

TCAGATCTCGAGCTATTACTTATCGTCGTC

Library 2:

TCAGATCTCGAGCTATTACTTATCGTCGTC

Library 3:

GAGATGGTTTCCACCTGCACTCCMNNMNNMNNMNNMNNMNNTCCAGTACC

AGAACCGCCTG

The PCR was run for 20 cycles with annealing temperature of 60 °C. The product DNA was gel-purified, and amplified by the following primers:

Forward primer:

Library 1:

CTGCAGCAAGGTCTGCAGGCTAGTGGTGGAGGAGGCTCTGGTGCTAGCGCAG

CGAATGATGAAAATTACTTC

Library 2:

CTGCAGCAAGGTCTGCAGGCTAGTGGTGGAGGAGGCTCTGGTGCTAGCGCAG

CGAATGATGAAAATTACTTC

Library 3:

GGAGTGCAGGTGGAAACCATCTC

Reverse primer:

All libraries:

TCAGATCTCGAGCTATTACTTATCGTCGTC

The template DNA was equally divided into 8 portions. Each portion was mixed with 100 μM forward and reverse primers, 1× Taq Reaction Buffer without magnesium chloride, 2 mM

magnesium chloride, 2 units of Taq Polymerase, 10 nmol dNTP (VWR) in a total volume of 50 µL. For the same library, 8 PCRs were gel-purified and combined.

Linearized vector was gel purified. We combined 2 µg of linearized vector with 8 µg of PCR fragment and concentrated using pellet paint (Millipore) following manufacturer's protocols on the day of electroporation. The precipitated DNA was resuspended in 20 µL of ultra-pure water (Thermo Fisher Scientific).

In parallel, fresh electrocompetent EBY100 yeast cells were prepared. Cells were passaged at least twice in YPD (yeast extract peptone dextrose media, 20 g dextrose, 20 g peptone and 10 g yeast extract in 1 L deionized water) prior to this procedure to ensure that cells were healthy. Saturated culture of yeast was inoculated to 200 mL of YPD to an initial OD_{600} of 0.3-0.4. Cells were grown at 30 °C with shaking at 220 r.p.m. for roughly 6 h until OD_{600} reached 1.8-2.2, and then centrifuged at 4 °C and 3,000 r.p.m. for 3 min. The cell pellet was resuspended and washed with ice-cold water, centrifuged again, and resuspended in 50 mL ice-cold sterile lithium acetate (100 mM in water). After this step, yeast was placed on ice whenever possible until after electroporation. DTT was added to a final concentration of 10 mM. The cells were then incubated at 30 °C and 220 r.p.m. for 20 min, centrifuged at 4 °C and 3,000 r.p.m. for 3 min, washed with 50 mL ice-cold water, centrifuged again, and resuspended in 0.8 mL of electroporation buffer (1M sorbital / 1 mM CaCl2).

Immediately after yeast cells preparation, 400 µL of the cells was mixed with the 20 µL concentrated DNA prepared as described above, and the mixture was transferred to an electroporation cuvette. Electroporation was done using a Bio-Rad Gene pulser XCell with the following settings: 500 V, 15 msec pulse duration, one pulse only, 2 mm cuvette. Cells were immediately rescued with 1 mL of 1:1 mixture of sorbitol and YPD media. The cuvette was washed

three times, each time with 1 mL of the fresh sorbitol and YPD mixture. All cells were combined and incubated at 30 °C for 30 min with no shaking, then 30 min with shaking at 220 r.p.m. 10 µL of the cells was plated onto three SDCAA plates with a serial dilution of 100x, 1,000x, and 10,000x, for determining library size. The rest of the cells was centrifuged at 3,000 r.p.m. for 2 min, resuspended in 5 mL SDCAA media, centrifuged again, and transferred to 200 mL SDCAA media supplemented with 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 µg/mL streptomycin, Gibco), and 30 µg/mL kanamycin (DOT Scientific). This culture was grown at 30 °C with shaking at 220 r.p.m. for 12-24 h until OD_{600} reached 15 (OD_{600} ~ 1 corresponds to roughly 1×10^7 yeast cells/mL). For each batch of library generation, a negative control was done with no plasmid DNA.

The combined library size for CapN and CapC (four libraries combined for CapN, three libraries combined for CapC) was each determined to be ~ 1×10^7.

Yeast labeling.

Same labeling procedures were used for both non-library yeast culture and yeast library. (Generation of non-library and library yeast culture is described above under "Yeast strain and non-library culture" and "Yeast library generation", respectively.) Yeast was freshly passaged in SDCAA media prior to experiment. To induce expression of the pCTCON2 plasmid, 500 µL of the overnight yeast in SDCAA media was added to 5 mL of SGCAA (synthetic galactose plus casein amino acid media, 2% galactose, 0.67% yeast nitrogen base without amino acids (BD Difco), 0.5% Casamino acids (BD Difco), 0.54% Disodium phosphate, 0.856% Monosodium phosphate)) media and let grow at 30 °C with shaking at 220 r.p.m. overnight. Prior to labeling, 250 µL (or 1 mL for the first round of library selection) overnight yeast in SGCAA media was

centrifuged at 8000 r.p.m. for 30 s, and the supernatant was discarded. The cell pellet was resuspended and washed twice, each time with 1 mL of PBSB (sterile phosphate-buffered saline supplemented with 0.1% bovine serum albumin).

Yeast that expressed TEV protease cleavage site (CapN constructs) was treated with TEVp prior to labeling with antibody-fluorophore conjugate. Samples were incubated with 200 µL PBSB containing TEVp (expressed and purified as described above, under "Expression and purification of TEV protease") and 10 µM shield-1(AOBIOUS INC) at 4 °C for 3 h with rotation. For negative control, either TEVp or shield-1 or both, was not present in PBSB. 1 mM DTT and 30 mM reduced and 3 mM oxidized of glutathione were added to all samples to keep TEVp under reducing conditions.

Yeast that expressed SsrA (CapC constructs) was subject to APEX2 labeling[401] prior to labeling with antibody-fluorophore and streptavidin-fluorophore conjugates. Samples were incubated with 100 µL of PBSB containing SspB-APEX2 (expressed and purified as described above, under "Expression and purification of SspB-APEX2") and 5 µM shield-1 at room temperature for 10 min with rotation. For negative control, either SspB-APEX2 or shield-1, or both, was not present in PBSB. After incubation, samples were washed twice, each time with 1 mL PBSB, and were resuspended in 950 µL PBSB with 1% BSA (bovine serum albumin). 1 µL biotin-phenol (1 mM in dimethyl sulfoxide) was added and thoroughly mixed with the sample. Then, 1 µL of hydrogen peroxide (0.5 mM in water, freshly prepared, EMD chemicals) was added and thoroughly mixed. After incubation for exactly 2 min, 200 µL of quenching solution 1 (30 mM Trolox (Thermo Fisher Scientific), 60 mM sodium ascorbate (Millipore Sigma), freshly prepared) was added. The sample was centrifuged at 8,000 r.p.m. for 30 s, and the supernatant was discarded. 400 µL of quenching solution 2 (5 mM Trolox, 10 mM sodium ascorbate, freshly

prepared) was then added to resuspend the cell pellet. After another centrifugation at 8,000 r.p.m. for 30 s, the supernatant was discarded, and the sample was washed twice, each time with 1 mL PBSB.

After TEVp or APEX2 labeling, samples were labeled with antibody-fluorophore and/or streptavidin-fluorophore conjugates. To label FLAG and HA epitope tags, primary anti-FLAG or anti-HA antibodies were used, followed by secondary antibodies conjugated with Alexa Fluor 568 or 647. To detect biotinylated proteins from APEX2 labeling, streptavidin conjugated with PE (phycoerythrin) was used. All antibodies were diluted to 1 µg/mL in PBSB, streptavidin-PE (Jackson Immuno Research) was diluted 200-fold, and each yeast sample was incubated with 100 µL of the mixture at room temperature for 15 min with rotation. Two washes with PBSB were done after each step of labeling. All samples were resuspended in PBSB and analyzed or sorted by FACS (fluorescence-activated cell sorting, following procedures described under "FACS analysis and library selection") within 24 h of labeling.

FACS analysis and library selection.

After labeling according to the procedures described above (under "Yeast labeling"), non-library yeast samples were analyzed with an LSRFortessa cell analyzer flow cytometer (BD Biosciences) equipped with 640 nm laser and 670/14 emission filter (for Alexa Fluor647) as well as 561 nm laser and 586/15 emission filter (for Alexa Fluor568 and PE). Library samples were sorted with a FACSAria III cell sorter flow cytometer (BD Biosciences) equipped with 633 nm laser and 660/20 emission filter (for Alexa Fluor647) as well as 561 nm laser and 582/15 emission filter (for Alexa Fluor568 and PE).

CapN library was generated and labeled as described above (under "Yeast library generation" and "Yeast labeling", respectively). For positive selection, both TEVp and shield-1 were added. For negative selection, only TEVp but not shield-1 was added. A total of four rounds of selections were performed. Number of cells collects for each round was as follows:

Round 1 (negative selection):

0.034% of the cells were collected from Library 1 (1.4×10^7 cells)

0.041% of the cells were collected from Library 2 (2.2×10^7 cells)

0.028% of the cells were collected from Library 3 (2.1×10^7 cells)

0.025% of the cells were collected from Library 4 (1.0×10^7 cells)

The collected cells are combined for further selection.

Round 2 (positive selection):

3.4% of the cells were collected (1.2×10^6 cells)

Round 3 (negative selection):

11.3% of the cells were collected (3.5×10^6 cells)

Round 4 (positive selection):

5.8% of the cells were collected (1.5×10^6 cells)

CapC library was also generated and labeled as described above (under "Yeast library generation" and "Yeast labeling", respectively). For positive selection, both SspB-APEX2 and shield-1 were added. For negative selection, only SspB-APEX2 but not shield-1 was added. A total of two rounds of selections were performed. Number of cells collected for each round was as follows:

Round 1 (negative selection):

0.3% of the cells were collected (1.5×10^4 cells)

Round 2 (positive selection):

0.1% of cells collected (3×10^3 cells)

All yeast cells were collected into 5 mL SDCAA media with 100 units/mL penicillin, 100 μg/mL streptomycin, and 30 μg/mL kanamycin. Immediately after sorting, cells were grown at 30 °C with shaking at 220 r.p.m. for 2-5 d until saturation. For the next round of sorting, cells were passaged at least once in SDCAA media and labeled according to the procedures described above, under "Yeast labeling". After the last round of sorting, plasmids were extracted by Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research) with modified manufacturer's protocol. Briefly, overnight yeast culture (500 μL) was transferred to 10 mL fresh SDCAA media and grew until OD_{600} = 1-2. Yeast cells were spun down and washed with phosphate buffered saline (PBS) once, followed by resuspension with 200 μL Solution I and 6 μL Zymolase solution provided in the kit. Vigorous vortexing was performed for > 1 minute. Yeast cells were placed in the 37 °C shaker overnight to degrade the cell wall. Following the overnight incubation, yeast cells were vigorous vortexed for > 5 minutes. Then, 200 μL Solution II was added to the yeast cells, followed by brief vortexing and incubation at room temperature for 5 min. 400 μL neutralizing solution was added and the cells were vortexed briefly. The cell lysate was spun down at 20,000 × *g* for 10 min and the supernatant was loaded to a DNA column (Epoch Life Science) to purify the plasmid DNA. The extracted plasmid DNA from the yeast library was heat shock transformed into XL1-Blue *Escherichia coli*. Individual clones were sequenced, transformed into EBY100 yeast cells (as described above under "Yeast strain and non-library culture"), labeled (as described above under "Yeast labeling"), and analyzed by FACS (as described earlier in this section).

All-atom molecular dynamics simulations.

We performed atomistic molecular dynamics simulations of FKBP and the RYSPNL hexapeptide with capped N- and C-termini (Ac-RYSPNL-NHMe) to analyze the binding pose of the peptide to its binding site. The simulations were performed with the GROMACS 2018.1 software package[410] with the AMBER99SB-ILDN protein force field[411] and TIP3P water[412].

We started the simulations based on a structure of the FKBP12 protein (PDB: 1NSG) in which we introduced the F36V mutation using DeepView/Swiss-Pdb Viewer[413]. The peptide was placed in an alpha-helical conformation with sidechains generated with the Scwrl4 program[414] outside of the putative FKBP binding site. The initial distance of the peptide center of mass relative to the center of mass of the V36 of FKBP was 16.7 Å. The smallest distance between any atom of the peptide and the FKBP protein was 5.75 Å, which allows for a separation by at least two hydration layers.

The system was placed in a cubic simulation box of 75 Å x 75 Å x 75 Å and solvated with 13444 water molecules in addition to 67 water molecules resolved in the crystal structure. 38 sodium and 40 chloride ions were added to approximate physiological salt concentrations of 150mM and to neutralize the charge of the protein (+1 e) and peptide (+1 e) at pH 7. The protonation states of the protein sidechains were estimated by the pdb2gmx tool in GROMACS. The simulations were performed in periodic boundary conditions with the particle-mesh Ewald[415] algorithm for the treatment of long-ranged electrostatic interactions using a 1.2 Å grid constant and fourth order interpolation. A 10 Å cutoff was used for short-ranged Lennard-Jones and electrostatic interactions with corrections for the pressure and total energy. The LINCS algorithm[416] was used to constrain bond lengths in the protein during dynamics simulations and the SETTLE algorithm[417] was used to constrain the geometry of water molecules.

After a steepest descent energy minimization for 1000 steps, the system was equilibrated in molecular dynamics simulations in the isothermal-isobaric ensemble at 300 K and 1 bar for 100 picoseconds using a simulation time step of 1 femtosecond and a Berendsen[418] thermostat and barostat with a 1 picosecond time constant. In this equilibration, the non-hydrogen atoms of the protein and peptide were constrained to their initial positions using isotropic position restraints with force constants of 10 k/(mol Å2). This was followed by a production simulation for 2 microseconds using a time step of 2 femtoseconds without position restraints in the isothermal-isobaric ensemble at 300 K and 1 bar with a Nosé-Hoover thermostat[419,420] with a 1 picosecond time constant and a Parrinello-Rahman barostat[421] with a time constant of 2 picoseconds.

The trajectories were analyzed using a simple clustering algorithm[422] using non-hydrogen atoms and a 1.5 Å cut-off. Further the distance between the center of mass for each sidechain of the peptide and the V36 of FKBP was monitored.

Fluorescence microscopy of cultured cells.

Confocal imaging was performed on a Nikon inverted confocal microscope with 10× air, 20× air, and 60× oil-immersion objectives, outfitted with a Yokogawa CSU-X1 5000RPM spinning disk confocal head, and Ti2-ND-P perfect focus system 4, a compact 4-line laser source: 405 nm (100 mW) 488 nm (100 mW), 561 nm (100 mW) and 640-nm (75 mW) lasers. The following combinations of laser excitation and emission filters were used for various fluorophores: DAPI (405 nm excitation; 455/50 emission), EGFP/Alexa Fluor 488 (488 nm excitation; 525/36 emission), mCherry/Alexa Fluor 568 (568 nm excitation; 605/52 emission), Alexa Fluor 647 (647 nm excitation; 705/72 emission), and differential interference contrast (DIC). ORCA-Flash 4.0

LT+sCMOS camera. Acquisition times ranged from 100 to 1000 msec. All images were collected and processed using Nikon NIS-Elements hardware control and analysis module.

HEK 293T cell culture and transfection.

Low passage HEK 293T cells (less than 20 passages) were cultured at 37 °C under 5% CO2 in T25 or T75 flasks in complete growth media, 1:1 DMEM (Dulbecco's Modified Eagle medium, Gibco): MEM (Eagle's minimal essential medium) supplemented with 10% FBS (Fetal Bovine Serum, Sigma), 20 mM HEPES (Gibco), and 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 µg/mL streptomycin, Gibco). For imaging experiments, 48-well plates were pretreated with 200 µL of 20 µg/mL human fibronectin (Millipore Sigma) for 10 min at 37°C. HEK 293T cells were then plated in 48-well plates at 60%-90% confluence. A mix of DNA was incubated with 1 µL 1 mg/mL PEI max solution in 10 µL serum-free DMEM media for 15 min at room temperature. Complete DMEM growth media (100 µL) was then mixed with the DNA-PEI max solution and added to the HEK 293T cells that were fully attached to well bottom and incubated for 18 h before further processing.

Production of lentivirus supernatant for HEK cell transduction.

New cell culture flasks were incubated with 20 µg/mL human fibronectin (HFN, Millipore Sigma) at 37 °C for at least 10 min. We found this to facilitate cells to attach to the surface and increase transfection efficiency. After incubation, HFN was aspirated, and HEK 293T cells were plated at 70-90% confluence. For a T25 flask, cells were grown at 37 °C for 1-3 h in 5 mL complete DMEM growth media, 1:1 DMEM (Dulbecco's Modified Eagle medium, Gibco): MEM (Eagle's minimal essential medium) supplemented with 10% FBS (Fetal Bovine Serum, Sigma), 20 mM

HEPES (Gibco), and 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 µg/mL streptomycin, Gibco). After incubation, 2.5 µg viral DNA, 0.25 µg pVSVG, and 2.25 µg delta8.9 lentiviral helper plasmid were combined with 250 µL of DMEM and thoroughly mixed. Then, 25 µL PEI max solution (polyethylenimine HCl Max, pH 7.3, 1 mg/mL, Polysciences) was added. The mixture was incubated at room temperature for at least 10 min, mixed with 1 mL complete media, and transferred to the T25 flask. Cells were incubated at 37 °C for 48 h, and the supernatant with virus was collected, flash frozen in liquid nitrogen, and stored at -80 °C for up to one year.

HEK 293T cell culture and infection.

Low passage HEK 293T cells (less than 20 passages) were cultured at 37 °C under 5% CO_2 in T25 or T75 flasks in complete growth media. For imaging experiments, 24-well glass-bottom plates (Cellvis) were pretreated with 350 µL 20 µg/mL human fibronectin (Millipore Sigma) for 10 min at 37 °C. HEK 293T cells were then plated in 24-well plates at 40%-60% confluence. For infection of a single well in a 24-well plate, 100-200 µL of each supernatant virus was added gently to the top of the media and incubated for 48 h before further processing.

HEK 293T cell stimulation, imaging, and data analysis for shield-1 induced protein translocation to the plasma membrane.

HEK 293T cells were plated in 24-well plates as described above at 80% confluence and then transfected with 200 ng of mCherry-CapN-SsrA-CapC-CAAX, and 200 ng of SspB-EGFP plasmid. Cells were incubated for 24 h at 37 °C before further processing. HEK 293T cells were imaged with 60× oil-immersion objective on the Nikon inverted confocal microscope. Shield-1 dissolved in complete growth media was added gently to the top of the media to 10 µM during

imaging. Intensity profile is acquired with Nikon NIS-Elements analysis module and plotted by GraphPad Prism 7.

HEK 293T cell stimulation, imaging, and data analysis for shield-1 induced delocalization from the plasma membrane to the cytosol.

HEK 293T cells were plated in 24-well plates at 40% confluence and then transduced with 200 μL of transmembrane domain lentivirus supernatant and 200 μL of the mCherry-TEV protease lentivirus supernatant, and incubated for 48 h at 37 °C before further processing. Two extra wells without any infection are also plated for background subtraction. Shield-1 dissolved in complete growth media was added gently to the top of the media to 10 μM. The two non-infection wells are treated with shield-1 and without shield-1, respectively. HEK 293T cells were incubated for 18 h at 37 °C before imaging.

HEK 293T cell stimulation, imaging, and data analysis for shield-1 controlled nuclear-cytoplasmic protein distribution.

HEK 293T cells were plated in 24-well plates at 40% confluence and then transduced with 50 μL of PKIt NES-EGFP-CapN-NLS-CapC lentivirus supernatant and 150 μL of NES-mCherry lentivirus supernatant. Cells were incubated for 24 h at 37 °C, and replated in 24-well plates. Shield-1 dissolved in complete growth media was added gently to the top of the media to 10 μM. HEK 293T cells were incubated for 18 h at 37 °C before imaging. HEK 293T cells were imaged with 60× oil-immersion objective on the Nikon inverted confocal microscope. Individual cells and the nuclei are determined by mCherry signal. Mean intensity and area are acquired with Nikon NIS-Elements analysis module. Mean intensities were subtracted by background mean intensity,

and the resulting number is multiplied with the area to give total intensities. The EGFP distribution ratio was calculated by the total intensity of EGFP in the cytosol to that in the whole cell. *P* value was determined by unpaired two-tailed *t*-tests.

HEK 293T cell stimulation, luminescence plate reading, and data analysis for shield-1 controlled opioid signaling.

For the split luciferase assay, HEK 293T cells were plated in 96-well white bottom plates as described above at 80% confluence and then transfected with 100 ng of Nb44-LgBit, and 100 ng of enkephalin-CapC-μOR-SmBit plasmid. Cells were incubated for 24 h at 37 °C. The supernatant media was aspirated and replaced with 25 μL NanoGlo reagent (Promega). Cells were treated with 20 μL of 20 μM different drugs including naloxone (Sigma-Aldrich), shield-1, and loperamide (Tocris Bioscience), and equilibrated for 5 min before the luminescence was measured by BioTek CYTATION 5 plate reader. Data was plotted by GraphPad Prism 7.

For the GloSensor assay, HEK 293T cells were plated in 96-well white bottom plates as described above at 80% confluence and then transfected with 50 ng of GloSensor, and 100 ng of enkephalin-CapC-μOR plasmid. Cells were incubated for 24 h at 37 °C. The supernatant media was aspirated and replaced with 100uL of 2 mM D-luciferin potassium salt (Gold Bio) in complete growth media (with 50 mM HEPES). Luminescence was measured by BioTek CYTATION 5 plate reader. Baseline luminescence were measured for 15 min to reach equilibration, and then cells were treated with 1 μL of 100 μM forskolin (Sigma-Aldrich) and measured for another 20 min until the signal become stable. At 45 min, cells were treated with 1 μL of 1 mM different drugs including naloxone (Sigma-Aldrich), shield-1, and DAMGO (Sigma-Aldrich). Data was plotted by GraphPad Prism 7.

HEK 293T cell stimulation, imaging, and data analysis for shield-1 dependent gene transcription activation.

HEK 293T cells were plated in 24-well plates at 40% confluence and then transduced with 100 μL of UAS-mCherry lentivirus supernatant, 100 μL of Gal4 DBD lentivirus supernatant, and 50 μL of VP16 lentivirus supernatant and incubated for 48 h at 37 °C before further processing. Two extra wells without any infection were also plated for background subtraction. Shield-1 dissolved in complete growth media was added gently to the top of the media to 10μM. The two non-infection wells were treated with shield-1 and without shield-1 respectively. HEK 293T cells were incubated for 18 h at 37 °C before intensity measurement. For **Figure 2-14**, **2-15**, **2-16** and **Appendix Figure A-7**, **A-8**, **A-9**, **A-10**, and **A-11**, HEK 293T cells were imaged with 20× air objective on a Nikon inverted confocal microscope. Twelve fields of view were acquired for each condition. For mCherry signal, mean intensity was acquired from each image with Nikon NIS-Elements analysis module. Mean intensities were subtracted by the average of the twelve background images' intensities and plotted by Prism 7. Several intensities were negative and were not shown in **Figure 2-14**, but they were still counted in the average of each condition. EGFP signal is analyzed in the same way as mCherry, except that the intensities were obtained within object area by setting a lower threshold, because the EGFP signal is relatively low. *P* values were determined by unpaired two-tailed *t*-tests. For **Figure 2-17**, HEK 293T cells were lifted by trypsin and resuspended into 1 mL complete media before analyzed by flow cytometry. mCherry mean intensity of cells expressing EGFP were calculated and plotted by Prism 7.

AAV supernatant production.

AAV supernatant was used for neuronal culture experiments. 6-well plate were pretreated with human fibronectin for 10 min at 37°C. HEK 293T cells were then plated in 6-well plates at 60–90% and transfected 2-3 h later. For each well, 0.35 μg viral DNA, 0.29 μg AAV1 serotype, 0.29 μg AAV2 serotype plasmid, and 0.7 μg helper plasmid pDF6 with 80 μL serum-free DMEM and 10 μL PEI max (PEI Max, pH 7.3 1 mg/mL, Polysciences) were mixed and incubated for 15 min at room temperature, and then 2 mL complete growth media was added and mixed. The DNA mix was added gently on the top of the cells. HEK 293T cells were incubated for 40 h at 37 °C and then the supernatant (containing secreted AAV) was collected. The virus supernatant was stored in sterile Eppendorf tubes (0.5 mL/tube), flash frozen by liquid nitrogen and stored at −80 °C.

Concentrated AAV production.

AAV was prepared for in vivo use as described previously. Three T150 flasks of HEK 293T cells with fewer than ten passages were transfected at 80% confluence. For each T150 flask, 5.2 μg construct plasmid, 4.35 μg AAV1 and 4.35 μg AAV2 serotype plasmids, 10.4 μg pDF6 adenovirus helper plasmid, and 130 μL PEI (PEI Max, pH 7.3 1 mg/mL, Polysciences) are mixed in 500 μL of serum-free DMEM for 10 min at room temperature. The DNA mixture are further suspended in 10 mL of complete media and added to cells. HEK 293T cells were incubated for 40 to 48 h at 37 °C under 5% CO_2. Cells are collected with a cell scraper, resuspended in 10 mL DPBS and then collected by centrifugation at 1200 rpm at room temperature for 5 min. The supernatant was discarded, and the pellet was resuspended in 20 mL 100 mM NaCl, 20 mM Tris (pH = 8.0). 1 mL of freshly prepared 10% sodium deoxycholate (Sigma-Aldrich) in water was added to the resuspended cells. Benzonase nuclease (Millipore-Sigma) was added to a final concentration of 50

units per mL. The solution was incubated in 37 °C water bath for 1 h and then centrifuged at 8000 rpm for 10 min. A heparin column was first equilibrated with 10 mL 100 mM NaCl, 20 mM Tris (pH = 8.0) using a peristaltic pump, and then loaded with the virus supernatant. The column was washed with 25 mL of 100 mM NaCl, 20 mM Tris (pH = 8.0), using the peristaltic pump, and then washed with 1 mL of 200 mM NaCl, 20mM Tris (pH=8.0) and 1mL of 300mM NaCl, 20mM Tris (pH=8.0) using a 5 mL syringe. For virus elution, 1.5 mL of 400 mM NaCl, 20 mM Tris (pH = 8.0); 3.0 mL of 450 mM NaCl, 20 mM Tris (pH = 8.0) and 1.5 mL of 500 mM NaCl, 20 mM Tris (pH = 8.0) was applied sequentially to the column by a 5 mL syringe. The eluted virus was concentrated using Amicon Ultra 15 mL centrifugal units with a 100,000 molecular weight cut off at 8000 r.p.m. for 6 min, to a final volume of 500 μL. For buffer exchange, 1 mL of sterile 20mM Tris, 150 mM NaCl, 0.05% PF68 solution was added to the filter unit and the column was centrifuged again until the virus volume was ~ 500 μL. The buffer exchange step was repeated two more times, and the final volume was ~ 100 μL. The concentrated AAV was aliquoted in 5 μL to the 0.6 mL low retention microcentrifuge tubes (Thermo Fisher Scientific) and stored at −80 °C.

Neuronal culture experiment.

For small scale experiments, rat cortical neurons (Thermo Fisher Scientific, Cat# A1084001) were plated according to the user protocol. Briefly, the half area 96-well glass plates (Corning, CLS4580-10EA) were coated with 50 μl 0.1 mg/ml of poly-D-lysine (Gibco) for 1 h, and then washed with ultrapure water twice. The frozen rat cortical neurons were quickly thawed in the 37 °C water bath until a small piece of ice was present. The cells were transferred to a 50 ml conical tube. To the cells, 1 ml pre-warmed 3:1 ratio of complete neurobasal media (NM) and glial enriching medium (GEM) mix was very slowly dropped in with gentle swirling. NM is composed

of neurobasal (Thermo Fisher Scientific) supplemented 2% B27 (Thermo Fisher Scientific), 50 mM HEPES (Thermo Fisher Scientific), 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 μg/mL streptomycin, Thermo Fisher Scientific), and 1% GlutaMAX (Thermo Fisher Scientific). GEM is composed of DMEM (Gibco) supplemented with 10% FBS (Fetal Bovine Serum, Sigma), 2% B27 (Thermo Fisher Scientific), 50 mM HEPES (Thermo Fisher Scientific), 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 μg/mL streptomycin, Thermo Fisher Scientific), and 1% GlutaMAX (Thermo Fisher Scientific). Additional 2 ml of complete neurobasal media was added to the cells. The viable cell density was determined by adding 10 μL of the cell suspension to 10 μl 0.4% Trypan blue, followed by cell counting using hemocytometer. 0.25×10^5 viable cells were plated on each well, and cells were grown at 37 °C with 5% CO_2. Half of the media was replaced with fresh complete neurobasal media within 4-24 h after plating. For maintaining the cells, half of the media was changed every three days.

For neuronal infection, either concentrated AAVes or supernatant AAVes were added the neurons at DIV5-DIV10 (days in vitro). Five days after infection, neurons were treated with 2 μM of shield-1 for 24 h and then imaged alive.

Mean intensities were subtracted by the average of five background images' intensities and plotted by Prism 7. *P* values were determined by unpaired two-tailed *t*-tests.

Animals.

All procedures were carried out with approval from the University Committee on Use and Care of Animals at the University of Michigan. C57BL/6 mice were maintained under a 12-hour light/dark cycle and were provided with food and water ad libitum. Adult mice with both sexes were used.

Stereotactic injection of AAV into the mouse brain.

The stereotactic injection procedure was performed as previously described[423]. Briefly, adult mice were anesthetized with isoflurane (5% for induction, 1.5% for maintenance), injected with 5 mg/kg of carprofen, and placed in a stereotactic apparatus. Body temperature was maintained at 35-37 °C. 400 nL of concentrated AAV encoding shield-1-dependent gene regulation constructs under the hSyn promoter were stereotactically injected into the lateral hypothalamic area (±0.95 mm lateral to midline, -1.40 mm posterior, and -5.25 mm ventral to bregma) at a rate of 50 nL/min. The pipette was left undisturbed in the brain for 10 min following injection to allow for pressure to equalize and prevent a vacuum effect as the pipette was removed. Mice were given subcutaneous 1 mL saline injections and were recovered from surgery. Additional 5 mg/kg subcutaneous administration of carprofen was provided the day after surgery.

AAV viral injection into the mouse liver.

Abdomen shaved adult mice were placed on a stereotaxic apparatus to maintain anesthetic state under 1.5% isoflurane. Abdomen was disinfected by povidone followed by alcohol prior to surgery. A ~ 2 cm midline incision was made in the abdomen to expose the liver. 1 μL of AAV encoding shield-1-dependent gene regulation constructs under the CMV promoter was delivered at a speed of 500 nL/min through a micropipette directly inserted into liver.

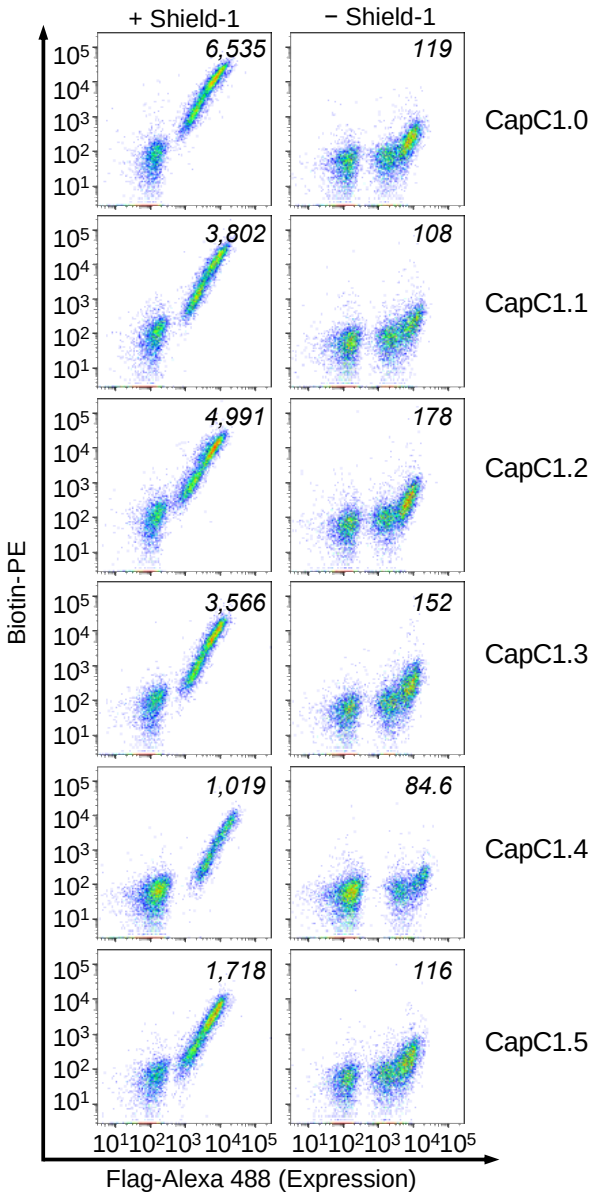
Shield-1 administration and histology.

For mouse brain, 7 days after the injection of the viral vectors, 1 μL of 1 mM aquashield-1 (Cheminpharma) or saline control was locally administered into LHA by stereotactic injection.

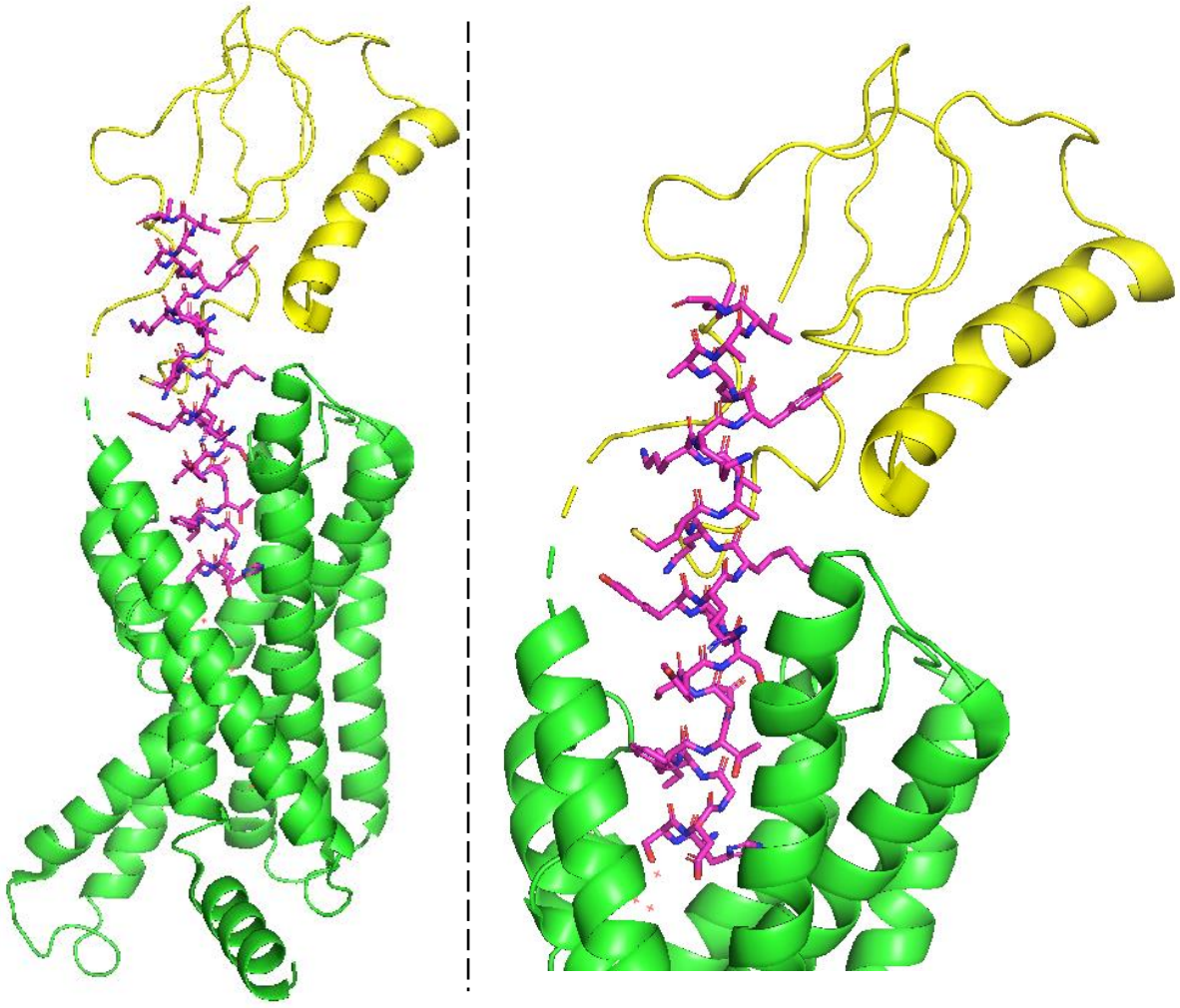
For mouse liver, 7 days after the expression of the viral vectors, animals were intraperitoneally injected with two dosages of 40 mg/kg aquashield-1 or saline with 24 h apart. 48 h after the first injection of aquashield-1 or saline, animals were euthanized and perfused with PBS and 4% paraformaldehyde (PFA). Brain and liver tissues were harvested and fixed overnight in 4% PFA then cryoprotected in 30% sucrose for 48 h at 4 °C. The fixed tissue was then embedded in optimum cutting temperature compound and sectioned at 30 µm. Sections were rinsed in 0.1% PBS Tween-20 and stained with DAPI (1:10,000, Invitrogen, D1306) for 10 min at room temperature. Sections were then rinsed again and mounted with Prolong Gold mounting media (Invitrogen, P36930). Confocal images were taken on a Nikon A1 Confocal microscope.

Appendix Method A-1 is largely adapted from: Shen, J., Geng, L., Li, X., Emery, C., Kroning, K., Shingles, G., Lee, K., Heyden, M., Li, P. and Wang, W. A general method for chemogenetic control of peptide function. *Nature Methods* (2023), 20, 112-122.

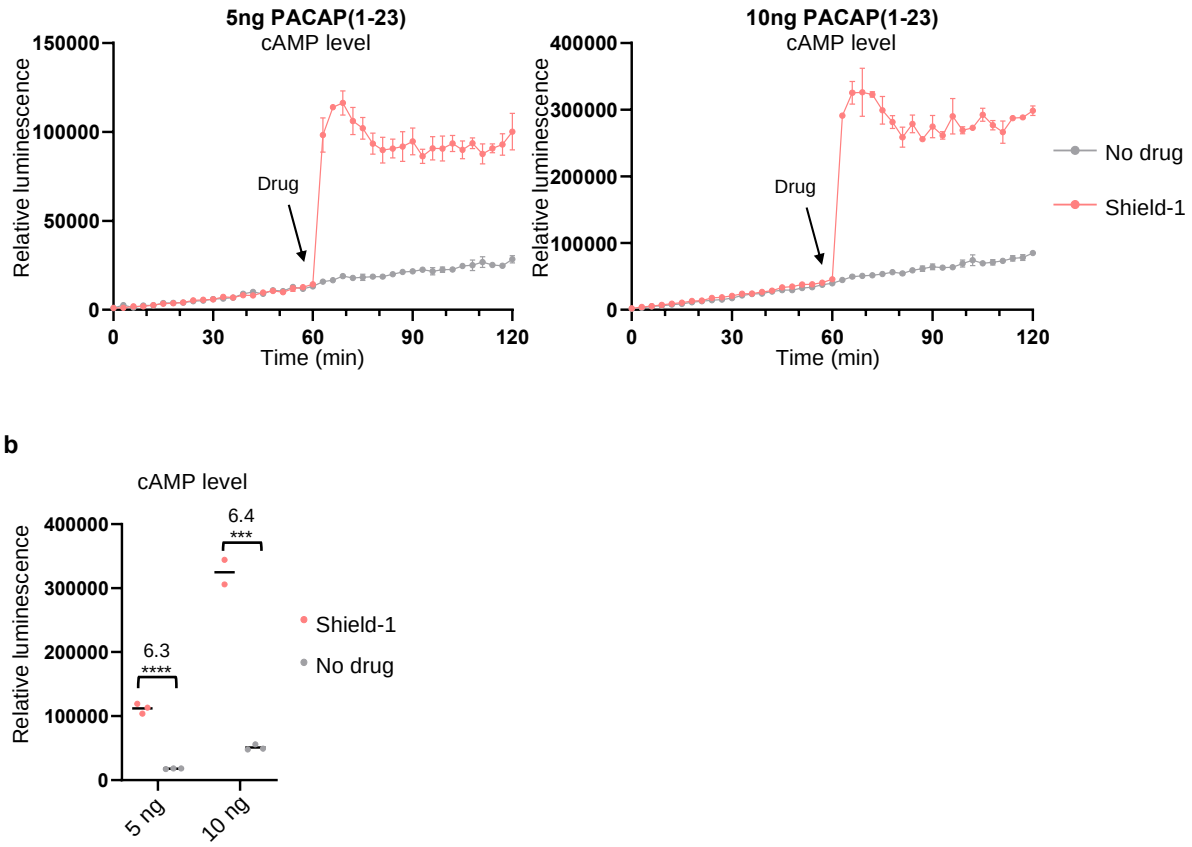
Appendix B: Appendix Figures and Methods Related to Chapter 3



Appendix Figure B-1 FACS analysis of the most enriched five clones from DuoSelect, corresponding to CapC1.1 ~ CapC1.5 shown in Table 3-1. Values are median PE intensity of FLAG-positive cells. All five clones showed shield-1 dependence.



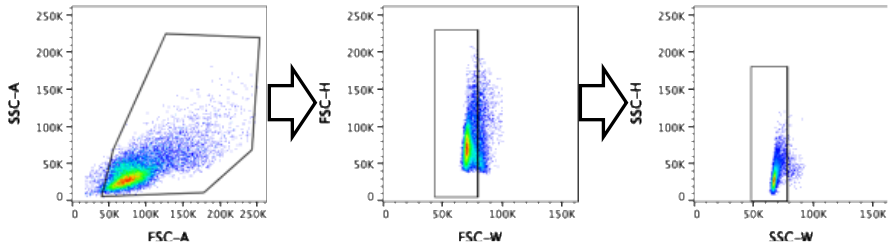
Appendix Figure B-2 Crystal structure of the binding pocket of PAC1R/PACAP(1-27) complex. Left: Full view of PAC1R-PACAP complex. Right: Zoomed in view of PACAP binding pocket. Green: PAC1R. Magenta: PACAP(1-27). Yellow: Extracellular domain of PAC1R. PDB: 8E3X.



Appendix Figure B-3 The cAMP assay for CapC1.1-controlled PACAP. **a,** The cAMP level with different transfection conditions. For a 96-well plate, 5 ng/well (left) or 10 ng/well (right) of PACAP plasmid was cotransfected with 50 ng/well PAC1R and GloSensor plasmid. Cells were stimulated with shield-1 (10 μM) or media at 60 min. n = 3 wells from one replicate for all conditions, except that n = 2 wells for 10 ng/well PACAP with "+ shield-1" condition. Errors, s.e.m. **b,** Shield-1 dependence for the experiment in **a**. Each dot stands for the mean luminescence signal of one well at the last three time points (114 min, 117 min, 120min). The values in the plot are the ratio of the mean luminescence of shield-1 to no drug conditions. *P* values are determined by unpaired two-tailed *t*-tests. ****P* < 0.001; *****P* < 0.0001. Experiment was conducted by Gwendolyn Shingles.

Appendix Method B-1 Analysis method for quantifying the activation/leakage signal ratio in the yeast surface display platform. This method was used for producing the statistics in **Figure 3-6**. Similar method was used to produce the statistic in **Figure 3-9**.

The raw flow cytometry ".fcs" data was first imported to FlowJo, and the single cell populations were gated according to the scheme below (**Appendix Figure B-4**).



Appendix Figure B-4 Consecutive gates used to select the single cell population for analysis.

The gated single cell populations were then exported as ".csv" files for further analysis in R. In brief, for each sample, the Flag negative cells were first filtered out (the threshold was determined based on the plots in FlowJo) and their mean PE and Alexa 647 signals were used for background subtraction. For Flag positive cells, the background corrected PE and Alexa 647 signals were calculated by the raw signal subtracted with that of the mean background signal. Note that each sample is corrected by its own Flag negative cell population. The Alexa 647/PE ratios of each Flag positive cell were calculated. The sample conditions and R code used are shown as followed (**Appendix Table B-1**).

Appendix Table B-1 Labeling and conditions of samples in the analysis.

Sample #	Sample
1	Library
2	Conventional Post 1st round
3	Conventional Post 2nd round
4	Conventional Post 3rd round
5	DuoSelect Post 1st round
6	DuoSelect Post 2nd round
7	DuoSelect Post 3rd round

```
library(dplyr)

temp = list.files(pattern="*.csv")

SsrA_dataset = lapply(temp, read.csv)

ratio <- c()

sd_ratio <- c()

sem_ratio <- c()

ratio_cor <- c()

sd_ratio_cor <- c()

sem_ratio_cor <- c()

PE_mean_bg <- c()

A647_mean_bg <- c()

n_pos <- c()

SsrA_dataset_filt <- list()

SsrA_dataset_ratio <- list()

ratio_median <- c()

ratio_cor_median <- c()


# use a for loop to calculate HA/FLAG ratio of each sample
for (i in c(1:7)) {

 #filter and calculate background
  SsrA_i <- SsrA_dataset[[i]]

  SsrA_i_filt <- filter(SsrA_dataset[[i]], Comp.530_30..488..A >
500)

  SsrA_i_neg <- filter(SsrA_dataset[[i]], Comp.530_30..488..A <=
500, Comp.586_15..561..A <= 250)

  SsrA_dataset_filt <- c(SsrA_dataset_filt, list(SsrA_i_filt))


  #calculate biotin/flag ratio
```

```
  PE_mean <- mean(SsrA_i_neg$Comp.586_15..561..A)

  A647_mean <- mean(SsrA_i_neg$Comp.670_14..640..A)

  PE_mean_bg <- c(PE_mean_bg, PE_mean)

  A647_mean_bg <- c(A647_mean_bg, A647_mean)


  #calculate statistics
  SsrA_i_ratio <- mutate(SsrA_i_filt, ratio = Comp.670_14..640..A
/   Comp.586_15..561..A,   ratio_cor   =   (Comp.670_14..640..A-
A647_mean) / (Comp.586_15..561..A-PE_mean))

  SsrA_dataset_ratio <- c(SsrA_dataset_ratio, list(SsrA_i_ratio))

  ratio <- c(ratio, mean(SsrA_i_ratio$ratio))

  ratio_cor <- c(ratio_cor, mean(SsrA_i_ratio$ratio_cor))

  sd_ratio <- c(sd_ratio, sd(SsrA_i_ratio$ratio))

  sd_ratio_cor <- c(sd_ratio_cor, sd(SsrA_i_ratio$ratio_cor))

  ratio_median <- c(ratio_median, median(SsrA_i_ratio$ratio))

  ratio_cor_median<-c(ratio_cor_median,
median(SsrA_i_ratio$ratio_cor))


# calculate the statistics
  n_i <- nrow(SsrA_i_filt)

  n_pos <- c(n_pos, n_i)

  sem_ratio_i <- sd(SsrA_i_ratio$ratio)/(sqrt(n_i))

  sem_ratio <- c(sem_ratio, sem_ratio_i)

  sem_ratio_cor_i <- sd(SsrA_i_ratio$ratio_cor)/(sqrt(n_i))

  sem_ratio_cor <- c(sem_ratio_cor, sem_ratio_i)

}


# export the statistics
res_cor <- data.frame(ratio_cor, ratio_cor_median, sem_ratio_cor,
n = n_pos)
```

Appendix Method B-2 Experiment methods and materials used in Chapter 3.

Cloning.

Constructs for yeast surface display were cloned into the pCTCON2 vector. Constructs for protein expression in HEK 293T cells were cloned into the pAAV viral vector for transfection or the pLX208 lentiviral vector for transduction.

FKBP for CapN and CapC was amplified from YFP-LID (Addgene plasmid #31767, Thomas Wandless laboratory). Codon optimized FKBP for CapC was synthesized by IDT.

For cloning, PCR fragments were amplified using Q5 or Taq DNA polymerase (New England Biolabs (NEB)). The vectors were double-digested with restriction enzymes (NEB), gel purified, and ligated to gel-purified PCR fragments by T4 ligation, Gibson assembly, or the In-Fusion HD Cloning Plus kit (Takara Bio). Ligated plasmid products were introduced into competent XL1-Blue *Escherichia coli* cells by heat shock transformation, or in the case of In-Fusion cloning, into the Stellar competent *Escherichia coli* cells from the kit following the corresponding protocol. For economical In-Fusion cloning, we used a modified protocol that proportionally decreased the amount of each reagent or competent cell by half or up to three quarters than the recommended amount.

Expression and purification of SspB-APEX2 fusion protein.

SspB-APEX2 in pYFJ16 vector was expressed with polyhistidine-tag. The plasmid was introduced into homemade competent BL21-CodonPlus (DE3)-RIPL *Escherichia coli* cells by heat shock transformation. Cells were cultured in 5 mL Miller's LB medium (Bio Basic) supplemented with 100 mg/L ampicillin at 37 °C with shaking at 220 r.p.m. for 6 h. Then, this saturated culture was transferred to 500 mL LB with 100 mg/L ampicillin, which was grown at 37 °C with shaking at 220 r.p.m. for roughly 2-3 h until OD_{600} = 0.4-0.8. IPTG (isopropyl β-D-1-

thiogalactopyranoside, EMD Millipore) was added to the culture to a final concentration of 1 mM, and the culture was grown at 16 °C with shaking at 220 r.p.m. overnight. All following procedures were done at 4 °C unless otherwise specified. Cells were harvested by centrifugation at 5,000 r.p.m. for 5 min. The cell pellet was lysed and resuspended with 15 mL ice-cold B-PER bacterial protein extraction reagent (Thermo Fisher Scientific). Benzonase nuclease (Millipore-Sigma) was added to a final concentration of ~ 100 units/ml. The mixture was incubated on ice for 5 min, and centrifuged at 10,000 r.p.m. for 15 min. The supernatant was incubated with 3 mL Ni-NTA resin (Thermo Fisher Scientific) for 10 min with rotation and then transferred to a gravity column. The resin was washed with 5 mL washing buffer (30 mM imidazole, 50 mM Tris, 300 mM NaCl, pH = 7.8), then protein was eluted with 3 mL elution buffer (200 mM imidazole, 50 mM Tris, 300 mM NaCl, pH = 7.8). The eluent was concentrated with a 15 mL 10,000 Da cutoff centrifugal unit (Millipore), flash frozen in liquid nitrogen, and stored at -80 °C.

Yeast strain and non-library culture.

Non-library yeast culture was generated by chemical transformation of the yeast surface display plasmid pCTCON2 into *Saccharomyces cerevisiae* strain EBY100 competent cells. Preparation of EBY100 competent cells has been described elsewhere[334]. To transform, 1 µg of the plasmid DNA was mixed with 5 µL competent cells. 200 µL of Frozen-EZ Yeast Solution 3 (Zymo Research) was added and thoroughly mixed. The mixture was incubated at 30 °C for 30 min to 2 h, and then transferred to 5 mL SDCAA (synthetic dextrose plus casein amino acid media, 2% dextrose, 0.67% yeast nitrogen base without amino acids (BD Difco), 0.5% Casamino acids (BD Difco), 0.54% Disodium phosphate, 0.856% Monosodium phosphate) lacking tryptophan for growth at 30 °C with shaking at 220 r.p.m. After the initial saturation (OD_{600} > 10) in 2-3 d, the

yeast culture was passaged at least once prior to experiment, or was stored at 4 °C for up to half a year. Passaging was done by adding 500 µL of saturated culture into 5 mL fresh SDCAA media and growing at 30 °C and 220 r.p.m. overnight. A negative control with no plasmid DNA accompanied each batch of transformation to ensure the media was selective.

Yeast library generation.

CapC library was the second CapC library in Chapter 2. CapN mutant libraries were generated by first producing plasmid libraries using targeted mutagenesis followed by transformation into EBY100 yeast competent cells by electroporation. Targeted mutagenesis was done by regular PCR (polymerase chain reaction) using primers with mixed bases (IDT). The first two bases in each codon corresponding to a mutant amino acid were designed to be an equal mix of A, C, G, T, and the third base was an equal mix of G and T. The PCR fragment was amplified such that it had forty extra bases beyond the two restriction sites used for linearizing the vector.

500 ng of the template DNA plasmid (Aga2p-FLAG-FKBP-binding sequence-SsrA) was mixed with 100 µmol forward and reverse primers annealing to the outside of FKBP gene, 1× Q5 High GC Enhancer, 1× Q5 Reaction Buffer, 1 unit of Q5 High-Fidelity Polymerase, 10 nmol dNTP (VWR) in a total volume of 50 µL (two reactions for each library).

CapN libraries:

Forward primer: CAGGCTAGTGGTGGAGGAGGCTCTGGTGCTAGCGACTAC

Reverse primer:

CTCGAGTTAGAAGTAATTTTCATCATTCGCTGCCAAATTAGGMNNMNNMNNMNNC

ACTTCCTCCACTCCACGCGTTTC

The PCR was run for 20 cycles with annealing temperature of 60 °C. The product DNA was gel-purified, and amplified by the following primers for 30 cycles:

Forward primer: CAGGCTAGTGGTGGAGGAGGCTCTGGTGCTAGCGACTAC

Reverse primer:

GTGGGAACAAAGTCGATTTTGTTACATCTACACTGTTGTTATCAGATCTCGAGTTAG

AAGTAATTTTCATCATTCGC

The template DNA was equally divided into 8 portions. Each portion was mixed with 100 μM forward and reverse primers, 1× Taq Reaction Buffer without magnesium chloride, 2 mM magnesium chloride, 2 units of Taq Polymerase, 10 nmol dNTP (VWR) in a total volume of 50 μL. For the same library, 8 PCRs were gel-purified and combined.

Linearized vector was gel purified. We combined 2 μg of linearized vector with 8 μg of PCR fragment and concentrated using pellet paint (Millipore) following manufacturer's protocols on the day of electroporation. The precipitated DNA was resuspended in 20 μL of ultra-pure water (Thermo Fisher Scientific).

In parallel, fresh electrocompetent EBY100 yeast cells were prepared. Cells were passaged at least twice in YPD (yeast extract peptone dextrose media, 20 g dextrose, 20 g peptone and 10 g yeast extract in 1 L deionized water) prior to this procedure to ensure that cells were healthy. Saturated culture of yeast was inoculated to 200 mL of YPD to an initial OD_{600} of 0.3-0.4. Cells were grown at 30 °C with shaking at 220 r.p.m. for roughly 6 h until OD_{600} reached 1.8-2.2, and then centrifuged at 4 °C and 3,000 r.p.m. for 3 min. The cell pellet was resuspended and washed with ice-cold water, centrifuged again, and resuspended in 50 mL ice-cold sterile lithium acetate (100 mM in water). After this step, yeast was placed on ice whenever possible until after electroporation. DTT was added to a final concentration of 10 mM. The cells were then incubated

at 30 °C and 220 r.p.m. for 20 min, centrifuged at 4 °C and 3,000 r.p.m. for 3 min, washed with 50 mL ice-cold water, centrifuged again, and resuspended in 0.8 mL of electroporation buffer (1M sorbital / 1 mM CaCl2).

Immediately after yeast cells preparation, 400 µL of the cells was mixed with the 20 µL concentrated DNA prepared as described above, and the mixture was transferred to an electroporation cuvette. Electroporation was done using a Bio-Rad Gene pulser XCell with the following settings: 500 V, 15 msec pulse duration, one pulse only, 2 mm cuvette. Cells were immediately rescued with 1 mL of 1:1 mixture of sorbitol and YPD media. The cuvette was washed three times, each time with 1 mL of the fresh sorbitol and YPD mixture. All cells were combined and incubated at 30 °C for 30 min with no shaking, then 30 min with shaking at 220 r.p.m. 10 µL of the cells was plated onto three SDCAA plates with a serial dilution of 100x, 1,000x, and 10,000x, for determining library size. The rest of the cells was centrifuged at 3,000 r.p.m. for 2 min, resuspended in 5 mL SDCAA media, centrifuged again, and transferred to 200 mL SDCAA media supplemented with 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 µg/mL streptomycin, Gibco), and 30 µg/mL kanamycin (DOT Scientific). This culture was grown at 30 °C with shaking at 220 r.p.m. for 12-24 h until OD_{600} reached 15 (OD_{600} ~ 1 corresponds to roughly 1×10^7 yeast cells/mL). For each batch of library generation, a negative control was done with no plasmid DNA.

The library size for CapN was each determined to be ~ 1×10^7.

One-color yeast labeling.

Same labeling procedures were used for both non-library yeast culture and yeast library. (Generation of non-library and library yeast culture is described above under “Yeast strain and

non-library culture" and "Yeast library generation", respectively.) Yeast was freshly passaged in SDCAA media prior to experiment. To induce expression of the pCTCON2 plasmid, 500 µL of the overnight yeast in SDCAA media was added to 5 mL of SGCAA (synthetic galactose plus casein amino acid media, 2% galactose, 0.67% yeast nitrogen base without amino acids (BD Difco), 0.5% Casamino acids (BD Difco), 0.54% Disodium phosphate, 0.856% Monosodium phosphate)) media and let grow at 30 °C with shaking at 220 r.p.m. overnight. Prior to labeling, 250 µL (or 1 mL for the first round of library selection) overnight yeast in SGCAA media was centrifuged at 8000 r.p.m. for 30 s, and the supernatant was discarded. The cell pellet was resuspended and washed twice, each time with 1 mL of PBSB (sterile phosphate-buffered saline supplemented with 0.1% bovine serum albumin).

Yeast that expressed SsrA was subject to APEX2 labeling[401] prior to labeling with antibody-fluorophore and streptavidin-fluorophore conjugates. Samples were incubated with 100 µL of PBSB containing SspB-APEX2 (expressed and purified as described above, under "Expression and purification of SspB-APEX2") and 5 µM shield-1 at room temperature for 10 min with rotation. For negative control, either SspB-APEX2 or shield-1, or both, was not present in PBSB. After incubation, samples were washed twice, each time with 1 mL PBSB, and were resuspended in 950 µL PBSB with 1% BSA (bovine serum albumin). 1 µL biotin-phenol (1 mM in dimethyl sulfoxide) was added and thoroughly mixed with the sample. Then, 1 µL of hydrogen peroxide (0.5 mM in water, freshly prepared, EMD chemicals) was added and thoroughly mixed. After incubation for exactly 2 min, 200 µL of quenching solution 1 (30 mM Trolox (Thermo Fisher Scientific), 60 mM sodium ascorbate (Millipore Sigma), freshly prepared) was added. The sample was centrifuged at 8,000 r.p.m. for 30 s, and the supernatant was discarded. 400 µL of quenching solution 2 (5 mM Trolox, 10 mM sodium ascorbate, freshly prepared) was then added to resuspend

the cell pellet. After another centrifugation at 8,000 r.p.m. for 30 s, the supernatant was discarded, and the sample was washed twice, each time with 1 mL PBSB.

After APEX2 labeling, samples were labeled with antibody-fluorophore and streptavidin-fluorophore conjugates. To label FLAG epitope tag, primary anti-FLAG antibodies were used, followed by secondary antibodies conjugated with Alexa Fluor 488. To detect biotinylated proteins from APEX2 labeling, streptavidin conjugated with PE (phycoerythrin) was used. All antibodies were diluted to 1 μg/mL in PBSB, streptavidin-PE (Jackson Immuno Research) was diluted 200-fold, and each yeast sample was incubated with 100 μL of the mixture at room temperature for 15 min with rotation. Two washes with PBSB were done after each step of labeling. All samples were resuspended in PBSB and analyzed or sorted by FACS (fluorescence-activated cell sorting, following procedures described under "FACS analysis and library selection") within 24 h of labeling.

DuoSelect yeast labeling.

Yeast that expressed SsrA was subject to two rounds of APEX2 labeling[401] with the same experimental procedure described in one-color yeast labeling. Yeast was first labeled under no shield-1 condition, followed by excess streptavidin-PE conjugate labeling (1:25 dilution, 200 μL). The yeast was then subjected to the second round of APEX2 labeling where shield-1 (10 μM) was supplied. After the second round of APEX2 labeling, samples were labeled with antibody-fluorophore and streptavidin-fluorophore conjugates. To label FLAG epitope tag, primary anti-FLAG antibodies were used, followed by secondary antibodies conjugated with Alexa Fluor 488. To detect biotinylated proteins from APEX2 labeling, streptavidin conjugated with Alexa Fluor 647 was used (1:50 dilution, 200 μL).

FACS analysis and library selection.

After labeling according to the procedures described above, non-library yeast samples were analyzed with an LSR Fortessa cell analyzer flow cytometer (BD Biosciences) equipped with 640 nm laser and 670/14 emission filter (for Alexa Fluor 647) as well as 561 nm laser and 586/15 emission filter (for PE). Library samples were sorted with a FACS Aria III cell sorter flow cytometer (BD Biosciences) equipped with 633 nm laser and 660/20 emission filter (for Alexa Fluor647) as well as 561 nm laser and 582/15 emission filter (for PE).

For CapC selection, the conventional method and the DuoSelect were performed side by side. In conventional method, both SspB-APEX2 and shield-1 were added in positive selection; only SspB-APEX2 but not shield-1 was added in negative selection. Number of cells collected for each round was as follows:

Conventional selection:

Round 1 (negative selection): 0.66% of the cells were collected (1.3×10^7 cells)

Round 2 (positive selection): 0.01% of cells collected (6.1×10^6 cells)

Round 3 (negative selection): 0.04% of cells collected (4.4×10^6 cells)

DuoSelect:

Round 1: 0.02% of the cells were collected (1.2×10^7 cells)

Round 2: 0.01% of the cells were collected (4.0×10^5 cells)

Round 3: 0.001% of the cells were collected (9.0×10^6 cells)

For CapN selection, only DuoSelect was performed. Number of cells collected for each round was as follows:

Round 1: 0.09% of the cells were collected (8.6×10^6 cells)

Round 2: 0.07% of the cells were collected (9.1×10^6 cells)

Round 3: 0.72% of the cells were collected (1.0×10^7 cells)

All yeast cells were collected into 5 mL SDCAA media with 100 units/mL penicillin, 100 μg/mL streptomycin, and 30 μg/mL kanamycin. Immediately after sorting, cells were grown at 30 °C with shaking at 220 r.p.m. for 2-5 d until saturation. For the next round of sorting, cells were passaged at least once in SDCAA media and labeled according to the procedures described above. After the last round of sorting, plasmids were extracted by Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research) with modified manufacturer's protocol. Briefly, overnight yeast culture (500 μL) was transferred to 10 mL fresh SDCAA media and grew until OD_{600} = 1-2. Yeast cells were spun down and washed with phosphate buffered saline (PBS) once, followed by resuspension with 200 μL Solution I and 6 μL Zymolase solution provided in the kit. Vigorous vortexing was performed for > 1 minute. Yeast cells were placed in the 37 °C shaker overnight to degrade the cell wall. Following the overnight incubation, yeast cells were vigorous vortexed for > 5 minutes. Then, 200 μL Solution II was added to the yeast cells, followed by brief vortexing and incubation at room temperature for 5 min. 400 μL neutralizing solution was added and the cells were vortexed briefly. The cell lysate was spun down at 20,000 × *g* for 10 min and the supernatant was loaded to a DNA column (Epoch Life Science) to purify the plasmid DNA. The extracted plasmid DNA from the yeast library was heat shock transformed into XL1-Blue *Escherichia coli*. Individual clones were sequenced, transformed into EBY100 yeast cells, and analyzed by FACS.

HEK 293T cell culture transfection.

Low passage HEK 293T cells (less than 20 passages) were cultured at 37 °C under 5% CO2 in T25 or T75 flasks in complete growth media, 1:1 DMEM (Dulbecco's Modified Eagle

medium, Gibco): MEM (Eagle's minimal essential medium) supplemented with 10% FBS (Fetal Bovine Serum, Sigma), 20 mM HEPES (Gibco), and 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 µg/mL streptomycin, Gibco). For imaging experiments, 48-well plates were pretreated with 200 µL of 20 µg/mL human fibronectin (Millipore Sigma) for 10 min at 37°C. HEK 293T cells were then plated in 48-well plates at 60%-90% confluence. A mix of DNA was incubated with 1 µL 1 mg/mL PEI max solution in 10 µL serum-free DMEM media for 15 min at room temperature. Complete DMEM growth media (100 µL) was then mixed with the DNA-PEI max solution and added to the HEK 293T cells that were fully attached to well bottom and incubated for 18 h before further processing.

Production of lentivirus supernatant for HEK cell transduction.

New cell culture flasks were incubated with 20 µg/mL human fibronectin (HFN, Millipore Sigma) at 37 °C for at least 10 min. We found this to facilitate cells to attach to the surface and increase transfection efficiency. After incubation, HFN was aspirated, and HEK 293T cells were plated at 70-90% confluence. For a T25 flask, cells were grown at 37 °C for 1-3 h in 5 mL complete DMEM growth media, 1:1 DMEM (Dulbecco's Modified Eagle medium, Gibco): MEM (Eagle's minimal essential medium) supplemented with 10% FBS (Fetal Bovine Serum, Sigma), 20 mM HEPES (Gibco), and 1% Penicillin-Streptomycin (50 units/mL penicillin and 50 µg/mL streptomycin, Gibco). After incubation, 2.5 µg viral DNA, 0.25 µg pVSVG, and 2.25 µg delta8.9 lentiviral helper plasmid were combined with 250 µL of DMEM and thoroughly mixed. Then, 25 µL PEI max solution (polyethylenimine HCl Max, pH 7.3, 1 mg/mL, Polysciences) was added. The mixture was incubated at room temperature for at least 10 min, mixed with 1 mL complete

media, and transferred to the T25 flask. Cells were incubated at 37 °C for 48 h, and the supernatant with virus was collected, flash frozen in liquid nitrogen, and stored at -80 °C for up to one year.

HEK 293T cell culture and infection.

Low passage HEK 293T cells (less than 20 passages) were cultured at 37 °C under 5% CO_2 in T25 or T75 flasks in complete growth media. For imaging experiments, 24-well glass-bottom plates (Cellvis) were pretreated with 350 µL 20 µg/mL human fibronectin (Millipore Sigma) for 10 min at 37 °C. HEK 293T cells were then plated in 24-well plates at 40%-60% confluence. For infection of a single well in a 24-well plate, 100-200 µL of each supernatant virus was added gently to the top of the media and incubated for 48 h before further processing.

Shield-1 dependent gene transcription activation.

HEK 293T cells were plated in 96-well plates at 40% confluence and then transduced with 33 µL of UAS-luciferase lentivirus supernatant, 33 µL of Gal4 DBD lentivirus supernatant, and 33 µL of VP16 lentivirus supernatant and incubated for 48 h at 37 °C before further processing. The media was replaced with 50 µL complete growth media with 10 µM shield-1. HEK 293T cells were incubated for 18 h at 37 °C before luminescence measurement. Cells were treated with 50 µL BrightGlo reagent (Promega), and equilibrated for 5 min. The luminescence was then measured by BioTek CYTATION 5 plate reader.

HEK 293T cell stimulation, luminescence plate reading, and data analysis for shield-1 controlled enkephalin signaling.

HEK 293T cells were plated in 96-well white bottom plates as described above at 80% confluence and then transfected with 50 ng of GloSensor, and 100 ng of enkephalin-CapC-μOR plasmid. Cells were incubated for 24 h at 37 °C. The supernatant media was aspirated and replaced with 100uL of 2 mM D-luciferin potassium salt (Gold Bio) in complete growth media (with 50 mM HEPES). Luminescence was measured by BioTek CYTATION 5 plate reader. Baseline luminescence were measured for 15 min to reach equilibration, and then cells were treated with 1 μL of 100 μM forskolin (Sigma-Aldrich) and measured for another 20 min until the signal become stable. At 45 min, cells were treated with 1 μL of 1 mM different drugs including naloxone (Sigma-Aldrich), shield-1, and DAMGO (Sigma-Aldrich). Data was plotted by GraphPad Prism 7.

HEK 293T cell stimulation, luminescence plate reading, and data analysis for shield-1 controlled PACAP signaling.

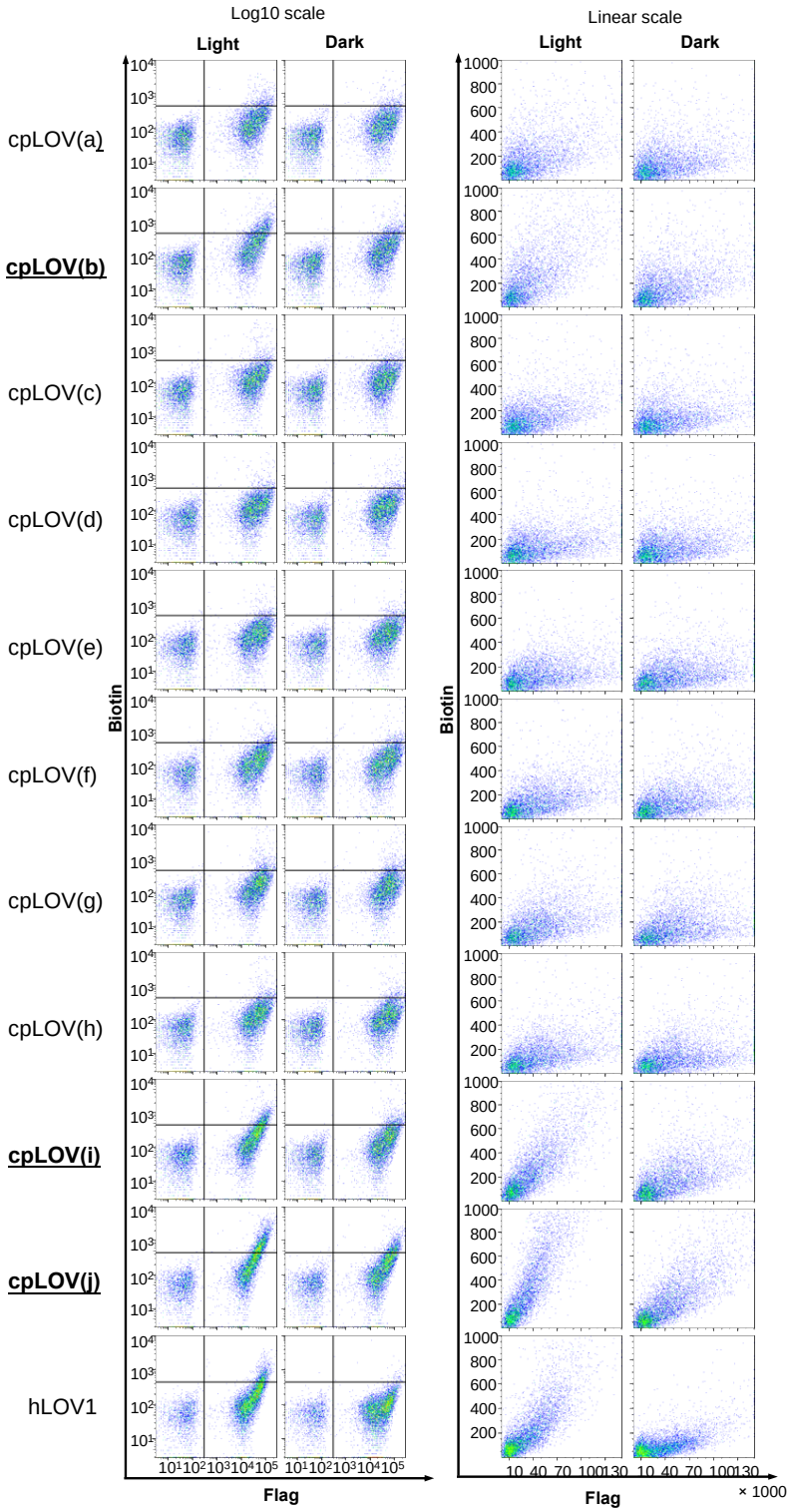
HEK 293T cells were plated in 96-well white bottom plates as described above at 80% confluence and then transfected with 50 ng of GloSensor, 50 ng of PAC1R, and 50 ng of PACAP-CapC-CD4 plasmid (later optimized to 5 ng). Cells were incubated for 24 h at 37 °C. The supernatant media was aspirated and replaced with 100uL of 2 mM D-luciferin potassium salt (Gold Bio) in complete growth media (with 50 mM HEPES). Luminescence was measured by BioTek CYTATION 5 plate reader. Baseline luminescence were first measured to reach equilibration, and then cells were treated with 1 μL of 1 mM different drugs including PACAP1-27 (Selleck Chemicals) and shield-1. Data was plotted by GraphPad Prism 7.

HEK 293T cell stimulation, luminescence plate reading, and data analysis for shield-1 controlled α-MSH signaling.

HEK 293T cells were plated in 96-well white bottom plates as described above at 80% confluence and then transfected with 50 ng of GloSensor, 50 ng of MC4R, and 50 ng of α-MSH-CapC-CD4 plasmid. Cells were incubated for 24 h at 37 °C. The supernatant media was aspirated and replaced with 100uL of 2 mM D-luciferin potassium salt (Gold Bio) in complete growth media (with 50 mM HEPES). Luminescence was measured by BioTek CYTATION 5 plate reader. Baseline luminescence were first measured to reach equilibration, and then cells were treated with 1 µL of 1 mM different drugs including Bio-Ahx-Melanotan I (Sigma-Aldrich), SNT-207707 (MedChem Express) and shield-1. Data was plotted by GraphPad Prism 7.

Appendix Method B-2 is largely adapted from: Shen, J., Geng, L., Li, X., Emery, C., Kroning, K., Shingles, G., Lee, K., Heyden, M., Li, P. and Wang, W. A general method for chemogenetic control of peptide function. *Nature Methods* (2023), 20, 112-122.

Appendix C: Appendix Figures and Methods Related to Chapter 4



Appendix Figure C-1 Related to **Figure 4-6**. Full sets of linear and log scale plots of the flow cytometry analysis of different SsrA-cpLOV fusion constructs. The quadrant was drawn based on the negative population in the bottom left corner in Q4. The percentage in Q1 indicates the ratio of the cell count in Q1 to that in (Q1 + Q2). The Light/Dark value is the ratio of the percentage of the light to that of the dark state. The cpLOV fusion constructs with the highest light dependence were bolded and underlined.

Appendix Method C-1 Analysis method for quantifying the SspB binding in the yeast surface display platform. This method was used for producing the statistics in **Figure 4-6**.

The raw flow cytometry ".fcs" data was first imported to FlowJo, and then exported as ".csv" files for further analysis in R. In brief, for each sample, the Flag negative cells were first filtered out (the threshold was determined based on the plots in FlowJo) and their mean Flag and biotin signals were used for background subtraction. For Flag positive cells, the background corrected biotin and Flag signals were calculated by the raw signal subtracted with that of the mean background signal. Note that each sample is corrected by its own Flag negative cell population. The biotin/Flag ratios of each Flag positive cell were calculated, and then applied to the significance test. The experiment conditions (**Appendix Table C-1**) and the R code are shown as followed.

Appendix Table C-1 Labeling and conditions of samples in the analysis.

Sample #	Construct	Stimulation	Sample #	Construct	Stimulation
1	cpLOV(a)	Light	12	cpLOV(f)	Dark
2	cpLOV(a)	Dark	13	cpLOV(g)	Light
3	cpLOV(b)	Light	14	cpLOV(g)	Dark
4	cpLOV(b)	Dark	15	cpLOV(h)	Light
5	cpLOV(c)	Light	16	cpLOV(h)	Dark
6	cpLOV(c)	Dark	17	cpLOV(i)	Light
7	cpLOV(d)	Light	18	cpLOV(i)	Dark
8	cpLOV(d)	Dark	19	cpLOV(j)	Light
9	cpLOV(e)	Light	20	cpLOV(j)	Dark
10	cpLOV(e)	Dark	21	hLOV1	Light
11	cpLOV(f)	Light	22	hLOV1	Dark

```
library(dplyr)
temp = list.files(pattern="*.csv")
SsrA_dataset = lapply(temp, read.csv)
ratio_cor <- c()
sd_ratio_cor <- c()
sd <- c()
mean_bg <- c()
PE_mean_bg <- c()
FLAG_mean_bg <- c()
```

```
sem <- c()
n_pos <- c()
SsrA_dataset_filt <- list()
SsrA_dataset_ratio <- list()
# use a for loop to calculate HA/FLAG ratio of each sample
for (i in c(1:22)) {
  #filter
  SsrA_i <- SsrA_dataset[[i]]
  SsrA_i_filt <- filter(SsrA_dataset[[i]], AF647 > 323)
  SsrA_i_neg <- filter(SsrA_dataset[[i]], AF647 <= 323)
  SsrA_dataset_filt <- c(SsrA_dataset_filt, list(SsrA_i_filt))

  #calculate background
  PE_mean <- mean(SsrA_i_neg$Comp.586_15..561..A)
  FLAG_mean <- mean(SsrA_i_neg$AF647)
  PE_mean_bg <- c(PE_mean_bg, PE_mean)
  FLAG_mean_bg <- c(FLAG_mean_bg, FLAG_mean)

  #calculate biotin/flag ratio
  SsrA_i_ratio <- mutate(SsrA_i_filt, ratio=Comp.586_15..561..A /
AF647,  ratio_cor  =  (Comp.586_15..561..A-PE_mean)  /  (AF647-
FLAG_mean))
  SsrA_dataset_ratio <- c(SsrA_dataset_ratio, list(SsrA_i_ratio))
  ratio_cor <- c(ratio_cor, mean(SsrA_i_ratio$ratio_cor))
  sd_ratio_cor <- c(sd_ratio_cor, sd(SsrA_i_ratio$ratio_cor))

  #calculate statistics
  median_bg                    <-                  c(median_bg,
median(SsrA_i_neg$Comp.586_15..561..A))
  mean_bg <- c(mean_bg, mean(SsrA_i_neg$Comp.586_15..561..A))
  n_i <- nrow(SsrA_i_filt)
  sd <- c(sd, sd(SsrA_i_filt$Comp.586_15..561..A))
  sem_i <- sd(SsrA_i_filt$Comp.586_15..561..A)/(sqrt(n_i))
  sem <- c(sem, sem_i)
  n_pos <- c(n_pos, n_i)
}

# export the statistics
res <- data.frame(ratio, ratio_cor, sd_ratio_cor)

# use a for loop to perform t-test on samples
p_val_ratio_cor <- c()
for (i in c(1:11)) {
  res_i_cor <- t.test(x = SsrA_dataset_ratio[[2*i]]$ratio_cor, y
= SsrA_dataset_ratio[[2*i-1]]$ratio_cor)
  p_val_ratio_cor <- c(p_val_ratio_cor, res_i_cor$p.value)
}
```

Appendix Method C-2 Experiment methods and materials used in Chapter 4.

Cloning.

Yeast surface display constructs were cloned into the pCTCON2 vector. Constructs for HEK293T cell experiments were cloned into the pLX208 lentiviral vector for lentiviruses production. Constructs for protein expression and purification in *Escherichia coli* were cloned into the pYFJ16 vector.

For cloning, PCR fragments were amplified using Q5 DNA polymerase (New England Biolabs (NEB)). The vectors were double-digested with restriction enzymes (NEB), gel purified, and ligated with gel-purified PCR fragments using T4 ligase (NEB), and Gibson assembly. Ligated DNA were heat-transformed into competent XL1-Blue *Escherichia coli* cells.

Expression and purification of TEV protease.

Full-length TEV protease (TEVp, S219V) was expressed as a fusion to maltose binding protein (MBP) with a polyhistidine-tag. His-tag-MBP-TEVp(S219V) transformed cells were cultured in 5 mL Miller's lysogeny broth (LB) medium (Bio Basic) supplemented with 100 mg/L ampicillin at 37 °C with shaking at 220 r.p.m. overnight for 12 h. This saturated culture was transferred to 500 mL LB with 100 mg/L ampicillin and grown at 37 °C with shaking at 220 r.p.m. for 2-3 h until OD_{600} reaches 0.4-0.8. Isopropyl β-D-1-thiogalactopyranoside (IPTG, MilliporeSigma) was added to the culture to a final concentration of 1 mM for protein expression induction. The culture was grown at 16 °C with shaking at 220 r.p.m. for additional 16-24 h. Cells were then harvested by centrifugation at 4,000 × *g* for 5 min. The cell pellet was resuspended with 15 mL ice-cold B-PER (Thermo Fisher Scientific) supplemented with 1 mM Dithiothreitol (DTT, Thermo Fisher Scientific) and 100 units/mL benzonase nuclease (MilliporeSigma). The mixture

was incubated on ice for 5-10 min, and centrifuged at 17,000 × *g* for 15 min to clarify the cell lysate. The clarified cell lysate was incubated with 3 mL Ni-NTA resin slurry (Thermo Fisher Scientific) for 10 min with rotation and then transferred to a gravity column. The resin was washed with 5 mL washing buffer (30 mM imidazole, 50 mM Tris, 300 mM sodium chloride, 1 mM DTT, pH = 7.8). The protein was eluted with 3 mL elution buffer (200 mM imidazole, 50 mM Tris, 300 mM sodium chloride, 1 mM DTT, pH = 7.8). The eluent was concentrated to about 1/10 of its original volume with a 15 mL 10,000 Da cutoff centrifugal unit (MilliporeSigma). The concentrated proteins were mixed with 80% (v/v) glycerol to a final concentration of 30% and then flash frozen in liquid nitrogen and stored at -80 °C.

Expression and purification of SspB-APEX2.

SspB-APEX2 was expressed with polyhistidine-tag in BL21 *Escherichia coli* cells following the same procedures as "Expression and purification of TEV protease" except that DTT was not added into the cell lysate, washing buffer, or elution buffer.

Yeast culture.

Yeast display plasmids in pCTCON2 were transformed into *Saccharomyces cerevisiae* strain EBY100 competent cells following the kit protocol by Zymo Research. Briefly, 1 µg of the DNA was mixed with 5 µL yeast competent cells, and 200 µL of Frozen-EZ Yeast Solution 3 (Zymo Research). The cells were incubated at 30 °C for 30 min to 2 h, and then transferred to 5 mL synthetic dextrose plus casein amino acid media (SDCAA, 20 g/L dextrose, 6.7 g/L yeast nitrogen base without amino acids (BD Difco), 5 g/L casamino acids (BD Difco), 5.4 g/L disodium phosphate, 8.56 g/L monosodium phosphate in deionized water). The cells were grown at 30 °C

with shaking at 220 r.p.m. After the initial saturation (OD_{600} > 10) in 2-3 days, the yeast culture was passaged at least once in SDCAA prior to protein expression induction. To induce expression of the pCTCON2 plasmid, 500 µL of the overnight yeast in SDCAA media was added to 5 mL of SGCAA (synthetic galactose plus casein amino acid media, 20 g/L galactose, 6.7 g/L yeast nitrogen base without amino acids (BD Difco), 5 g/L Casamino acids (BD Difco), 5.4 g/L Disodium phosphate, 8.56 g/L Monosodium phosphate)) media and let grow at 30 °C with shaking at 220 r.p.m. overnight.

Yeast labeling.

250 µL of yeast cells induced in SGCAA overnight were mixed with 1 mL PBSB (sterile phosphate-buffered saline supplemented with 1 g/L bovine serum albumin), and then centrifuged at 6,000 × *g* for 30 s. The cell pellet was resuspended and washed once with 1 mL PBSB.

Yeast expressing SsrA was subject to APEX2 labeling[1] prior to antibody labeling. For the light condition, the yeast samples were irradiated with blue LED light all the time except during centrifugation before antibody labeling. For the dark condition, experiments are performed in a dark room with a red light. Yeast samples were incubated with 100 µL of SspB-APEX2 solution (expressed and purified as described above, under "Expression and purification of SspB-APEX2") in the light or in the dark at room temperature for 10 min with rotation. After incubation, samples were washed twice with 1 mL PBSB, and further resuspended in 950 µL PBSB with 1% bovine serum albumin (BSA). 1 µL biotin-phenol (1 mM in dimethyl sulfoxide) was added and thoroughly mixed with the sample by vortexing. Then, 1 µL of hydrogen peroxide (0.5 mM in water, freshly prepared, MilliporeSigma) was added and thoroughly mixed by vortexing. After exactly 2 min, 200 µL of quenching solution 1 (30 mM Trolox (Thermo Fisher Scientific), 60 mM sodium

ascorbate (MilliporeSigma), freshly prepared) was added to quench the reaction. The samples were centrifuged at 6,000 × *g* for 30 s, and the supernatant was discarded. 400 µL of quenching solution 2 (5 mM Trolox, 10 mM sodium ascorbate, freshly prepared) was then added. After another centrifugation at 6,000 × *g* for 30 s, the supernatant was discarded, and the sample was washed twice with 1 mL PBSB. Samples were then incubated with 100uL PBSB solution with mouse anti-FLAG antibody (2.5 µM, Sigma) and streptavidin-phycoerythrin (PE) (200-fold dilution, Jackson ImmunoResearch) at room temperature for 15 min with rotation, washed twice with 1 mL PBSB, and incubated with Alexa Fluor 647 anti-mouse antibody at 2.5 µM in 100 µL PBSB) at room temperature for 15 min with rotation. Samples were washed twice and resuspended in 1 mL PBSB for FACS within 24 h as described under "FACS analysis and library selection".

Yeast cells that expressed TEV protease cleavage site (TEVcs) was first incubated with 200 µL of PBSB containing TEVp (expressed and purified as described above, under "Expression and purification of TEV protease") in the light or in the dark for 3 h with rotation. For negative control, TEVp were omitted. 30 mM reduced and 3 mM oxidized glutathione (MilliporeSigma) were added to all samples to keep TEVp under reducing conditions. After incubation with TEVp, samples were washed twice with PBSB and subsequently incubated with primary antibodies (mouse anti-FLAG and rabbit anti-HA antibodies at 2.5 µM each in 100 µL PBSB) and secondary antibodies (Alexa Fluor 647 anti-mouse and Alexa Fluor 568 anti-rabbit antibodies at 2.5 µM each in 100 µL PBSB) at room temperature for 15 min with rotation. Two washes with PBSB were performed after each step of labeling. Samples were resuspended in PBSB for FACS within 24 h as described under "FACS analysis and library selection".

FACS analysis.

Labeled yeast cells were analyzed with an LSRFortessa cell analyzer flow cytometer (BD Biosciences) equipped with 640 nm laser and 670/14 nm emission filter (for Alexa Fluor 647) as well as 561 nm laser and 586/15 nm emission filter (for Alexa Fluor 568 and PE). Library samples were sorted with a FACSAria III cell sorter flow cytometer (BD Biosciences) equipped with 633 nm laser and 660/20 nm emission filter (for Alexa Fluor 647) as well as 561 nm laser and 582/15 nm emission filter (for Alexa Fluor 568 and PE).

FACS data are analyzed by FlowJo and R. For cpLOV caging SsrA experiment (**Appendix Method C-1**), FLAG tag negative cells were filtered out according to the log10 scale flow cytometry plots (Supplementary Figure 1). For background subtraction, biotin and flag signals are subtracted by the mean signal of the flag tag negative cells. To eliminate the influence of difference in expression level and antibody labeling, biotin/flag signal ratios were taken for each data point to indicate the binding of SspB. *P* values were determined by unpaired two-tailed t-test. Bar plots were made by GraphPad Prism 8. For cpLOV caging TEVcs experiment, a similar analysis method as the caging SsrA experiment was performed. HA/Flag signal ratios were calculated to indicate uncleaved TEVcs level.

HEK 293T cell culture.

HEK293T cells less than 20 passages were cultured at 37 °C under 5% CO_2 in T25 or T75 flasks in complete growth media (1:1 Dulbecco's Modified Eagle Medium (DMEM, Gibco): Minimum Essential Media (MEM, Gibco) supplemented with 10% fetal bovine serum (FBS, by volume, MilliporeSigma), 50 mM HEPES (Gibco), 50 units/mL penicillin, and 50 μg/mL streptomycin).

Production of lentivirus supernatant for HEK 293T cell transduction.

New cell culture flasks were incubated with 20 µg/mL human fibronectin (HFN, MilliporeSigma) at 37 °C for at least 10 min. After incubation, HFN was aspirated, and HEK293T cells (less than 20 passages) were plated at 70-90% confluence. For a T25 flask, cells were incubated at 37 °C for 1-3 h. 2.5 µg viral DNA, 0.25 µg pVSVG, and 2.25 µg delta8.9 lentiviral helper plasmid were mixed and diluted with 250 µL of DMEM. Then, 25 µL PEI MAX solution was added to the DNA mixture. The mixture was incubated at room temperature for at least 10 min, mixed with 1 mL complete growth media, and transferred to the T25 flask. Cells were incubated at 37 °C for 48 h, and the supernatant solution was collected, flash frozen in liquid nitrogen, and stored at -80 °C before use.

HEK 293T cell lentiviral transduction for transcriptional assay.

HEK 293T cells less than 20 passages were cultured at 37 °C under 5% CO_2 in T25 or T75 flasks in complete growth media. 48-well plates were pretreated with 200 µL 20 µg/mL HFN for 10 min at 37 °C. HEK 293T cells were then plated at 40%-60% confluence. For transduction of a single well in a 24-well plate, 100 µL of each supernatant virus (UAS-mCherry reporter gene, DRD1-TEVcs-cpLOV-TF, Arrestin-TEVp) was added gently to the top of the media. The plates were wrapped with aluminum foil and incubated at 37 °C under 5% CO_2 for 48 h before stimulation. The culture media was aspirated, and the cells were stimulated with 200 µL 100 µM dopamine hydrochloride (Alfa Aesar) in complete growth media or white light for 10 or 30 min. The dopamine solution was then aspirated, and the cells were washed with 200 µL complete growth media for three times. For the cells that were not treated with dopamine, complete growth media was used in stimulation step. The plates were wrapped with aluminum foil and incubated at

37 °C under 5% CO_2 for 24 h (cpLOV single caged TEVcs in **Figure 4-12**) or 48 h (double caged TEVcs in **Figure 4-14 and Figure 4-15**) before fluorescence microscope imaging.

Fluorescence microscopy of cultured cells.

Confocal imaging was performed on a Nikon inverted confocal microscope with 20× air objectives, outfitted with a Yokogawa CSU-X1 5000RPM spinning disk confocal head, and Ti2-ND-P perfect focus system 4, a compact 4-line laser source: 405 nm (100 mW), 488 nm (100 mW), 561 nm (100 mW) and 640 nm (75 mW) lasers. The following combinations of laser excitation and emission filters were used for various fluorophores: DAPI (405 nm excitation; 455/50 nm emission), EGFP/Alexa Fluor 488 (488 nm excitation; 525/36 nm emission), mCherry/Alexa Fluor 568 (568 nm excitation; 605/52 nm emission), Alexa Fluor 647 (647 nm excitation; 705/72 nm emission). ORCA-Flash 4.0 LT+sCMOS camera. Acquisition times ranged from 100 to 1000 msec. All images were collected and processed using Nikon NIS-Elements hardware control and analysis module.

Fluorescence microscopy image analysis.

For confocal images, 10-12 fields of view per well were taken. Confocal fluorescence microscopy images were analyzed using the General Analysis 3 module on the Nikon NIS-Elements AR Analysis software. For the transcriptional assay reporter gene channel, a threshold was set to be just above the background fluorescence. Signal above the threshold from each field of view was integrated, background corrected, and plotted as dot plots using GraphPad Prism 8. Mean intensity was calculated for each condition. *P* values were determined by Wilcoxon-Mann-Whitney test.

Appendix Method C-2 is largely adapted from: Geng, L., Shen, J. and Wang, W. Circularly permuted AsLOV2 as an optogenetic module for engineering photoswitchable peptides. *Chemical Communications* (2021), 57(65), 8051-8054.

Appendix D: Designing a Chemogenetic Inhibitor for CBP/p300 Transcriptional Coactivator

Background and Significance

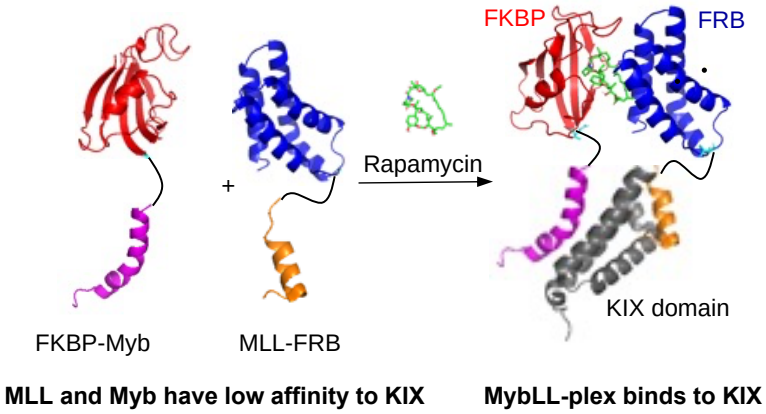
Gene transcription is an essential and ubiquitous process in biological systems. Dysregulation of gene transcription can cause a wide range of diseases, including cancer, autoimmunity, neurological disorders, etc[424]. The protein-protein interaction (PPI) between transcriptional activators and coactivators is an important mechanism contributing to gene transcription, and therefore serves as excellent targets for gene expression modulation and potential therapeutic development.

CBP/p300 is a transcriptional coactivator regulating gene transcription by interacting with multiple transcriptional activators. KIX domain is one of the activator binding domain of CBP/p300, which interacts with over fifteen different transcriptional activators through two binding surfaces[425]. The dysregulation of the PPI between the KIX domain and its binding activators leads to many human diseases. For example, c-Myb is an oncoprotein vital for the proliferation of immature hematopoietic cells[426]; MLL is involved in the maintenance of blood cell production and the initiation of various leukemias[427]. Therefore, KIX domain is an important target for studying cancer and could potentially serve as a therapeutic target.

To study the biological role of KIX domain, KIX specific inhibitors are desired. Although many inhibitors have been developed, their selectivity towards CBP/p300 KIX over other KIX motifs (e.g., Med15 and RECQL5) is underexplored and the off-target profiling is not completed[428]. A dual-targeting CBP/p300 KIX inhibiting peptide, MybLL-tide[429], was developed by covalently

connecting the binding peptides derived from c-Myb and MLL, so that both binding surfaces of KIX were targeted to achieve high specificity. MybLL-tide has picomolar affinity towards CBP/p300 KIX and shows 5600-fold higher selectivity towards other KIX motifs and activator-binding domains.

Despite that MybLL-tide exhibits excellent in vitro affinity towards KIX, its bioactivity is limited by peptide's low cell permeability. To achieve higher inhibition efficiency as well as faster temporal control of the inhibition, we sought to develop a chemogenetically-controlled MybLL-tide-like complex (MybLL-plex) (**Appendix Figure D-1**). MybLL-plex consists of two components, where the MLL and Myb are each fused to a CIP component. At the basal state, MLL and Myb both have lower affinity towards KIX. The addition of the small molecule dimerizer brings MLL and Myb in proximity, mimicking the MybLL-tide, and therefore generates the MybLL-plex with higher affinity towards KIX domain.

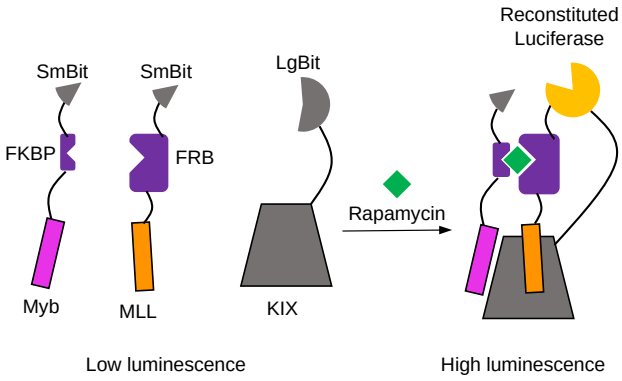


Appendix Figure D-1 Design of MybLL-plex. Myb and MLL are fused to FKBP and FRB, respectively. Addition of rapamycin induces dimerization of FKBP and FRB and forms MybLL-plex with higher binding affinity towards KIX.

MybLL-plex can be genetically delivered to cells to reach high intracellular concentration. In addition, MybLL-plex allows fast inhibition of KIX domain by using a small molecule as the signal input. With these advantages, MybLL-plex could be a powerful tool to investigate the biological role of KIX domain.

Design and development of MybLL-plex

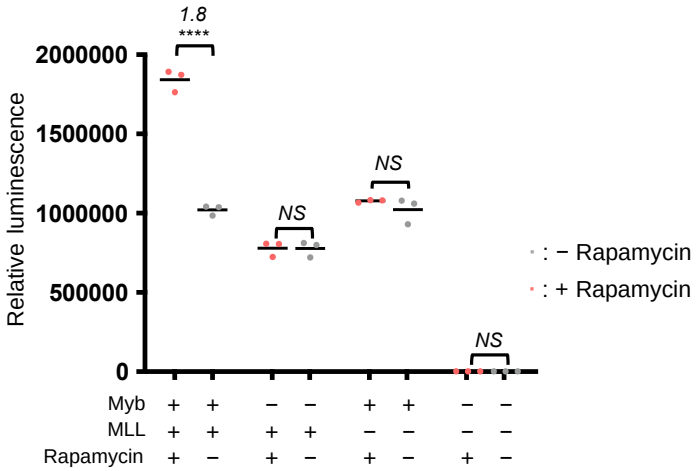
Based on previous studies[429], the chimeric MybLL-tide peptide has a strong picomolar affinity towards KIX, while the individual peptides (MLL and Myb) have much weaker micromolar affinities. Therefore, we propose to fuse MLL and Myb to a pair of chemically induced dimerization protein domains, FKBP and FRB[44] (**Appendix Figure D-1**). At basal state, MLL and Myb will bind relatively poorly to KIX due to their low micromolar affinity. Addition of a chemical, rapamycin, will induce the heterodimerization of FKBP and FRB, bringing MLL and Myb into proximity, forming the MybLL-plex and mimicking MybLL-tide with stronger affinity towards KIX.



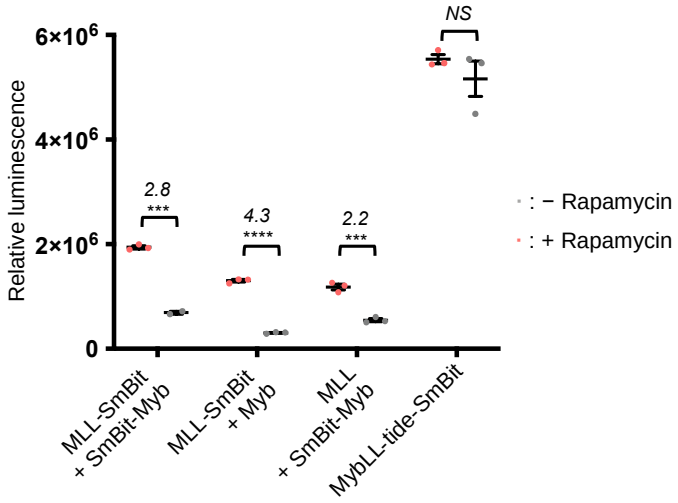
Appendix Figure D-2 Scheme of the split luciferase assay. The binding of any MybLL-plex component can reconstitute the split luciferase.

To quantify the binding of MybLL-plex to KIX, we employed the split luciferase assay, NanoGlo[370] (**Appendix Figure D-2**). The luminescence under no rapamycin condition represents the background binding of Myb or MLL to KIX. The luminescence under rapamycin condition represents the MybLL-plex binding to KIX. As shown in **Appendix Figure D-3**, rapamycin induced 1.8-fold luminescence increase for MybLL-plex design (both Myb and MLL are present), while no significant luminescence change was observed when only one component is present. Also,

the "+ rapamycin" condition for MybLL-plex is much higher than those lacking any of the component. This result suggests that rapamycin can induce the formation of MybLL-plex and achieve higher affinity towards KIX domain.

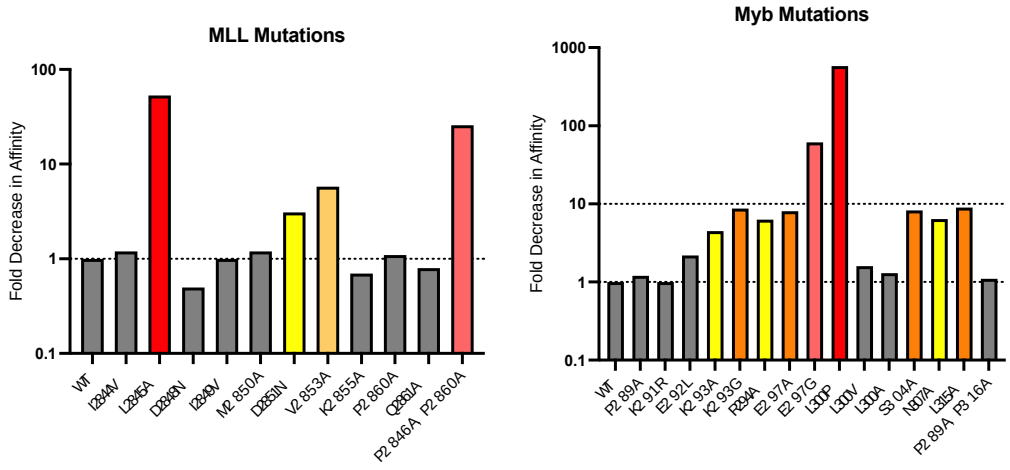


Appendix Figure D-3 Quantification of luminescence in the assay shown in **Appendix Figure D-2**. Values on the plot are the ratio of mean luminescence of + rapamycin to that of − rapamycin conditions for each construct. The center lines indicate mean values of luminescence. *P* values are determined by unpaired two-tailed *t*-tests. ****$P < 0.0001$; *NS*, not significant.



Appendix Figure D-4 Quantification of luminescence of the control study. For the first four columns, SmBit exists on either one or both components. In the fifth construct, SmBit is fused to MybLL-tide as a positive control. Values on the plot are the ratio of mean luminescence of + rapamycin to that of − rapamycin conditions for each construct. The center lines indicate mean values of luminescence. *P* values are determined by unpaired two-tailed *t*-tests. ***$P < 0.001$; ****$P < 0.0001$; *NS*, not significant.

To confirm that the increased luminescence signal yielded from the increased binding affinity of MybLL-plex, we further investigated if the two copies of SmBit component have an impact on the luminescence readout. As shown in **Appendix Figure D-4**, similar rapamycin responses were observed regardless of the number of SmBit copies existing, suggesting that the increased luminescence is mainly due to the increased affinity instead of the multivalence effect. Further, in a control condition, where SmBit is fused to the MybLL-tide, the luminescence is significantly higher than MybLL-plex systems, suggesting that MybLL-plex has lower binding affinity towards KIX compared to MybLL-tide.

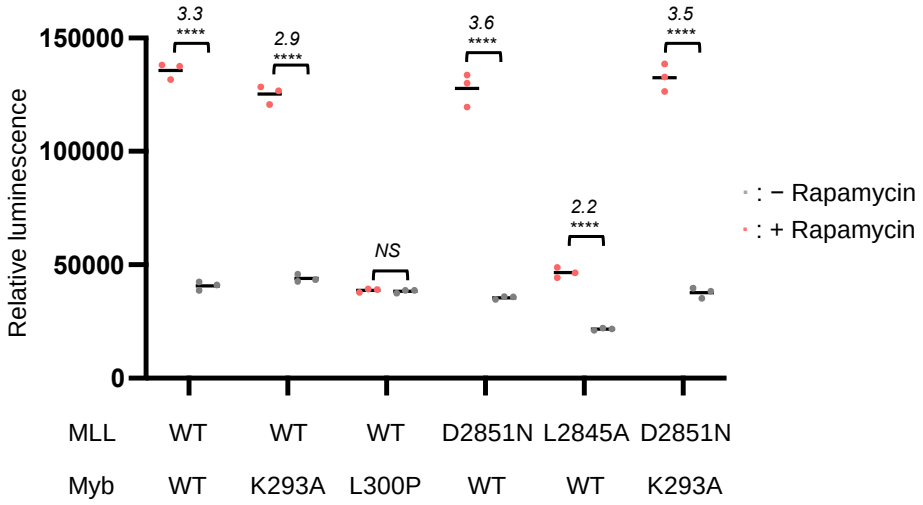


Appendix Figure D-5 MLL and Myb variants with decreased affinity.

Although a rapamycin-dependent activation was observed for MybLL-plex, non-negligible background binding activity exists. We hypothesized that such background activity could be reduced by using lower affinity MLL and Myb mutants (**Appendix Figure D-5**). MLL(D2851N) and MLL(L2845A) reduce the MLL affinity by 3-fold and 53-fold. Myb(K293A) and Myb(L300P) reduce the Myb affinity by 5-fold and 580-fold respectively. We then sought to test several combinations of these mutants.

As shown in **Appendix Figure D-6**, the mutants which mildly altered the affinity, MLL(D2851N) and Myb(K293A), did not significantly change the rapamycin response, nor the

background activity. The two mutants MLL(L2845A) and Myb(L300P), which drastically reduces the affinity, also reduced MybLL-plex activity, even totally abolished the rapamycin dependence. Therefore, this approach might not be feasible to tune the MybLL-plex background activity.



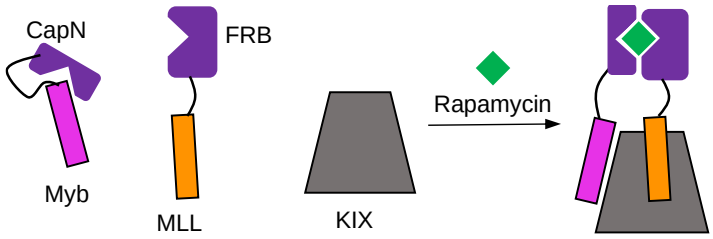
Appendix Figure D-6 MybLL-plex with different MLL and Myb mutations. Values on the plot are the ratio of mean luminescence of + rapamycin to that of − rapamycin conditions for each construct. The center lines indicate mean values of luminescence. *P* values are determined by unpaired two-tailed *t*-tests. ****$P < 0.0001$; *NS*, not significant.

Conclusions and future directions

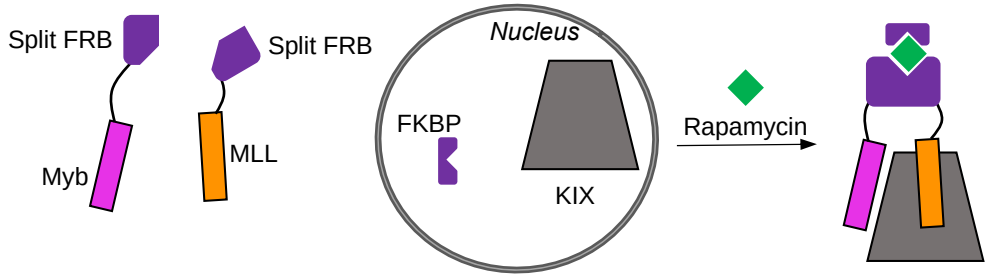
We designed and developed a chemogenetically controlled inhibitor for KIX domain from CBP/p300 transcriptional coactivator, MybLL-plex. MybLL-plex designed by splitting the KIX specific inhibitor MybLL-tide into two fragments with lower binding affinity towards KIX. The two fragments can be reconstituted by using CIP. In a cell culture based split luciferase assay, we determined the MybLL-plex can bind to KIX domain with a rapamycin dependence up to 3.6-fold.

The current MybLL-plex still exhibits non-negligible background activity. To further optimize this system, several possible approaches could be taken. 1) Replace the FKBP domain with CapN (more details are in Chapter 2), so that the mechanism of steric hinderance and proximity are combined (**Appendix Figure D-7**). The extra caging from CapN can potentially

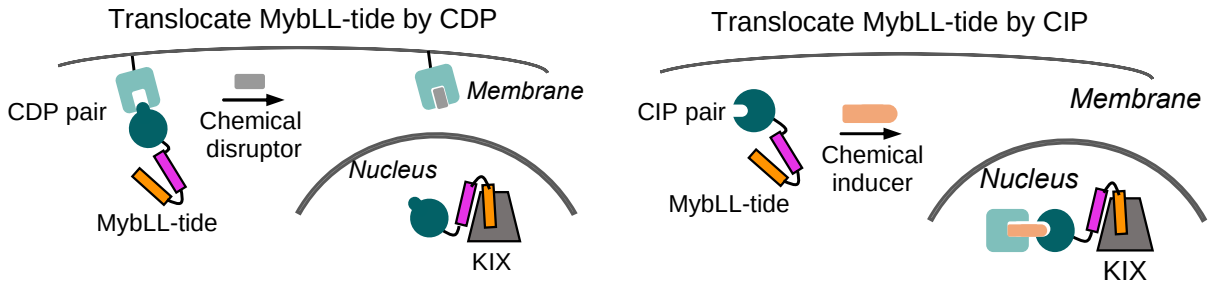
further reduce the background activity. 2) Since CBP/p300 primarily function in nucleus, Myb and MLL can be excluded from the nucleus to reduce background and translocated into nucleus upon stimulation (**Appendix Figure D-8**). To efficiently induce this translocation, the chemically induced trimerization system can be used[12]. 3) Use CIP to translocate the whole MybLL-tide into nucleus, or use CDP to delocalize MybLL-tide from cell membrane (**Appendix Figure 9**).



Appendix Figure D-7 Using CapN instead of FKBP for MybLL-plex. The use of CapN could potentially reduce background leakage.



Appendix Figure D-8 Using chemically induced trimerization to translocate Myb and MLL, as well as to reconstitute MybLL-plex. FKBP can be anchored in nucleus by using NLS.



Appendix Figure D-9 Using CIP and CDP to translocate MybLL-tide into nucleus.

Appendix D is partially adapted from the LSI Cubed proposal: Shen, J., Liu, Y. and Rodriguez, J. Development of an intracellular chemogenetic system to inhibit CBP/p300 KIX.

Dr. Stephen Joy, a previous postdoctoral researcher from the Mapp lab, University of Michigan, significantly contributed to the proposal writing and project conception.

Bibliography

1. Spencer, D.M., Wandless, T.J., Schreiber, S.L. & Crabtree, G.R. Controlling signal transduction with synthetic ligands. *Science* **262**, 1019-1024 (1993).
2. Liu, J., Farmer, J.D., Jr., Lane, W.S., Friedman, J., Weissman, I. & Schreiber, S.L. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* **66**, 807-815 (1991).
3. Clipstone, N.A. & Crabtree, G.R. Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation. *Nature* **357**, 695-697 (1992).
4. Michnick, S.W., Rosen, M.K., Wandless, T.J., Karplus, M. & Schreiber, S.L. Solution structure of FKBP, a rotamase enzyme and receptor for FK506 and rapamycin. *Science* **252**, 836-839 (1991).
5. Zheng, X.F., Florentino, D., Chen, J., Crabtree, G.R. & Schreiber, S.L. TOR kinase domains are required for two distinct functions, only one of which is inhibited by rapamycin. *Cell* **82**, 121-130 (1995).
6. Choi, J., Chen, J., Schreiber, S.L. & Clardy, J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **273**, 239-242 (1996).
7. Amara, J.F., Clackson, T., Rivera, V.M., Guo, T., Keenan, T., Natesan, S., Pollock, R., Yang, W., Courage, N.L., Holt, D.A. & Gilman, M. A versatile synthetic dimerizer for the regulation of protein-protein interactions. *Proc Natl Acad Sci U S A* **94**, 10618-10623 (1997).
8. Clackson, T., Yang, W., Rozamus, L.W., Hatada, M., Amara, J.F., Rollins, C.T., Stevenson, L.F., Magari, S.R., Wood, S.A., Courage, N.L., Lu, X., Cerasoli, F., Jr., Gilman, M. & Holt, D.A. Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc Natl Acad Sci U S A* **95**, 10437-10442 (1998).
9. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G. & Wandless, T.J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004 (2006).
10. Liberles, S.D., Diver, S.T., Austin, D.J. & Schreiber, S.L. Inducible gene expression and protein translocation using nontoxic ligands identified by a mammalian three-hybrid screen. *Proc Natl Acad Sci U S A* **94**, 7825-7830 (1997).
11. Bayle, J.H., Grimley, J.S., Stankunas, K., Gestwicki, J.E., Wandless, T.J. & Crabtree, G.R. Rapamycin analogs with differential binding specificity permit orthogonal control of protein activity. *Chem Biol* **13**, 99-107 (2006).
12. Wu, H.D., Kikuchi, M., Dagliyan, O., Aragaki, A.K., Nakamura, H., Dokholyan, N.V., Umehara, T. & Inoue, T. Rational design and implementation of a chemically inducible heterotrimerization system. *Nat Methods* **17**, 928-936 (2020).
13. Liang, F.S., Ho, W.Q. & Crabtree, G.R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci Signal* **4**, rs2 (2011).

14. Miyamoto, T., DeRose, R., Suarez, A., Ueno, T., Chen, M., Sun, T.P., Wolfgang, M.J., Mukherjee, C., Meyers, D.J. & Inoue, T. Rapid and orthogonal logic gating with a gibberellin-induced dimerization system. *Nat Chem Biol* **8**, 465-470 (2012).
15. Kopytek, S.J., Standaert, R.F., Dyer, J.C. & Hu, J.C. Chemically induced dimerization of dihydrofolate reductase by a homobifunctional dimer of methotrexate. *Chem Biol* **7**, 313-321 (2000).
16. Belshaw, P.J., Ho, S.N., Crabtree, G.R. & Schreiber, S.L. Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins. *Proc Natl Acad Sci U S A* **93**, 4604-4607 (1996).
17. Czlapinski, J.L., Schelle, M.W., Miller, L.W., Laughlin, S.T., Kohler, J.J., Cornish, V.W. & Bertozzi, C.R. Conditional glycosylation in eukaryotic cells using a biocompatible chemical inducer of dimerization. *J Am Chem Soc* **130**, 13186-13187 (2008).
18. Liu, P., Calderon, A., Konstantinidis, G., Hou, J., Voss, S., Chen, X., Li, F., Banerjee, S., Hoffmann, J.E., Theiss, C., Dehmelt, L. & Wu, Y.W. A bioorthogonal small-molecule-switch system for controlling protein function in live cells. *Angew Chem Int Ed Engl* **53**, 10049-10055 (2014).
19. Gutnick, A., Banghart, M.R., West, E.R. & Schwarz, T.L. The light-sensitive dimerizer zapalog reveals distinct modes of immobilization for axonal mitochondria. *Nat Cell Biol* **21**, 768-777 (2019).
20. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H. & Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **21**, 86-89 (2003).
21. Mollwitz, B., Brunk, E., Schmitt, S., Pojer, F., Bannwarth, M., Schiltz, M., Rothlisberger, U. & Johnsson, K. Directed evolution of the suicide protein O(6)-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* **51**, 986-994 (2012).
22. Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D.H., Bulleit, R.F. & Wood, K.V. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **3**, 373-382 (2008).
23. Gendreizig, S., Kindermann, M. & Johnsson, K. Induced protein dimerization in vivo through covalent labeling. *J Am Chem Soc* **125**, 14970-14971 (2003).
24. Ballister, E.R., Aonbangkhen, C., Mayo, A.M., Lampson, M.A. & Chenoweth, D.M. Localized light-induced protein dimerization in living cells using a photocaged dimerizer. *Nat Commun* **5**, 5475 (2014).
25. Zimmermann, M., Cal, R., Janett, E., Hoffmann, V., Bochet, C.G., Constable, E., Beaufils, F. & Wymann, M.P. Cell-permeant and photocleavable chemical inducer of dimerization. *Angew Chem Int Ed Engl* **53**, 4717-4720 (2014).
26. Hill, Z.B., Martinko, A.J., Nguyen, D.P. & Wells, J.A. Human antibody-based chemically induced dimerizers for cell therapeutic applications. *Nat Chem Biol* **14**, 112-117 (2018).
27. Guo, Z., Smutok, O., Johnston, W.A., Walden, P., Ungerer, J.P.J., Peat, T.S., Newman, J., Parker, J., Nebl, T., Hepburn, C., Melman, A., Suderman, R.J., Katz, E. & Alexandrov, K. Design of a methotrexate-controlled chemical dimerization system and its use in bio-electronic devices. *Nat Commun* **12**, 7137 (2021).

28. Kang, S., Davidsen, K., Gomez-Castillo, L., Jiang, H., Fu, X., Li, Z., Liang, Y., Jahn, M., Moussa, M., DiMaio, F. & Gu, L. COMBINES-CID: An Efficient Method for De Novo Engineering of Highly Specific Chemically Induced Protein Dimerization Systems. *J Am Chem Soc* **141**, 10948-10952 (2019).
29. Glasgow, A.A., Huang, Y.M., Mandell, D.J., Thompson, M., Ritterson, R., Loshbaugh, A.L., Pellegrino, J., Krivacic, C., Pache, R.A., Barlow, K.A., Ollikainen, N., Jeon, D., Kelly, M.J.S., Fraser, J.S. & Kortemme, T. Computational design of a modular protein sense-response system. *Science* **366**, 1024-1028 (2019).
30. Foight, G.W., Wang, Z., Wei, C.T., Jr Greisen, P., Warner, K.M., Cunningham-Bryant, D., Park, K., Brunette, T.J., Sheffler, W., Baker, D. & Maly, D.J. Multi-input chemical control of protein dimerization for programming graded cellular responses. *Nat Biotechnol* **37**, 1209-1216 (2019).
31. Rivera, V.M., Wang, X., Wardwell, S., Courage, N.L., Volchuk, A., Keenan, T., Holt, D.A., Gilman, M., Orci, L., Cerasoli, F., Jr., Rothman, J.E. & Clackson, T. Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. *Science* **287**, 826-830 (2000).
32. Barrero, J.J., Papanikou, E., Casler, J.C., Day, K.J. & Glick, B.S. An improved reversibly dimerizing mutant of the FK506-binding protein FKBP. *Cell Logist* **6**, e1204848 (2016).
33. Boncompain, G., Divoux, S., Gareil, N., de Forges, H., Lescure, A., Latreche, L., Mercanti, V., Jollivet, F., Raposo, G. & Perez, F. Synchronization of secretory protein traffic in populations of cells. *Nat Methods* **9**, 493-498 (2012).
34. Cunningham-Bryant, D., Dieter, E.M., Foight, G.W., Rose, J.C., Loutey, D.E. & Maly, D.J. in Journal of the American Chemical Society, Vol. 141 3352-3355 (2019).
35. Giordano-Attianese, G., Gainza, P., Gray-Gaillard, E., Cribioli, E., Shui, S., Kim, S., Kwak, M.J., Vollers, S., Corria Osorio, A.J., Reichenbach, P., Bonet, J., Oh, B.H., Irving, M., Coukos, G. & Correia, B.E. A computationally designed chimeric antigen receptor provides a small-molecule safety switch for T-cell therapy. *Nat Biotechnol* **38**, 426-432 (2020).
36. Karginov, A.V., Zou, Y., Shirvanyants, D., Kota, P., Dokholyan, N.V., Young, D.D., Hahn, K.M. & Deiters, A. Light regulation of protein dimerization and kinase activity in living cells using photocaged rapamycin and engineered FKBP. *J Am Chem Soc* **133**, 420-423 (2011).
37. Karginov, A.V., Ding, F., Kota, P., Dokholyan, N.V. & Hahn, K.M. Engineered allosteric activation of kinases in living cells. *Nat Biotechnol* **28**, 743-747 (2010).
38. Umeda, N., Ueno, T., Pohlmeyer, C., Nagano, T. & Inoue, T. A photocleavable rapamycin conjugate for spatiotemporal control of small GTPase activity. *J Am Chem Soc* **133**, 12-14 (2011).
39. Brown, K.A., Zou, Y., Shirvanyants, D., Zhang, J., Samanta, S., Mantravadi, P.K., Dokholyan, N.V. & Deiters, A. Light-cleavable rapamycin dimer as an optical trigger for protein dimerization. *Chem Commun (Camb)* **51**, 5702-5705 (2015).
40. Wright, C.W., Guo, Z.F. & Liang, F.S. Light control of cellular processes by using photocaged abscisic acid. *Chembiochem* **16**, 254-261 (2015).
41. Schelkle, K.M., Griesbaum, T., Ollech, D., Becht, S., Buckup, T., Hamburger, M. & Wombacher, R. Light-induced protein dimerization by one- and two-photon activation of gibberellic acid derivatives in living cells. *Angew Chem Int Ed Engl* **54**, 2825-2829 (2015).

42. Chang, D., Feng, S., Girik, V., Riezman, H. & Winssinger, N. Luciferase Controlled Protein Interactions. *J Am Chem Soc* **143**, 3665-3670 (2021).
43. Hiblot, J., Yu, Q., Sabbadini, M.D.B., Reymond, L., Xue, L., Schena, A., Sallin, O., Hill, N., Griss, R. & Johnsson, K. Luciferases with Tunable Emission Wavelengths. *Angew Chem Int Ed Engl* **56**, 14556-14560 (2017).
44. Banaszynski, L.A., Liu, C.W. & Wandless, T.J. Characterization of the FKBP.rapamycin.FRB ternary complex. *J Am Chem Soc* **127**, 4715-4721 (2005).
45. Haruki, H., Nishikawa, J. & Laemmli, U.K. The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol Cell* **31**, 925-932 (2008).
46. Lin, Y.C., Nihongaki, Y., Liu, T.Y., Razavi, S., Sato, M. & Inoue, T. Rapidly reversible manipulation of molecular activity with dual chemical dimerizers. *Angew Chem Int Ed Engl* **52**, 6450-6454 (2013).
47. Feng, S., Laketa, V., Stein, F., Rutkowska, A., MacNamara, A., Depner, S., Klingmuller, U., Saez-Rodriguez, J. & Schultz, C. A rapidly reversible chemical dimerizer system to study lipid signaling in living cells. *Angew Chem Int Ed Engl* **53**, 6720-6723 (2014).
48. Bottone, S., Joliot, O., Cakil, Z.V., El Hajji, L., Rakotoarison, L.M., Boncompain, G., Perez, F. & Gautier, A. A fluorogenic chemically induced dimerization technology for controlling, imaging and sensing protein proximity. *Nat Methods* (2023).
49. Komatsu, T., Kukelyansky, I., McCaffery, J.M., Ueno, T., Varela, L.C. & Inoue, T. Organelle-specific, rapid induction of molecular activities and membrane tethering. *Nat Methods* **7**, 206-208 (2010).
50. Nihongaki, Y., Matsubayashi, H.T. & Inoue, T. A molecular trap inside microtubules probes luminal access by soluble proteins. *Nat Chem Biol* **17**, 888-895 (2021).
51. Miyamoto, T., Rho, E., Sample, V., Akano, H., Magari, M., Ueno, T., Gorshkov, K., Chen, M., Tokumitsu, H., Zhang, J. & Inoue, T. Compartmentalized AMPK signaling illuminated by genetically encoded molecular sensors and actuators. *Cell Rep* **11**, 657-670 (2015).
52. Klemm, J.D., Beals, C.R. & Crabtree, G.R. Rapid targeting of nuclear proteins to the cytoplasm. *Curr Biol* **7**, 638-644 (1997).
53. Kohler, J.J. & Bertozzi, C.R. Regulating cell surface glycosylation by small molecule control of enzyme localization. *Chem Biol* **10**, 1303-1311 (2003).
54. Kohler, J.J., Czlapinski, J.L., Laughlin, S.T., Schelle, M.W., de Graffenried, C.L. & Bertozzi, C.R. Directing flux in glycan biosynthetic pathways with a small molecule switch. *Chembiochem* **5**, 1455-1458 (2004).
55. de Graffenried, C.L., Laughlin, S.T., Kohler, J.J. & Bertozzi, C.R. A small-molecule switch for Golgi sulfotransferases. *Proc Natl Acad Sci U S A* **101**, 16715-16720 (2004).
56. Ballister, E.R., Ayloo, S., Chenoweth, D.M., Lampson, M.A. & Holzbaur, E.L.F. Optogenetic control of organelle transport using a photocaged chemical inducer of dimerization. *Curr Biol* **25**, R407-R408 (2015).
57. Kapitein, L.C., Schlager, M.A., van der Zwan, W.A., Wulf, P.S., Keijzer, N. & Hoogenraad, C.C. Probing intracellular motor protein activity using an inducible cargo trafficking assay. *Biophys J* **99**, 2143-2152 (2010).
58. Karpova, A.Y., Tervo, D.G., Gray, N.W. & Svoboda, K. Rapid and reversible chemical inactivation of synaptic transmission in genetically targeted neurons. *Neuron* **48**, 727-735 (2005).
59. Nakamura, H., Lee, A.A., Afshar, A.S., Watanabe, S., Rho, E., Razavi, S., Suarez, A., Lin, Y.C., Tanigawa, M., Huang, B., DeRose, R., Bobb, D., Hong, W., Gabelli, S.B., Goutsias,

J. & Inoue, T. Intracellular production of hydrogels and synthetic RNA granules by multivalent molecular interactions. *Nat Mater* **17**, 79-89 (2018).
60. Miyazaki, Y., Mizumoto, K., Dey, G., Kudo, T., Perrino, J., Chen, L.C., Meyer, T. & Wandless, T.J. A method to rapidly create protein aggregates in living cells. *Nat Commun* **7**, 11689 (2016).
61. Belshaw, P.J., Spencer, D.M., Crabtree, G.R. & Schreiber, S.L. Controlling programmed cell death with a cyclophilin-cyclosporin-based chemical inducer of dimerization. *Chem Biol* **3**, 731-738 (1996).
62. Luik, R.M., Wang, B., Prakriya, M., Wu, M.M. & Lewis, R.S. Oligomerization of STIM1 couples ER calcium depletion to CRAC channel activation. *Nature* **454**, 538-542 (2008).
63. Luo, Z., Tzivion, G., Belshaw, P.J., Vavvas, D., Marshall, M. & Avruch, J. Oligomerization activates c-Raf-1 through a Ras-dependent mechanism. *Nature* **383**, 181-185 (1996).
64. MacCorkle, R.A., Freeman, K.W. & Spencer, D.M. Synthetic activation of caspases: artificial death switches. *Proc Natl Acad Sci U S A* **95**, 3655-3660 (1998).
65. Camacho-Soto, K., Castillo-Montoya, J., Tye, B. & Ghosh, I. Ligand-gated split-kinases. *J Am Chem Soc* **136**, 3995-4002 (2014).
66. Spencer, D.M., Graef, I., Austin, D.J., Schreiber, S.L. & Crabtree, G.R. A general strategy for producing conditional alleles of Src-like tyrosine kinases. *Proc Natl Acad Sci U S A* **92**, 9805-9809 (1995).
67. Holsinger, L.J., Spencer, D.M., Austin, D.J., Schreiber, S.L. & Crabtree, G.R. Signal transduction in T lymphocytes using a conditional allele of Sos. *Proc Natl Acad Sci U S A* **92**, 9810-9814 (1995).
68. Graef, I.A., Holsinger, L.J., Diver, S., Schreiber, S.L. & Crabtree, G.R. Proximity and orientation underlie signaling by the non-receptor tyrosine kinase ZAP70. *EMBO J* **16**, 5618-5628 (1997).
69. Suh, B.C., Inoue, T., Meyer, T. & Hille, B. Rapid chemically induced changes of PtdIns(4,5)P2 gate KCNQ ion channels. *Science* **314**, 1454-1457 (2006).
70. Burslem, G.M. & Crews, C.M. Small-Molecule Modulation of Protein Homeostasis. *Chem Rev* **117**, 11269-11301 (2017).
71. Zhao, L., Zhao, J., Zhong, K., Tong, A. & Jia, D. Targeted protein degradation: mechanisms, strategies and application. *Signal Transduct Target Ther* **7**, 113 (2022).
72. Liu, Z., Hu, M., Yang, Y., Du, C., Zhou, H., Liu, C., Chen, Y., Fan, L., Ma, H., Gong, Y. & Xie, Y. An overview of PROTACs: a promising drug discovery paradigm. *Mol Biomed* **3**, 46 (2022).
73. Takahashi, D., Moriyama, J., Nakamura, T., Miki, E., Takahashi, E., Sato, A., Akaike, T., Itto-Nakama, K. & Arimoto, H. AUTACs: Cargo-Specific Degraders Using Selective Autophagy. *Mol Cell* **76**, 797-810 e710 (2019).
74. Li, Z., Zhu, C., Ding, Y., Fei, Y. & Lu, B. ATTEC: a potential new approach to target proteinopathies. *Autophagy* **16**, 185-187 (2020).
75. Ji, C.H., Kim, H.Y., Lee, M.J., Heo, A.J., Park, D.Y., Lim, S., Shin, S., Ganipisetti, S., Yang, W.S., Jung, C.A., Kim, K.Y., Jeong, E.H., Park, S.H., Bin Kim, S., Lee, S.J., Na, J.E., Kang, J.I., Chi, H.M., Kim, H.T., Kim, Y.K., Kim, B.Y. & Kwon, Y.T. The AUTOTAC chemical biology platform for targeted protein degradation via the autophagy-lysosome system. *Nat Commun* **13**, 904 (2022).

76. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**, 917-922 (2009).
77. Morawska, M. & Ulrich, H.D. An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* **30**, 341-351 (2013).
78. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., Scott, T.G., Vittori, S., Perry, J.A., Qi, J., Winter, G.E., Wong, K.K., Gray, N.S. & Bradner, J.E. The dTAG system for immediate and target-specific protein degradation. *Nat Chem Biol* **14**, 431-441 (2018).
79. Buckley, D.L., Raina, K., Darricarrere, N., Hines, J., Gustafson, J.L., Smith, I.E., Miah, A.H., Harling, J.D. & Crews, C.M. HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem Biol* **10**, 1831-1837 (2015).
80. Etersque, J.M., Lee, I.K., Sharma, N., Xu, K., Ruff, A., Northrup, J.D., Sarkar, S., Nguyen, T., Lauman, R., Burslem, G.M. & Sellmyer, M.A. Regulation of eDHFR-tagged proteins with trimethoprim PROTACs. *Nat Commun* **14**, 7071 (2023).
81. Tomoshige, S., Naito, M., Hashimoto, Y. & Ishikawa, M. Degradation of HaloTag-fused nuclear proteins using bestatin-HaloTag ligand hybrid molecules. *Org Biomol Chem* **13**, 9746-9750 (2015).
82. Nabet, B., Ferguson, F.M., Seong, B.K.A., Kuljanin, M., Leggett, A.L., Mohardt, M.L., Robichaud, A., Conway, A.S., Buckley, D.L., Mancias, J.D., Bradner, J.E., Stegmaier, K. & Gray, N.S. Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nat Commun* **11**, 4687 (2020).
83. Mootz, H.D., Blum, E.S., Tyszkiewicz, A.B. & Muir, T.W. Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. *J Am Chem Soc* **125**, 10561-10569 (2003).
84. Mootz, H.D., Blum, E.S. & Muir, T.W. Activation of an autoregulated protein kinase by conditional protein splicing. *Angew Chem Int Ed Engl* **43**, 5189-5192 (2004).
85. Schwartz, E.C., Saez, L., Young, M.W. & Muir, T.W. Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat Chem Biol* **3**, 50-54 (2007).
86. Pratt, M.R., Schwartz, E.C. & Muir, T.W. Small-molecule-mediated rescue of protein function by an inducible proteolytic shunt. *Proc Natl Acad Sci U S A* **104**, 11209-11214 (2007).
87. Ho, S.N., Biggar, S.R., Spencer, D.M., Schreiber, S.L. & Crabtree, G.R. Dimeric ligands define a role for transcriptional activation domains in reinitiation. *Nature* **382**, 822-826 (1996).
88. Biggar, S.R. & Crabtree, G.R. Chemically regulated transcription factors reveal the persistence of repressor-resistant transcription after disrupting activator function. *J Biol Chem* **275**, 25381-25390 (2000).
89. Buhrlage, S.J., Brennan, B.B., Minter, A.R. & Mapp, A.K. Stereochemical promiscuity in artificial transcriptional activators. *J Am Chem Soc* **127**, 12456-12457 (2005).
90. Rowe, S.P., Casey, R.J., Brennan, B.B., Buhrlage, S.J. & Mapp, A.K. Transcriptional up-regulation in cells mediated by a small molecule. *J Am Chem Soc* **129**, 10654-10655 (2007).
91. Zetsche, B., Volz, S.E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33**, 139-142 (2015).

92. Gao, Y., Xiong, X., Wong, S., Charles, E.J., Lim, W.A. & Qi, L.S. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* **13**, 1043-1049 (2016).
93. Bao, Z., Jain, S., Jaroenpuntaruk, V. & Zhao, H. Orthogonal Genetic Regulation in Human Cells Using Chemically Induced CRISPR/Cas9 Activators. *ACS Synth Biol* **6**, 686-693 (2017).
94. Braun, S.M.G., Kirkland, J.G., Chory, E.J., Husmann, D., Calarco, J.P. & Crabtree, G.R. Rapid and reversible epigenome editing by endogenous chromatin regulators. *Nat Commun* **8**, 560 (2017).
95. Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.F., Hoffman, A.R., Huang, C.C., Pitteri, S.J. & Wang, K.C. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat Commun* **8**, 15993 (2017).
96. Jullien, N., Sampieri, F., Enjalbert, A. & Herman, J.P. Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res* **31**, e131 (2003).
97. Strader, C.D., Gaffney, T., Sugg, E.E., Candelore, M.R., Keys, R., Patchett, A.A. & Dixon, R.A. Allele-specific activation of genetically engineered receptors. *J Biol Chem* **266**, 5-8 (1991).
98. Coward, P., Wada, H.G., Falk, M.S., Chan, S.D., Meng, F., Akil, H. & Conklin, B.R. Controlling signaling with a specifically designed Gi-coupled receptor. *Proc Natl Acad Sci U S A* **95**, 352-357 (1998).
99. Urban, D.J. & Roth, B.L. DREADDs (designer receptors exclusively activated by designer drugs): chemogenetic tools with therapeutic utility. *Annu Rev Pharmacol Toxicol* **55**, 399-417 (2015).
100. Roth, B.L. DREADDs for Neuroscientists. *Neuron* **89**, 683-694 (2016).
101. Claes, M., De Groef, L. & Moons, L. The DREADDful Hurdles and Opportunities of the Chronic Chemogenetic Toolbox. *Cells* **11** (2022).
102. Miura, Y., Senoo, A., Doura, T. & Kiyonaka, S. Chemogenetics of cell surface receptors: beyond genetic and pharmacological approaches. *RSC Chem Biol* **3**, 269-287 (2022).
103. Armbruster, B.N., Li, X., Pausch, M.H., Herlitze, S. & Roth, B.L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc Natl Acad Sci U S A* **104**, 5163-5168 (2007).
104. Dong, S., Rogan, S.C. & Roth, B.L. Directed molecular evolution of DREADDs: a generic approach to creating next-generation RASSLs. *Nat Protoc* **5**, 561-573 (2010).
105. Alexander, G.M., Rogan, S.C., Abbas, A.I., Armbruster, B.N., Pei, Y., Allen, J.A., Nonneman, R.J., Hartmann, J., Moy, S.S., Nicolelis, M.A., McNamara, J.O. & Roth, B.L. Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. *Neuron* **63**, 27-39 (2009).
106. Zhu, H., Pleil, K.E., Urban, D.J., Moy, S.S., Kash, T.L. & Roth, B.L. Chemogenetic inactivation of ventral hippocampal glutamatergic neurons disrupts consolidation of contextual fear memory. *Neuropsychopharmacology* **39**, 1880-1892 (2014).
107. Guettier, J.M., Gautam, D., Scarselli, M., Ruiz de Azua, I., Li, J.H., Rosemond, E., Ma, X., Gonzalez, F.J., Armbruster, B.N., Lu, H., Roth, B.L. & Wess, J. A chemical-genetic approach to study G protein regulation of beta cell function in vivo. *Proc Natl Acad Sci U S A* **106**, 19197-19202 (2009).

108. Nakajima, K. & Wess, J. Design and functional characterization of a novel, arrestin-biased designer G protein-coupled receptor. *Mol Pharmacol* **82**, 575-582 (2012).
109. Vardy, E., Robinson, J.E., Li, C., Olsen, R.H.J., DiBerto, J.F., Giguere, P.M., Sassano, F.M., Huang, X.P., Zhu, H., Urban, D.J., White, K.L., Rittiner, J.E., Crowley, N.A., Pleil, K.E., Mazzone, C.M., Mosier, P.D., Song, J., Kash, T.L., Malanga, C.J., Krashes, M.J. & Roth, B.L. A New DREADD Facilitates the Multiplexed Chemogenetic Interrogation of Behavior. *Neuron* **86**, 936-946 (2015).
110. Slimko, E.M., McKinney, S., Anderson, D.J., Davidson, N. & Lester, H.A. Selective electrical silencing of mammalian neurons in vitro by the use of invertebrate ligand-gated chloride channels. *J Neurosci* **22**, 7373-7379 (2002).
111. Frazier, S.J., Cohen, B.N. & Lester, H.A. An engineered glutamate-gated chloride (GluCl) channel for sensitive, consistent neuronal silencing by ivermectin. *J Biol Chem* **288**, 21029-21042 (2013).
112. Lynagh, T. & Lynch, J.W. An improved ivermectin-activated chloride channel receptor for inhibiting electrical activity in defined neuronal populations. *J Biol Chem* **285**, 14890-14897 (2010).
113. Banghart, M., Borges, K., Isacoff, E., Trauner, D. & Kramer, R.H. Light-activated ion channels for remote control of neuronal firing. *Nat Neurosci* **7**, 1381-1386 (2004).
114. Berlin, S., Szobota, S., Reiner, A., Carroll, E.C., Kienzler, M.A., Guyon, A., Xiao, T., Trauner, D. & Isacoff, E.Y. A family of photoswitchable NMDA receptors. *Elife* **5** (2016).
115. Sandoz, G., Levitz, J., Kramer, R.H. & Isacoff, E.Y. Optical control of endogenous proteins with a photoswitchable conditional subunit reveals a role for TREK1 in GABA(B) signaling. *Neuron* **74**, 1005-1014 (2012).
116. Levitz, J., Pantoja, C., Gaub, B., Janovjak, H., Reiner, A., Hoagland, A., Schoppik, D., Kane, B., Stawski, P., Schier, A.F., Trauner, D. & Isacoff, E.Y. Optical control of metabotropic glutamate receptors. *Nat Neurosci* **16**, 507-516 (2013).
117. Lin, W.C., Tsai, M.C., Davenport, C.M., Smith, C.M., Veit, J., Wilson, N.M., Adesnik, H. & Kramer, R.H. A Comprehensive Optogenetic Pharmacology Toolkit for In Vivo Control of GABA(A) Receptors and Synaptic Inhibition. *Neuron* **88**, 879-891 (2015).
118. Broichhagen, J., Damijonaitis, A., Levitz, J., Sokol, K.R., Leippe, P., Konrad, D., Isacoff, E.Y. & Trauner, D. Orthogonal Optical Control of a G Protein-Coupled Receptor with a SNAP-Tethered Photochromic Ligand. *ACS Cent Sci* **1**, 383-393 (2015).
119. Levitz, J., Broichhagen, J., Leippe, P., Konrad, D., Trauner, D. & Isacoff, E.Y. Dual optical control and mechanistic insights into photoswitchable group II and III metabotropic glutamate receptors. *Proc Natl Acad Sci U S A* **114**, E3546-E3554 (2017).
120. Shields, B.C., Kahuno, E., Kim, C., Apostolides, P.F., Brown, J., Lindo, S., Mensh, B.D., Dudman, J.T., Lavis, L.D. & Tadross, M.R. Deconstructing behavioral neuropharmacology with cellular specificity. *Science* **356** (2017).
121. Donthamsetti, P.C., Broichhagen, J., Vyklicky, V., Stanley, C., Fu, Z., Visel, M., Levitz, J.L., Javitch, J.A., Trauner, D. & Isacoff, E.Y. Genetically Targeted Optical Control of an Endogenous G Protein-Coupled Receptor. *J Am Chem Soc* **141**, 11522-11530 (2019).
122. Donthamsetti, P., Winter, N., Hoagland, A., Stanley, C., Visel, M., Lammel, S., Trauner, D. & Isacoff, E. Cell specific photoswitchable agonist for reversible control of endogenous dopamine receptors. *Nat Commun* **12**, 4775 (2021).
123. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547-5551 (1992).

124. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. & Bujard, H. Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-1769 (1995).
125. Mansuy, I.M. & Bujard, H. Tetracycline-regulated gene expression in the brain. *Curr Opin Neurobiol* **10**, 593-596 (2000).
126. Das, A.T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr Gene Ther* **16**, 156-167 (2016).
127. Urlinger, S., Baron, U., Thellmann, M., Hasan, M.T., Bujard, H. & Hillen, W. Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci U S A* **97**, 7963-7968 (2000).
128. Yang, L., Yin, J., Wu, J., Qiao, L., Zhao, E.M., Cai, F. & Ye, H. Engineering genetic devices for in vivo control of therapeutic T cell activity triggered by the dietary molecule resveratrol. *Proc Natl Acad Sci U S A* **118** (2021).
129. Wang, Y., Liao, S., Guan, N., Liu, Y., Dong, K., Weber, W. & Ye, H. A versatile genetic control system in mammalian cells and mice responsive to clinically licensed sodium ferulate. *Sci Adv* **6**, eabb9484 (2020).
130. Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179-186 (1986).
131. Varshavsky, A. The N-end rule pathway of protein degradation. *Genes Cells* **2**, 13-28 (1997).
132. Levy, F., Johnston, J.A. & Varshavsky, A. Analysis of a conditional degradation signal in yeast and mammalian cells. *Eur J Biochem* **259**, 244-252 (1999).
133. Stankunas, K., Bayle, J.H., Gestwicki, J.E., Lin, Y.M., Wandless, T.J. & Crabtree, G.R. Conditional protein alleles using knockin mice and a chemical inducer of dimerization. *Mol Cell* **12**, 1615-1624 (2003).
134. Stankunas, K., Bayle, J.H., Havranek, J.J., Wandless, T.J., Baker, D., Crabtree, G.R. & Gestwicki, J.E. Rescue of degradation-prone mutants of the FK506-rapamycin binding (FRB) protein with chemical ligands. *Chembiochem* **8**, 1162-1169 (2007).
135. Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D. & Wandless, T.J. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem Biol* **17**, 981-988 (2010).
136. Miyazaki, Y., Imoto, H., Chen, L.C. & Wandless, T.J. Destabilizing domains derived from the human estrogen receptor. *J Am Chem Soc* **134**, 3942-3945 (2012).
137. Navarro, R., Chen, L.C., Rakhit, R. & Wandless, T.J. A Novel Destabilizing Domain Based on a Small-Molecule Dependent Fluorophore. *ACS Chem Biol* **11**, 2101-2104 (2016).
138. Weber, E.W., Parker, K.R., Sotillo, E., Lynn, R.C., Anbunathan, H., Lattin, J., Good, Z., Belk, J.A., Daniel, B., Klysz, D., Malipatlolla, M., Xu, P., Bashti, M., Heitzeneder, S., Labanieh, L., Vandris, P., Majzner, R.G., Qi, Y., Sandor, K., Chen, L.C., Prabhu, S., Gentles, A.J., Wandless, T.J., Satpathy, A.T., Chang, H.Y. & Mackall, C.L. Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling. *Science* **372** (2021).
139. Sando, R., 3rd, Baumgaertel, K., Pieraut, S., Torabi-Rander, N., Wandless, T.J., Mayford, M. & Maximov, A. Inducible control of gene expression with destabilized Cre. *Nat Methods* **10**, 1085-1088 (2013).
140. Miyamae, Y., Chen, L.C., Utsugi, Y., Farrants, H. & Wandless, T.J. A Method for Conditional Regulation of Protein Stability in Native or Near-Native Form. *Cell Chem Biol* **27**, 1573-1581 e1573 (2020).

141. Chung, H.K., Jacobs, C.L., Huo, Y., Yang, J., Krumm, S.A., Plemper, R.K., Tsien, R.Y. & Lin, M.Z. Tunable and reversible drug control of protein production via a self-excising degron. *Nat Chem Biol* **11**, 713-720 (2015).
142. Sahillioglu, A.C., Toebes, M., Apriamashvili, G., Gomez, R. & Schumacher, T.N. CRASH-IT Switch Enables Reversible and Dose-Dependent Control of TCR and CAR T-cell Function. *Cancer Immunol Res* **9**, 999-1007 (2021).
143. Neklesa, T.K., Tae, H.S., Schneekloth, A.R., Stulberg, M.J., Corson, T.W., Sundberg, T.B., Raina, K., Holley, S.A. & Crews, C.M. Small-molecule hydrophobic tagging-induced degradation of HaloTag fusion proteins. *Nat Chem Biol* **7**, 538-543 (2011).
144. Tae, H.S., Sundberg, T.B., Neklesa, T.K., Noblin, D.J., Gustafson, J.L., Roth, A.G., Raina, K. & Crews, C.M. Identification of hydrophobic tags for the degradation of stabilized proteins. *Chembiochem* **13**, 538-541 (2012).
145. Vishweshwaraiah, Y.L., Chen, J. & Dokholyan, N.V. Engineering an Allosteric Control of Protein Function. *J Phys Chem B* **125**, 1806-1814 (2021).
146. Fauser, J., Leschinsky, N., Szynal, B.N. & Karginov, A.V. Engineered Allosteric Regulation of Protein Function. *J Mol Biol* **434**, 167620 (2022).
147. Karginov, A.V., Tsygankov, D., Berginski, M., Chu, P.H., Trudeau, E.D., Yi, J.J., Gomez, S., Elston, T.C. & Hahn, K.M. Dissecting motility signaling through activation of specific Src-effector complexes. *Nat Chem Biol* **10**, 286-290 (2014).
148. Chu, P.H., Tsygankov, D., Berginski, M.E., Dagliyan, O., Gomez, S.M., Elston, T.C., Karginov, A.V. & Hahn, K.M. Engineered kinase activation reveals unique morphodynamic phenotypes and associated trafficking for Src family isoforms. *Proc Natl Acad Sci U S A* **111**, 12420-12425 (2014).
149. Klomp, J.E., Huyot, V., Ray, A.M., Collins, K.B., Malik, A.B. & Karginov, A.V. Mimicking transient activation of protein kinases in living cells. *Proc Natl Acad Sci U S A* **113**, 14976-14981 (2016).
150. Dagliyan, O., Shirvanyants, D., Karginov, A.V., Ding, F., Fee, L., Chandrasekaran, S.N., Freisinger, C.M., Smolen, G.A., Huttenlocher, A., Hahn, K.M. & Dokholyan, N.V. Rational design of a ligand-controlled protein conformational switch. *Proc Natl Acad Sci U S A* **110**, 6800-6804 (2013).
151. Iwakura, M. & Nakamura, T. Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase. *Protein Eng* **11**, 707-713 (1998).
152. Farrants, H., Tarnawski, M., Muller, T.G., Otsuka, S., Hiblot, J., Koch, B., Kueblbeck, M., Krausslich, H.G., Ellenberg, J. & Johnsson, K. Chemogenetic Control of Nanobodies. *Nat Methods* **17**, 279-282 (2020).
153. Goreshnik, I. & Maly, D.J. A small molecule-regulated guanine nucleotide exchange factor. *J Am Chem Soc* **132**, 938-940 (2010).
154. Rose, J.C., Huang, P.S., Camp, N.D., Ye, J., Leidal, A.M., Goreshnik, I., Trevillian, B.M., Dickinson, M.S., Cunningham-Bryant, D., Debnath, J., Baker, D., Wolf-Yadlin, A. & Maly, D.J. A computationally engineered RAS rheostat reveals RAS-ERK signaling dynamics. *Nat Chem Biol* **13**, 119-126 (2017).
155. Huang, P.S., Ban, Y.E., Richter, F., Andre, I., Vernon, R., Schief, W.R. & Baker, D. RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **6**, e24109 (2011).

156. Rose, J.C., Stephany, J.J., Valente, W.J., Trevillian, B.M., Dang, H.V., Bielas, J.H., Maly, D.J. & Fowler, D.M. Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nat Methods* **14**, 891-896 (2017).
157. Rose, J.C., Stephany, J.J., Wei, C.T., Fowler, D.M. & Maly, D.J. Rheostatic Control of Cas9-Mediated DNA Double Strand Break (DSB) Generation and Genome Editing. *ACS Chem Biol* **13**, 438-442 (2018).
158. Wei, C.T., Popp, N.A., Peleg, O., Powell, R.L., Borenstein, E., Maly, D.J. & Fowler, D.M. A chemically controlled Cas9 switch enables temporal modulation of diverse effectors. *Nat Chem Biol* (2023).
159. Bonger, K.M., Chen, L.C., Liu, C.W. & Wandless, T.J. Small-molecule displacement of a cryptic degron causes conditional protein degradation. *Nature Chemical Biology* **7**, 531-537 (2011).
160. Richman, S.A., Wang, L.C., Moon, E.K., Khire, U.R., Albelda, S.M. & Milone, M.C. Ligand-Induced Degradation of a CAR Permits Reversible Remote Control of CAR T Cell Activity In Vitro and In Vivo. *Mol Ther* **28**, 1600-1613 (2020).
161. Kolar, K., Knobloch, C., Stork, H., Znidaric, M. & Weber, W. OptoBase: A Web Platform for Molecular Optogenetics. *ACS Synth Biol* **7**, 1825-1828 (2018).
162. Repina, N.A., Rosenbloom, A., Mukherjee, A., Schaffer, D.V. & Kane, R.S. At Light Speed: Advances in Optogenetic Systems for Regulating Cell Signaling and Behavior. *Annu Rev Chem Biomol Eng* **8**, 13-39 (2017).
163. McCue, A.C. & Kuhlman, B. Design and engineering of light-sensitive protein switches. *Curr Opin Struct Biol* **74**, 102377 (2022).
164. Klewer, L. & Wu, Y.W. Light-Induced Dimerization Approaches to Control Cellular Processes. *Chemistry* **25**, 12452-12463 (2019).
165. Yamada, M., Nagasaki, S.C., Ozawa, T. & Imayoshi, I. Light-mediated control of Gene expression in mammalian cells. *Neurosci Res* **152**, 66-77 (2020).
166. Manoilov, K.Y., Verkhusha, V.V. & Shcherbakova, D.M. A guide to the optogenetic regulation of endogenous molecules. *Nat Methods* **18**, 1027-1037 (2021).
167. Zoltowski, B.D., Schwerdtfeger, C., Widom, J., Loros, J.J., Bilwes, A.M., Dunlap, J.C. & Crane, B.R. Conformational switching in the fungal light sensor Vivid. *Science* **316**, 1054-1057 (2007).
168. Schwerdtfeger, C. & Linden, H. VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation. *EMBO J* **22**, 4846-4855 (2003).
169. Zoltowski, B.D. & Crane, B.R. Light activation of the LOV protein vivid generates a rapidly exchanging dimer. *Biochemistry* **47**, 7012-7019 (2008).
170. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nature Communications* **6** (2015).
171. Benedetti, L., Marvin, J.S., Falahati, H., Guillen-Samander, A., Looger, L.L. & De Camilli, P. Optimized Vivid-derived Magnets photodimerizers for subcellular optogenetics in mammalian cells. *Elife* **9** (2020).
172. Bugaj, L.J., Choksi, A.T., Mesuda, C.K., Kane, R.S. & Schaffer, D.V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nature Methods* **10**, 249-252 (2013).
173. Wang, Q. & Lin, C. Mechanisms of Cryptochrome-Mediated Photoresponses in Plants. *Annu Rev Plant Biol* **71**, 103-129 (2020).

174. Liu, H., Yu, X., Li, K., Klejnot, J., Yang, H., Lisiero, D. & Lin, C. Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in Arabidopsis. *Science* **322**, 1535-1539 (2008).
175. Kennedy, M.J., Hughes, R.M., Peteya, L.A., Schwartz, J.W., Ehlers, M.D. & Tucker, C.L. Rapid blue-light-mediated induction of protein interactions in living cells. *Nature Methods* **7**, 973-975 (2010).
176. Taslimi, A., Zoltowski, B., Miranda, J.G., Pathak, G.P., Hughes, R.M. & Tucker, C.L. Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nature Chemical Biology* **12**, 425-430 (2016).
177. Mahmoudi, P., Veladi, H. & Pakdel, F.G. Optogenetics, Tools and Applications in Neurobiology. *J Med Signals Sens* **7**, 71-79 (2017).
178. Shimizu-Sato, S., Huq, E., Tepperman, J.M. & Quail, P.H. A light-switchable gene promoter system. *Nature Biotechnology* **20**, 1041-1044 (2002).
179. 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).
180. Muller, K., Engesser, R., Metzger, S., Schulz, S., Kampf, M.M., Busacker, M., Steinberg, T., Tomakidi, P., Ehrbar, M., Nagy, F., Timmer, J., Zubriggen, M.D. & Weber, W. A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. *Nucleic Acids Res* **41**, e77 (2013).
181. Noda, N. & Ozawa, T. Light-controllable Transcription System by Nucleocytoplasmic Shuttling of a Truncated Phytochrome B. *Photochem Photobiol* **94**, 1071-1076 (2018).
182. Hiltbrunner, A., Viczian, A., Bury, E., Tscheuschler, A., Kircher, S., Toth, R., Honsberger, A., Nagy, F., Fankhauser, C. & Schafer, E. Nuclear accumulation of the phytochrome A photoreceptor requires FHY1. *Curr Biol* **15**, 2125-2130 (2005).
183. Hiltbrunner, A., Tscheuschler, A., Viczian, A., Kunkel, T., Kircher, S. & Schafer, E. FHY1 and FHL act together to mediate nuclear accumulation of the phytochrome A photoreceptor. *Plant Cell Physiol* **47**, 1023-1034 (2006).
184. Tang, K., Beyer, H.M., Zurbriggen, M.D. & Gartner, W. The Red Edge: Bilin-Binding Photoreceptors as Optogenetic Tools and Fluorescence Reporters. *Chem Rev* **121**, 14906-14956 (2021).
185. Zhou, Y., Kong, D., Wang, X., Yu, G., Wu, X., Guan, N., Weber, W. & Ye, H. A small and highly sensitive red/far-red optogenetic switch for applications in mammals. *Nat Biotechnol* **40**, 262-272 (2022).
186. Kyriakakis, P., Catanho, M., Hoffner, N., Thavarajah, W., Hu, V.J., Chao, S.S., Hsu, A., Pham, V., Naghavian, L., Dozier, L.E., Patrick, G.N. & Coleman, T.P. Biosynthesis of Orthogonal Molecules Using Ferredoxin and Ferredoxin-NADP(+) Reductase Systems Enables Genetically Encoded PhyB Optogenetics. *ACS Synth Biol* **7**, 706-717 (2018).
187. Uda, Y., Goto, Y., Oda, S., Kohchi, T., Matsuda, M. & Aoki, K. Efficient synthesis of phycocyanobilin in mammalian cells for optogenetic control of cell signaling. *Proc Natl Acad Sci U S A* **114**, 11962-11967 (2017).
188. Kaberniuk, A.A., Shemetov, A.A. & Verkhusha, V.V. A bacterial phytochrome-based optogenetic system controllable with near-infrared light. *Nat Methods* **13**, 591-597 (2016).
189. Redchuk, T.A., Omelina, E.S., Chernov, K.G. & Verkhusha, V.V. Near-infrared optogenetic pair for protein regulation and spectral multiplexing. *Nature Chemical Biology* **13**, 633-639 (2017).

190. Redchuk, T.A., Karasev, M.M., Omelina, E.S. & Verkhusha, V.V. Near-Infrared Light-Controlled Gene Expression and Protein Targeting in Neurons and Non-neuronal Cells. *Chembiochem* **19**, 1334-1340 (2018).
191. Rizzini, L., Favory, J.J., Cloix, C., Faggionato, D., O'Hara, A., Kaiserli, E., Baumeister, R., Schafer, E., Nagy, F., Jenkins, G.I. & Ulm, R. Perception of UV-B by the Arabidopsis UVR8 protein. *Science* **332**, 103-106 (2011).
192. Favory, J.J., Stec, A., Gruber, H., Rizzini, L., Oravecz, A., Funk, M., Albert, A., Cloix, C., Jenkins, G.I., Oakeley, E.J., Seidlitz, H.K., Nagy, F. & Ulm, R. Interaction of COP1 and UVR8 regulates UV-B-induced photomorphogenesis and stress acclimation in Arabidopsis. *EMBO J* **28**, 591-601 (2009).
193. Wu, D., Hu, Q., Yan, Z., Chen, W., Yan, C., Huang, X., Zhang, J., Yang, P., Deng, H., Wang, J., Deng, X. & Shi, Y. Structural basis of ultraviolet-B perception by UVR8. *Nature* **484**, 214-219 (2012).
194. 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).
195. Konold, P.E., Mathes, T., Weienborn, J., Groot, M.L., Hegemann, P. & Kennis, J.T.M. Unfolding of the C-Terminal Jα Helix in the LOV2 Photoreceptor Domain Observed by Time-Resolved Vibrational Spectroscopy. *Journal of Physical Chemistry Letters* **7**, 3472-3476 (2016).
196. Peter, E., Dick, B. & Baeurle, S.A. Mechanism of signal transduction of the LOV2-Jα photosensor from Avena sativa. *Nature Communications* **1** (2010).
197. Wang, H., Vilela, M., Winkler, A., Tarnawski, M., Schlichting, I., Yumerefendi, H., Kuhlman, B., Liu, R., Danuser, G. & Hahn, K.M. LOVTRAP: An optogenetic system for photoinduced protein dissociation. *Nature Methods* **13**, 755-758 (2016).
198. Metz, S., Jager, A. & Klug, G. Role of a short light, oxygen, voltage (LOV) domain protein in blue light- and singlet oxygen-dependent gene regulation in Rhodobacter sphaeroides. *Microbiology (Reading)* **158**, 368-379 (2012).
199. 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).
200. Andresen, M., Stiel, A.C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M.C., Hell, S.W. & Jakobs, S. Structural basis for reversible photoswitching in Dronpa. *Proc Natl Acad Sci U S A* **104**, 13005-13009 (2007).
201. 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).
202. Crefcoeur, R.P., Yin, R., Ulm, R. & Halazonetis, T.D. Ultraviolet-B-mediated induction of protein-protein interactions in mammalian cells. *Nature Communications* **4** (2013).
203. Chen, D., Gibson, E.S. & Kennedy, M.J. A light-triggered protein secretion system. *Journal of Cell Biology* **201**, 631-640 (2013).
204. Duan, L., Che, D., Zhang, K., Ong, Q., Guo, S. & Cui, B. Optogenetic control of molecular motors and organelle distributions in cells. *Chemistry and Biology* **22**, 671-682 (2015).
205. Adrian, M., Nijenhuis, W., Hoogstraaten, R.I., Willems, J. & Kapitein, L.C. A Phytochrome-Derived Photoswitch for Intracellular Transport. *ACS Synth Biol* **6**, 1248-1256 (2017).
206. Shin, Y., Berry, J., Pannucci, N., Haataja, M.P., Toettcher, J.E. & Brangwynne, C.P. Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-171 e114 (2017).

207. Taslimi, A., Vrana, J.D., Chen, D., Borinskaya, S., Mayer, B.J., Kennedy, M.J. & Tucker, C.L. An optimized optogenetic clustering tool for probing protein interaction and function. *Nature Communications* **5** (2014).
208. Lee, S., Park, H., Kyung, T., Kim, N.Y., Kim, S., Kim, J. & Heo, W.D. Reversible protein inactivation by optogenetic trapping in cells. *Nat Methods* **11**, 633-636 (2014).
209. Bugaj, L.J., Spelke, D.P., Mesuda, C.K., Varedi, M., Kane, R.S. & Schaffer, D.V. Regulation of endogenous transmembrane receptors through optogenetic Cry2 clustering. *Nat Commun* **6**, 6898 (2015).
210. Endo, M., Hattori, M., Toriyabe, H., Ohno, H., Kamiguchi, H., Iino, Y. & Ozawa, T. Optogenetic activation of axon guidance receptors controls direction of neurite outgrowth. *Sci Rep* **6**, 23976 (2016).
211. Kyung, T., Lee, S., Kim, J.E., Cho, T., Park, H., Jeong, Y.M., Kim, D., Shin, A., Kim, S., Baek, J., Kim, J., Kim, N.Y., Woo, D., Chae, S., Kim, C.H., Shin, H.S., Han, Y.M., Kim, D. & Heo, W.D. Optogenetic control of endogenous Ca(2+) channels in vivo. *Nat Biotechnol* **33**, 1092-1096 (2015).
212. Chang, K.Y., Woo, D., Jung, H., Lee, S., Kim, S., Won, J., Kyung, T., Park, H., Kim, N., Yang, H.W., Park, J.Y., Hwang, E.M., Kim, D. & Heo, W.D. Light-inducible receptor tyrosine kinases that regulate neurotrophin signalling. *Nat Commun* **5**, 4057 (2014).
213. Liu, R., Yang, J., Yao, J., Zhao, Z., He, W., Su, N., Zhang, Z., Zhang, C., Zhang, Z., Cai, H., Zhu, L., Zhao, Y., Quan, S., Chen, X. & Yang, Y. Optogenetic control of RNA function and metabolism using engineered light-switchable RNA-binding proteins. *Nat Biotechnol* **40**, 779-786 (2022).
214. Kainrath, S., Stadler, M., Reichhart, E., Distel, M. & Janovjak, H. Green-Light-Induced Inactivation of Receptor Signaling Using Cobalamin-Binding Domains. *Angewandte Chemie - International Edition* **56**, 4608-4611 (2017).
215. Hongdusit, A., Liechty, E.T. & Fox, J.M. Analysis of Three Architectures for Controlling PTP1B with Light. *ACS Synth Biol* **11**, 61-68 (2022).
216. Mansouri, M., Hussherr, M.D., Strittmatter, T., Buchmann, P., Xue, S., Camenisch, G. & Fussenegger, M. Smart-watch-programmed green-light-operated percutaneous control of therapeutic transgenes. *Nat Commun* **12**, 3388 (2021).
217. Li, X., Zhang, C., Xu, X., Miao, J., Yao, J., Liu, R., Zhao, Y., Chen, X. & Yang, Y. A single-component light sensor system allows highly tunable and direct activation of gene expression in bacterial cells. *Nucleic Acids Res* **48**, e33 (2020).
218. Dietler, J., Schubert, R., Krafft, T.G.A., Meiler, S., Kainrath, S., Richter, F., Schweimer, K., Weyand, M., Janovjak, H. & Moglich, A. A Light-Oxygen-Voltage Receptor Integrates Light and Temperature. *J Mol Biol* **433**, 167107 (2021).
219. Wang, X., Chen, X. & Yang, Y. Spatiotemporal control of gene expression by a light-switchable transgene system. *Nature Methods* **9**, 266-269 (2012).
220. Liu, H., Gomez, G., Lin, S., Lin, S. & Lin, C. Optogenetic control of transcription in zebrafish. *PLoS One* **7**, e50738 (2012).
221. Yamada, M., Suzuki, Y., Nagasaki, S.C., Okuno, H. & Imayoshi, I. Light Control of the Tet Gene Expression System in Mammalian Cells. *Cell Rep* **25**, 487-500 e486 (2018).
222. Konermann, S., Brigham, M.D., Trevino, A.E., Hsu, P.D., Heidenreich, M., Cong, L., Platt, R.J., Scott, D.A., Church, G.M. & Zhang, F. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-476 (2013).

223. Ruess, J., Parise, F., Milias-Argeitis, A., Khammash, M. & Lygeros, J. Iterative experiment design guides the characterization of a light-inducible gene expression circuit. *Proc Natl Acad Sci U S A* **112**, 8148-8153 (2015).
224. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nature Biotechnology* **33**, 755-760 (2015).
225. Nihongaki, Y., Otabe, T., Ueda, Y. & Sato, M. A split CRISPR-Cpf1 platform for inducible genome editing and gene activation. *Nat Chem Biol* **15**, 882-888 (2019).
226. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. CRISPR-Cas9-based photoactivatable transcription system. *Chemistry and Biology* **22**, 169-174 (2015).
227. Polstein, L.R. & Gersbach, C.A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology* **11**, 198-200 (2015).
228. Nihongaki, Y., Furuhata, Y., Otabe, T., Hasegawa, S., Yoshimoto, K. & Sato, M. CRISPR-Cas9-based photoactivatable transcription systems to induce neuronal differentiation. *Nature Methods* **14**, 963-966 (2017).
229. Schindler, S.E., McCall, J.G., Yan, P., Hyrc, K.L., Li, M., Tucker, C.L., Lee, J.M., Bruchas, M.R. & Diamond, M.I. Photo-activatable Cre recombinase regulates gene expression in vivo. *Sci Rep* **5**, 13627 (2015).
230. 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).
231. Yao, S., Yuan, P., Ouellette, B., Zhou, T., Mortrud, M., Balaram, P., Chatterjee, S., Wang, Y., Daigle, T.L., Tasic, B., Kuang, X., Gong, H., Luo, Q., Zeng, S., Curtright, A., Dhaka, A., Kahan, A., Gradinaru, V., Chrapkiewicz, R., Schnitzer, M., Zeng, H. & Cetin, A. RecV recombinase system for in vivo targeted optogenomic modifications of single cells or cell populations. *Nat Methods* **17**, 422-429 (2020).
232. Jung, H., Kim, S.W., Kim, M., Hong, J., Yu, D., Kim, J.H., Lee, Y., Kim, S., Woo, D., Shin, H.S., Park, B.O. & Heo, W.D. Noninvasive optical activation of Flp recombinase for genetic manipulation in deep mouse brain regions. *Nat Commun* **10**, 314 (2019).
233. Rost, B.R., Schneider-Warme, F., Schmitz, D. & Hegemann, P. Optogenetic Tools for Subcellular Applications in Neuroscience. *Neuron* **96**, 572-603 (2017).
234. Tichy, A.M., Gerrard, E.J., Sexton, P.M. & Janovjak, H. Light-activated chimeric GPCRs: limitations and opportunities. *Curr Opin Struct Biol* **57**, 196-203 (2019).
235. Rozenberg, A., Inoue, K., Kandori, H. & Beja, O. Microbial Rhodopsins: The Last Two Decades. *Annu Rev Microbiol* **75**, 427-447 (2021).
236. Pedersen, N.P. & Gross, R.E. in Neuromodulation (Second Edition). (eds. E.S. Krames, P.H. Peckham & A.R. Rezai) 487-500 (Academic Press, 2018).
237. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nat Neurosci* **18**, 1213-1225 (2015).
238. Schneider, F., Grimm, C. & Hegemann, P. Biophysics of Channelrhodopsin. *Annu Rev Biophys* **44**, 167-186 (2015).
239. Zhou, X.E., Melcher, K. & Xu, H.E. Structure and activation of rhodopsin. *Acta Pharmacol Sin* **33**, 291-299 (2012).
240. Chow, B.Y., Han, X., Dobry, A.S., Qian, X., Chuong, A.S., Li, M., Henninger, M.A., Belfort, G.M., Lin, Y., Monahan, P.E. & Boyden, E.S. High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98-102 (2010).

241. Han, X., Chow, B.Y., Zhou, H., Klapoetke, N.C., Chuong, A., Rajimehr, R., Yang, A., Baratta, M.V., Winkle, J., Desimone, R. & Boyden, E.S. A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. *Front Syst Neurosci* **5**, 18 (2011).
242. Mattis, J., Tye, K.M., Ferenczi, E.A., Ramakrishnan, C., O'Shea, D.J., Prakash, R., Gunaydin, L.A., Hyun, M., Fenno, L.E., Gradinaru, V., Yizhar, O. & Deisseroth, K. Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. *Nat Methods* **9**, 159-172 (2011).
243. Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K. & Kandori, H. A light-driven sodium ion pump in marine bacteria. *Nat Commun* **4**, 1678 (2013).
244. Han, X. & Boyden, E.S. Multiple-color optical activation, silencing, and desynchronization of neural activity, with single-spike temporal resolution. *PLoS One* **2**, e299 (2007).
245. Gradinaru, V., Thompson, K.R., Zhang, F., Mogri, M., Kay, K., Schneider, M.B. & Deisseroth, K. Targeting and readout strategies for fast optical neural control in vitro and in vivo. *J Neurosci* **27**, 14231-14238 (2007).
246. Gradinaru, V., Thompson, K.R. & Deisseroth, K. eNpHR: a Natronomonas halorhodopsin enhanced for optogenetic applications. *Brain Cell Biol* **36**, 129-139 (2008).
247. Gradinaru, V., Zhang, F., Ramakrishnan, C., Mattis, J., Prakash, R., Diester, I., Goshen, I., Thompson, K.R. & Deisseroth, K. Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154-165 (2010).
248. Nagel, G., Ollig, D., Fuhrmann, M., Kateriya, S., Musti, A.M., Bamberg, E. & Hegemann, P. Channelrhodopsin-1: a light-gated proton channel in green algae. *Science* **296**, 2395-2398 (2002).
249. Nagel, G., Szellas, T., Huhn, W., Kateriya, S., Adeishvili, N., Berthold, P., Ollig, D., Hegemann, P. & Bamberg, E. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc Natl Acad Sci U S A* **100**, 13940-13945 (2003).
250. Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* **8**, 1263-1268 (2005).
251. Ishizuka, T., Kakuda, M., Araki, R. & Yawo, H. Kinetic evaluation of photosensitivity in genetically engineered neurons expressing green algae light-gated channels. *Neurosci Res* **54**, 85-94 (2006).
252. Li, X., Gutierrez, D.V., Hanson, M.G., Han, J., Mark, M.D., Chiel, H., Hegemann, P., Landmesser, L.T. & Herlitze, S. Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin. *Proc Natl Acad Sci U S A* **102**, 17816-17821 (2005).
253. Nagel, G., Brauner, M., Liewald, J.F., Adeishvili, N., Bamberg, E. & Gottschalk, A. Light activation of channelrhodopsin-2 in excitable cells of Caenorhabditis elegans triggers rapid behavioral responses. *Curr Biol* **15**, 2279-2284 (2005).
254. Berndt, A., Schoenenberger, P., Mattis, J., Tye, K.M., Deisseroth, K., Hegemann, P. & Oertner, T.G. High-efficiency channelrhodopsins for fast neuronal stimulation at low light levels. *Proc Natl Acad Sci U S A* **108**, 7595-7600 (2011).
255. Dawydow, A., Gueta, R., Ljaschenko, D., Ullrich, S., Hermann, M., Ehmann, N., Gao, S., Fiala, A., Langenhan, T., Nagel, G. & Kittel, R.J. Channelrhodopsin-2-XXL, a powerful optogenetic tool for low-light applications. *Proc Natl Acad Sci U S A* **111**, 13972-13977 (2014).

256. Berndt, A., Yizhar, O., Gunaydin, L.A., Hegemann, P. & Deisseroth, K. Bi-stable neural state switches. *Nat Neurosci* **12**, 229-234 (2009).
257. Bamann, C., Gueta, R., Kleinlogel, S., Nagel, G. & Bamberg, E. Structural guidance of the photocycle of channelrhodopsin-2 by an interhelical hydrogen bond. *Biochemistry* **49**, 267-278 (2010).
258. Yizhar, O., Fenno, L.E., Prigge, M., Schneider, F., Davidson, T.J., O'Shea, D.J., Sohal, V.S., Goshen, I., Finkelstein, J., Paz, J.T., Stehfest, K., Fudim, R., Ramakrishnan, C., Huguenard, J.R., Hegemann, P. & Deisseroth, K. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477**, 171-178 (2011).
259. Gunaydin, L.A., Yizhar, O., Berndt, A., Sohal, V.S., Deisseroth, K. & Hegemann, P. Ultrafast optogenetic control. *Nat Neurosci* **13**, 387-392 (2010).
260. Klapoetke, N.C., Murata, Y., Kim, S.S., Pulver, S.R., Birdsey-Benson, A., Cho, Y.K., Morimoto, T.K., Chuong, A.S., Carpenter, E.J., Tian, Z., Wang, J., Xie, Y., Yan, Z., Zhang, Y., Chow, B.Y., Surek, B., Melkonian, M., Jayaraman, V., Constantine-Paton, M., Wong, G.K. & Boyden, E.S. Independent optical excitation of distinct neural populations. *Nat Methods* **11**, 338-346 (2014).
261. Govorunova, E.G., Sineshchekov, O.A., Li, H., Janz, R. & Spudich, J.L. Characterization of a highly efficient blue-shifted channelrhodopsin from the marine alga Platymonas subcordiformis. *J Biol Chem* **288**, 29911-29922 (2013).
262. Zhang, F., Prigge, M., Beyriere, F., Tsunoda, S.P., Mattis, J., Yizhar, O., Hegemann, P. & Deisseroth, K. Red-shifted optogenetic excitation: a tool for fast neural control derived from Volvox carteri. *Nat Neurosci* **11**, 631-633 (2008).
263. Lin, J.Y., Knutsen, P.M., Muller, A., Kleinfeld, D. & Tsien, R.Y. ReaChR: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nat Neurosci* **16**, 1499-1508 (2013).
264. Prigge, M., Schneider, F., Tsunoda, S.P., Shilyansky, C., Wietek, J., Deisseroth, K. & Hegemann, P. Color-tuned channelrhodopsins for multiwavelength optogenetics. *J Biol Chem* **287**, 31804-31812 (2012).
265. Kleinlogel, S., Feldbauer, K., Dempski, R.E., Fotis, H., Wood, P.G., Bamann, C. & Bamberg, E. Ultra light-sensitive and fast neuronal activation with the Ca(2)+-permeable channelrhodopsin CatCh. *Nat Neurosci* **14**, 513-518 (2011).
266. Kato, H.E., Zhang, F., Yizhar, O., Ramakrishnan, C., Nishizawa, T., Hirata, K., Ito, J., Aita, Y., Tsukazaki, T., Hayashi, S., Hegemann, P., Maturana, A.D., Ishitani, R., Deisseroth, K. & Nureki, O. Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **482**, 369-374 (2012).
267. Wietek, J., Wiegert, J.S., Adeishvili, N., Schneider, F., Watanabe, H., Tsunoda, S.P., Vogt, A., Elstner, M., Oertner, T.G. & Hegemann, P. Conversion of channelrhodopsin into a light-gated chloride channel. *Science* **344**, 409-412 (2014).
268. Berndt, A., Lee, S.Y., Ramakrishnan, C. & Deisseroth, K. Structure-guided transformation of channelrhodopsin into a light-activated chloride channel. *Science* **344**, 420-424 (2014).
269. Wietek, J., Beltramo, R., Scanziani, M., Hegemann, P., Oertner, T.G. & Wiegert, J.S. An improved chloride-conducting channelrhodopsin for light-induced inhibition of neuronal activity in vivo. *Sci Rep* **5**, 14807 (2015).
270. Berndt, A., Lee, S.Y., Wietek, J., Ramakrishnan, C., Steinberg, E.E., Rashid, A.J., Kim, H., Park, S., Santoro, A., Frankland, P.W., Iyer, S.M., Pak, S., Ahrlund-Richter, S., Delp, S.L., Malenka, R.C., Josselyn, S.A., Carlen, M., Hegemann, P. & Deisseroth, K. Structural

foundations of optogenetics: Determinants of channelrhodopsin ion selectivity. *Proc Natl Acad Sci U S A* **113**, 822-829 (2016).
271. Govorunova, E.G., Sineshchekov, O.A., Janz, R., Liu, X. & Spudich, J.L. NEUROSCIENCE. Natural light-gated anion channels: A family of microbial rhodopsins for advanced optogenetics. *Science* **349**, 647-650 (2015).
272. Govorunova, E.G., Sineshchekov, O.A. & Spudich, J.L. Proteomonas sulcata ACR1: A Fast Anion Channelrhodopsin. *Photochem Photobiol* **92**, 257-263 (2016).
273. Wietek, J., Broser, M., Krause, B.S. & Hegemann, P. Identification of a Natural Green Light Absorbing Chloride Conducting Channelrhodopsin from Proteomonas sulcata. *J Biol Chem* **291**, 4121-4127 (2016).
274. Berglund, K., Birkner, E., Augustine, G.J. & Hochgeschwender, U. Light-emitting channelrhodopsins for combined optogenetic and chemical-genetic control of neurons. *PLoS One* **8**, e59759 (2013).
275. Berglund, K., Clissold, K., Li, H.E., Wen, L., Park, S.Y., Gleixner, J., Klein, M.E., Lu, D., Barter, J.W., Rossi, M.A., Augustine, G.J., Yin, H.H. & Hochgeschwender, U. Luminopsins integrate opto- and chemogenetics by using physical and biological light sources for opsin activation. *Proc Natl Acad Sci U S A* **113**, E358-367 (2016).
276. Tung, J.K., Gutekunst, C.A. & Gross, R.E. Inhibitory luminopsins: genetically-encoded bioluminescent opsins for versatile, scalable, and hardware-independent optogenetic inhibition. *Sci Rep* **5**, 14366 (2015).
277. Cosentino, C., Alberio, L., Gazzarrini, S., Aquila, M., Romano, E., Cermenati, S., Zuccolini, P., Petersen, J., Beltrame, M., Van Etten, J.L., Christie, J.M., Thiel, G. & Moroni, A. Engineering of a light-gated potassium channel. *Science* **348**, 707-710 (2015).
278. Plugge, B., Gazzarrini, S., Nelson, M., Cerana, R., Van Etten, J.L., Derst, C., DiFrancesco, D., Moroni, A. & Thiel, G. A potassium channel protein encoded by chlorella virus PBCV-1. *Science* **287**, 1641-1644 (2000).
279. Alberio, L., Locarno, A., Saponaro, A., Romano, E., Bercier, V., Albadri, S., Simeoni, F., Moleri, S., Pelucchi, S., Porro, A., Marcello, E., Barsotti, N., Kukovetz, K., Boender, A.J., Contestabile, A., Luo, S., Moutal, A., Ji, Y., Romani, G., Beltrame, M., Del Bene, F., Di Luca, M., Khanna, R., Colecraft, H.M., Pasqualetti, M., Thiel, G., Tonini, R. & Moroni, A. A light-gated potassium channel for sustained neuronal inhibition. *Nat Methods* **15**, 969-976 (2018).
280. Karunarathne, W.K., Giri, L., Kalyanaraman, V. & Gautam, N. Optically triggering spatiotemporally confined GPCR activity in a cell and programming neurite initiation and extension. *Proc Natl Acad Sci U S A* **110**, E1565-1574 (2013).
281. Bailes, H.J., Zhuang, L.Y. & Lucas, R.J. Reproducible and sustained regulation of Galphas signalling using a metazoan opsin as an optogenetic tool. *PLoS One* **7**, e30774 (2012).
282. Spoida, K., Eickelbeck, D., Karapinar, R., Eckhardt, T., Mark, M.D., Jancke, D., Ehinger, B.V., Konig, P., Dalkara, D., Herlitze, S. & Masseck, O.A. Melanopsin Variants as Intrinsic Optogenetic On and Off Switches for Transient versus Sustained Activation of G Protein Pathways. *Curr Biol* **26**, 1206-1212 (2016).
283. Kim, J.M., Hwa, J., Garriga, P., Reeves, P.J., RajBhandary, U.L. & Khorana, H.G. Light-driven activation of beta 2-adrenergic receptor signaling by a chimeric rhodopsin containing the beta 2-adrenergic receptor cytoplasmic loops. *Biochemistry* **44**, 2284-2292 (2005).

284. Airan, R.D., Thompson, K.R., Fenno, L.E., Bernstein, H. & Deisseroth, K. Temporally precise in vivo control of intracellular signalling. *Nature* **458**, 1025-1029 (2009).
285. Oh, E., Maejima, T., Liu, C., Deneris, E. & Herlitze, S. Substitution of 5-HT1A receptor signaling by a light-activated G protein-coupled receptor. *J Biol Chem* **285**, 30825-30836 (2010).
286. Siuda, E.R., Copits, B.A., Schmidt, M.J., Baird, M.A., Al-Hasani, R., Planer, W.J., Funderburk, S.C., McCall, J.G., Gereau, R.W.t. & Bruchas, M.R. Spatiotemporal control of opioid signaling and behavior. *Neuron* **86**, 923-935 (2015).
287. Barish, P.A., Xu, Y., Li, J., Sun, J., Jarajapu, Y.P. & Ogle, W.O. Design and functional evaluation of an optically active mu-opioid receptor. *Eur J Pharmacol* **705**, 42-48 (2013).
288. Gunaydin, L.A., Grosenick, L., Finkelstein, J.C., Kauvar, I.V., Fenno, L.E., Adhikari, A., Lammel, S., Mirzabekov, J.J., Airan, R.D., Zalocusky, K.A., Tye, K.M., Anikeeva, P., Malenka, R.C. & Deisseroth, K. Natural neural projection dynamics underlying social behavior. *Cell* **157**, 1535-1551 (2014).
289. Morri, M., Sanchez-Romero, I., Tichy, A.M., Kainrath, S., Gerrard, E.J., Hirschfeld, P.P., Schwarz, J. & Janovjak, H. Optical functionalization of human Class A orphan G-protein-coupled receptors. *Nat Commun* **9**, 1950 (2018).
290. Xu, Y., Hyun, Y.M., Lim, K., Lee, H., Cummings, R.J., Gerber, S.A., Bae, S., Cho, T.Y., Lord, E.M. & Kim, M. Optogenetic control of chemokine receptor signal and T-cell migration. *Proc Natl Acad Sci U S A* **111**, 6371-6376 (2014).
291. Li, P., Rial, D., Canas, P.M., Yoo, J.H., Li, W., Zhou, X., Wang, Y., van Westen, G.J., Payen, M.P., Augusto, E., Goncalves, N., Tome, A.R., Li, Z., Wu, Z., Hou, X., Zhou, Y., AP, I.J., Boyden, E.S., Cunha, R.A., Qu, J. & Chen, J.F. Optogenetic activation of intracellular adenosine A2A receptor signaling in the hippocampus is sufficient to trigger CREB phosphorylation and impair memory. *Mol Psychiatry* **20**, 1339-1349 (2015).
292. van Wyk, M., Pielecka-Fortuna, J., Lowel, S. & Kleinlogel, S. Restoring the ON Switch in Blind Retinas: Opto-mGluR6, a Next-Generation, Cell-Tailored Optogenetic Tool. *PLoS Biol* **13**, e1002143 (2015).
293. McGregor, K.M., Becamel, C., Marin, P. & Andrade, R. Using melanopsin to study G protein signaling in cortical neurons. *J Neurophysiol* **116**, 1082-1092 (2016).
294. Tichy, A.M., So, W.L., Gerrard, E.J. & Janovjak, H. Structure-guided optimization of light-activated chimeric G-protein-coupled receptors. *Structure* **30**, 1075-1087 e1074 (2022).
295. Kapolka, N.J. & Roth, B.L. Structural insights from G-protein-coupled receptor complexes enable the rational engineering of improved light-activated designer receptors. *Structure* **30**, 1043-1045 (2022).
296. Schmidt, D., Tillberg, P.W., Chen, F. & Boyden, E.S. A fully genetically encoded protein architecture for optical control of peptide ligand concentration. *Nat Commun* **5**, 3019 (2014).
297. Hoffmann, M.D., Mathony, J., Upmeier Zu Belzen, J., Harteveld, Z., Aschenbrenner, S., Stengl, C., Grimm, D., Correia, B.E., Eils, R. & Niopek, D. Optogenetic control of Neisseria meningitidis Cas9 genome editing using an engineered, light-switchable anti-CRISPR protein. *Nucleic Acids Res* **49**, e29 (2021).
298. Gil, A.A., Carrasco-Lopez, C., Zhu, L., Zhao, E.M., Ravindran, P.T., Wilson, M.Z., Goglia, A.G., Avalos, J.L. & Toettcher, J.E. Optogenetic control of protein binding using light-switchable nanobodies. *Nat Commun* **11**, 4044 (2020).

299. Carrasco-Lopez, C., Zhao, E.M., Gil, A.A., Alam, N., Toettcher, J.E. & Avalos, J.L. Development of light-responsive protein binding in the monobody non-immunoglobulin scaffold. *Nat Commun* **11**, 4045 (2020).
300. Dagliyan, O., Tarnawski, M., Chu, P.H., Shirvanyants, D., Schlichting, I., Dokholyan, N.V. & Hahn, K.M. Engineering extrinsic disorder to control protein activity in living cells. *Science* **354**, 1441-1444 (2016).
301. Gehrig, S., Macpherson, J.A., Driscoll, P.C., Symon, A., Martin, S.R., MacRae, J.I., Kleinjung, J., Fraternali, F. & Anastasiou, D. An engineered photoswitchable mammalian pyruvate kinase. *FEBS J* **284**, 2955-2980 (2017).
302. Lee, J., Natarajan, M., Nashine, V.C., Socolich, M., Vo, T., Russ, W.P., Benkovic, S.J. & Ranganathan, R. Surface sites for engineering allosteric control in proteins. *Science* **322**, 438-442 (2008).
303. Lee, S.Y., Cheah, J.S., Zhao, B., Xu, C., Roh, H., Kim, C.K., Cho, K.F., Udeshi, N.D., Carr, S.A. & Ting, A.Y. Engineered allostery in light-regulated LOV-Turbo enables precise spatiotemporal control of proximity labeling in living cells. *Nat Methods* **20**, 908-917 (2023).
304. Shaaya, M., Fauser, J., Zhurikhina, A., Conage-Pough, J.E., Huyot, V., Brennan, M., Flower, C.T., Matsche, J., Khan, S., Natarajan, V., Rehman, J., Kota, P., White, F.M., Tsygankov, D. & Karginov, A.V. Light-regulated allosteric switch enables temporal and subcellular control of enzyme activity. *Elife* **9** (2020).
305. Wu, Y.I., Frey, D., Lungu, O.I., Jaehrig, A., Schlichting, I., Kuhlman, B. & Hahn, K.M. A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461**, 104-108 (2009).
306. Winkler, A., Barends, T.R., Udvarhelyi, A., Lenherr-Frey, D., Lomb, L., Menzel, A. & Schlichting, I. Structural details of light activation of the LOV2-based photoswitch PA-Rac1. *ACS Chem Biol* **10**, 502-509 (2015).
307. Wang, X., He, L., Wu, Y.I., Hahn, K.M. & Montell, D.J. Light-mediated activation reveals a key role for Rac in collective guidance of cell movement in vivo. *Nat Cell Biol* **12**, 591-597 (2010).
308. Yoo, S.K., Deng, Q., Cavnar, P.J., Wu, Y.I., Hahn, K.M. & Huttenlocher, A. Differential regulation of protrusion and polarity by PI3K during neutrophil motility in live zebrafish. *Dev Cell* **18**, 226-236 (2010).
309. Hayashi-Takagi, A., Yagishita, S., Nakamura, M., Shirai, F., Wu, Y.I., Loshbaugh, A.L., Kuhlman, B., Hahn, K.M. & Kasai, H. Labelling and optical erasure of synaptic memory traces in the motor cortex. *Nature* **525**, 333-338 (2015).
310. Dietz, D.M., Sun, H., Lobo, M.K., Cahill, M.E., Chadwick, B., Gao, V., Koo, J.W., Mazei-Robison, M.S., Dias, C., Maze, I., Damez-Werno, D., Dietz, K.C., Scobie, K.N., Ferguson, D., Christoffel, D., Ohnishi, Y., Hodes, G.E., Zheng, Y., Neve, R.L., Hahn, K.M., Russo, S.J. & Nestler, E.J. Rac1 is essential in cocaine-induced structural plasticity of nucleus accumbens neurons. *Nat Neurosci* **15**, 891-896 (2012).
311. Strickland, D., Moffat, K. & Sosnick, T.R. Light-activated DNA binding in a designed allosteric protein. *Proc Natl Acad Sci U S A* **105**, 10709-10714 (2008).
312. Mills, E., Chen, X., Pham, E., Wong, S. & Truong, K. Engineering a photoactivated caspase-7 for rapid induction of apoptosis. *ACS Synth Biol* **1**, 75-82 (2012).
313. Hongdusit, A., Zwart, P.H., Sankaran, B. & Fox, J.M. Minimally disruptive optical control of protein tyrosine phosphatase 1B. *Nat Commun* **11**, 788 (2020).

314. Hongdusit, A. & Fox, J.M. Optogenetic Analysis of Allosteric Control in Protein Tyrosine Phosphatases. *Biochemistry* **60**, 254-258 (2021).
315. Duplus-Bottin, H., Spichty, M., Triqueneaux, G., Place, C., Mangeot, P.E., Ohlmann, T., Vittoz, F. & Yvert, G. A single-chain and fast-responding light-inducible Cre recombinase as a novel optogenetic switch. *Elife* **10** (2021).
316. Ma, G., He, L., Liu, S., Xie, J., Huang, Z., Jing, J., Lee, Y.T., Wang, R., Luo, H., Han, W., Huang, Y. & Zhou, Y. Optogenetic engineering to probe the molecular choreography of STIM1-mediated cell signaling. *Nat Commun* **11**, 1039 (2020).
317. Paonessa, F., Criscuolo, S., Sacchetti, S., Amoroso, D., Scarongella, H., Pecoraro Bisogni, F., Carminati, E., Pruzzo, G., Maragliano, L., Cesca, F. & Benfenati, F. Regulation of neural gene transcription by optogenetic inhibition of the RE1-silencing transcription factor. *Proc Natl Acad Sci U S A* **113**, E91-100 (2016).
318. Stone, O.J., Pankow, N., Liu, B., Sharma, V.P., Eddy, R.J., Wang, H., Putz, A.T., Teets, F.D., Kuhlman, B., Condeelis, J.S. & Hahn, K.M. Optogenetic control of cofilin and alphaTAT in living cells using Z-lock. *Nat Chem Biol* **15**, 1183-1190 (2019).
319. Ju, J., Lee, H.N., Ning, L., Ryu, H., Zhou, X.X., Chun, H., Lee, Y.W., Lee-Richerson, A.I., Jeong, C., Lin, M.Z. & Seong, J. Optical regulation of endogenous RhoA reveals selection of cellular responses by signal amplitude. *Cell Rep* **40**, 111080 (2022).
320. Zhou, X.X., Zou, X., Chung, H.K., Gao, Y., Liu, Y., Qi, L.S. & Lin, M.Z. A Single-Chain Photoswitchable CRISPR-Cas9 Architecture for Light-Inducible Gene Editing and Transcription. *ACS Chemical Biology* **13**, 443-448 (2018).
321. Jones, T.t., Liu, A. & Cui, B. Light-Inducible Generation of Membrane Curvature in Live Cells with Engineered BAR Domain Proteins. *ACS Synth Biol* **9**, 893-901 (2020).
322. Strickland, D., Lin, Y., Wagner, E., Hope, C.M., Zayner, J., Antoniou, C., Sosnick, T.R., Weiss, E.L. & Glotzer, M. TULIPs: Tunable, light-controlled interacting protein tags for cell biology. *Nature Methods* **9**, 379-384 (2012).
323. Lee, H.J. & Zheng, J.J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal* **8**, 8 (2010).
324. Lungu, O.I., Hallett, R.A., Choi, E.J., Aiken, M.J., Hahn, K.M. & Kuhlman, B. Designing Photoswitchable Peptides Using the AsLOV2 Domain. *Chemistry and Biology* **19**, 507-517 (2012).
325. Guntas, G., Hallett, R.A., Zimmerman, S.P., Williams, T., Yumerefendi, H., Bear, J.E. & Kuhlman, B. Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 112-117 (2015).
326. Van Bergeijk, P., Adrian, M., Hoogenraad, C.C. & Kapitein, L.C. Optogenetic control of organelle transport and positioning. *Nature* **518**, 111-114 (2015).
327. Yu, D., Lee, H., Hong, J., Jung, H., Jo, Y., Oh, B.H., Park, B.O. & Heo, W.D. Optogenetic activation of intracellular antibodies for direct modulation of endogenous proteins. *Nat Methods* **16**, 1095-1100 (2019).
328. Bracha, D., Walls, M.T., Wei, M.T., Zhu, L., Kurian, M., Avalos, J.L., Toettcher, J.E. & Brangwynne, C.P. Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175**, 1467-1480 e1413 (2018).
329. Cleveland, J.D. & Tucker, C.L. Photo-SNAP-tag, a Light-Regulated Chemical Labeling System. *ACS Chem Biol* **15**, 2212-2220 (2020).

330. Hallett, R.A., Zimmerman, S.P., Yumerefendi, H., Bear, J.E. & Kuhlman, B. Correlating in Vitro and in Vivo Activities of Light-Inducible Dimers: A Cellular Optogenetics Guide. *ACS Synth Biol* **5**, 53-64 (2016).
331. Hartzell, E.J., Terr, J. & Chen, W. Engineering a Blue Light Inducible SpyTag System (BLISS). *J Am Chem Soc* **143**, 8572-8577 (2021).
332. Yi, J.J., Wang, H., Vilela, M., Danuser, G. & Hahn, K.M. Manipulation of endogenous kinase activity in living cells using photoswitchable inhibitory peptides. *ACS Synthetic Biology* **3**, 788-795 (2014).
333. Garcia-Marcos, M., Parag-Sharma, K., Marivin, A., Maziarz, M., Luebbers, A. & Nguyen, L.T. Optogenetic activation of heterotrimeric G-proteins by LOV2GIVe, a rationally engineered modular protein. *Elife* **9** (2020).
334. Wang, W., Wildes, C.P., Pattarabanjird, T., Sanchez, M.I., Glober, G.F., Matthews, G.A., Tye, K.M. & Ting, A.Y. A light- and calcium-gated transcription factor for imaging and manipulating activated neurons. *Nat Biotechnol* **35**, 864-871 (2017).
335. Lee, D., Creed, M., Jung, K., Stefanelli, T., Wendler, D.J., Oh, W.C., Mignocchi, N.L., Lüscher, C. & Kwon, H.B. Temporally precise labeling and control of neuromodulatory circuits in the mammalian brain. *Nature Methods* **14**, 495-503 (2017).
336. Lee, D., Hyun, J.H., Jung, K., Hannan, P. & Kwon, H.B. A calcium- and light-gated switch to induce gene expression in activated neurons. *Nat Biotechnol* **35**, 858-863 (2017).
337. Kim, M.W., Wang, W., Sanchez, M.I., Coukos, R., von Zastrow, M. & Ting, A.Y. Time-gated detection of protein-protein interactions with transcriptional readout. *Elife* **6** (2017).
338. Sanchez, M.I. & Ting, A.Y. Directed evolution improves the catalytic efficiency of TEV protease. *Nat Methods* **17**, 167-174 (2020).
339. Niopek, D., Benzinger, D., Roensch, J., Draebing, T., Wehler, P., Eils, R. & Di Ventura, B. Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nature Communications* **5** (2014).
340. Yumerefendi, H., Dickinson, D.J., Wang, H., Zimmerman, S.P., Bear, J.E., Goldstein, B., Hahn, K. & Kuhlman, B. Control of protein activity and cell fate specification via light-mediated nuclear translocation. *PLoS ONE* **10** (2015).
341. Lerner, A.M., Hepperla, A.J., Keele, G.R., Meriesh, H.A., Yumerefendi, H., Restrepo, D., Zimmerman, S., Bear, J.E., Kuhlman, B., Davis, I.J. & Strahl, B.D. An optogenetic switch for the Set2 methyltransferase provides evidence for transcription-dependent and -independent dynamics of H3K36 methylation. *Genome Res* **30**, 1605-1617 (2020).
342. Niopek, D., Wehler, P., Roensch, J., Eils, R. & Di Ventura, B. Optogenetic control of nuclear protein export. *Nature Communications* **7** (2016).
343. Wong, S., Mosabbir, A.A. & Truong, K. An engineered split intein for photoactivated protein trans-splicing. *PLoS ONE* **10** (2015).
344. Bonger, K.M., Rakhit, R., Payumo, A.Y., Chen, J.K. & Wandless, T.J. General method for regulating protein stability with light. *ACS Chemical Biology* **9**, 111-115 (2014).
345. Renicke, C., Schuster, D., Usherenko, S., Essen, L.O. & Taxis, C. A LOV2 domain-based optogenetic tool to control protein degradation and cellular function. *Chemistry and Biology* **20**, 619-626 (2013).
346. Stevens, L.M., Kim, G., Koromila, T., Steele, J.W., McGehee, J., Stathopoulos, A. & Stein, D.S. Light-dependent N-end rule-mediated disruption of protein function in Saccharomyces cerevisiae and Drosophila melanogaster. *PLoS Genet* **17**, e1009544 (2021).

347. Zheng, C. & Baum, B.J. Evaluation of promoters for use in tissue-specific gene delivery. *Methods Mol Biol* **434**, 205-219 (2008).
348. Kim, H., Kim, M., Im, S.K. & Fang, S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res* **34**, 147-159 (2018).
349. Chen, X., Ravindra Kumar, S., Adams, C.D., Yang, D., Wang, T., Wolfe, D.A., Arokiaraj, C.M., Ngo, V., Campos, L.J., Griffiths, J.A., Ichiki, T., Mazmanian, S.K., Osborne, P.B., Keast, J.R., Miller, C.T., Fox, A.S., Chiu, I.M. & Gradinaru, V. Engineered AAVs for non-invasive gene delivery to rodent and non-human primate nervous systems. *Neuron* **110**, 2242-2257 e2246 (2022).
350. Goertsen, D., Flytzanis, N.C., Goeden, N., Chuapoco, M.R., Cummins, A., Chen, Y., Fan, Y., Zhang, Q., Sharma, J., Duan, Y., Wang, L., Feng, G., Chen, Y., Ip, N.Y., Pickel, J. & Gradinaru, V. AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nat Neurosci* **25**, 106-115 (2022).
351. Jendryka, M., Palchaudhuri, M., Ursu, D., van der Veen, B., Liss, B., Katzel, D., Nissen, W. & Pekcec, A. Pharmacokinetic and pharmacodynamic actions of clozapine-N-oxide, clozapine, and compound 21 in DREADD-based chemogenetics in mice. *Sci Rep* **9**, 4522 (2019).
352. Gomez, J.L., Bonaventura, J., Lesniak, W., Mathews, W.B., Sysa-Shah, P., Rodriguez, L.A., Ellis, R.J., Richie, C.T., Harvey, B.K., Dannals, R.F., Pomper, M.G., Bonci, A. & Michaelides, M. Chemogenetics revealed: DREADD occupancy and activation via converted clozapine. *Science* **357**, 503-507 (2017).
353. Brunello, N., Masotto, C., Steardo, L., Markstein, R. & Racagni, G. New insights into the biology of schizophrenia through the mechanism of action of clozapine. *Neuropsychopharmacology* **13**, 177-213 (1995).
354. van den Pol, A.N. Neuropeptide transmission in brain circuits. *Neuron* **76**, 98-115 (2012).
355. Song, H.K. & Eck, M.J. Structural basis of degradation signal recognition by SspB, a specificity-enhancing factor for the ClpXP proteolytic machine. *Molecular Cell* **12**, 75-86 (2003).
356. Matthew, A.N., Zephyr, J., Hill, C.J., Jahangir, M., Newton, A., Petropoulos, C.J., Huang, W., Kurt-Yilmaz, N., Schiffer, C.A. & Ali, A. Hepatitis C Virus NS3/4A Protease Inhibitors Incorporating Flexible P2 Quinoxalines Target Drug Resistant Viral Variants. *Journal of Medicinal Chemistry* **60**, 5699-5716 (2017).
357. Dingwall, C. & Laskey, R.A. Nuclear targeting sequences - a consensus? *Trends in Biochemical Sciences* **16**, 478-481 (1991).
358. Xu, D., Farmer, A., Collett, G., Grishin, N.V. & Chook, Y.M. Sequence and structural analyses of nuclear export signals in the NESdb database. *Molecular Biology of the Cell* **23**, 3677-3693 (2012).
359. So, W.H., Wong, C.T.T. & Xia, J. Peptide photocaging: A brief account of the chemistry and biological applications. *Chinese Chemical Letters* **29**, 1058-1062 (2018).
360. Renner, C. & Moroder, L. Azobenzene as conformational switch in model peptides. *ChemBioChem* **7**, 868-878 (2006).
361. Stanton, B.Z., Chory, E.J. & Crabtree, G.R. Chemically induced proximity in biology and medicine. *Science* **359** (2018).

362. Wu, H.D., Kikuchi, M., Dagliyan, O., Aragaki, A.K., Nakamura, H., Dokholyan, N.V., Umehara, T. & Inoue, T. Rational design and implementation of a chemically inducible heterotrimerization system. *Nature Methods* **17**, 928-936 (2020).
363. Dagliyan, O., Dokholyan, N.V. & Hahn, K.M. Engineering proteins for allosteric control by light or ligands. *Nat Protoc* **14**, 1863-1883 (2019).
364. Clackson, T., Yang, W., Rozamus, L.W., Hatada, M., Amara, J.F., Rollins, C.T., Stevenson, L.F., Magari, S.R., Wood, S.A., Courage, N.L., Lu, X., Cerasoli Jr, F., Gilman, M. & Holt, D.A. Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10437-10442 (1998).
365. Banaszynski, L.A., Sellmyer, M.A., Contag, C.H., Wandless, T.J. & Thorne, S.H. Chemical control of protein stability and function in living mice. *Nature Medicine* **14**, 1123-1127 (2008).
366. Wacker, D., Stevens, R.C. & Roth, B.L. How Ligands Illuminate GPCR Molecular Pharmacology. *Cell* **170**, 414-427 (2017).
367. Mansour, A., Hoversten, M.T., Taylor, L.P., Watson, S.J. & Akil, H. The cloned μ, δ and κ receptors and their endogenous ligands: Evidence for two opioid peptide recognition cores. *Brain Research* **700**, 89-98 (1995).
368. Koehl, A., Hu, H., Maeda, S., Zhang, Y., Qu, Q., Paggi, J.M., Latorraca, N.R., Hilger, D., Dawson, R., Matile, H., Schertler, G.F.X., Granier, S., Weis, W.I., Dror, R.O., Manglik, A., Skiniotis, G. & Kobilka, B.K. Structure of the μ-opioid receptor-Gi protein complex. *Nature* **558**, 547-552 (2018).
369. Guan, X.M., Tong Sun, K. & Kobilka, B.K. Enhancement of membrane insertion and function in a type IIIb membrane protein following introduction of a cleavable signal peptide. *Journal of Biological Chemistry* **267**, 21995-21998 (1992).
370. Dixon, A.S., Schwinn, M.K., Hall, M.P., Zimmerman, K., Otto, P., Lubben, T.H., Butler, B.L., Binkowski, B.F., Machleidt, T., Kirkland, T.A., Wood, M.G., Eggers, C.T., Encell, L.P. & Wood, K.V. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem Biol* **11**, 400-408 (2016).
371. Fan, F., Binkowski, B.F., Butler, B.L., Stecha, P.F., Lewis, M.K. & Wood, K.V. Novel genetically encoded biosensors using firefly luciferase. *ACS Chemical Biology* **3**, 346-351 (2008).
372. Che, T., Agnihotri, H.D., Shukla, A.K. & Roth, B.L. Biased ligands at opioid receptors: Current status and future directions. *Science Signaling* **14** (2021).
373. Shen, J., Geng, L., Li, X., Emery, C., Kroning, K., Shingles, G., Lee, K., Heyden, M., Li, P. & Wang, W. A general method for chemogenetic control of peptide function. *Nat Methods* **20**, 112-122 (2023).
374. McCafferty, J., Griffiths, A.D., Winter, G. & Chiswell, D.J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554 (1990).
375. Hanes, J. & Pluckthun, A. In vitro selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci U S A* **94**, 4937-4942 (1997).
376. Packer, M.S. & Liu, D.R. Methods for the directed evolution of proteins. *Nat Rev Genet* **16**, 379-394 (2015).
377. Osbourn, J.K. in Comprehensive Medicinal Chemistry II. (eds. J.B. Taylor & D.J. Triggle) 431-447 (Elsevier, Oxford; 2007).

378. Boder, E.T. & Wittrup, K.D. Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology* **15**, 553-557 (1997).
379. Mahdavi, S.Z.B., Oroojalian, F., Eyvazi, S., Hejazi, M., Baradaran, B., Pouladi, N., Tohidkia, M.R., Mokhtarzadeh, A. & Muyldermans, S. An overview on display systems (phage, bacterial, and yeast display) for production of anticancer antibodies; advantages and disadvantages. *Int J Biol Macromol* **208**, 421-442 (2022).
380. Cherf, G.M. & Cochran, J.R. Applications of Yeast Surface Display for Protein Engineering. *Methods Mol Biol* **1319**, 155-175 (2015).
381. Wang, Y., Xue, P., Cao, M., Yu, T., Lane, S.T. & Zhao, H. Directed Evolution: Methodologies and Applications. *Chem Rev* **121**, 12384-12444 (2021).
382. Livingston, K.E., Mahoney, J.P., Manglik, A., Sunahara, R.K. & Traynor, J.R. Measuring ligand efficacy at the mu-opioid receptor using a conformational biosensor. *Elife* **7** (2018).
383. Liao, C., de Molliens, M.P., Schneebeli, S.T., Brewer, M., Song, G., Chatenet, D., Braas, K.M., May, V. & Li, J. Targeting the PAC1 Receptor for Neurological and Metabolic Disorders. *Curr Top Med Chem* **19**, 1399-1417 (2019).
384. Dias, B.G. & Ressler, K.J. PACAP and the PAC1 receptor in post-traumatic stress disorder. *Neuropsychopharmacology* **38**, 245-246 (2013).
385. Piper, S.J., Deganutti, G., Lu, J., Zhao, P., Liang, Y.L., Lu, Y., Fletcher, M.M., Hossain, M.A., Christopoulos, A., Reynolds, C.A., Danev, R., Sexton, P.M. & Wootten, D. Understanding VPAC receptor family peptide binding and selectivity. *Nat Commun* **13**, 7013 (2022).
386. Bourgault, S., Vaudry, D., Guilhaudis, L., Raoult, E., Couvineau, A., Laburthe, M., Segalas-Milazzo, I., Vaudry, H. & Fournier, A. Biological and structural analysis of truncated analogs of PACAP27. *J Mol Neurosci* **36**, 260-269 (2008).
387. Anderson, E.J., Cakir, I., Carrington, S.J., Cone, R.D., Ghamari-Langroudi, M., Gillyard, T., Gimenez, L.E. & Litt, M.J. 60 YEARS OF POMC: Regulation of feeding and energy homeostasis by alpha-MSH. *J Mol Endocrinol* **56**, T157-174 (2016).
388. Deisseroth, K., Feng, G., Majewska, A.K., Miesenböck, G., Ting, A. & Schnitzer, M.J. Next-generation optical technologies for illuminating genetically targeted brain circuits. *Journal of Neuroscience* **26**, 10380-10386 (2006).
389. Miesenböck, G. The optogenetic catechism. *Science* **326**, 395-399 (2009).
390. Hoffmann, M.D., Bubeck, F., Eils, R. & Niopek, D. Controlling Cells with Light and LOV. *Advanced Biosystems* **2** (2018).
391. 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).
392. Strickland, D., Yao, X., Gawlak, G., Rosen, M.K., Gardner, K.H. & Sosnick, T.R. Rationally improving LOV domaiǧ€"based photoswitches. *Nature Methods* **7**, 623-626 (2010).
393. 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).
394. Pham, E., Mills, E. & Truong, K. A synthetic photoactivated protein to generate local or global Ca 2+ signals. *Chemistry and Biology* **18**, 880-890 (2011).
395. Baarlink, C., Wang, H. & Grosse, R. Nuclear actin network assembly by formins regulates the SRF coactivator MAL. *Science* **340**, 864-867 (2013).

396. Spiltoir, J.I., Strickland, D., Glotzer, M. & Tucker, C.L. Optical Control of Peroxisomal Trafficking. *ACS Synthetic Biology* **5**, 554-560 (2016).
397. He, L., Tan, P., Zhu, L., Huang, K., Nguyen, N.T., Wang, R., Guo, L., Li, L., Yang, Y., Huang, Z., Huang, Y., Han, G., Wang, J. & Zhou, Y. Circularly permuted LOV2 as a modular photoswitch for optogenetic engineering. *Nature Chemical Biology* **17**, 915-923 (2021).
398. Lu, H., Mazumder, M., Jaikaran, A.S.I., Kumar, A., Leis, E.K., Xu, X., Altmann, M., Cochrane, A. & Woolley, G.A. A Yeast System for Discovering Optogenetic Inhibitors of Eukaryotic Translation Initiation. *ACS Synth Biol* **8**, 744-757 (2019).
399. Peter, D., Igreja, C., Weber, R., Wohlbold, L., Weiler, C., Ebertsch, L., Weichenrieder, O. & Izaurralde, E. Molecular architecture of 4E-BP translational inhibitors bound to eIF4E. *Mol Cell* **57**, 1074-1087 (2015).
400. Yumerefendi, H., Lerner, A.M., Zimmerman, S.P., Hahn, K., Bear, J.E., Strahl, B.D. & Kuhlman, B. Light-induced nuclear export reveals rapid dynamics of epigenetic modifications. *Nature Chemical Biology* **12**, 399-401 (2016).
401. Lam, S.S., Martell, J.D., Kamer, K.J., Deerinck, T.J., Ellisman, M.H., Mootha, V.K. & Ting, A.Y. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods* **12**, 51-54 (2014).
402. Shi, Y., Chen, Y., Deng, L., Du, K., Lu, S. & Chen, T. Structural Understanding of Peptide-Bound G Protein-Coupled Receptors: Peptide-Target Interactions. *J Med Chem* **66**, 1083-1111 (2023).
403. Thom, C., Ehrenmann, J., Vacca, S., Waltenspuhl, Y., Schoppe, J., Medalia, O. & Pluckthun, A. Structures of neurokinin 1 receptor in complex with G(q) and G(s) proteins reveal substance P binding mode and unique activation features. *Sci Adv* **7**, eabk2872 (2021).
404. Sun, W., Yuan, Q., Zhang, H., Yang, F., Ling, S., Luo, Y., Lv, P., Eric Xu, H., Tian, C., Yin, W. & Shi, P. Structural insights into the activation of neurokinin 2 receptor by neurokinin A. *Cell Discov* **8**, 72 (2022).
405. Hong, C., Byrne, N.J., Zamlynny, B., Tummala, S., Xiao, L., Shipman, J.M., Partridge, A.T., Minnick, C., Breslin, M.J., Rudd, M.T., Stachel, S.J., Rada, V.L., Kern, J.C., Armacost, K.A., Hollingsworth, S.A., O'Brien, J.A., Hall, D.L., McDonald, T.P., Strickland, C., Brooun, A., Soisson, S.M. & Hollenstein, K. Structures of active-state orexin receptor 2 rationalize peptide and small-molecule agonist recognition and receptor activation. *Nat Commun* **12**, 815 (2021).
406. Liang, Y.L., Belousoff, M.J., Zhao, P., Koole, C., Fletcher, M.M., Truong, T.T., Julita, V., Christopoulos, G., Xu, H.E., Zhang, Y., Khoshouei, M., Christopoulos, A., Danev, R., Sexton, P.M. & Wootten, D. Toward a Structural Understanding of Class B GPCR Peptide Binding and Activation. *Mol Cell* **77**, 656-668 e655 (2020).
407. Jiang, W. & Zheng, S. Structural insights into galanin receptor signaling. *Proc Natl Acad Sci U S A* **119**, e2121465119 (2022).
408. Josephs, T.M., Belousoff, M.J., Liang, Y.L., Piper, S.J., Cao, J., Garama, D.J., Leach, K., Gregory, K.J., Christopoulos, A., Hay, D.L., Danev, R., Wootten, D. & Sexton, P.M. Structure and dynamics of the CGRP receptor in apo and peptide-bound forms. *Science* **372** (2021).
409. McCauley, J.A. & Rudd, M.T. Hepatitis C virus NS3/4a protease inhibitors. *Current Opinion in Pharmacology* **30**, 84-92 (2016).

410. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. & Lindah, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
411. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. & Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics* **78**, 1950-1958 (2010).
412. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).
413. Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723 (1997).
414. Krivov, G.G., Shapovalov, M.V. & Dunbrack, R.L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77**, 778-795 (2009).
415. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092 (1993).
416. Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).
417. Miyamoto, S. & Kollman, P.A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry* **13**, 952-962 (1992).
418. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. & Haak, J.R. in The Journal of Chemical Physics, Vol. 81 3684-3690 (1984).
419. Nosé, S. in Molecular Physics, Vol. 52 255-268 (1984).
420. Hoover, W.G. in Physical Review A (1985).
421. Parrinello, M. & Rahman, A. in Journal of Applied Physics, Vol. 52 7182-7190 (1981).
422. Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W.F. & Mark, A.E. in Angewandte Chemie International Edition, Vol. 38 236-240 (1999).
423. Li, P., Li, S.B., Wang, X., Phillips, C.D., Schwarz, L.A., Luo, L., de Lecea, L. & Krasnow, M.A. Brain Circuit of Claustrophobia-like Behavior in Mice Identified by Upstream Tracing of Sighing. *Cell Reports* **31** (2020).
424. Lee, T.I. & Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251 (2013).
425. Thakur, J.K., Yadav, A. & Yadav, G. Molecular recognition by the KIX domain and its role in gene regulation. *Nucleic Acids Res* **42**, 2112-2125 (2014).
426. Ramsay, R.G., Barton, A.L. & Gonda, T.J. Targeting c-Myb expression in human disease. *Expert Opin Ther Targets* **7**, 235-248 (2003).
427. Ernst, P., Wang, J. & Korsmeyer, S.J. The role of MLL in hematopoiesis and leukemia. *Curr Opin Hematol* **9**, 282-287 (2002).
428. Yadav, A., Thakur, J.K. & Yadav, G. KIXBASE: A comprehensive web resource for identification and exploration of KIX domains. *Sci Rep* **7**, 14924 (2017).
429. Joy, S.T., Henley, M.J., De Salle, S.N., Beyersdorf, M.S., Vock, I.W., Huldin, A.J.L. & Mapp, A.K. A Dual-Site Inhibitor of CBP/p300 KIX is a Selective and Effective Modulator of Myb. *J Am Chem Soc* **143**, 15056-15062 (2021).