Development and application of genetic networks for engineering photo-controlled proteins

by

Katherine Emily Brechun

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Chemistry
University of Toronto

As part of a Joint Education Placement with the
Institute for Biochemistry and Biology
University of Potsdam

Published online at the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-43092
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-430924

Development and application of genetic networks for engineering photo-controlled proteins

Katherine Emily Brechun

Doctor of Philosophy

Department of Chemistry, University of Toronto
Institute for Biochemistry and Biology, University of Potsdam

2019

Abstract

Light-switchable proteins are being used increasingly to understand and manipulate complex molecular systems. The success of this approach has fueled the development of tailored photo-switchable proteins, to enable targeted molecular events to be studied using light. The development of novel photo-switchable tools has to date largely relied on rational design. Complementing this approach with directed evolution would be expected to facilitate these efforts. Directed evolution, however, has been relatively infrequently used to develop photo-switchable proteins due to the challenge presented by high-throughput evaluation of switchable protein activity. This thesis describes the development of two genetic circuits that can be used to evaluate libraries of switchable proteins, enabling optimization of both the on- and off-states. A screening system is described, which permits detection of DNA-binding activity based on conditional expression of a fluorescent protein. In addition, a tunable selection system is presented, which allows for the targeted selection of protein-protein interactions of a desired affinity range. This thesis additionally describes the development and characterization of a synthetic protein that was designed to investigate chromophore reconstitution in photoactive yellow protein (PYP), a promising scaffold for engineering photo-controlled protein tools.

Acknowledgments

I would like to thank my supervisors, Prof. Andrew Woolley and Prof. Katja Arndt, for the opportunity to become involved in such an interesting area of research. I am extremely grateful for their invaluable expertise and guidance, as well as their encouragement and optimism in the face of setbacks. I thank them as well for the extra time and effort that was required to enable me to undertake this joint degree, and for their support along the way.

I am grateful for the support I received from the members of my committee, Prof. Voula Kanelis, Prof. Radhakrishnan Mahadevan and Prof. Kristian Müller. I thank them for their thoughtful comments, scientific guidance and enthusiasm. I thank Prof. Deborah Zamble for agreeing to be a member of my examination committee, and Prof. Shohei Koide from New York University for serving as my external examiner.

I gratefully acknowledge the funding I've received. Without the financial support from NSERC and the DAAD, it would not have been possible for me to participate in this joint degree. The opportunity to divide my time between two research groups in different countries has inspired me, and the challenges associated with the arrangement have helped me grow as a person.

I am very grateful to have had the opportunity to work with so many dedicated, enthusiastic and good-natured researchers. I've learned a lot from many people.

For my time spent at the University of Toronto, I thank past and present members of the Woolley lab. I am especially grateful to Anna for training me when I first joined the lab and to Anil for his patient guidance in chromophore synthesis. I thank Carrie for her mentorship, as well as Jakeb, Amir and Lulu for the fun times we've shared over the years as graduate students. I also thank Gero and Ryan for their support while writing this dissertation. The UofT Chemistry Department has a very strong community and I am grateful to have been a part of it. I thank squash partners, volleyball enthusiasts and climbing friends for all the good times we've shared.

For my time at the University of Potsdam, I thank past and present members of the Arndt lab. Thank you for taking me under your wing and helping me settle in a foreign country. From lending me a bike to helping me outfit my kitchen, I don't think I could bake enough cakes to express my gratitude. I thank Sebastian and Stefan as well for their friendship; I am very happy

that we were able to spend so much time working together. In addition, I thank the members of the gymnastics club in Potsdam, USV Potsdam Turnen, for helping me keep my life well-balanced.

Daniel, I thank you for your patience, perspective and support when dealing with the ups and downs of research. Thank you for helping me combat any frustration with a solid game of squash or badminton. I thank my family for their support and encouragement during course of this degree and always. Thank you for always being there for me. I dedicate this work to my grandmother, Marie Neeb, who passed away during my degree.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

5FdU	2′-deoxy-5-fluorouridine
Abs.	absorbance
Amp	ampicillin
aM-Pp	1,6-anhydroMurNAc-pentapeptide
amu	atomic mass unit
AP-1	activator protein 1
A. thaliana	*Arabidopsis thaliana*
A. sativa	*Avena sativa*
aTc	anhydrotetracycline
a.u.	arbitrary units
aZIP	acidic leucine zipper
BAD	Bcl-2-associated agonist of cell death
BAX	Bcl-2-associated X protein
Bcl-2	B cell lymphoma 2
BFP	blue fluorescent protein
BH3	Bcl-2 homology domain 3
BLA or βla	β-lactamase
BLIP	β-lactamase inhibitor peptide
BSA	bovine serum albumen
bZIP	basic leucine zipper
CAT	chloramphenicol acetyltransferase
cAMP	cyclic adenosine monophosphate
C. freundii	*Citrobacter freundii*
cLOV	circularly permutated LOV
Cm	chloramphenicol
cPCR	colony PCR
cPYP	circularly permutated PYP
CAT	chloramphenicol acetyltransferase
CoA	coenzyme A

CREB	cAMP response element binding protein
Cry2	cryptochrome 2
CV	column volume
DHFR	dihydrofolate reductase
DMSO	dimethyl sulfoxide
DN	dominant negative
DNA	deoxyribonucleic acid
DOG	2-deoxy-galactose
Dox	doxycycline
dTMP	deoxythymidine monophosphate
DTT	dithiothreitol
DYT	double yeast tryptone
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
em.	emission wavelength
EMSA	electrophoretic mobility shift assay
epPCR	error prone PCR
Eq.	equation
ESI-MS	electrospray ionization mass spectrometry
ex.	excitation wavelength
Fab	antigen-binding fragment
FACS	fluorescence-activated cell sorting
FAD	flavin adenine dinucleotide
FDA	U.S. Food & Drug Administration
Flp	flippase
FMN	flavin mononucleotide
FPLC	fast protein liquid chromatography
FRET	Förster resonance energy transfer
FRT	Flp recognition target
GEFs	guanine nucleotide exchange factors
GFP	green fluorescent protein

GnHCl	guanidine hydrochloride
GPCR	G-protein coupled receptor
G-protein	guanine nucleotide binding protein
GRK2ct	C-terminal domain of GPCR kinase 2
GDP	guanosine diphosphate
GTP	guanosine triphosphate
H. halophila	*Halorhodospira halophila*
hsvTK	herpes simplex virus thymidine kinase
I. loihiensis	*Iodomarina loihiensis*
IM	inner membrane
IPTG	isopropyl-β-D-1-thiogalactopyranoside
Kan	kanamycin
K_d	equilibrium dissociation constant
k_f	forwards rate constant
k_r	reverse rate constant
LB	Luria-Bertani
LED	light emitting diode
LOV	light-oxygen-voltage
LPS	lipopolysaccharide
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
NES	nuclear export sequence
Ni-NTA	nickel-charged affinity resin
NLS	nuclear localization sequence
NMR	nuclear magnetic resonance
MIC	minimal inhibitory concentration
MWCO	molecular weight cut-off
OD_{600}	optical density at 600 nm
OM	outer membrane
PAS	Per-Arnt-Sim domain

PBS	phosphate buffered saline
pCL or 4CL	*p*-coumaroyl-CoA ligase
PCR	polymerase chain reaction
PDB	protein databank
phot1	phototropin 1
PhyB	phytochrome B
PIF 3	phytochrome interacting factor 3
PIF 6	phytochrome interacting factor 6
p/o	promoter and operator
PVDF	polyvinylidene difluoride
PYP	photoactive yellow protein
RBS	ribosome binding site
R. capsulatus	*Rhodobacter capsulatus*
rcf	relative centrifugal force
R. sphaeroides	*Rhodobacter sphaeroides*
ROS	reactive oxygen species
rpm	revolutions per minute
SD	standard deviation
SDS PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
Spec	spectinomycin
ssTorA	signal sequence from trimethylamine *N*-oxide reductase
Strep	streptomycin
SUMO	small ubiquitin-like modifier
TAE	Tris-acetate EDTA
TAL	tyrosine ammonia lyase
Tat	twin arginine translocase
tdT	tdTomato
Tet	tetracycline
TetA	a Tet efflux pump (Tet/H^+ antiporter)
TetC	a Tet efflux pump (Tet/H^+ antiporter)
TetR	Tet repressor

TF	transcription factor
TSS	transcription start site
T. tepidum	*Thermochromatium tepidum*
UV	ultraviolet
vis	visible
wt	wild-type
Y2H	yeast-2-hybrid

1 General Introduction

Note: Sections of this chapter were reproduced with permission from Elsevier from the following review article: K.E. Brechun, K.M. Arndt, G.A. Woolley. Strategies for photo-control of endogenous protein activity. 2017. *Current Opinion in Structural Biology*. 45: 53-58. The text has been modified for consistency.

The original article can be accessed via the following link:
https://doi.org/10.1016/j.sbi.2016.11.014

The ability to control protein function with light enables complex, dynamic systems to be probed with high precision, permitting questions to be tackled that traditional techniques are not sensitive enough to address. Traditional genetic techniques, such as gene silencing or overexpression methods, have poor temporal resolution, and permanent genetic changes can yield misleading phenotypes due to compensation by other related genes.[1] Biochemical techniques, such as diffusible enzyme agonists or antagonists, suffer from poor spatial resolution, off-target effects and irreversibility. In contrast, light as an effector can enable specificity, high spatiotemporal resolution, tunable dynamics and reversibility. Photo-controlled proteins have proven to be valuable tools, facilitating efforts to elucidate processes ranging from gene expression in neurons[2] to cellular motility.[3,4] Some naturally occurring photo-controlled proteins may be used directly;[5,6] however, most require modification to serve as useful light-responsive tools. The desire for novel photo-controlled protein tools has intensified a challenging and exciting avenue in protein engineering: the development proteins with conditional function.

This dissertation focuses on protein engineering to develop synthetic photo-controlled proteins. To support a directed evolution engineering methodology, genetic circuits were developed enabling screening and selection of photo-controlled protein activity. In addition, chromophore reconstitution in the photosensor photoactive yellow protein (PYP) was studied, to facilitate the use of PYP as a scaffold for photo-switchable protein development. What follows is an introduction to protein engineering, with a focus on the development of photo-controlled proteins; a brief review of photo-controlled protein tool development is provided, and the available methodologies for library construction and evaluation are discussed.

1.1 Protein engineering

Protein engineering is the process of modifying a protein with the intention of altering its physical or biochemical properties. This field is both an active area in basic research, enabling scientists to examine structure-function relationships, protein folding and evolutionary principles in natural proteins, and a central component of biotechnology, allowing proteins to be fine-tuned or repurposed for a target application. There are two general approaches used in protein engineering: rational design and directed evolution. In rational design, targeted modifications are made to a protein sequence based on sequence alignments, structural information and/or computational modelling. Directed evolution, in contrast, mimics protein development in nature, where random genetic diversity is incorporated into a protein sequence and the resulting variants are evaluated for a desired outcome. Frequently, these two approaches are merged as complementary strategies.[7]

Protein engineering initiatives to develop photo-controlled proteins have to date largely relied on rational design strategies. Over the last decade, protein engineers have successfully developed a multitude of synthetic light-switchable tools. These light-switchable proteins have been engineered using two main design approaches: 1) caging strategies, using photo-controlled protein conformational changes, and 2) protein relocalization strategies, using photo-controlled protein–protein interactions.

1.1.1 Engineering photo-controlled proteins using caging strategies

A conceptually straightforward strategy to design a photo-switchable protein tool is to fuse an effector (*i.e.,* an inhibitor or activator) to a photo-switchable protein in order to control the availability of the effector. In this approach, light-induced structural changes of the photo-switchable protein are used to control the accessibility of the effector by creating or alleviating steric hindrance – a strategy termed ‘caging’. This concept is depicted in Figure 1-1.

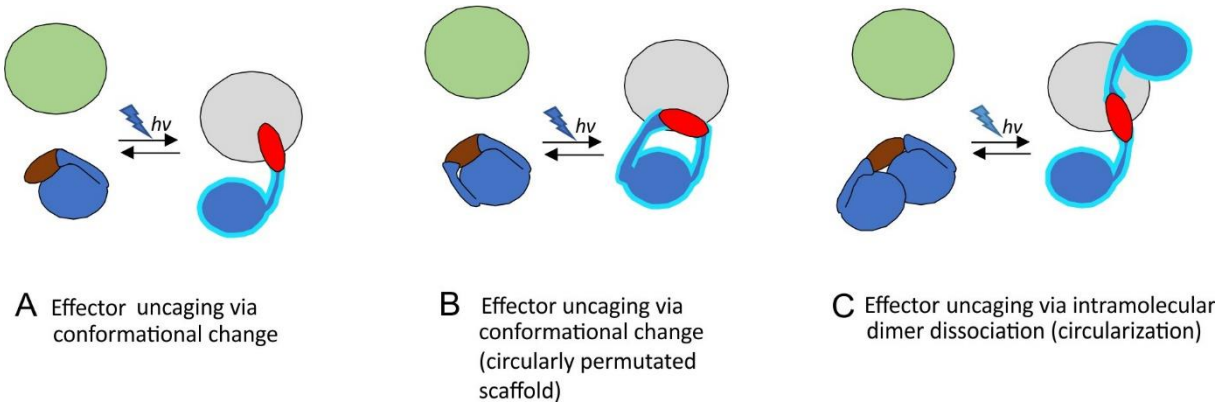


Figure 1-1: Strategies for the photo-control of protein activity based on caging.
The concept of using caging to develop a photo-controlled protein tool is demonstrated using an inhibitor as an example effector. (A) Light-induced structural rearrangements in the N- or C-terminus of a photoactive protein can be used to cage an inhibitor. Irradiation of the fusion protein leads to inhibition of a target protein. (B) Similarly, circularly permutated photoactive scaffolds have been introduced that enable peptides to be caged with both termini tethered to the photoactive protein. (C) An inhibitor may also be caged via tethering to photoactive proteins that undergo light-induced dimer dissociation. Blue: photoactive domain, Green: active endogenous protein, Grey: inhibited endogenous protein, Red: inhibitor (Dark red: unavailable inhibitor). A blue glow indicates the light-activated state of the photoactive protein domain.

Photo-caging of an effector is typically achieved using photo-switchable proteins that exhibit a photocycle with large light-dependent conformational changes. The domains most commonly used for this approach are the light-oxygen-voltage (LOV) 2 domain of phototropin 1 from *Avena sativa*[8] (*As*LOV2) and *Arabidopsis thaliana,*[9] as well as PYP from *Halorhodospira halophila*.[10,11] The photocycle of the *As*LOV2 domain is triggered by a flavin mononucleotide (FMN) chromophore that forms a thiol-adduct with a proximal cysteine when irradiated with blue light.[12] This causes a signal to propagate through the protein core, leading to the undocking and partial unfolding of the C-terminal helix, termed the Jα helix.[8] In the absence of blue light, the thiol-adduct decays, and the Jα helix re-folds and docks on the protein core. The photocycle of PYP is triggered by a covalently bound *p*-coumaric acid-based chromophore,[13] which likewise absorbs blue light. This results in chromophore isomerization, leading to a protein structural rearrangement involving the N-terminal cap.[10] PYP re-folds spontaneously in the dark. Recently, a circularly permuted PYP scaffold[14] and a circularly permutated LOV scaffold[15] were developed, expanding the potential for caging peptides using these domains[15–18] (Figure 1-1B). An alternative strategy for caging was introduced using an engineered variant of the fluorescent

protein Dronpa, which undergoes light-induced dimer dissociation (Figure 1-1C). Lin and colleagues demonstrated that a target effector could be caged by fusing both ends of the effector to a pair of photo-switchable proteins based on Dronpa that dimerize in the dark, thereby circularizing the target.[19] Irradiation leads to dimer dissociation and enhanced availability of a binding interface. It may also be possible to implement this strategy using the LOV domain from *Rhodobacter sphaeroides*, which was shown to likewise undergo light-induced dimer dissociation.[20]

Light-induced unfolding of the *As*LOV Jα helix was first used to control endogenous protein activity in a ground-breaking study by Wu and colleagues.[3] A photo-controlled version of the small GTPase Rac1 was developed by creating fusions of Rac1 and the Jα helix that were designed to sterically block the Rac1 binding surface in the dark-state, and expose it in the unfolded light-state.[3] To optimize the design, a small library of fusion variants (composed of several tens of members) was constructed and screened manually.[21] One construct, termed PA-Rac1, showed a ten-fold increase (2 – 0.2 μM) in affinity for the target PAK upon blue light irradiation.[3] The change in Rac1 accessibility was sufficient to produce dramatic effects on cytoskeletal dynamics; strikingly, blue light activation of PA-Rac1 at the edge of a mouse embryonic fibroblast cell was shown to produce local protrusions, together with retraction on the opposite side of the cell.[3]

The concept of caging an effector using a photo-switchable protein has since been advanced by several groups. Kuhlman and colleagues designed a chimeric fusion that embedded a peptide derived from the *Shigella* protein ipaA in the *As*LOV Jα helix.[22] Photo-controlled kinase inhibitors were created by two groups by caging inhibitor peptides also using an *As*LOV scaffold. The Hahn group developed photo-controlled inhibitors of cAMP-dependent kinase and myosin light chain kinase by designing fusions of known peptide inhibitors with the Jα helix.[23] A different strategy was employed by Cunniff *et al.,* who developed a photo-controlled inhibitor of AMP-activated protein kinase by inserting an inhibitory peptide in the linker between the LOV core and the Jα helix.[24] Optogenetic inhibitors of mammalian transcription factors have been created by caging effectors using PYP[15] as well as *As*LOV[16] scaffolds. Ali and colleagues caged a dominant negative inhibitor of the mammalian transcription factor CREB by creating a chimeric fusion with the N-terminal domain of PYP.[25] Irradiation of opto-DN-CREB exposed a

hydrophobic patch on PYP that trapped the binding interface of the dominant negative inhibitor. Paonessa and colleagues created photo-controlled inhibitors for the mammalian transcription factor REST, by designing fusions between inhibitory domains and the *As*LOV2 Jα helix based on computational modelling.[2] These examples illustrate the wide range of engineering strategies used to successfully achieve effector caging. Clearly, no single strategy can be used to cage all targets.

A generalizable strategy for photo-controlled protein inhibition based on photo-caging was introduced by Renicke and colleagues, who created a photo-controlled degradation tag. To do so, the murine ornithine decarboxylase-like degradation sequence was caged via fusion to the Jα helix of the *A. thaliana* LOV domain.[26] Blue light irradiation uncaged the degradation tag, leading to protein degradation. This tag was subsequently optimized and shown to function in yeast and higher eukaryotes.[27] Bonger *et al.*[28] created a similar photo-controllable degradation tag, termed B-LID (blue-light induced degradation), that has been used to confer photo-control on the widely used Tet-OFF gene regulatory system.[29] Although a very generalizable strategy for protein inhibition, photo-induced degradation requires replacement of the endogenous protein with the photo-controlled variant, and the protein inhibition kinetics are dictated by the protein synthesis and degradation machinery.

1.1.2 Photo-controlled relocalization design strategies

The second main approach being used to design light-switchable tools is photo-controlled protein relocalization. Living cells are highly structured environments and endogenous protein activity can be controlled by altering the local concentration of effectors, without necessarily altering intrinsic binding affinities. Photo-control of endogenous protein activity can be achieved via (i) recruitment of an effector, (ii) relocalization (*i.e.* removal) of an effector, or (iii) protein sequestering. These strategies are depicted in Figure 1-2.

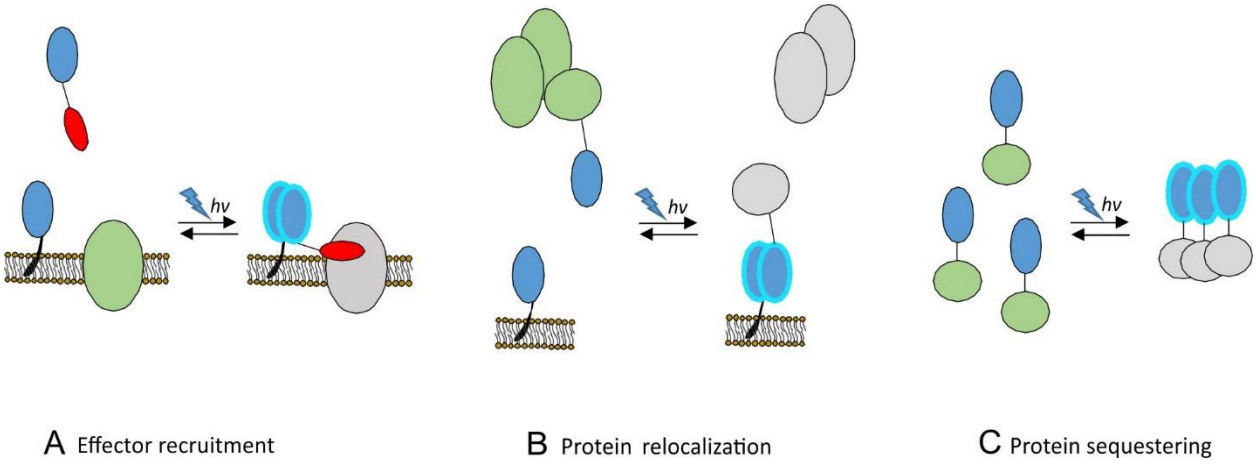


Figure 1-2: Strategies for the photo-control of protein activity based on relocalization. (A) Light-induced recruitment of a photoactive protein domain tagged to an effector, such as an inhibitor as depicted above, can be used to control the function a localized target protein. (B) Alternatively, an endogenous protein fused to a photoactive protein can be relocalized via photo-induced recruitment to an anchored photoactive protein in order to disrupt and inactivate a protein complex. (C) Light-induced oligomerization of photoactive protein domains can be used to inhibit a tagged endogenous protein through sequestering. Blue: photoactive domain, Green: active endogenous protein, Grey: inhibited endogenous protein, Red: inhibitor (Dark red: unavailable inhibitor), Black: membrane anchoring tag. A blue glow indicates light-activated state of the photoactive protein domain.

Relocalization is most frequently accomplished using photo-controlled protein-protein interactions. A variety of naturally occurring photo-controlled protein–protein interactions are known, and these have been used to make optogenetic tools. In addition to LOV-based systems,[30] phytochromes and cryptochromes have been widely used.[31] Phytochromes are red-light sensing domains from plants that contain bilin-based chromophores. When irradiated with red light, *Arabidopsis* phytochrome B (PhyB) binds to phytochrome interacting factors 3 and 6 (PIF3, PIF6), whereas irradiation with far-red light causes dissociation of the PhyB:PIF complexes. Cryptochromes are a family of blue-light responsive photoreceptors that contain a flavin adenine dinucleotide (FAD) chromophore. The *Arabidopsis* cryptochrome 2 (Cry2) has been used extensively to photo-control protein–protein interactions. Cry2 homo-oligomerizes and also heterodimerizes with its binding partner, the basic helix–loop–helix protein CIB1.[32] Numerous synthetic systems based on these scaffolds have been developed.[33–35] For a full discussion of the available optical dimerization systems, readers are directed to a recent review.[36]

1.1.2.1 Effector recruitment

Photo-controlled effector recruitment employs optical dimerization systems to recruit an activator or inhibitor to an area of interest (Figure 1-2A). An impressive example of this approach was demonstrated by developing optogenetic tools to inactivate endogenous GPCR subunits.[37] The interaction between Cry2 and CIBN, a truncated version of CIB1,[32] was used to recruit inhibitors to GPCR, by anchoring CIBN to the cell membrane and fusing Cry2 to a target effector. In the first approach, Cry2 was fused to the C-terminal domain of GPCR kinase 2 (GRK2ct). Upon irradiation, GRK2ct was recruited to the membrane where it sequestered G-protein βγ subunits, preventing their interaction with their native targets. In the second approach, Cry2 was fused to a GTPase-accelerating protein, RGS4. RGS4 accelerates the conversion of active αGTP to inactive αGDP, which rebinds βγ and prevents the interaction of βγ with downstream targets. Light-triggered recruitment of these GPCR signaling inhibitors enabled signaling gradients to be created in single cells and the resulting migratory responses to be studied.[37]

1.1.2.2 Subcellular protein relocalization

Rather than recruiting an effector, an endogenous protein may be relocated away from its site of action to photo-control its activity (Figure 1-2B). This approach relies on being able to replace the endogenous protein with a modified version, without affecting its behavior.[38] Two notable examples of this strategy are the DeLight[39] (depletion with light) system and the LOVTRAP[40] system. DeLight uses the photo-controlled interaction of the photochromes PhyB and PIF6 to inactivate target proteins by fusing PIF6 to a protein of interest and PhyB to an anchoring motif. Nine different anchoring motifs were used, allowing target proteins to be relocated to various locations in the cell, including the cell membrane, endosome and nucleus. Upon irradiation, the target protein associates with the anchored protein within seconds. LOVTRAP is a conceptually similar strategy that uses light-induced dissociation to de-sequester proteins.[40] Wang *et al.* used mRNA selection to find a small protein (Zdk) that binds *As*LOV2 in the dark, developing a synthetic photo-controlled protein-protein interaction. Either Zdk or LOV can be linked to an anchoring motif, and the protein of interest is fused to the binding partner. The ability to fuse these sub-cellular photo-controlled relocalization tags to a variety of target proteins makes this strategy highly appealing for its generalizability.

Another interesting subcellular relocalization strategy is the development of photo-controlled nuclear shuttles. Several methods have been developed to photo-control protein import or export from the nucleus using phytochromes[41] and LOV scaffolds.[42,43] For example, Niopek *et al.* created a nuclear export sequence (NES)/*As*LOV2 Jα helix hybrid and fused this construct to histone H2B, thereby preventing its exit from the nucleus. Irradiation allows the NES sequence to bind and sequester the major nuclear export receptor, CRM1, thereby blocking nuclear export of CRM1 cargoes.[43,44] Yumerefendi *et al*.[42] caged a NES using a novel dimeric LOV-based design in which the NES peptide was fused with the Jα helix of one LOV protein and the A'α helix of a second LOV protein. Photo-controlled nuclear shuttles are attractive due to their rapid kinetics, reversibility and generalizability; however, these tools require cell-type optimization. To ensure reversibility, the constructs contain localization motifs opposing the caged motif (*i.e.* a nuclear localization sequence (NLS) to oppose a NES), but the relative strengths of the motifs change in different cell types.

1.1.2.3 Light-induced protein sequestering

Photo-controlled inhibition of a protein can be accomplished by light-induced formation of artificial soluble aggregates (Figure 1-2C). Lee and colleagues developed a light-induced clustering method called LARIAT (light-activated reversible inhibition by assembled trap)[45] that uses Cry2–CIB1 interactions. Cry2 was fused to a protein of interest and CIB1 was fused to the C-terminal domain of CaM-KIIα, which forms a soluble oligomer of 12 identical subunits. Blue light-induced association of Cry2–CIB1 resulted in rapid (~30 s) formation of soluble aggregates. Upon removal of the light, the aggregates dissociated with a half-life of ~5 min. This approach was used to sequester Vav2, an activator of Rac1. In cells over-expressing Cry2-Vav2, global lamellipodia formation could be attenuated by irradiating cells with blue light, inducing clustering of Cry2-Vav2. The applicability of the system was extended by fusing a GFP-binding single domain antibody to Cry2, allowing in principle any GFP-tagged protein to be sequestered.[45]

The Tucker group discovered a point mutant of Cry2, designated Cry2olig, that exhibited greatly enhanced blue light-induced homo-oligomerization,[46] obviating the need for a scaffold protein to induce sequestration. In this single component system, a protein of interest is fused to Cry2olig

to enable blue light-triggered clustering of the protein of interest. The clusters dissociate in minutes ($\tau_{1/2}$ = 23 min) in the dark. Here too, presumably fusion to an affinity tag (*e.g.* a GFP-binding antibody) could be used as a general handle to permit clustering of tagged targets.

1.1.3 Challenges associated with photo-switchable protein development

The design and engineering of photo-controlled proteins has seen impressive recent success; however, despite this success, developing and optimizing photo-controlled proteins is time-consuming, case-specific, and can yield unexpected results. Numerous caged variants of nuclear export sequences had to be screened to find ones with suitable switching behavior.[43] Cunniff and colleagues[24] developed a photo-controlled kinase inhibitor by inserting an inhibitor in the linker between the *As*LOV2 core and the Jα helix; however, introducing other sequences in this location proved unsuccessful.[23] The rationally designed opto-DN-CREB inhibitor of the transcription factor CREB, while effective, had activity opposite to the original design.[25] Efforts to develop general methods to engineer photo-controlled proteins are a current focus.

Although proven to be highly successful, photo-controlled caging strategies have several limitations. To date, most caged constructs are designed as Jα helix chimeras, which involves extensive engineering, or as C-terminal fusions, where the linker design has a major impact on success.[47] Furthermore, not all proteins may be amenable to caging in this manner; efforts to develop alternative caging scaffolds are important to pursue. The caging by circularization approach is appealing for its generality, although it may lead to oligomer formation under certain conditions, and is currently limited by the number of well-characterized dimerizing systems that undergo light-induced dissociation rather than association.

Protein relocalization strategies have the potential to be highly generalizable, enabling photo-control of a variety of targets with minimal engineering efforts; however, this approach also has limitations. Not all protein activities would be expected to be inhibited through sequestration and relocalization approaches. For example, an enzyme that acts on a diffusible substrate would still be expected to function when sequestered, albeit with altered kinetics. Furthermore, relocalization strategies that require the expression of two components require appropriate matching of the expression levels of the constructs, which is a complication not present with

single-component designs.[40] Expression level, as well as affinity, influences the interaction strength of partners in a given application.

The design and engineering of proteins in general has greatly benefited from the development of effective library-based screening and selection approaches. The application of directed evolution to engineer photo-switchable proteins, however, is hampered by a lack of methods available for evaluating switchable protein activity. In order to find sequences that are functional in the light, but not in the dark (or *vice versa*), switchable protein variants must be evaluated both for a particular activity and the lack thereof. Despite this challenge, there have been several recent reports of photo-switchable protein development using directed evolution with very successful outcomes.[15,16,33,40,48] For example, the Kuhlmann group used directed evolution to optimize the dynamic range of a photo-switchable construct that they had previously developed using rational design. The rationally designed interaction pair had an eight-fold change in light/dark affinity;[22] optimization of the construct using directed evolution yielded a sequence with a ~36-fold change in affinity.[33]

Protein engineering by directed evolution is an iterative two-step process involving library generation, followed by high-throughput analysis of the resulting variants to identify sequences with improved function. Strategies for library generation are discussed below, followed by an overview of the methods available for high-throughput evaluation of switchable proteins.

1.2 Library generation

The development of high-quality libraries of randomized protein-encoding DNA sequences is a field unto itself, with continual development and refinement of techniques (for an interesting review see ref. 49). The vastness of protein sequence space (*i.e.,* the concept of all theoretical possible sequences encoded by a protein[50]) poses a fundamental challenge in library design. For example, a polypeptide of six amino acids in length has a sequence space of 64 million possible amino acid sequences ($20^6 = 6.4 \times 10^7$). While obviously many of the variants would not encode useful proteins, the example illustrates the vast number of combinatorial possibilities attainable with a set of 20 amino acids. Furthermore, due to the degeneracy of the genetic code, these 64 million protein sequences could be encoded by over 68 billion ($64^6 = \sim 6.9 \times 10^{10}$) unique DNA sequences. The phenomenal number of possible sequences highlights the need to develop

intelligent libraries, in order to encode maximum functionally rich diversity with the minimum library size.[51]

1.2.1 Error-prone polymerase chain reaction

Error-prone polymerase chain reaction (epPCR)[52] is a pioneering method for the incorporation of random sequence variation into a protein-encoding DNA sequence. A conceptionally simple strategy, this method is based on the PCR amplification of a gene under conditions designed to decrease polymerase fidelity. Due to mutational bias of the polymerase and the redundancy of the genetic code, epPCR only samples a very limited fraction of the protein sequence space. However, despite this limitation the method enjoys wide-spread popularity due to its simplicity, and has been proven to be very effective, in particular when used in combination with other randomization methods.[51,53]

1.2.2 Shuffling methods

DNA shuffling is another pioneering method for library generation. Introduced by Stemmer in 1994,[54,55] this method generates genetic diversity by recombining related genes in a process mimicking meiotic recombination. The ability to recombine homologous genes, or mutated copies of a single gene, makes it possible to combine beneficial mutations from multiple templates or to revert deleterious mutations back to the wild-type sequence. In the original protocol, homologous genes are fragmented through a controlled digest with DNaseI and subsequently reassembled using a primer-less PCR reaction dependent on self-priming of the homologous fragments. Numerous methods have been developed based on this idea,[53,56] often proposing alternatives to DNaseI digestion, which is notoriously difficult to control.

1.2.3 Degenerate codons

A third important strategy for library generation is the use of oligonucleotides with degenerate codons.[57] This method enables the development of focused libraries, where randomization is targeted to a set of residues selected through structural data, computational modelling and/or sequence homology. The use of degenerate codons decreases library size, removes unwanted mutational bias (*i.e.* from synonymous codons) and enables stop codons to be avoided. An impressive example of this approach was demonstrated by Sidhu and colleagues, who used

degenerate codon schemes to develop antigen-binding fragments (Fab) with affinities of 2-10 nM.[58] Interestingly, the authors observed a striking enrichment in Tyr residues mediating contacts with the antigen, and they drew attention to the fact that Tyr is particularly well-suited to mediate protein-protein interactions, due to its intermediate hydrophilicity and ability to participate in a variety of interactions.[58] Some examples of degenerate codon schemes are shown in Table 1-1.

Table 1-1: Degenerate codon schemes.

Codon	**Type**	**Amino acids**	**No. of stop codons**	**No. of codons**	**No. of residues**
NNN	All 20 residues	All	3	64	20
NNK	All 20 residues	All	1	32	20
NDT	Broad mixture	R/H/D/C/G/V/L/F/I/S/N/Y	None	12	12
DVT	Hydrophilic	C/D/G/N/S/T/Y/A	None	9	8
NTT	Hydrophobic	F/I/L/V	None	4	4
DSC	Small	A/C/G/S/T	None	5	5

Where N = A/C/T/G, K = G/T, D = A/G/T, V = A/C/G, S = G/C

1.2.4 Libraries altering protein architecture

In addition to mutagenesis strategies targeting amino acid identity, strategies have been developed to create libraries producing more substantial changes in protein architecture, such as randomized circular permutations and domain insertions. These techniques have been used, often in conjunction, to develop switchable protein biosensors,[59 & ref. therein] and have recently seen increased use in photo-switchable protein development, with several examples of constructs designed using circular permutations[14–17] and domain insertions.[48,60]

While circular permutations can be rationally designed, by using structural data to guide the creation of new N- and C-termini, methods to develop circular permutation libraries have been introduced, enabling a high-throughput approach. Guntas and colleagues created libraries by treating a circularized coding sequence of a gene of interest with dilute DNaseI to create randomized double-stranded DNA breaks to linearize the template (Figure 1-3A).[61] This method, however, was reported to be very challenging due to its reliance on DNaseI activity.[62] An alternative method was introduced that uses PCR to amplify fragments of a tandem-duplicated gene of interest using an array of primer pairs (Figure 1-3B).[62] Although only semi-randomized,

this method is technically much less challenging, and the increased experimental control allows library construction to be rationally guided if desired.

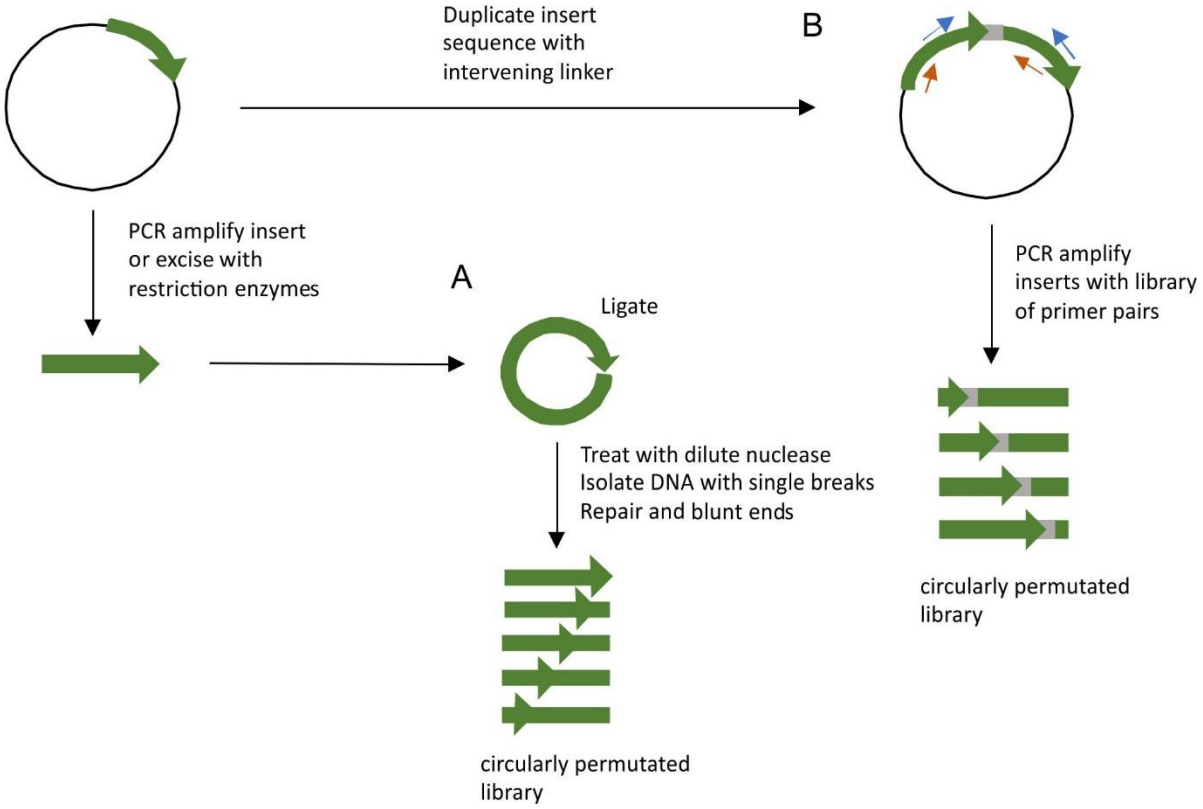


Figure 1-3: Strategies to create circularly permutated gene libraries.
(A) Circularly permutated gene libraries have been created by subjecting a circularized gene template to a controlled nuclease digest, in order to randomly linearize the coding sequence.[61] (B) Alternatively, a circularly permutated gene library can be created by PCR amplifying gene segments from a tandem-duplicated gene template, using an array of primer pairs.[62] For simplicity, only two primer pairs are shown. In this example, the tandem duplicated genes are connected via a linker sequence (grey).

It is difficult to predict *a priori* how domain insertions should be constructed to yield optimal coupling between domains, making a high-throughput library-based approach very attractive. Domain insertion libraries have been created using controlled nuclease digests (using DNaseI[63] or S1 nuclease[64]) to create randomized double-stranded breaks in a DNA template, followed by blunt-ended ligation with the sequence for domain insertion (Figure 1-4A). An alternative method was demonstrated using the engineered transposon MuDel, a transposon with low target sequence preference that contains flanking terminal restriction enzyme sites to enable transposon

excision.[65] MuDel can be used to create domain insertion libraries by subjecting a plasmid encoding a gene of interest to random transposition, followed by excision of the transposon and blunt-ended ligation with the sequence for domain insertion (Figure 1-4B).[66] Semi-randomized domain insertion libraries can also be created using multiplex inverse PCR, where an array of divergent primer pairs are used to amplify and thus linearize a plasmid, preparing it for ligation with the domain insertion sequence (Figure 1-4C).[48,62] The use of PCR to linearize the template sequence allows for simultaneous incorporation of insertions, deletions or linker libraries at the insertion site,[62] which has been shown to be important in conveying allostery.[66]

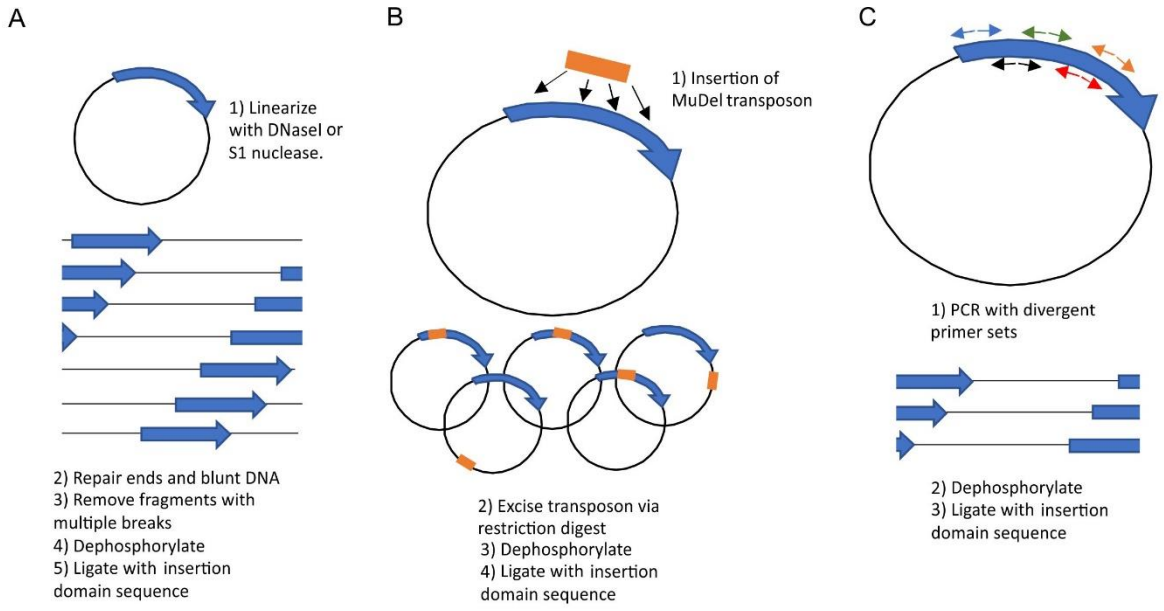


Figure 1-4: Strategies to generate randomized domain insertion libraries.
(A) A plasmid encoding a gene of interest is linearized via random double-stranded DNA breaks, achieved through a controlled nuclease digest. The prepared linear plasmids are then re-circularized via ligation with the sequence of the domain chosen for insertion. (B) A plasmid encoding the gene of interest is treated with MuDel transposon, to achieve randomized transposon insertion (orange bar). Subsequent excision of MuDel using restriction enzymes results in template linearization. Ligation of the insertion domain sequence results in plasmid re-circularization. (C) The gene of interest is linearized using divergent primer pairs, followed by ligation with the insertion domain sequence to re-circularize the parent plasmid. In all cases, the insertion domain sequence may be ligated in the form of the native coding sequence, or as a library of circularly permutated variants (see Figure 1-3), a strategy which has been found to be highly effective.[67]

The domain insertion libraries dependent on randomized double-stranded DNA breaks (through controlled nuclease activity or transposition) produce plasmid linearization that could occur anywhere in the plasmid; as a result, many fusions with the sequence for domain insertion occur outside of the target coding sequence. Furthermore, since the DNA cleavage could occur in any reading frame, only 33% of the insertions within the target gene would be expected to occur in-frame. Identification of useful fusions, despite this large background, requires a robust screening or selection assay to enable analysis of the resulting protein variants.

The ability to select for useful variants from a background of non-functional proteins is a requirement of any directed evolution project. Recovering improved protein variants from a library can be likened to searching for a needle in haystack.[68] To address this problem, numerous creative strategies exist to efficiently search libraries for useful sequences.

1.3 Evaluating protein libraries

A critical component for evaluating libraries is the ability to link a phenotype (*i.e.* a protein function or characteristic) with the corresponding genotype (*i.e.* the DNA sequence). This can be accomplished *in vivo*, using bacterial cells, or through *in vitro* means, where a protein is tethered to its genetic sequence or coupled as part of a viral particle. What follows is a discussion of available library evaluation methodologies, including protein display methods and bacterial genetic circuits.

1.3.1 Protein display methods

Protein display methods couple a protein phenotype with the corresponding genotype using a single cell or viral particle, or by tethering the protein to a form of the genetic coding sequence (such as mRNA). These display methods are most frequently used for affinity selection; tethering an interaction partner to a solid support enables a library of displayed protein variants to be evaluated for binding to the interaction partner using an affinity purification-based approach.[69] The recovered protein variants are frequently subjected to subsequent rounds of affinity-based selection with increasing stringency and further rounds of mutagenesis.[69] These methods are amenable to development of switchable binding interactions, because panning rounds can be performed using on or off conditions (*i.e.* light/dark or in the presence/absence of a ligand).

Recently, phage display was used to develop synthetic binding partners for selected photoactive proteins.[16,33] In phage display, a protein of interest is expressed as a fusion with a viral coat protein, such that upon assembly of the viral particles the foreign protein is displayed on the surface of the phage.[69] Reis and colleagues used phage display to evaluate a library constructed using the GA domain as a protein scaffold, to recover synthetic GA domain variants binding selectively to the light- or dark-state of two different photoactive proteins.[16] Likewise, the Kuhlman group used phage display to evaluate computationally designed, focused *As*LOV2 libraries, in order to develop photo-controlled effectors with enhanced dynamic range (~35-fold and ~60-fold changes in K_d).[33]

In a related study, Wang *et al.* generated a binding partner for *As*LOV2 using mRNA display.[40] mRNA display is an *in vitro* display method where a newly translated protein is covalently attached to its encoding mRNA via an adaptor molecule, creating the required phenotype-genotype link to enable rounds of affinity-based selections and mutagenesis.[70] The *in vitro* nature of this method eliminates the need to transform cells, which enables the analysis of larger libraries.[70] Wang and colleagues evaluated a library of 5×10^{13} unique protein variants and recovered a variant with an affinity that is >150 times higher for the dark-state of LOV versus the light-state.[40]

Display methods have proven to be very powerful techniques and are expected to be used increasingly in the development of photo-switchable protein tools. However, not all proteins are amenable to all types of display. For example, difficulties have been encountered in the display of PYP and LOV on the surface of phage (S. Hoffmann [PYP] and R. Woloschuk [PYP and LOV], personal communication). Kuhlman *et al.* displayed *As*LOV2 on the surface of phage but evaluated a relatively small library (5×10^7 variants), potentially also due to poor display efficiency.[33]

Protein libraries can alternatively be evaluated using genetic circuits in yeast and bacteria. Although the *in vivo* nature of these methods restricts the practical library size limits, evaluating protein variants in the context of a host proteome can be beneficial as it imposes additional pressure for solubility, stability and selectivity of the library variants.

1.3.2 Genetic circuits

Genetic circuits enabling protein evaluation are often classified as screens or selections. In a genetic circuit designed as a screen, a protein function of interest is coupled to a quantifiable output to permit identification, such as the production of pigment, fluorescence or luminescence. A screen does not enhance or inhibit the growth of any particular phenotype; it merely offers a rapid form of detection. A selection system in contrast, is a genetic circuit where a desired protein function is coupled to the survivability of a cell. The ability to select for cells with a phenotype of interest facilitates the evaluation of larger libraries, since members without the function of interest are eliminated. The practical size limits for agar plate screens and selections have been estimated to be ~10^5 and ~10^9 colonies respectively.[71]

1.3.2.1 Genetic circuits to screen for desired protein function

Genetic screens in general can be used in the development of switchable proteins, because library evaluations can be performed under conditions supporting the on- or off-state of the protein, with each state yielding either the presence or absence of the reporter gene. In a classic example of a screen, Ohlendorf and colleagues evaluated a library of light-regulated histidine kinase constructs by designing a system where catalytically active protein led to the transcription of a red fluorescent protein.[47] The ability to rapidly assess protein activity via fluorescence in light and dark conditions, enabled the evaluation of a large number of library variants, yielding both a light-regulated kinase as well as important insights regarding linker architecture in signal transmission. A fluorescent screen was also used to support the development of a switchable dCas9.[48] Here, a guide RNA targeting the promoter sequence of a red fluorescent protein was used, enabling fluorescence-based detection of dCas9 DNA-binding activity. In this study, fluorescence-activated cell sorting (FACS) was used to facilitate evaluation of the library, allowing the screen to be performed with higher throughput.

1.3.2.2 Genetic circuits to select for desired protein function

The majority of established selection systems cannot be used in the development of switchable proteins, because the circuits cannot control for a lack of protein activity in the off-state. It should be acknowledged that it would be possible to overcome this using replica plating; however, this labor-intensive step decreases the throughput of the circuit, which eliminates the

advantage of a selection over a screen. Despite this, there have been a few selection systems that have been creatively designed to enable evaluation of switchable activity (*vide infra*).

Selection systems are most commonly designed as positive selections, where a target protein function is linked to cell survivability such that cells survive only when the desired function is above a set threshold (Figure 1-5A). This is commonly achieved using antibiotic resistance markers. A classic example of a positive selection is the yeast-two-hybrid (Y2H) assay, which is used to select for protein-protein interactions.[72] In this method, a DNA-binding domain and a transcriptional activator are each fused to one half of a putatively interacting protein pair. Interaction of the two test proteins localizes the transcriptional activator to a target DNA-binding site, leading to transcription of a chosen reporter gene. Due to its flexibility and ease of use, Y2H has been used extensively in the detection and development of protein-protein interactions,[73] with numerous protocol modifications enabling use in different hosts[74] and detection of alternative interaction complexes.[75,76] Notably, bacterial reverse two-hybrid approaches have been introduced, where the association of interacting protein pairs reconstitutes a transcriptional repressor.[77,78]

Selection systems can also be designed as negative selections, where cell survivability is linked to the protein activity such that cells only survive when the target activity is below a selection threshold (Figure 1-5B). In other words, the assay is designed to select against a particular activity. Often, the activity threshold in negative selections is very low, which has earned negative selectable markers the nickname ‘suicide genes.’ An example of a negative selection agent is levansucrase (encoded by the gene *sacB*) which confers sensitivity to sucrose.[79,80] Levansucrase converts sucrose to levan, which accumulates in the periplasm of Gram-negative bacteria and leads to cell lysis.[80]

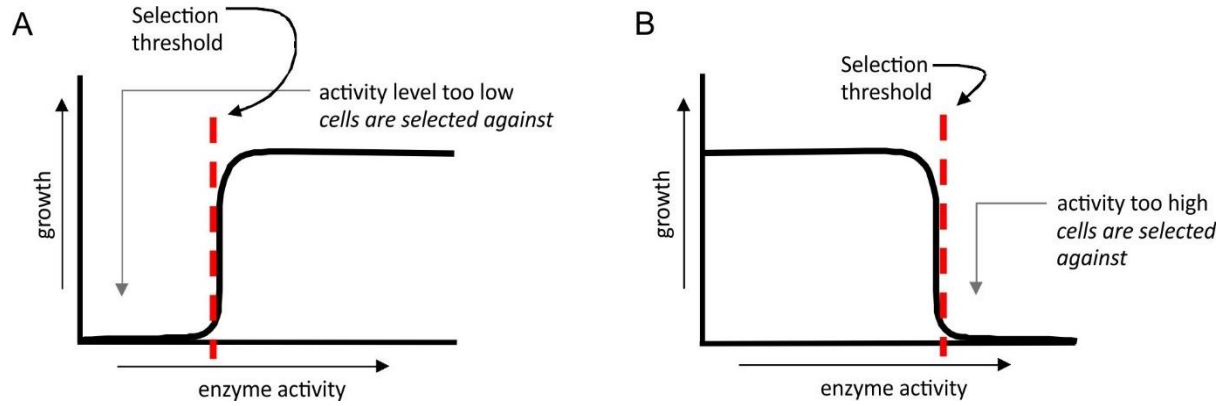


Figure 1-5: Schematic illustrating positive and negative selection.
In positive selection (A) an activity of interest (*e.g.* enzyme activity) is selected for. In such a design, only cells with an enzyme variant that has an activity level above the selection threshold are able to grow. This is the more common design used for selection systems. In negative selection (B) the activity of interest is selected against. In this design, only cells with an enzyme variant with an activity level below the selection threshold are able to grow.

1.3.2.3 Methods to select for switchable proteins

Positive and negative selection markers can be combined to enable bi-directional selection, where selection can be imposed either for or against gene expression. This was demonstrated by Arnold and colleagues through co-expression of a chloramphenicol (Cm) resistance gene and β-lactamase inhibitor peptide (BLIP) in a synthetic operon, using cells with constitutive β-lactamase expression (Figure 1-6).[81] Positive selection for the operon activity could be imposed using Cm pressure, while negative selection for the operon activity could be imposed using ampicillin (Amp) pressure. This genetic circuit was used to evaluate an epPCR library of a transcription factor, in order to identify a variant that responded specifically to a non-natural ligand.[81] Engineering specificity presents a problem analogous to that faced in developing switchable proteins; a lack of activity must be ensured unless in the appropriate context.

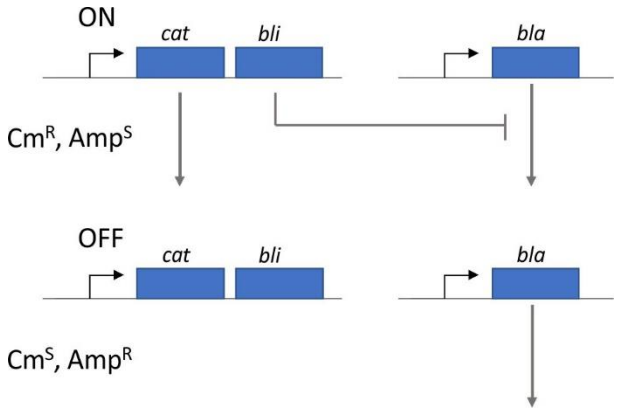


Figure 1-6: Combined use of positive and negative selectable markers.
The combined use of positive and negative selectable markers allows for bi-directional selection of operon activity.[81] In cells with constitutive expression of a β-lactamase gene (*bla*), selection for expression of an operon encoding a gene for Cm resistance (chloramphenicol acetyltransferase, *cat*) and β-lactamase inhibitor peptide (*bli*) was achieved using Cm pressure. Selection against operon expression was achieved using Amp pressure, to prevent expression of BLIP and inhibition of β-lactamase. A subscript 'R' indicates resistance and a subscript 'S' indicates sensitivity.

Expression of negative selectable markers has been reported to suffer from a high rate of tolerance development, since the circuit inherently selects for cells with mutations inactivating the negative selectable marker.[79,82] In one study, negative selective pressure was applied using two negative selectable markers, which dramatically decreased the occurrence of false positive colonies.[83] In another example, the gene sequences for a negative selectable marker and a positive selectable marker were fused together via a flexible linker (with the positive selectable marker sequence on the 3′ end) to reduce the recovery of false positives.[84]

Several dual selection schemes have been reported that function with a single marker. TetA, which encodes a membrane channel functioning as a tetracycline (Tet) / H^+ antiporter,[85] can be used as a dual selectable marker,[83,86,87] because it conveys resistance to Tet but sensitizes cells to toxic metals and lipophilic chelators (such as fusaric acid). Similar to this example, the outer membrane channel TolC has also been shown to function as a dual selectable marker.[88] TolC is an essential component in bacterial multidrug efflux pumps, which confer resistance to various small molecules, such as sodium dodecyl sulfate (SDS); SDS pressure therefore creates a positive selection for TolC expression. The expression of TolC, however, renders cells susceptible to the bacteriocin colicin E1, which is dependent on TolC for access to the

cytoplasm. Several other dual selectable markers have been reported using auxotrophic strains. For example, in strains deficient in thymidine kinase activity (*tk*-), herpes simplex virus thymidine kinase (hsvTK) can be used as a dual selectable marker. Thymidine kinase is the first enzyme in the salvage pathway for synthesis of the nucleotide dTMP. When the *de novo* dTMP pathway is inhibited using 2′-deoxy-5-fluorouridine (5FdU), cells are dependent on the salvage pathway; hsvTK activity rescues *tk*- cells from thymidine starvation.[89] Negative selection pressure can be applied using mutagenic nucleosides, such as ganciclovir; hsvTK activity leads to promiscuous incorporation of the nucleosides into cellular DNA and cell death. A similar strategy has been used with *galK,* encoding galactokinase; expression of *galK* rescues galactose auxotrophs but makes cells susceptible to the toxic galactose analog, 2-deoxy-galactose (DOG).[90] Table 1-2 shows a summary of some available dual selectable markers.

Table 1-2: Summary of dual selectable markers for Gram-negative bacteria.

Gene	**Gene product**	**Positive selection**	**Negative selection**	**Notes**	**References**
tetA	Tet efflux pump	Tet	toxic metals or fusaric acid		Refs. 83,86,87
tolC	outer membrane channel	SDS	colicin E1	Requires Δ*tolC* strain	Ref. 88
galK	galactokinase	galactose minimal media	DOG	Requires Δ*galK* strain	Ref. 90
hsv-*tk*	hsvTK	5FdU	ganciclovir	Requires Δ*tk* strain	Ref. 89

Dual selection was recently demonstrated using two positive selectable markers.[91] Hoffmann *et al.* developed a circuit containing two converging antibiotic resistance genes, encoding Cm resistance and spectinomycin (Spec) resistance, with expression controlled by a strong and a weak promoter respectively (Figure 1-7). In this circuit, expression of the Cm resistance gene from the strong promoter prevents expression of the Spec resistance gene from the weak promoter through transcription interference. Placing the DNA-binding site for the Lac repressor (LacI) in the promoter sequence of the strong promoter, allows for bi-directional selection of LacI DNA-binding activity; Cm pressure selects for non-binding, while Spec pressure selects for

binding. This circuit was used to evaluate an epPCR library for LacI variants with an inverted response to the natural ligand, isopropyl β-D-1-thiogalactopyranoside (IPTG).[91]

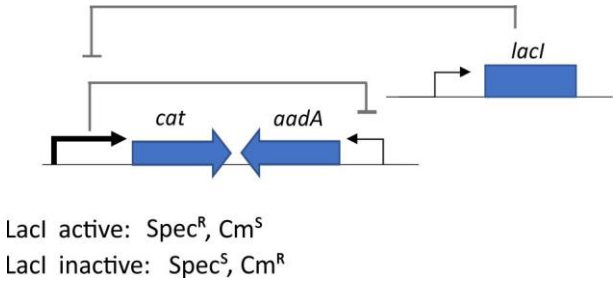


Figure 1-7: A dual selection system with two positive selectable markers.
A dual selection system using two positive selectable markers was developed to enable selection for switchable DNA binding.[91] Convergent genes for Cm resistance (encoded by *cat*) and Spec resistance (aminoglycoside adenylyltransferase, encoded by *aadA*) are expressed using a strong (large arrow) and weak (small arrow) promoter respectively. Expression of the Cm resistance gene blocks expression of the opposing Spec resistance gene. The Lac repressor (encoded by *lacI*), binds to the strong promoter, inhibiting expression of *cat* and allowing expression of *aadA*. A subscript 'R' indicates resistance and a subscript 'S' indicates sensitivity.

By imposing both positive and negative selection pressure for β-lactamase activity, Ostermeier and colleagues developed a tunable bandpass circuit, allowing for selection of various levels of β-lactamase activity using different concentrations of Amp (Figure 1-8).[92] Growth in the presence of Amp creates positive selection pressure for β-lactamase activity. Negative selection pressure for β-lactamase activity was imposed by exposing cells additionally to Tet pressure and supplying cells with a synthetic operon containing a Tet resistance gene with expression induced by a cell wall breakdown product, produced through the presence of Amp. Cells therefore only are Tet and Amp resistant when their β-lactamase activity balances the Amp in the media, such that they survive the Amp pressure, yet sustain a non-lethal degree of cell wall damage, inducing expression of the Tet resistance gene. The Ostermeier group used the ability to select for high or low levels of β-lactamase activity (using correspondingly high or low concentrations of Amp) to develop switchable β-lactamase biosensors.[92–94]

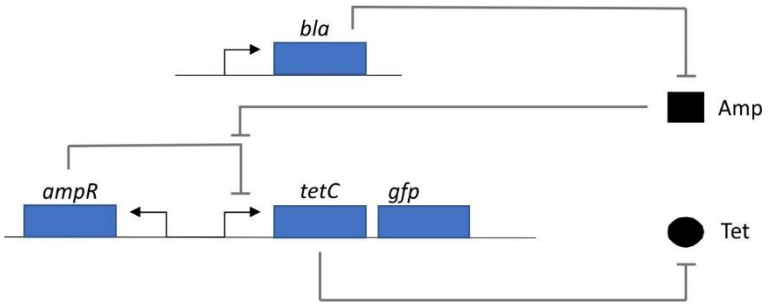


Figure 1-8: A bandpass assay for β-lactamase activity.
A bandpass assay was created by imposing positive and negative selection pressure for β-lactamase activity.[92] Growth in the present of Amp creates positive selection pressure for β-lactamase activity. A synthetic operon with a Tet resistance gene (*tetC*) under the control of a promoter induced by a cell wall breakdown product, imposed negative selection pressure for β-lactamase activity in the presence of Tet. A gene for green fluorescent protein (*gfp*) was included in the synthetic operon to enable screening for operon activity in the absence of Tet selection pressure. *bla* is the gene for β-lactamase, *ampR* is the gene for a transcriptional regulator controlled by a cell wall breakdown product.

Despite the circuits described above, overall there is a paucity of selection systems available for high-throughput evaluation of switchable protein activity. The reported dual selection markers, and combined positive and negative markers circuits, have predominantly been applied in the development of novel genetic circuits[82,84,86,87,89] or have been used in genome engineering to facilitate selection of unmarked modifications.[79,83,88,90] These tools would also be useful for engineering switchable protein activity. The array of available reporter genes and useful genetic circuits set the foundation for the development of additional circuits enabling selection for switchable protein activity.

1.4 Research outline

This thesis presents the development of novel screening and selection approaches to support the development of photo-controlled proteins. In Chapter 2, the development of a fluorescence-based screening system for a photo-switchable DNA-binding protein is discussed. This genetic circuit was designed to support a directed evolution approach towards the optimization of a photo-switchable DNA-binding protein previously developed using rational design by members of the Woolley research group.[95,96] This is followed by the introduction of a tunable selection system for protein-protein interactions, termed the ‘hitchhiker-bandpass’ assay, which allows for the

targeted recovery of protein interactions with a desired affinity range (Chapter 3). The unique ability to select for protein-protein interactions of a targeted affinity could be useful for numerous applications. In Chapter 4, the use of the hitchhiker-bandpass circuit to evaluate small molecule modulators of protein-protein interactions is discussed. In Chapter 5, on-going efforts are outlined towards the development of photo-controlled inhibitors of mammalian transcription factors using the hitchhiker-bandpass circuit. Finally, a synthetic fusion construct enabling *in vivo* analysis of the photocycle of PYP is presented in Chapter 6. This construct was used to study biosynthetic chromophore reconstitution in PYP, which is an important consideration for the use of PYP as a photoactive protein scaffold to develop photo-switchable protein tools.

2 Development of a fluorescence-based screening system for directed evolution of photo-controlled DNA-binding proteins

This chapter was reproduced with permission from Oxford University Press from the following article: M. Mazumder, K.E. Brechun, Y.B. Kim, S.A. Hoffmann, Y.Y. Chen, C.-L. Keiski, K.M. Arndt, D.R. McMillen, G.A. Woolley. An *Escherichia coli* system for evolving improved light-controlled DNA-binding proteins. 2015. *Protein Engineering Design and Selection.* 28(9): 293-302.

Author contributions: MM, KEB, YBK and GAW devised the genetic circuit. MM performed the cloning and statistical analysis. MM, KEB and YBK optimized the circuit and screened samples. KEB, YBK, YYC, C-LK performed purifications and biochemical characterizations. KEB and SAH constructed the library. KMA, DRM and GAW supervised the project. The main text has been modified for consistency.

The article may be accessed via the following link: https://doi.org/10.1093/protein/gzv033

2.1 Introduction

Controlling gene expression remotely with light is becoming a very attractive method in synthetic biology.[97–99] Optical control permits precise manipulation of the timing of transcriptional processes without the complexities associated with time-varying extracellular effector molecules. Optical control of gene expression can be used to explore natural physiological processes with an unparalleled degree of spatiotemporal resolution.[98,100] In some cases, optical control of transcription has been achieved by co-opting natural light-responsive dimerization elements to bring together a DNA-binding protein with a transcriptional activator domain.[98,101,102] Alternatively, a structure-guided approach may be taken in which a light-induced conformational change produced in a photoreceptor protein is propagated in a manner that affects the activity of a fused transcription factor.[95,96,100,103] For example, we reported the structure-guided development of a light-controlled DNA-binding protein, GCN4-Δ25PYP, based on a fusion between PYP and the yeast transcription factor, GCN4.[96] This type of structure-guided approach can also be adapted to the control of endogenous transcription factors.[2,25]

A light-controlled DNA-binding protein needs to recognize a specific DNA sequence and bind it

with high affinity in one state (usually the light-state) and much lower affinity in the other state (dark-state). Often, initially designed constructs have less than ideal degrees of switching, *i.e.* the difference in affinity between the two states is small. High-throughput methods for evolving proteins with light-switchable activity have only recently begun to be developed.[33,47] Although a large number of methods for detecting protein/DNA binding have been developed, only some of these are amenable to high-throughput screening.[104,105] An additional requirement for the screening or selection for light-switchable activity is that side-by-side assays must be performed under light and dark conditions. If the screening can be done *in vivo*, other factors, such as protein stability, can be simultaneously screened. *In vivo* screening methods may also be adapted as selection methods in which survival of the bacterium is tied to the function of the switchable transcription factor, thereby allowing even larger library sizes to be assayed.

Here we report the development of a bacterial system in which a light-switchable DNA-binding protein affects the production of a fluorescent reporter protein. Optimization of a number of parameters allowed the successful recovery of light-switchable sequences from a small test library. The system is amenable to high-throughput screening using FACS. In addition, the reporter system is general enough that it could be adapted to the directed evolution of essentially any DNA-binding protein.

2.2 Materials and Methods

2.2.1 Strains, media and reagents

The *Escherichia coli* strain DH5αZ1 (Expressys[106]) was used for cloning and experiments with the tdTomato circuit. This strain has the genes for the lactose operon repressor (LacI) and the tetracycline resistance operon repressor (TetR) integrated into the genome, allowing regulation of genes expressed with *pLlacO-1* and *pLtetO-1* promoters, respectively. *E. coli* BL21*(DE3) was used for protein expression for purification. A standard electroporation protocol[107] was used to transform all plasmids. MOPS EZ Rich Defined Medium kit (Teknova) was used for all light-responsive experiments. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. PfuTurbo Hotstart PCR Master Mix was purchased from Stratagene. Doxycycline (Dox) was purchased from Sigma-Aldrich. Luria-Bertani (LB) agar, LB broth and IPTG were purchased from Bioshop Canada Inc. The 4-hydroxycinnamic acid S-phenyl ester

activated chromophore was synthesized as described previously.[95] Addition of this chromophore (30 μg/mL) leads to complete conversion of apo protein to holo protein with all PYP constructs tested here.

2.2.2 Plasmid construction

The plasmid backbones, pZE12luc and pZA31luc (Expressys), were used to construct respectively, the reporter plasmid pZ12AP1-tdT, and different variants of the PYP expression plasmid (denoted pZA31-PYP). A plasmid with the gene for the red fluorescent protein tdTomato was purchased from Clontech and this gene was cloned into pZE12luc by amplifying via PCR with primers containing *KpnI* and *XbaI* restriction sites. The pZE12luc vector was cut with KpnI and XbaI to remove the luciferase gene and insert the tdTomato gene. The resulting plasmid was then digested with XhoI and KpnI for cloning the modified promoter. The *pLtetO-1* promoter sequence with the AP-1 site and flanking *XhoI* and *KpnI* restriction sites was purchased as oligonucleotides from Eurofins. The strands were hybridized and then digested with XhoI and KpnI. The modified promoter was ligated into pZE12luc vector with tdTomato, resulting in the plasmid pZ12AP1-tdT (see Figure 2-1). The *pLlacO-1* promoter sequence with flanking *AatII* and *PacI* restriction sites was purchased as oligonucleotides from Eurofins. The strands were hybridized and inserted into the pZA31luc plasmid backbone between the *ActII* and *PacI* restriction sites. All GCN4-Δ25PYP gene variants were cloned into the pZA31luc plasmid backbone between *KpnI* and *PacI* restriction sites, resulting in the pZA31-PYP plasmid series (see Figure 2-1). All plasmids were confirmed by sequencing.

2.2.3 Screening protocol and statistical analysis

We prepared glycerol stocks of *E. coli* DH5αZ1 transformed with the reporter plasmid and each of the variants of the PYP plasmid. For each experiment, a fresh LB agar plate was streaked from the glycerol stock. A single colony was picked from the plate for each version of GCN4-Δ25PYP for overnight growth in LB media. Fresh colourless MOPS EZ Rich Defined Media was prepared and supplemented with 100 μg/mL Amp, 35 μg/mL Cm, 30 μg/mL chromophore, 1 mM IPTG and 250 ng/mL Dox. Each of the overnight cultures was diluted 100-fold into 10 mL of the supplemented colourless MOPS EZ Rich Defined Media. Six replicates of each variant of GCN4-Δ25PYP were pipetted into two identical 96-well plates (200 μL/well). One plate was

placed under blue light and the other in the darkness, with all other conditions (temperature, agitation, humidity, oxygenation) kept identical. Cells were grown in a custom-built incubator that maintained identical conditions for both plates. Plates were incubated at 30°C with 430 rpm shaking using an LSE digital microplate shaker (Corning) for 5 - 6 h. Light-exposed cells were grown under blue LED arrays (450 nm) mounted 16.5 cm above the plates to give an intensity of 0.33 - 0.45 mW/cm^2 at the surface of the plates. An Infinite M1000 Pro plate reader (Tecan) was used to measure both bulk tdTomato fluorescence intensity (excitation 554 nm and emission 581 nm) and optical density at 600 nm (OD_{600}) for each well after 5 - 6 h of growth. All fluorescence values were normalized using OD_{600} values to compensate for small differences in cell growth rates between wells. Two-tailed *t*-tests ($P < 0.05$) were performed to test for the differences between the light and dark results for library screening experiments. Average deviation was used for indicating error bars where the sample size was ≤ 3; otherwise standard deviation was used.

2.2.4 Fluorescence-activated cell sorting

Single colonies of *E. coli* DH5αZ1 containing a switchable photo-controlled DNA-binding construct, a light-insensitive binding-control or a light-insensitive non-binding control (GCN4-Δ25PYP-v2 K143F, GCN4-PYP or Δ25PYP respectively) were grown overnight in LB media containing Cm (35 µg/mL) and Amp (100 µg/mL). A 100 µL volume of each of these cultures was sub-cultured into 10 mL MOPS EZ Rich Defined Media containing the above antibiotics, as well as chromophore (30 µg/mL) and IPTG (1 mM). After growing for 30 min at 30°C, Dox (200 ng/mL) was added to induce tdTomato expression. The samples were then split into two: one set of samples was placed under blue light and the other was protected from light. The samples were grown until the OD_{600} reached 0.3 - 0.5 (~4 h post-Dox induction), at which point an 850 µL culture volume was pelleted and re-suspended in 1.5 mL of sterile phosphate buffered saline (PSB) solution. The tdTomato fluorescence of each sample was measured using a BD FACSAria II™ instrument (BD Biosciences).

2.2.5 *In vitro* characterization of mutants

2.2.5.1 Protein mutagenesis and expression

Site-specific mutations to the GCN4-Δ25PYP construct were made using the Stratagene QuikChange protocol as described in Fan *et al.*[96] The gene for the PFY mutant that was found

via the *in vivo* screen (discussed later) was synthesized by Bio Basic. Proteins were expressed in *E. coli* BL21*(DE3), reconstituted with chromophore and purified as described previously.[96] The purity of the samples was confirmed using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) and electrospray ionization mass spectrometry (ESI-MS).

2.2.5.2 *In vitro* DNA-binding assays

All samples for electrophoretic mobility shift assays (EMSA) were prepared under red light. Cy-3 labelled AP-1 DNA (10 nM) was mixed with a GCN4-Δ25PYP mutant protein at final concentrations of 1 nM, 5 nM, 10 nM, 25 nM, 50 nM, 75 nM, 100 nM, 125 nM, 250 nM, 375 nM, 500 nM, 750 nM, 1 μM and 3 μM (unless otherwise indicated) in a total volume of 100 μ L EMSA buffer (20 mM Tris/HCl, pH 7.5; 100 mM KCl, 3 mM $MgCl_2$, 0.1% Triton, 5% glycerol, 100 μg/mL bovine serum albumen [BSA] and 1 mg/mL sheared salmon testes DNA). Samples were incubated overnight in the dark at 4 or 20°C. The following day 20 μL of the samples were run in the dark on an 8% native polyacrylamide gel containing 1× Tris-acetate-EDTA (TAE), pH 7.5 buffer which was also used as the running buffer. Gels were run at 4°C for 105 min at 300 V using an Emperor Penguin Water cooled dual-gel electrophoresis system. The gel was scanned with a green laser (532 nm) on a GE Healthcare Typhoon 9400 or Bio-Rad PharosFX Fluoimager. The remaining samples were irradiated with a Luxeon III Star LED Royal Blue (455 nm) for 5 min and then another set of 20 μL aliquots were loaded onto another 8% native polyacrylamide gel prepared as above. The gel was run and imaged as described above except that two Luxeon III Star LED Royal Blue (455 nm) each at 340 mW (700 mA) were used to irradiate the gel continuously during the run. Heat from the LEDs was dissipated by the temperature-regulated gel tank buffer. To determine the reversibility of DNA binding, the remaining samples were incubated in the dark overnight and then 20 μL aliquots were run in the dark on another 8% native polyacrylamide as detailed above. DNA binding was quantified by analysis with ImageQuant 5.0 or ImageLab (Bio-Rad) software. Using Igor Pro software, each dataset was fit to the Hill equation to determine Hill coefficients and the apparent K_d values, the concentration of protein required for half-maximal binding to the AP-1 site. Hill coefficients were found to vary between 1.6 and 2.6 when left unconstrained in the fitting. Hill coefficients were then constrained to 2 and apparent K_d values were recalculated. Hill coefficients near 2 are

expected since GCN4 binds as a dimer to DNA. The average and standard deviation from fits of experiments performed in triplicate are reported.

2.2.5.3 Thermal relaxation rates measured by UV/vis spectroscopy

UV/vis spectra of the PFY mutant were recorded in 1× TAE, 100 mM NaCl (pH 7.5) at 20°C in a 1.0 cm pathlength quartz cuvette using a PE Lambda 35 spectrophotometer. Protein concentrations were determined using an extinction coefficient of 45 mM^{-1} cm^{-1} at λ_{max} (~446 nm).[95] Samples for thermal relaxation were prepared at a final protein concentration of 5 μM in 1× TAE, 100 mM NaCl (pH 7.5). The protein samples with non-specific DNA consisted of the same 5 μM protein in 1× TAE, 100 mM NaCl (pH 7.5) but with sheared salmon testes DNA added to a final concentration of 1 mg/mL. Samples with target DNA contained in addition to the salmon testes DNA, an equimolar (*i.e.* 5 μM) amount of annealed AP-1 DNA. Irradiation of the protein sample was conducted using a Luxeon III Star LED Royal Blue (455 nm) Lambertian operating at ~340 mW (700 mA). Relaxation rates at 20°C were measured by recording the change in absorbance at 350 nm. Data were fit to single-exponential functions to determine rate constants.

2.2.5.4 Library generation

Using the PoPMuSiC-2.0 software (http://dezyme.com/), three sites in GCN4-Δ25PYP-v2 were found that were predicted to have a strong influence on the dark-state stability of the protein: E45, F143 and K144. These sites were predicted to behave interdependently; F143 and K144 are located on the surface of PYP and form a salt bridge with E45 in the dark-adapted protein. To generate the library, the plasmid pZA31-PYP containing GCN4-Δ25PYP-v2 (K143F) was amplified in two fragments using mutagenic primers that annealed to the sites targeted for mutagenesis. The primers contained degenerate NNC nucleotides at the targeted codons, where N represents an equal mixture of A, T, C and G. The primers used are shown in Table 2-1, with the degenerate codons underlined.

Table 2-1: Primers used for GCN4-Δ25PYP-v2 library generation.

Site and direction	Sequence
Site 45 forward	GTAACGAACACCTGNNCGAAGAGCTGGCTCG
Site 45 reverse	CGAGCCAGCTCTTCGNNCAGGTGTTCGTTAC
Site 143, 144 forward	CGACCAAAGTACGTGTCTTCATGNNCNNCTC CAAAACTGGCGACTCCTATT
Site 143, 144 reverse	AATAGGAGTCGCCAGTTTTGGAGNNGNNCA TGAAGACACGTACTTTGGTCG

2.3 Results and Discussion

2.3.1 Design of the reporter system

We designed the reporter system based on earlier work concerning the construction of synthetic promoters for the regulation of gene expression.[106,108,109] The system is shown schematically in Figure 2-1. A synthetic promoter designated *pLtetO-AP1* was placed upstream of the gene for the fluorescent protein tdTomato, which was designed to function as a reporter for DNA binding of the photo-controlled constructs. TdTomato was chosen for use since its excitation and emission wavelengths do not overlap with wavelengths used for photo-control of PYP. The *pLtetO-AP1* promoter incorporates the binding sequence for the light-controlled DNA-binding protein GCN4-Δ25PYP (AP-1),[96] as well as the binding site for the Tet repressor (TetR). TetR regulation of the system permits us to turn 'on' tdTomato production using anhydrotetracycline (aTc) or Dox. Inducible expression of tdTomato, as opposed to constitutive expression, avoids baseline production of the reporter protein, which could obscure effects of the light-dependent binding proteins on transcription of the reporter. The reporter plasmid has a *ColE1* origin of replication, which has an expected copy number of 50 - 70 per cell.[106] Placing the reporter system on a relatively high copy plasmid backbone resulted in larger fluorescence signals for use in monitoring of light-responsive repression of tdTomato.

The gene for the light-switchable PYP-based protein was cloned into the plasmid pZA31, such that expression of the gene was under the control of the IPTG-inducible *lac* promoter, *pLlacO-1* (Figure 2-1). The Lac repressor is chromosomally encoded in *E. coli* DH5αZ1. The pZA31 plasmid has a *p15A* origin, which has an expected copy number of 15–30 copies per cell.[106]

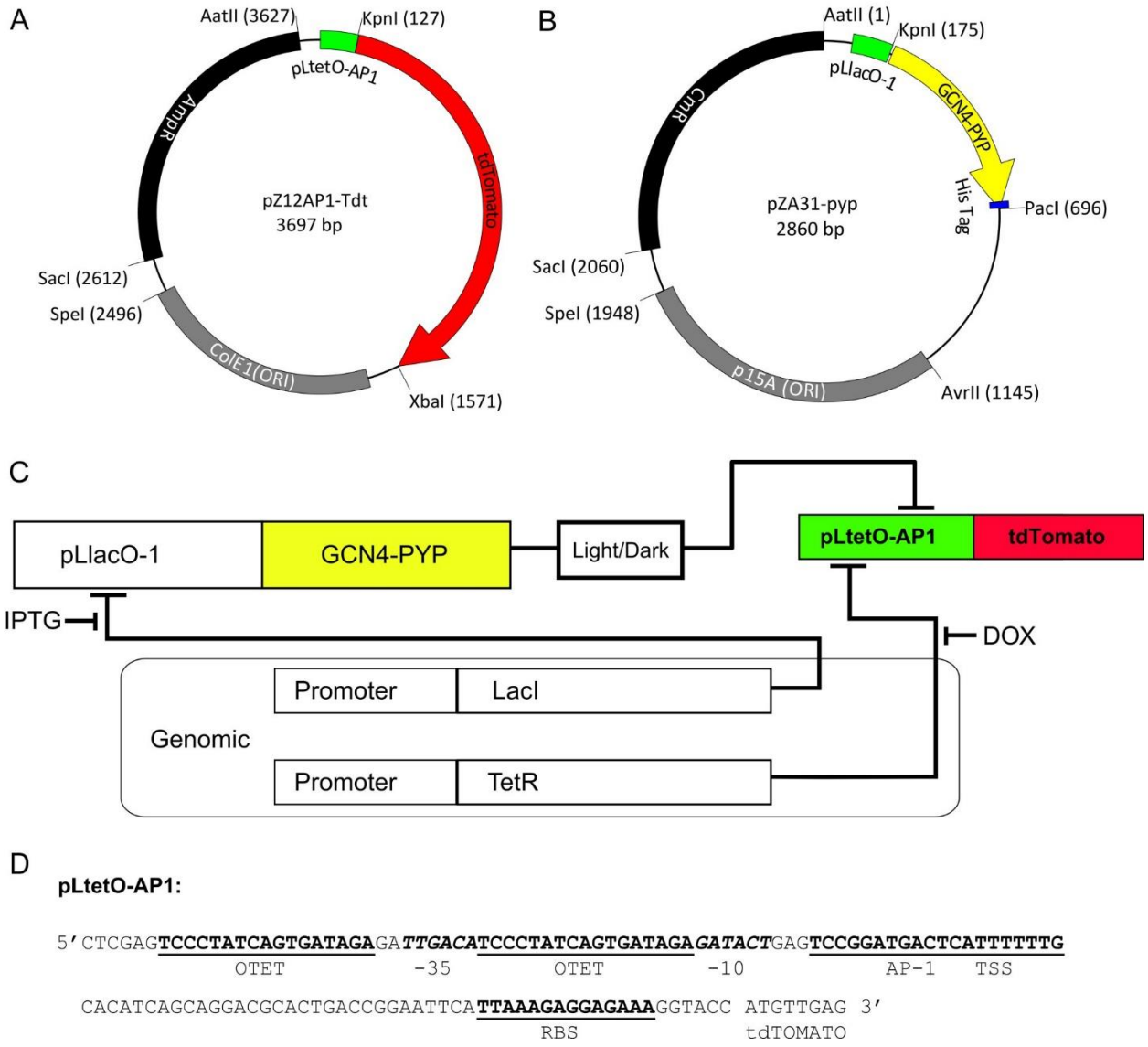


Figure 2-1: Design of the fluorescent reporter circuit.
Design of the fluorescent reporter circuit. (A) The tdTomato reporter plasmid has an Amp resistance gene for selection and *colE1* origin of replication for maintaining a plasmid copy number of 50-70 per cell. This plasmid expresses the reporter tdTomato from a synthetic promoter *pLtetO-AP1* that is controlled by both TetR and the light-responsive GCN4-Δ25PYP. (B) The light-responsive DNA-binding protein plasmid has a Cm resistance gene for selection and a *p15A* origin of replication for maintaining a plasmid copy number of 15-30 per cell. This plasmid expresses the GCN4-Δ25PYP protein from a LacI-controlled promoter *pLlacO-1.* (C) Schematic of the genetic network showing genomic expression of the repressor proteins LacI and TetR in *E. coli* strain DH5αZ1; IPTG and Dox are inducers that relieve the repression from LacI and TetR, respectively. Blue light (450 nm) was used to repress the expression of the reporter protein tdTomato. (D) Sequence of the synthetic promoter *pLtetO-AP1* showing the AP-1

binding site (the scrambled control replaced TCCGGATGACTCATTTTTTG with TCCGGGCTTAAATCTTTTTG). TSS and RBS refer to the original transcription start site and the ribosome binding site, respectively. Because the AP-1 site was inserted between the -10 box and the original TSS, the transcription will now start elsewhere, likely at one of the first G's of the AP-1 site. TetR binding sites are denoted as "OTET".

In order to validate light-responsive DNA-binding activity, we constructed several versions of the PYP containing plasmid. We used a positive control protein for DNA binding, GCN4-PYP, which had a linker between the GCN4-based DNA-binding domain and the PYP domain. This construct was designed such that the light-switchable domain has no influence on the DNA-binding activity, *i.e.* the protein always binds to its target AP-1 site independent of light conditions.[95] The protein Δ25PYP served as a negative control for DNA binding. This construct lacks the DNA-binding GCN4 domain and was consequently incapable of binding to DNA under any conditions. These control constructs were used to determine the dynamic range of the genetic circuit. Next, we constructed several different point mutant versions of the photo-switching GCN4-Δ25PYP protein that were previously characterized *in vitro,*[96] in order to test if these variants showed light-switchable repression of tdTomato expression in our circuit.

2.3.2 Optimization of experimental conditions for light-dependent reporter activity

In vivo implementation of this reporter system required close attention to the environment of the growing cells. An overview of the experimental protocol is shown in Figure 2-2. A customized incubator for growing cells in microplates was constructed so that light exposure could be varied while keeping all other conditions identical. Humidity, oxygen levels, shaking rate, temperature and light intensity were all found to be important variables and each of these was carefully controlled. In addition, we observed that it was important to: (i) test and control for direct effects of light on *E. coli* growth, (ii) optimize inducer concentrations, and (iii) test and control for effects of light on the inducers.

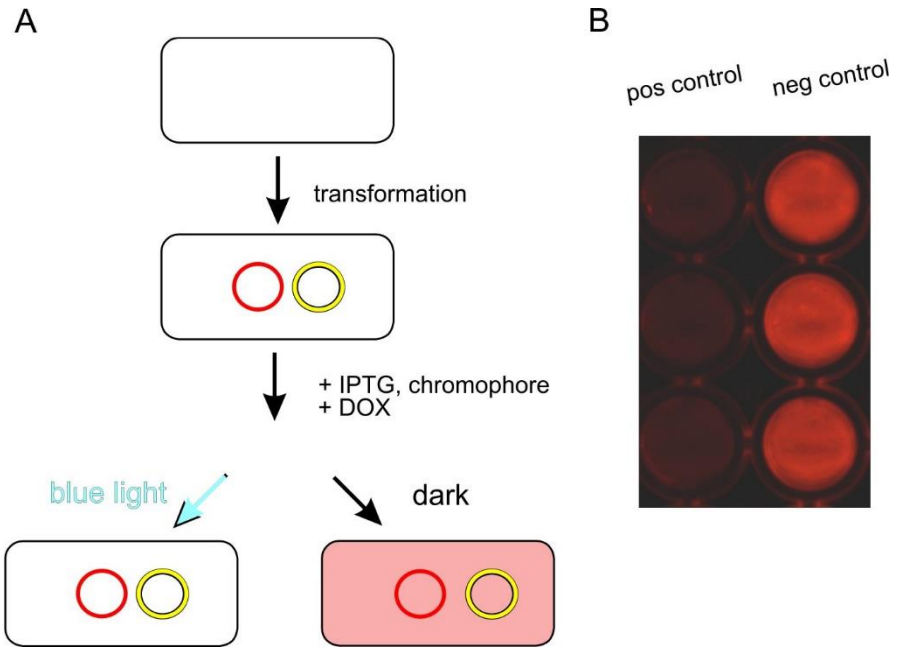


Figure 2-2: Schematic of experimental workflow.
(A) Overview of the reporter system. Two plasmids (red and yellow) were transformed into *E. coli* DH5αZ1. Red circles represent reporter plasmids (expressing the red fluorescent protein tdTomato) and yellow circles represent different versions of the GCN4-Δ25PYP plasmid. Cells were picked from plates and grown overnight. IPTG, chromophore and Dox were added in freshly prepared media to dilute the cells from the overnight culture before dividing them into two different 96-well plates for growth under light and dark conditions. (B) Fluorescence image of a microplate showing behavior of cultures with the control proteins (in triplicate). The positive control for DNA binding, GCN4-PYP, produced strong repression of tdTomato (left). The negative control for DNA binding, Δ25PYP, produced no repression of tdTomato (right).

It has been previously reported that light can have substantial effects on *E. coli* growth[110,111] and metabolism.[112] In this work, cells were exposed to blue light (450 nm) at an intensity of 0.33–0.45 mW/cm^2 delivered from a custom-built LED array. This light intensity was adjusted so that it was sufficient to cause isomerization of the light-switchable protein (measured by monitoring absorbance changes with purified proteins), without having strong effects on growth. Figure 2-3A shows a growth curve obtained under light and dark conditions. The calculated doubling time was essentially the same in both cultures (52.4 min in the light and 51.3 min in the dark, n = 3 each); however, there was a small effect on the cell density observed in stationary phase. The effect was more pronounced for overnight cultures, so fluorescence measurements of reporter expression were taken at times ≤ 6 h. Variation in cell density was taken into account when

calculating light-responsive repression of tdTomato expression by normalizing the fluorescence with the OD_{600} value for all samples.

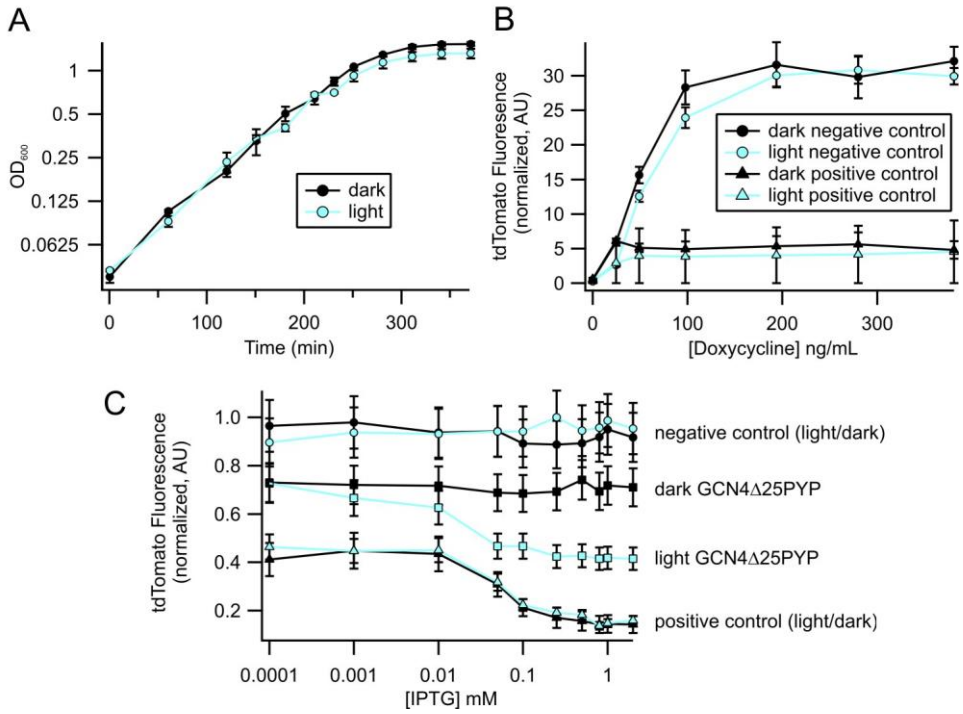


Figure 2-3: Optimization of light intensity and inducer concentrations.
(A) *E. coli* growth curve in light and dark conditions. The error bars indicate standard deviation from three biological samples. (B) Dose response curve for Dox induction. Negative and positive control constructs for were tested with the reporter plasmid expressing tdTomato. The negative control has no DNA-binding activity regardless of light, while the positive control always has DNA-binding activity regardless of light. The curves indicate both the maximum level of repression achievable and the Dox levels at which induction saturates. The error bars indicate standard deviation from three biological samples. Normalized fluorescence intensity with respect to OD_{600} was expressed in arbitrary units (a.u.). (C) Dose response curve showing IPTG-dependent repression of tdTomato by control proteins and the light-switchable construct GCN4-Δ25PYP-v2 (K143F E35L). Light blue lines and symbols represent samples exposed to light. Black lines and symbols represent samples maintained in the dark. The error bars indicate standard deviation from three biological samples.

We initially used aTc to induce expression of the tdTomato reporter protein from the synthetic promoter *pLtetO-AP1*, however, we found significant sensitivity of aTc to blue light. We therefore replaced aTc with Dox, which showed significantly lower sensitivity to blue light. Dox

was used in all further experiments. A Dox concentration of 250 ng/mL was found to saturate the response of the reporter (Figure 2-3B).

We optimized the expression level of the photo-controlled DNA-binding proteins by testing the circuit with various IPTG concentrations. For this analysis we used the light-switchable construct GCN4-Δ25PYP-v2 (K143F, E35L). A light-dependent response was seen when using IPTG concentrations of 0.5 mM and greater (Figure 2-3C). The degree of switching was sub-optimal; the light-state was not as strong a repressor of tdTomato expression as the light-insensitive DNA-binding control, and the dark-state showed more DNA-binding activity (detected as a decrease in tdTomato expression) than the non-binding control.

To test whether light-switchable repression could be detected in a high-throughput manner, we also analyzed cultures using FACS. Figure 2-4A shows that bacteria can be analyzed and sorted according to the degree of tdTomato expression. The difference in reporter expression levels between the positive and negative controls is easily detectable. Although small, the difference between the light and dark cultures of GCN4-Δ25PYP-v2 (K143F, E35L) is also detectable using FACS indicating that in principle, light-switchable proteins with a greater degree of switching could be selected using FACS.

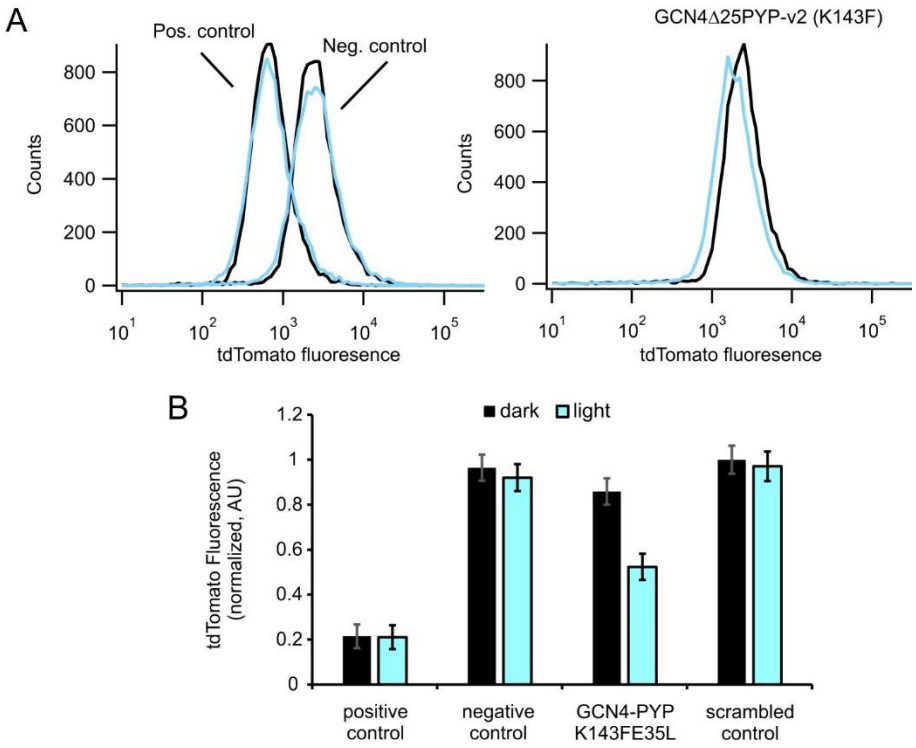


Figure 2-4: Analysis of circuit dynamic range and specificity.
(A) FACS analysis of reporter protein expression levels. *E. coli* cells were sorted according to red fluorescence intensity levels. (left) The light-insensitive control proteins show distinct differences in fluorescence levels (light – blue lines, dark – black lines). (right) The light-switchable protein GCN4-Δ25PYP-v2 (K143F), shows a small but distinct difference in light and dark conditions (light – blue line, dark – black line). (B) To test specificity of the DNA binding, the native DNA target sequence AP-1 was replaced with a scrambled control (see Figure 2-1 for the sequences). With this scrambled sequence, the positive control protein was no longer able to repress tdTomato expression. The error bars indicate average deviation from three biological samples. Fluorescence responses were compared at one IPTG concentration (0.5 mM).

To confirm changes in tdTomato expression were due to specific binding of the constructs to the AP-1 site in the tdTomato promoter, the system was tested using a tdTomato plasmid with a scrambled sequence in place of the AP-1 site. When tested with the light-insensitive DNA-binding positive control, no repression of the reporter was seen, indicating that binding is specific for the AP-1 site. Figure 2-4B shows the fluorescence responses observed. The scrambled AP-1 sequence is reported in the legend of Figure 2-1.

2.3.3 Detecting the extent of switching

Having established that the system could distinguish between light- and dark-states of a known light-switchable protein, we tested if it could distinguish between proteins exhibiting different degrees of switching. Previous structure-guided mutagenesis identified several sites in GCN4-PYP that altered the degree of switching (*e.g.* K143, E35, E45).[96] We expressed and purified a panel of six proteins with mutations at these sites and assayed their light-switchable DNA-binding ability using EMSAs. Figure 2-5 shows representative data.

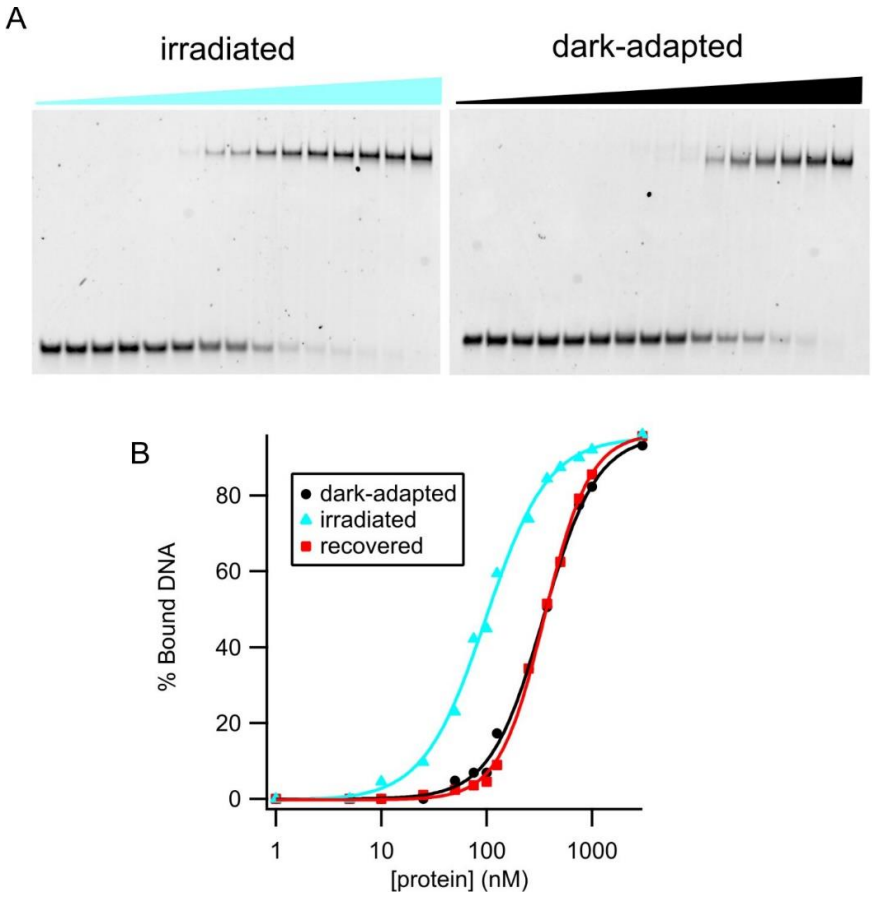


Figure 2-5: *In vitro* characterization of point mutants.
(A) Representative EMSA showing AP-1 binding by the photo-switchable protein GCN4-Δ25PYP-v2 (K143F V153I). Increasing concentrations of either irradiated (blue wedge, left) or dark-adapted (black wedge, right) protein (0, 1, 5, 10, 25, 50, 75, 100, 125, 250, 375, 500, 750, 1000, 3000 nM from left to right; the first lane is free DNA) were mixed with target DNA. (B) EMSA data were fit with the Hill equation to determine the apparent DNA affinity for each condition. Fitted curves are shown for GCN4-Δ25PYP-v2 (K143F E35L) binding under blue light irradiated (light blue line, triangles), dark-adapted (black line, circles) and dark recovered (thermally relaxed after blue light exposure) (red line, squares) conditions.

EMSA gels were analyzed to determine the amount of DNA bound *versus* the amount of free DNA at each protein concentration. Fitting these data to the Hill equation gave apparent K_d values for dark-adapted and blue light irradiated conditions. Table 2-2 shows the dissociation constants for the panel of mutants analyzed in this manner.

Table 2-2: *In vitro* apparent K_d values for different variants of the photo-switchable GCN4-Δ25PYP protein under light and dark conditions.

Variant of fusion PYP	**K_d in dark (nM)**	**K_d in light (nM)**	***In vitro* K_d ratio**
K143F, E35L	370 ± 25	90 ± 20	4.1 ± 0.5
K143F, V153I	360 ± 25	130 ± 15	2.8 ± 0.4
K143F	500 ± 35	200 ± 20	2.5 ± 0.3
K143F, E35F	600 ± 45	350 ± 25	1.7 ± 0.2
K143F, E45L	440 ± 30	190 ± 20	2.3 ± 0.3
K143F, E45F	500 ± 35	230 ± 20	2.2 ± 0.2
Positive control	13 ± 5	13 ± 5	
PFY mutant	1650 ± 200	800 ± 150	2.1 ± 0.5

Data were obtained by fitting to the Hill equation as described in the Materials and Methods section. Three EMSA gels were used to determine each K_d value. The PFY mutant was found using the screen, as described later in the chapter.

We next cloned all these variants into the pZA31 plasmid and transformed them into *E. coli* DH5αZ1 along with the reporter plasmid (pZ12AP1-tdT). These variants showed different levels of DNA binding upon light exposure, measured via the fluorescence output of tdTomato (Figure 2-6A). As before, the fluorescence measurements of light and dark samples were normalized using OD_{600}, and a two-tailed *t*-test ($P < 0.05$) was carried out for each mutant to verify that the light and dark fluorescence values were significantly different. We then compared these light and dark fluorescence ratios with the K_d value ratios obtained using EMSA analysis under light and dark conditions (Table 2-2). In general, a greater degree of switching *in vitro* corresponded to a greater degree of switching *in vivo*, but the quantitative correlation was weak (Figure 2-6B). Different degrees of protein expression, folding and/or turnover among the variants *in vivo* would be expected to diminish the correlation with *in vitro* binding data. Nevertheless, these data indicate that proteins with different degrees of switching ability can be distinguished using the *in vivo* tdTomato reporter assay.

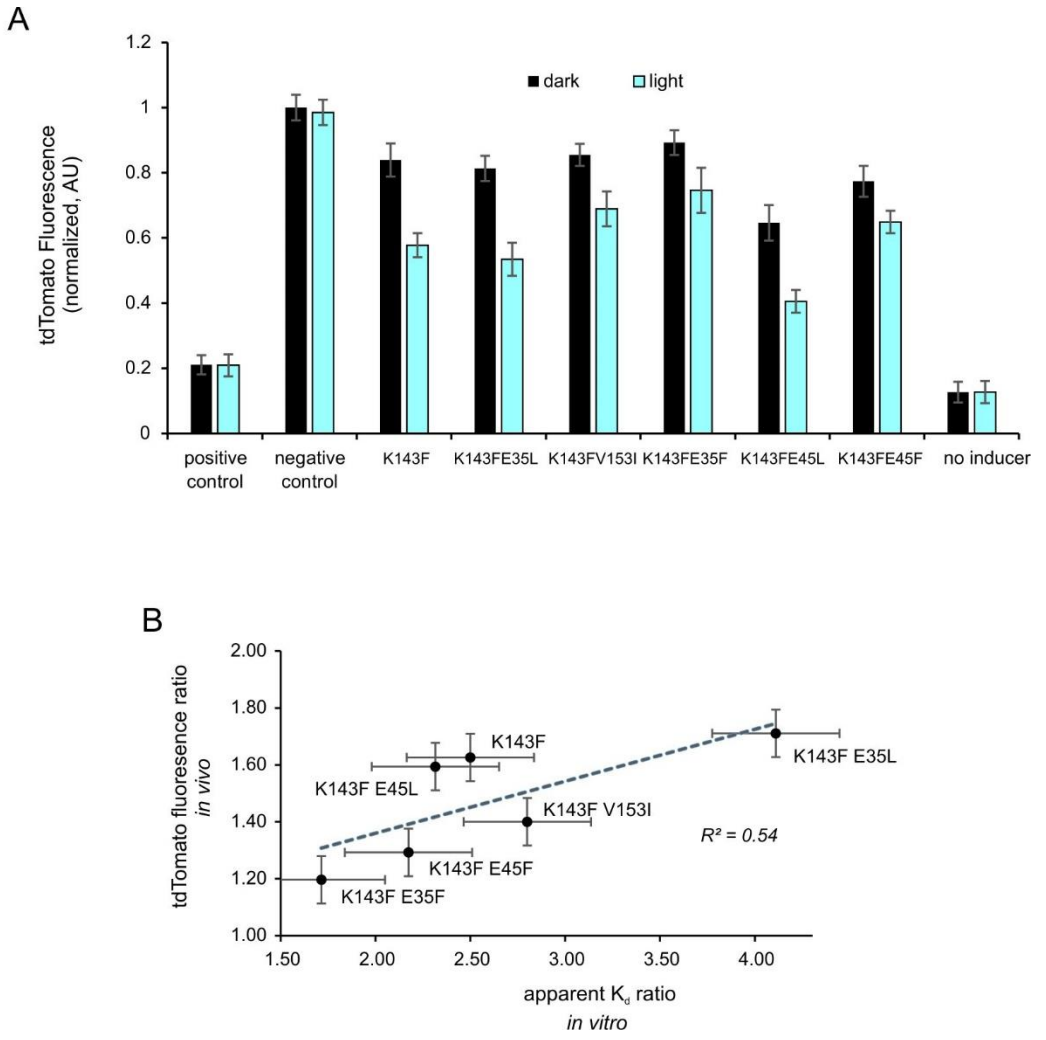
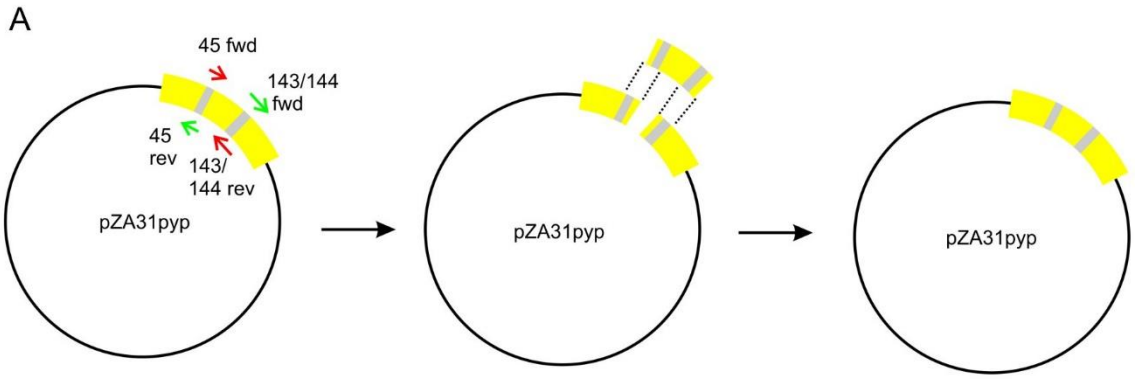


Figure 2-6: *In vivo* analysis of point mutants.
(A) Six variants of the photo-switchable GCN4-Δ25PYP fusion protein with various mutations respond to light to varying degrees. The error bars indicate standard deviation from six biological samples. Normalized fluorescence intensity is expressed in arbitrary units (a.u.). (B) Correlation plot showing the *in vivo* light *versus* dark fluorescence ratio versus the apparent K_d ratio *in vitro* for different variants of the photo-switchable GCN4-Δ25PYP fusion protein.

2.3.4 Library-based selection

Since the tdTomato reporter assay can detect differences in degrees of light-switching ability among a known set of proteins, it should therefore be useful for screening libraries of sequences to discover mutants showing greater degrees of switching. To explore whether this was feasible, we created a small library of mutants focused on sites E45, F143 and K144 in the protein sequence, which are known to affect the degree of switching.[96] To incorporate sequence

variation at these sites, primers with NNC codons (where N represents an equal mixture of A, T, C, and G) were used, as shown in Figure 2-7A.



B

switching non-switching

dark/light ratio	1.5	1.4	1.3	1.3	1.3	1.3		1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.1	1.1	1.1	1.1	1.1	1	1	1	1	1	0.9	0.9
seq. res. **45**	P	***E***	P	***E***	***E***	P		T	T	P	H	G	G	Y	R	P	P	G	G	P	R	A	L	A	P	G	T	T	T
seq. res. **143**	F	***F***	I	***F***	***F***	Y		I	S	T	T	R	F	A	C	L	R	D	F	G	D	C	A	R	F	P	G	N	G
seq. res. **144**	Y	***K***	D	***K***	***K***	V		P	H	G	P	A	F	P	D	H	G	R	F	V	A	N	L	G	D	S	L	D	L

Figure 2-7: Library design.
(A) The pZA31 GCN4-Δ25PYP-v2 (K143F) vector was amplified in two pieces with mutagenic primers annealing to the sites targeted for randomization (grey zones). One piece was produced with the set of primers coloured red. The second piece was produced using the set of primers coloured green. The primers contained 'NNC' codons to incorporate randomization during the amplification step. The resulting two fragments containing the randomized target codons were then re-assembled using Gibson assembly[113] to obtain the full vector. (B) Table of sequenced clones with those showing switching shaded green. The parent sequence (E, F, K) is shown in bold italic.

This mutagenic strategy produces a library with 15 possible amino acids encoded at each of the NNC sites (all except for Trp, Met, Glu, Gln and Lys).[114] The library size was 4096 distinct protein sequences. The library also contained the parent unmutated sequence E45, F143, K144, known to be an effective light-switchable sequence (*i.e.* GCN4-Δ25PYP-v2 E45, F143, K144) (Figure 2-3–2-5). Because NNC codons cannot code for Glu (E) or Lys (K), sequences with E45 and K144 cannot arise from mutagenesis and must come from the parent plasmid.

E. coli DH5αZ1 containing the tdTomato reporter plasmid were transformed with this library, and approximately 150 colonies were picked at random and grown using the microplate-based assay (with three repeats). Cultures were then sorted according to the degree of switching observed, measured as the ratio of tdTomato fluorescence observed when the cells were grown in the dark divided by the fluorescence observed under blue light conditions. Most cultures showed little or no switching, however, a subset (~5–10%) of cultures showed switching, arbitrarily set as a fluorescence ratio of 1.3 or higher. Plasmids from both the non-switching and the switching groups were then sequenced (Figure 2-7B). Among the switching group, the parent sequence is heavily represented (E45, F143, K144). This sequence never occurred in the non-switching group. This result demonstrates that the tdTomato reporter assay can recover an effective switcher from a random library of potential light-switchable proteins. Not surprisingly, this small library screen did not uncover any sequences with significantly better switching ability than the parent sequence. However, a few sequences showed behaviour comparable to the parent protein. We expressed and purified one of these, GCN4-Δ25PYP-v2 (E45P, K143F, K144Y) (designated the 'PFY mutant') and characterized the ability of the protein to bind target AP-1 DNA *in vitro* using EMSA (Figure 2-8). Binding to target DNA was indeed enhanced in blue light. The apparent K_d values observed were significantly weaker than those of the parent protein (Figure 2-8, Table 2-2), perhaps reflecting the destabilizing influence of the Pro residue on the coiled-coil. Interestingly, the binding of the PFY mutant to target DNA under blue-light irradiation showed a high degree of cooperativity; however, only a single bound state was observed by EMSA with a mobility similar to that observed for the parent DNA-protein complex. To confirm a specific interaction between the PFY mutant and AP-1 DNA, we measured thermal relaxation rates by monitoring the decrease of absorbance at 350 nm as a function of time after blue light irradiation (Figure 2-8C). Thermal relaxation rates for the PFY mutant were significantly faster than the parent protein (half-lives were on the order of 1 s compared with 90 s for the parent[96]). As was observed previously for the parent protein, the thermal relaxation rate is slowed in the presence of target DNA, confirming a specific interaction with the AP-1 sequence.

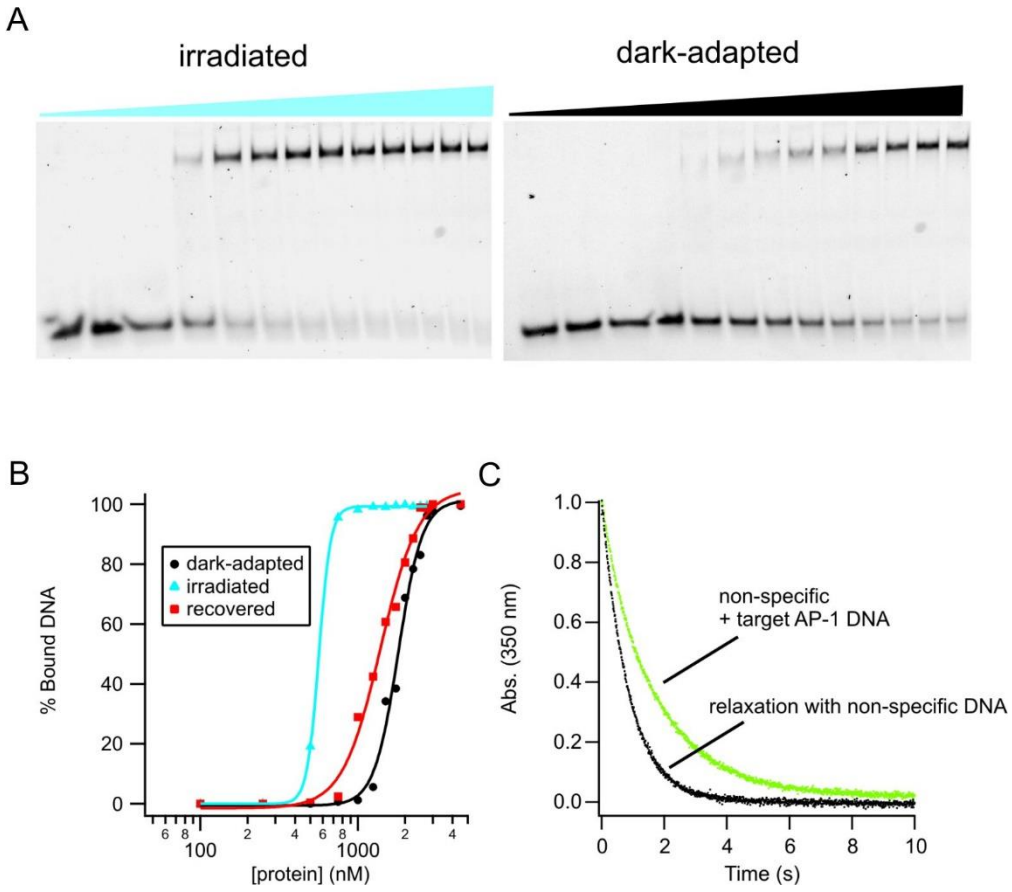


Figure 2-8: *In vitro* characterization of a recovered library variant.
(A) Representative EMSA showing AP-1 binding by the protein GCN4Δ25PYP-v2 (E45P, K143F, K144Y) found through the screening assay. Increasing concentrations of either irradiated (light blue wedge, left) or dark-adapted (black wedge, right) protein (0, 100, 250, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750 nM from left to right) were mixed with target DNA. (B) EMSA data were fit with the Hill equation to determine the apparent DNA affinity for each condition. Fitted curves are shown for GCN4-Δ25PYP-v2 (E45P, K143F, K144) binding under blue light irradiated (light blue line, triangles), dark-adapted (black line, circles) and dark recovered (thermally relaxed after blue light exposure) (red line, squares) conditions. Apparent K_d values and Hill coefficients are: blue-light irradiated K_d 800 nM, n = 10; dark-adapted K_d = 1650 nM, n = 4; recovered K_d = 1300 nM, n = 3) (C) Thermal relaxation of the protein GCN4-Δ25PYP-v2 (E45P, K143F, K144Y) after blue light irradiation in the presence of non-specific DNA (1 mg/mL) (black line) ($t_{½}$ = 0.65 ± 0.05s) with the addition of target AP-1 DNA (green line) ($t_{½}$ = 1.2 ± 0.05s) (TAE buffer, 20°C).

In our original structure-based approach to improving the degree of photo-switching of this protein, we deliberately avoided using Pro residues in the coiled-coil forming region of the sequence.[96] The *in vivo* screen of a mutant library reported here demonstrates that a Pro-

containing sequence can indeed function as a photo-controlled AP-1 DNA-binding protein. We expect that with larger libraries and automation, this microplate approach could be used directly to screen for mutants exhibiting higher degrees of switching. Alternatively, a FACS-based selection step could be used to choose clones that exhibited high fluorescence in the dark and low fluorescence when grown under blue light. One could then carry out directed evolution by using repeating cycles of introducing diversity and selection of the fittest. The approach is sufficiently general that it could be applied to the directed evolution of any light-switchable DNA-binding protein whose DNA target could be placed at a site analogous to the AP-1 site in the reporter plasmid described here.

3 Development of a tunable selection system for the selection of proteins with tuned interaction affinities

Note: This chapter was reproduced with permission from the following article: K.E. Brechun, K.M. Arndt, G.A. Woolley. Selection of protein-protein interactions of desired affinities with a bandpass circuit. 2019. *Journal of Molecular Biology*. 431: 391-400.

Author contributions: KEB performed the experiments, analyzed data and wrote the manuscript. KMA and GAW supervised the project and wrote the manuscript.

The article can be accessed under the following link: https://doi.org/10.1016/j.jmb.2018.11.011

3.1 Introduction

Protein–protein interactions are fundamental mediators of biological processes ranging from gene expression to immune responses.[115] In numerous cases, it has been observed that, to be functional, a protein–protein interaction should be neither too weak nor too tight, but just right (this observation has been termed the Goldilocks principle).[116,117] For example, several studies have found that when the affinity of a peptide/major histocompatibility complex for a T-cell receptor is either too low or too high, immune responses can be impaired.[117,118] Likewise, basic leucine zipper (bZIP) transcription factors homo- and heterodimerize with a range of affinities leading to complex interaction profiles, thereby controlling gene expression with great combinatorial regulatory potential.[119] Because binding affinities set bounds on binding and dissociation rates, precisely tuned binding affinities are important for the dynamic interplay of protein partners and temporal responsiveness to changes in the environment.[120,121] Perturbation of finely balanced protein–protein interactions is central to numerous diseased states.[122,123]

An ability to develop new protein–protein interactions and modify existing ones is foundational for numerous applications in biotechnology and medicine. For example, moderate-affinity, rather than high-affinity, T-cell receptor interactions have been proposed to be less likely to lead to unwanted toxic responses in T-cell-based immunotherapies.[117,124] Likewise, protein-based drug delivery systems require tuning of carrier protein affinity in order to control drug release rates and achieve a desired *in vivo* release profile.[125] Systems that enable selection of a specific level

of activity may also be used to evolve switchable proteins by changing the target level of activity in response to a given set of external cues.[92]

Current approaches for *in vivo* selection of interacting protein partners focus on the recovery of interactions with maximal binding strength, eliminating interactions below a selection threshold.[69,126] Selection methods for discovering and developing protein–protein interactions do uncover interactions of different affinity during the course of a multiple-round selection by varying (or relaxing) stringency, and it has been observed that yeast two-hybrid (Y2H) studies show some correlation between growth rates and protein–protein interaction affinity.[126] Also, by reprogramming sexual agglutination of yeast to allow library on library screening for protein–protein interactions, Younger *et al.*[127] showed that next-generation sequencing of the resulting diploid yeast strains could allow for quantitative evaluation of thousands of pairwise interactions. While these methods do link growth to the strength of a protein–protein interaction, they do not offer a way of specifically selecting for an intermediate affinity while avoiding higher-affinity clones. Using yeast surface display together with a defined binding target, a library of binding partners can be sorted using flow cytometry.[128] Reich and colleagues[129] have used this approach together with next-generation sequencing to rank hundreds of yeast-displayed peptides according to their affinities for a target protein. In this manner, strong, medium and weak affinity binders can be recovered, however, because it is a screening system rather than a selection system, this approach is more limited in the numbers of variants that can be assayed.

Here, we introduce a genetic circuit in *E. coli* that can act as a bandpass filter for protein–protein interactions and show, using proof-of-principle test cases, that the circuit allows one to select for protein–protein interactions of different strengths by changing the antibiotic concentrations in the media. To create this system, we combined two previously established sub-circuits: a hitchhiker circuit that links protein–protein interaction strength to β-lactamase activity,[130,131] and a bacterial bandpass filter that allows for selection for cells with specific, tunable levels of β-lactamase activity[92] (Figure 3-1). The hitchhiker circuit uses the endogenous twin-arginine translocation (Tat) pathway, a bacterial export system across the cytoplasmic membrane that accepts only folded substrates,[132,133] to select for interacting proteins. It has been shown previously that a protein lacking a signal sequence can be transported to the periplasmic space via the Tat pathway by interacting with a protein bearing a Tat-targeting signal sequence - a process termed

"hitchhiking".[132–134] The hitchhiker circuit uses this process to detect protein–protein interactions by fusing one half of an interacting protein pair to a Tat-targeting signal sequence, and the second half to a selectable marker, namely, β-lactamase devoid of its native signal sequence. β-Lactamase hydrolyzes β-lactam antibiotics, such as Amp, protecting bacteria from cell wall damage.[135] To fulfill this role, β-lactamase must be transported to the periplasm. While several factors may influence the β-lactamase activity observed, including Tat-transport efficiency and protein expression levels, the interaction strength of the pair of hitchhiker proteins has been shown to be a dominant factor for controlling the amount of β-lactamase transported and thereby for controlling cellular Amp resistance levels.[131] We expand on the hitchhiker translocation system by incorporating an inhibitor targeting the hitchhiker pair (Figure 3-1, inset). We show the binding strength of the inhibitor is inversely correlated with the cellular Amp resistance levels.

The hitchhiker circuit, with or without an inhibitor, achieves the high-pass component of the bandpass filter by allowing for cell growth only when cellular β-lactamase activity is high enough to counteract the growth inhibitory effect of Amp. To complete the bandpass filter, cells are simultaneously subjected to negative selection pressure for β-lactamase; that is, cells with β-lactamase activity above a certain level also do not survive. This is achieved by co-opting an endogenous transcriptional response to β-lactam antibiotics present in nearly all members of the *Enterobacteriaceae* family.[135] The presence of a β-lactam antibiotic leads to accumulation of cell wall breakdown products, which are imported into the cytoplasm and metabolized as part of the peptidoglycan recycling pathway.[135] One of the resulting fragments, 1,6-anhydroMurNAc-pentapeptide (aM-Pp), has been shown to bind to the transcriptional regulator AmpR, leading to the transcription of a genomic β-lactamase (AmpC) via the *ampC* promoter.[136] In the present work, the *ampC* promoter and AmpR from *Citrobacter freundii*[137] are used to control the production of a Tet resistance gene, as well as a fluorescent marker (green fluorescent protein (GFP), mCherry or Topaz) in response to cell wall breakdown (Figure 3-1). If β-lactamase levels are too high, all the Amp is hydrolyzed, the inducer aM-Pp does not accumulate, Tet resistance is not induced, and cells are sensitive to Tet. This completes the low-pass side of a bandpass filter for protein–protein interactions.

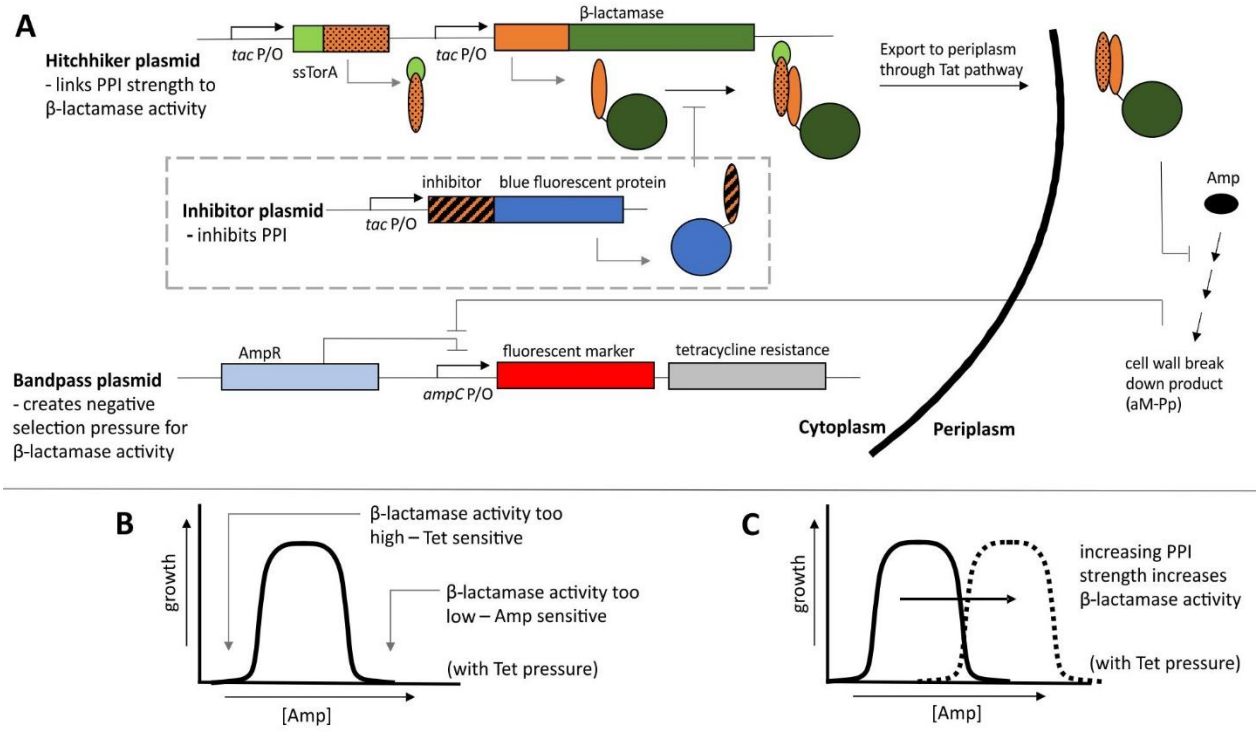


Figure 3-1: Schematic of the hitchhiker-bandpass genetic circuit.
(A) Selection of protein–protein interactions of desired affinities with a bandpass circuit. The hitchhiker plasmid (top) creates a circuit linking protein–protein interaction strength to cellular Amp resistance levels. When the two hitchhiker constructs interact (orange dotted and solid orange), a complex is formed that non-covalently links the Tat-targeting signal sequence (light green) to a β-lactamase devoid of its native signal sequence (dark green). This protein–protein interaction allows the β-lactamase to "hitchhike" to the periplasm via the Tat transport pathway where it hydrolyzes Amp (black oval). Co-expression of an inhibitor (orange stripes) targeting the hitchhiker pair impedes hitchhiker complex formation, lowering cellular Amp resistance levels. This component is optional (dashed box) and is discussed in the second half of the chapter. The bandpass plasmid (bottom) produces a circuit that leads to negative selection pressure for β-lactamase activity in the presence of Tet. The plasmid encodes a gene for Tet resistance that is under the control of the *ampC* promoter, which is constitutively repressed by the AmpR transcriptional regulator. When bound by the cell wall breakdown product aM-Pp, AmpR becomes an activator for the *ampC* promoter.[138] (B) The combined positive and negative selection pressure for β-lactamase activity limits cell growth to a specific range of Amp concentrations. High concentrations of Amp prevent growth because the β-lactamase activity is insufficient. Low concentrations of Amp prevent growth because the β-lactamase activity is too high; there is no cell wall damage and the Tet resistance gene is not expressed (*i.e.*, cells are Tet sensitive). At intermediate concentrations of Amp, the β-lactamase activity balances the Amp concentration, such that cells survive the pressure, yet sustain non-lethal cell wall damage,

leading to expression of the Tet resistance gene. (C) The Amp concentration permitting growth is dictated by the β-lactamase activity level, which is linked to protein-protein interaction strength. Increasing the interaction strength therefore changes the concentration of Amp permitting growth.

3.2 Materials and Methods

3.2.1 Strains

All cloning was performed using in *E. coli* XL1-Blue (Agilent Technologies, *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* (*F' proAB lacIq ZΔM15 Tn10* (*Tetr*))). Hitchhiker-bandpass experiments were performed in *E. coli* SNO301 (*ampD1, ampA1, ampC8, pyrB, recA, rpsL*).[139,140]

E. coli SNO301 is a K-12 strain derivative. The mutations *ampA1* and *ampC8* reduce genomic β-lactamase activity and the mutation *ampD1* inhibits the murine peptidase AmpD responsible for recycling cell wall breakdown products, including aM-Pp. The *ampD* mutation leads to the build-up of aM-Pp,[139] resulting in hyper-inducibility of the ampC promoter. The strain has an Amp minimal inhibitory concentration (MIC) of 1 μg/mL,[140] and with a plasmid copy of the *C. freundii ampR* gene and *ampC* promoter, sufficient cell wall damages occurs to activate the *ampC* promoter at Amp concentrations of 0.05 μg/mL - 1.0 μg/mL.[141]

3.2.2 Plasmid construction

Plasmids were constructed using classical restriction enzyme-based cloning, Gibson assembly[113] and SLiCE.[142]

The bandpass plasmid was developed using pTS1 (*CloDF13* origin of replication, Spec resistance). The original GFP construct was obtained from Prof. Marc Ostermeier. The Topaz mutant was created by making point mutations S65G, K79R, T203Y on the *gfp-mut2* gene in pTS1-GFP.[143] The mCherry version was created by exchanging the *gfp* gene for the *mCherry* gene.[144] The hitchhiker constructs were developed using pACYCDuet-1 (*p15A* origin of replication, Cm selectable marker). The *T7* promoters were exchanged with the *tac* promoter from pDIMC8-TEM1[92] to allow expression in the SNO301 strain.

The hitchhiker expression constructs were built using hitchhiker sequences obtained from Speck *et al.*[145] and cloned into pACYCDuet-1. The β-lactamase used was TEM-1 (originally from pBR322). The interacting peptides were the WinZip series,[146,147] the CREB homodimer,[25] truncated versions of cJun and cFos,[148] and full-length bZIP cJun and cFos.[149] In each case, an 11 residue Ser-Gly linker was used between the coil and β-lactamase fusion.

The inhibitor constructs were developed using a pET24b backbone (*ColE1/pMB1* origin of replication, Kan selectable marker). The *T7* promoter for expression was exchanged with the *tac* promoter from pDIMC8-TEM-1.[92] The inhibitors used were the CREB leucine zipper[150] and the Fos leucine zipper[149] with various lengths of an acidic extension.[149,150] The gene for a blue fluorescent protein, mTagBFP (with the point mutations I174A for increased stability and brightness[151], as well as C222S from other work [see Chapter 6]) was fused to each inhibitor sequence using a Ser-Gly linker. This blue fluorescent protein variant will be referred to as BFP.

All three plasmids encode the LacI repressor. A plasmid map for each type of plasmid (bandpass, hitchhiker and inhibitor) is shown in Figure 3-2.

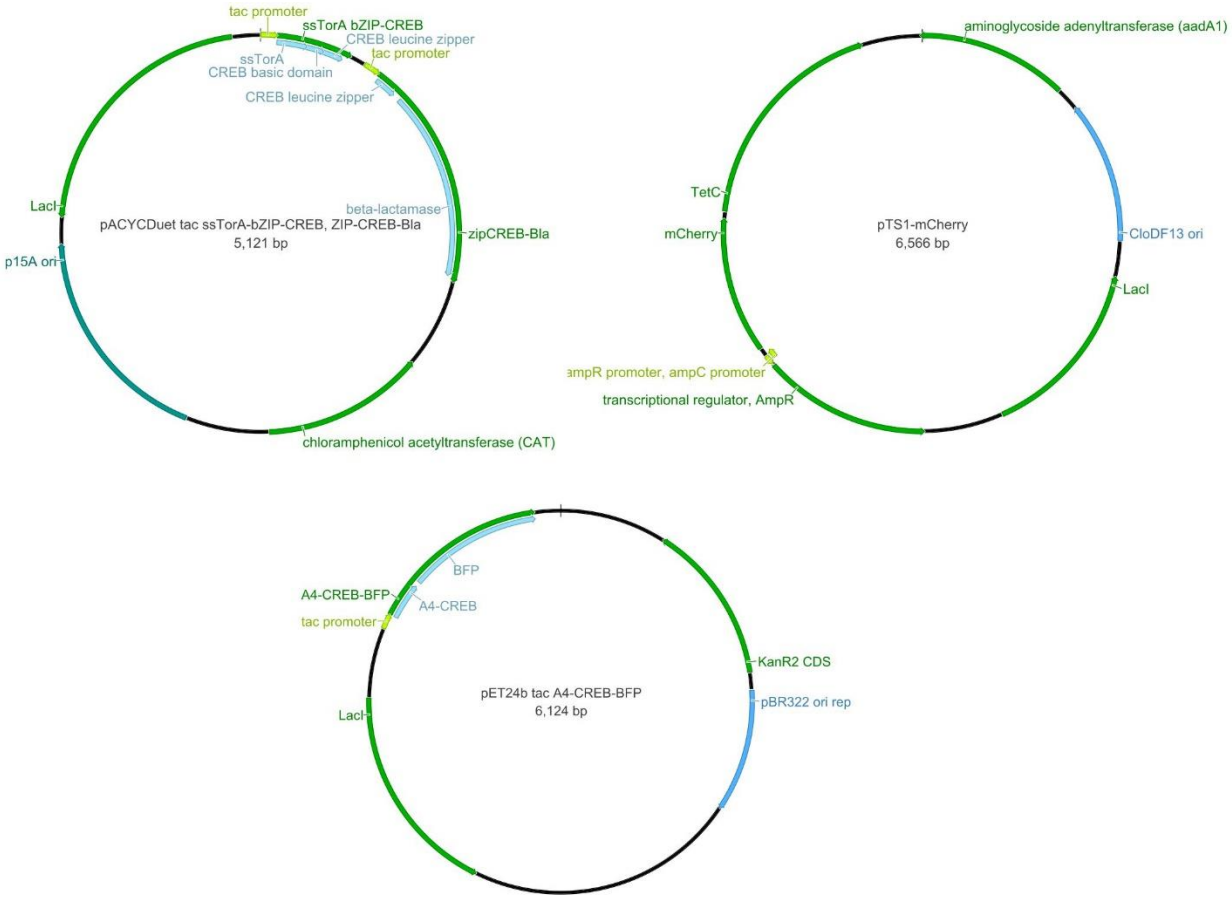


Figure 3-2: Plasmids used in the hitchhiker-bandpass circuit.
The plasmid with a pACYCDuet-1 backbone encodes the hitchhiker portion of the circuit. The plasmid with a pTS1 backbone encodes the bandpass portion of the circuit. The plasmid with the pET24b backbone encodes an inhibitor. The plasmid maps were generated with Geneious software.[152]

3.2.3 Culture preparations

E. coli SNO301 was electroporated with a bandpass plasmid, a hitchhiker plasmid and, when required, an inhibitor plasmid. A single colony was used to make an overnight culture in 5 mL of DYT supplemented with Cm (25 µg/mL), Spec (100 µg/mL), (Kan (50 µg/mL) for three-plasmid experiments) and 1% glucose. The following day, the culture was diluted to an OD_{660} of 0.5 in DYT and glycerol (15% final concentration), aliquoted into 100 µL samples and stored at -80°C.

3.2.4 Agar plate experiments

Petri dishes were prepared with 15 mL of LB agar containing Cm (25 μg/mL), Spec (100 μg/mL), Tet (20 μg/mL) and IPTG (300 μM). After solidification, plates were dried for 10 min in a laminar flow hood to decrease moisture content. A sterile disk of filter paper was placed in the center of the plate and spotted with Amp (2 μL, 100 mg/mL). The plates were left for 15 min to allow an Amp gradient to form. Glycerol stock cultures of each strain were thawed on ice, mixed in equal proportions and diluted 20-fold in DYT. The Amp disk was removed, and 150 μL of the mixed culture was spread across the entire plate and left to dry for 15 min. A second disk of filter paper was placed in the center of the plate and spotted with Amp (2 μL, 100 mg/mL). The plates were incubated for 90 min at room temperature, then at 37°C for 20 h. After incubation, plates were stored at 4°C for 20 h to allow for fluorescent chromophore maturation. The plates were imaged using a Fusion SL (Vilber Lourmat) imaging system with white light, and with EpiBlue LED lighting (with an F-595Y3 filter) for the GFP marker, and EpiRed LED lighting (with an F-695 Y5 filter) for the mCherry marker. The greyscale fluorescence images were false colored, the contrast adjusted, and overlaid using ImageJ software.

3.2.5 Liquid culture experiments

Round-bottom 96-well plates were prepared with 90 μL of DYT media containing Cm (25 μg/mL), Spec (100 μg/mL), (Kan (50 μg/mL) for three-plasmid experiments), Tet (20 μg/mL), IPTG (500 μM) and an Amp gradient created by serial dilutions. A glycerol starter culture was thawed on ice, diluted in prepared DYT media, and 10 μL of the diluted culture was added to each well to an inoculation OD_{660} of 0.001 for two-plasmid experiments, 0.005 for three-plasmid experiments, and 0.007 for three-plasmid experiments co-expressing A4-CREB-BFP or A4-CREB-(2xGly *f*)-BFP (due to apparent toxicity of the expressed proteins). When plating a mixed culture, equal volumes of each starter culture were combined. Experiments were often performed in parallel for replicates. For Figure 3-3C, the IPTG concentration used was 300 μM, and for Figure 3-9A and 3-9B, the IPTG concentration was increased to 1 mM to achieve better separation of the strains. Plates were incubated with a lid at 32°C with orbital shaking at 400 rpm for 18–20 h. Following incubation, plates were centrifuged for 15 min at 4000 rpm at 4°C, the supernatant was removed, and cells were re-suspended in 100 μL of PBS and transferred to a flat-bottom plate for measurements. Optical density (660 nm) and fluorescence (GFP ex. 486 nm,

em. 509 nm; mCherry ex. 587 nm, em. 610 nm; Topaz ex. 515 nm, em. 527 nm; BFP ex. 402 nm, em. 457 nm) were measured using an infinite M1000Pro plate reader (Tecan). In the GFP and Topaz mixed culture, GFP was excited at 466 nm to minimize background from Topaz. For the genetic analysis, streak plates were made to isolate single colonies, hitchhiker genes were amplified using colony PCR, and the amplified fragments were digested with *Pst*I and analyzed on an agarose gel.

3.3 Results and Discussion

3.3.1 Characterization of the hitchhiker-bandpass circuit

As a test case, to determine if the hitchhiker circuit could be integrated into the bandpass circuit, a series of hitchhiker vectors were constructed containing a set of synthetic leucine zipper coiled-coils with characterized interaction strengths (Table 3-1).[146,147]

Table 3-1: WinZip series of hitchhiker constructs.

Protein pair*	**K_d**	**Tm °C**	**Source**
WinZipB1/B1	(73 μM)**	(27.6)**	Ref. 146
WinZipA2/A2	17 nM	54.2	Ref. 147
WinZipA2/B1	4.5 nM	63.2	Ref. 147

*The first protein in the pair is fused to ssTorA, and the second protein is fused to β-lactamase
**Value is an estimate due to low stability of the complex.[146]

These sequences were originally developed using an *in vivo* dihydrofolate reductase protein fragment complementation assay and varying degrees of stringency.[146] Although hitchhiking has been demonstrated with a wide range of proteins,[130,131,133,134,153] we focused our attention on coiled-coil sequences since well-established guidelines exist for the manipulation of their binding affinities.[148,154,155] We used the Tat-targeting signal sequence from *E. coli* trimethylamine N-oxide reductase (ssTorA) and TEM-1 β-lactamase without its native signal sequence. When cells containing a bandpass plasmid and a hitchhiker plasmid were exposed to Tet and a range of Amp concentrations, growth and fluorescence were only observed in a narrow range of Amp concentrations (Figure 3-3). The Amp concentration at which cell growth occurred increased with the interaction strength of the expressed hitchhiker pair (Figure 3-3A). We then tested the ability to tune the circuit using external stimuli. Increasing the concentration of IPTG

is expected to increase the expression of the *tac* promoter-driven hitchhiker constructs, ultimately resulting in an increase in β-lactamase transport. Using the heterodimerizing hitchhiker construct WinZip A2/B1, we found that cells grew at higher concentrations of Amp in response to increasing concentrations of IPTG (Figure 3-3B). Saturation of this effect was observed at approximately 500 μM IPTG, presumably due to maximal expression levels being reached. Titrating the concentration of Tet adjusts the low-pass filter in the bandpass circuit. In the absence of Tet pressure, the low-pass filter is removed, allowing bacterial growth in low concentrations of Amp. Increasing Tet resulted in a corresponding narrowing of the band of bacterial growth (Figure 3-3C).

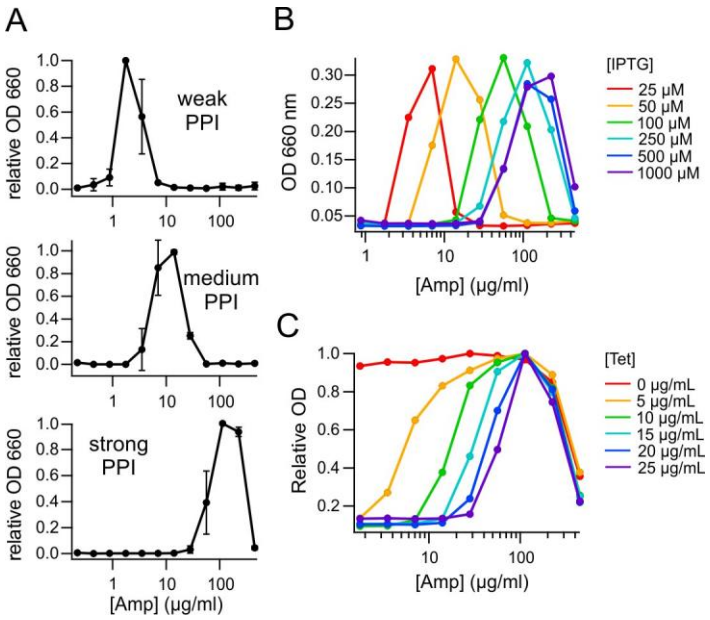


Figure 3-3: Characterization of the hitchhiker-bandpass circuit.
E. coli with the hitchhiker-bandpass circuit only grow in a specific range of Amp concentrations that depend on the interaction strength of the expressed hitchhiker pairs and can be tuned by external stimuli. (A) Cultures with an mCherry bandpass plasmid and a hitchhiker plasmid expressing (from top to bottom) a weakly interacting hitchhiker pair (WinZip B1/B1), a medium-strength hitchhiker pair (WinZip A2/A2) or a strongly interacting pair (WinZip A2/B1) grow only in specific concentration ranges of Amp. Data show the mean OD_{660} (n = 3) and error bars reflect the standard deviation (SD). (B) Increasing the concentration of IPTG used to induce a culture expressing WinZip A2/B1 (with an mCherry bandpass plasmid) resulted in a corresponding increase in Amp required for growth. The concentrations of IPTG used are indicated in the graph. (C) Titration of the Tet concentration controls the position of the low-pass filter. The same strain described in part (B) was exposed to a range of Tet concentrations, as indicated in the graph.

3.3.2 Separation of a mixed culture based on protein-protein interaction strength

The bandpass circuit has been used previously to separate strains based on β-lactamase activity levels.[92,156] In our circuit, β-lactamase activity levels are linked to the protein interaction strength of the hitchhiker pair, so it should therefore be possible to separate strains based on protein interaction strength. We developed different fluorescent versions of the bandpass plasmid to create bandpass plasmids with green, red or yellow fluorescent markers (Figure 3-1) so that the separation could be observed using fluorescence imaging. A mixed culture was created containing an equal mixture of cells expressing a weakly interacting hitchhiker pair (WinZip B1/B1) with the GFP bandpass plasmid, and cells expressing a strongly interacting hitchhiker pair (WinZip A2/B1) with the mCherry bandpass plasmid. The mixed culture was spread evenly on an LB agar plate containing IPTG, Tet, plasmid maintenance antibiotics (Spec and Cm), and a radial diffusion gradient of Amp, created by spotting a disk of filter paper in the center of the plate with Amp. After incubation, growth appeared in two isolated rings (Figure 3-4). Fluorescence imaging showed the culture with the red fluorescent marker was localized in the inner ring. These cells were expressing the strongly interacting hitchhiker pair, WinZip A2/B1, and required high levels of Amp for growth. The cells with the green fluorescent marker were expressing the weakly interacting hitchhiker pair, WinZip B1/B1, and grew where the concentration of Amp was low, forming the outer ring.

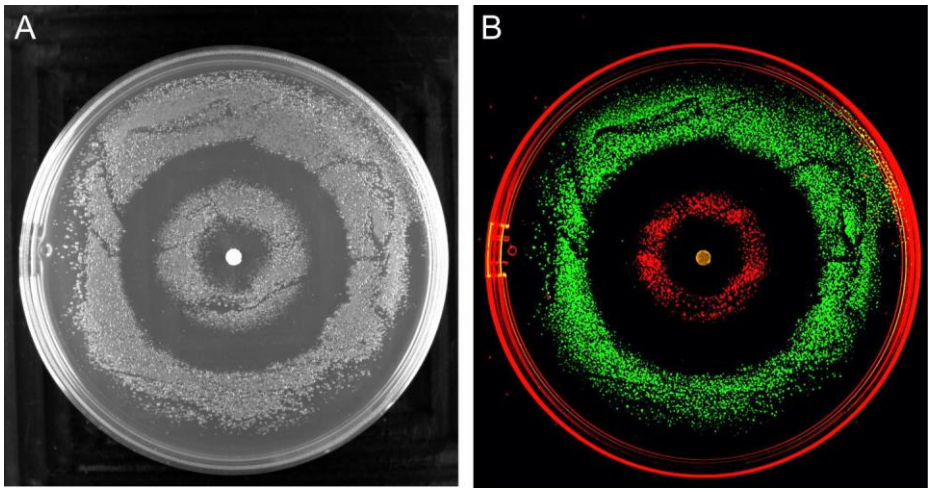


Figure 3-4: Separation of a mixed culture on an agar plate.
A mixed culture of *E. coli* was separated on an agar plate containing Tet with a radial diffusion gradient of Amp. The mixed culture contained a strain with the WinZip A2/B1 hitchhiker pair and the mCherry bandpass plasmid, and a second strain with the WinZip B1/B1 hitchhiker pair and the GFP bandpass plasmid. (A) Plating the mixed culture over the entire surface of the plate resulted in growth in two distinct rings. Fluorescence imaging (B) showed the two strains formed separate rings of growth.

3.3.3 Liquid culture separations with multiple fluorescent markers

The multiple fluorescent bandpass markers allowed fluorescence measurements to be used to analyze the results of mixed culture separations in liquid culture, to test whether cells could obtain Amp resistance passively due to leaked β-lactamase from neighboring cells, or conversely, whether cell wall breakdown products from cells with low β-lactamase activity could activate Tet resistance in cells with high β-lactamase activity. A mixed culture, containing cells co-transformed with the WinZip A2/A2 hitchhiker plasmid and the Topaz bandpass plasmid, and cells co-transformed with the WinZip A2/B1 hitchhiker plasmid and the GFP bandpass plasmid, was exposed to a gradient of Amp concentrations in a 96-well plate. After incubation there were two peaks of growth, with GFP and Topaz fluorescence each isolated to one independent peak corresponding to the interaction strength of the expressed hitchhiker pair (Figure 3-5A). We repeated this separation with cells that had the opposite coupling of the bandpass and hitchhiker plasmids. Once again, the fluorescence of each marker was isolated to only one of the two peaks of growth corresponding to the interaction strength of the expressed hitchhiker pair (Figure 3-5B).

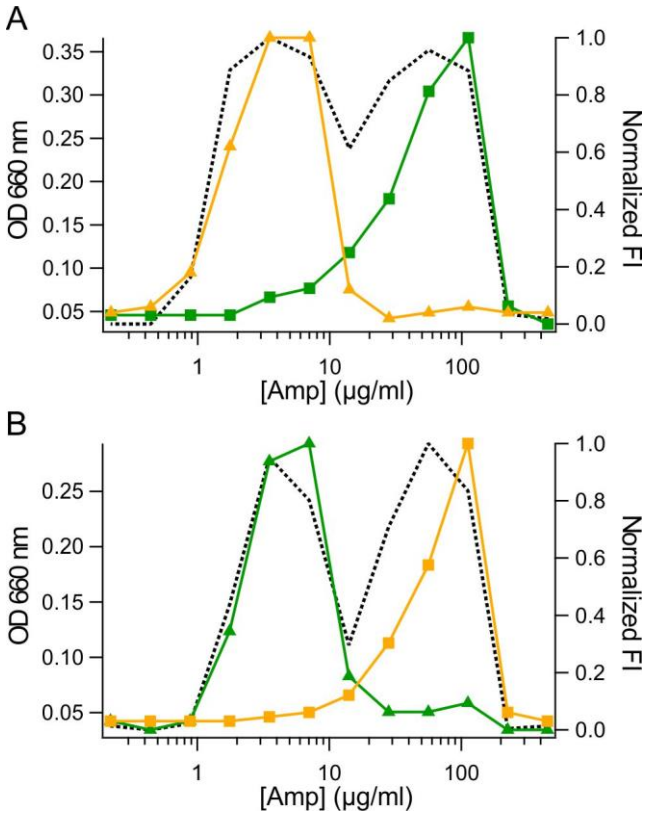


Figure 3-5: Fluorescence analysis of liquid culture separations.
(A) Separation of a mixed culture of WinZip A2/A2 homodimer (Topaz bandpass vector) and WinZip A2/B1 heterodimer (GFP bandpass vector), showed that the fluorescence of each culture was limited to the expected Amp concentrations. (B) This experiment was repeated with the opposite combination of hitchhiker and bandpass plasmids (WinZip A2/A2 with the GFP bandpass plasmid, and WinZip A2/B1 with the Topaz bandpass plasmid), showing once again complete separation of the two cultures. The graphs show optical density measurements (dotted black line) as well as fluorescence measurements for GFP (green line) and Topaz (yellow line). The graphs show the result of one representative experiment. This experiment was repeated 3 times.

We repeated the separations, this time using increasing starting inoculation densities of cells to see if passive Amp resistance becomes a problem at higher cell density (Figure 3-6). For this analysis, we used the mCherry bandpass plasmid to avoid fluorescence crosstalk between Topaz and GFP, which have similar excitation and emission spectra. We found that separation was complete, indicating cells do not acquire resistance to Amp or Tet from neighboring cells in the liquid medium at the tested cell densities.

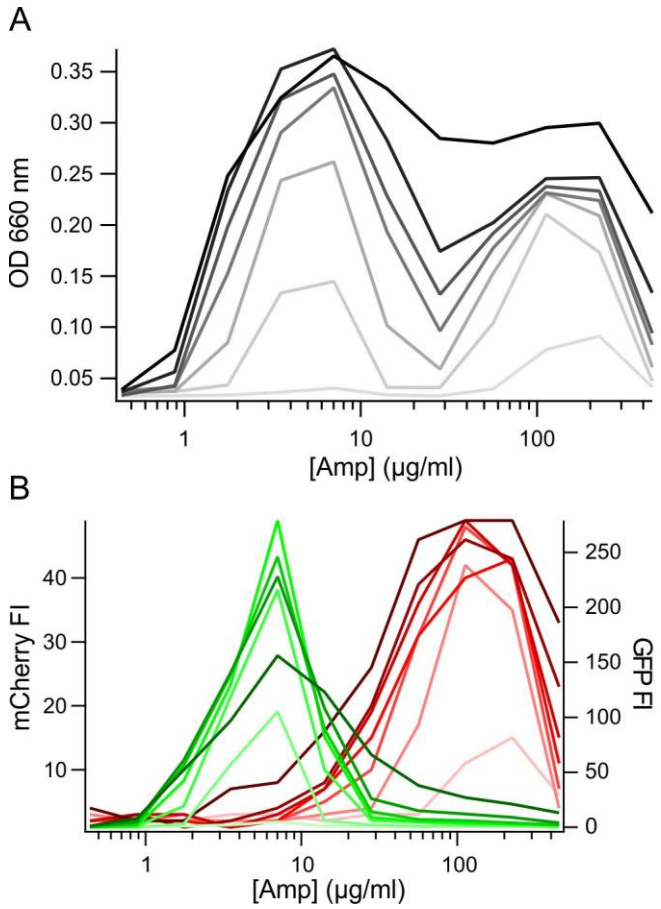


Figure 3-6: Mixture culture separations at higher cell densities.
A separation of a mixed culture, containing cells with the WinZip A2/A2 hitchhiker plasmid and the GFP bandpass plasmid (green), and cells with the WinZip A2/B1 hitchhiker plasmid and the mCherry bandpass plasmid (red), was performed using increasing starting inoculation concentrations of cells to determine whether passive Amp resistance occurs at high cell densities. The results show optical density measurements (A), and the fluorescence intensity of GFP and mCherry (B). Even at elevated concentrations of cells, the growth of each culture was restricted to one permissive concentration of Amp which corresponded to the interaction strength of the expressed hitchhiker pair. The cells were inoculated to a starting OD_{660} of 0.001, 0.0015, 0.003, 0.005, 0.007, 0.01, 0.015 (coloured from lightest to darkest shades). The graphs show the result of one representative experiment. This experiment was repeated 3 times.

3.3.4 Investigating the effect of changing the bandpass fluorescent marker on cellular Amp resistance levels

We found that hitchhiker-bandpass cultures with the mCherry bandpass plasmid all grew at slightly higher concentrations of Amp than the corresponding cultures with the GFP bandpass plasmid (Figure 3-7).

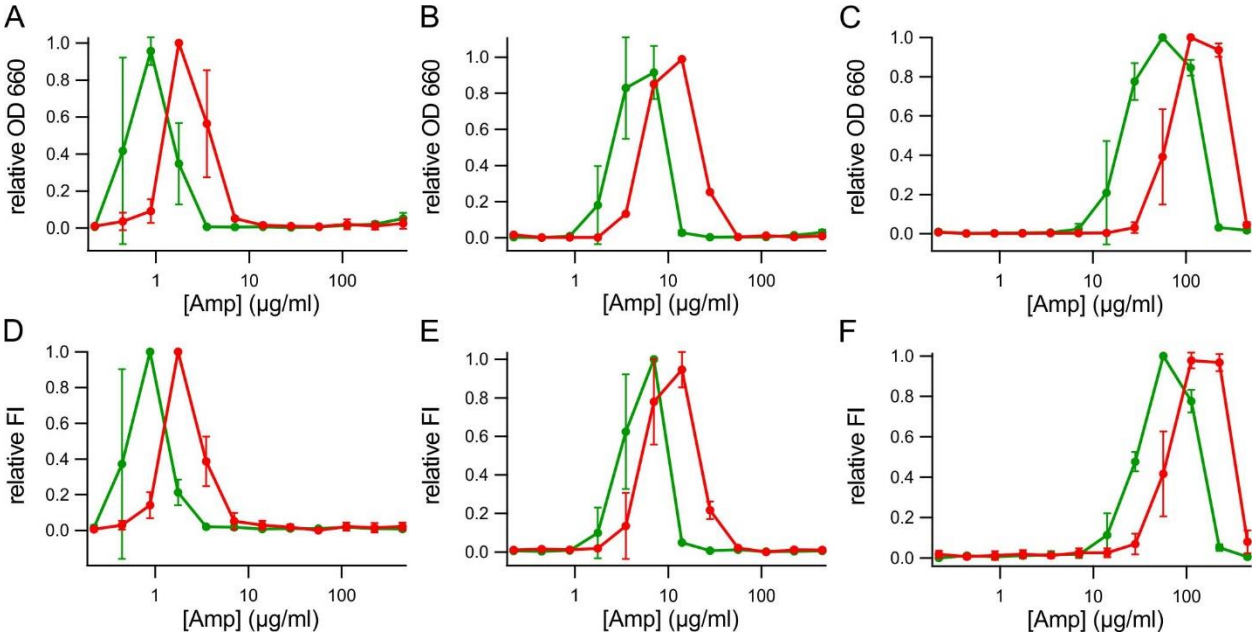


Figure 3-7: Comparison of hitchhiker-bandpass cultures with the GFP and mCherry markers. Cultures expressing the hitchhiker pairs WinZip B1/B1, WinZip A2/A2 and WinZip A2/B1 co-transformed with either the GFP marker (green) or the mCherry marker (red) were found to have shifted permissive concentrations of Amp. Graphs A - C show the normalized OD_{660} values for WinZip B1/B1, WinZip A2/A2 and WinZip A2/B1 respectively. Graphs D - F show the normalized mCherry or GFP fluorescence intensity values for WinZip B1/B1, WinZip A2/A2 and WinZip A2/B1 respectively. The graphs show mean values (n = 3) and error bars show the SD.

We made point mutations in GFP known to red-shift its fluorescence spectrum,[157] to make a bandpass plasmid with a Topaz fluorescent marker. Hitchhiker-bandpass cultures with the Topaz or GFP bandpass plasmids (differing in 3 spectral-tuning point mutations in the fluorescent marker) seemed to behave identically (Figure 3-8).

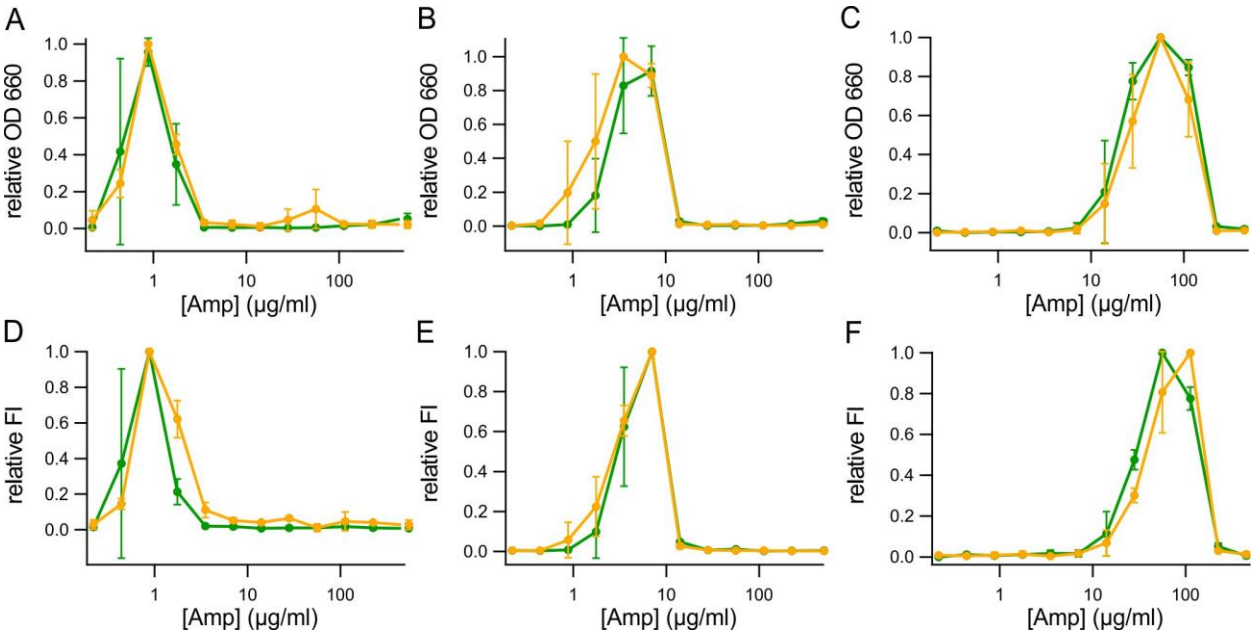


Figure 3-8: Comparison of hitchhiker-bandpass cultures with the GFP and Topaz markers. The WinZip hitchhiker constructs were each co-expressed with a bandpass plasmid with a GFP marker (green traces) or a Topaz marker (yellow traces). Graphs A – C show normalized OD_{660} data for WinZip B1/B1, WinZip A2/A2 and WinZip A2/B1 respectively. Graphs D – F show normalized GFP or Topaz fluorescence intensity for WinZip B1/B1, WinZip A2/A2 and WinZip A2/B1 respectively. The graphs show mean values (n = 3) and error bars show the SD.

We do not have an explanation for the difference seen when GFP is exchanged for mCherry. To avoid any influence from the nature of the fluorescent marker in the bandpass plasmid, hitchhiker-bandpass experiments that focused on comparing protein interaction strength were always performed using the same bandpass plasmid.

3.3.5 Separations of increased complexity in liquid culture

Precise control of the concentration of Amp in liquid cultures enabled separations of higher complexity. We separated a mixed culture of cells expressing three different WinZip peptides, with each strain co-expressing the GFP bandpass plasmid (Figure 3-9). The mixed culture showed growth maxima in three distinct Amp concentration ranges. Based on the Amp concentration observed to permit growth of pure cultures of each of the strains, we selected three wells from the mixed culture that were each expected to contain a pure culture after incubation. Cells were isolated by making streak plates (5 per well), and a portion of the hitchhiker plasmid was amplified by colony PCR and digested with restriction enzymes to genotype each isolated colony. Each well was found to contain a pure culture of the expected strain (Figure 3-9C).

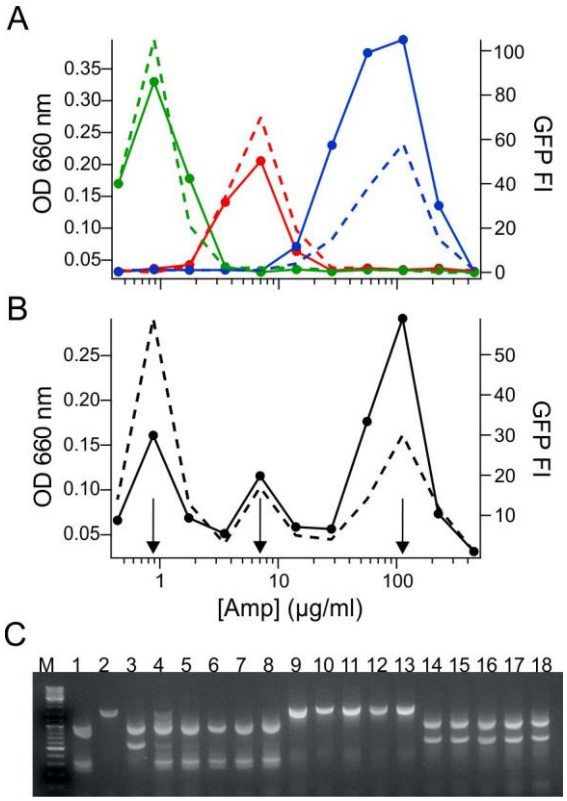


Figure 3-9: Separation of a mixed culture in 96-well plate format.
(A) Pure cultures of cells expressing weakly (WinZip B1/B1, green), medium (WinZip A2/A2, red) or strongly (WinZip A2/B1, blue) interacting hitchhiker pairs, containing the GFP bandpass plasmid, each grew in a distinct range of Amp concentrations. (B) An equal mixture of these three strains produced growth maxima in three distinct ranges of Amp concentrations. The graphs show optical density (solid lines) and GFP fluorescence intensity (dashed lines). Bacteria from selected wells (indicated by arrows) were genotyped. The results are shown on an agarose gel (C). Pure cultures of each WinZip B1/B1, WinZip A2/A2 and WinZip A2/B1 (lanes 1–3, respectively) were used as controls. Lanes 4–8 show cells isolated from the low Amp well, lanes 9–13 show cells isolated from the intermediate Amp well, and lanes 14–18 show cells isolated from the high Amp well. M indicates the marker.

This experiment was repeated three more times with genetic analysis of up to 10 isolated colonies per well. Complete separation was observed in all cases (Figure 3-10).

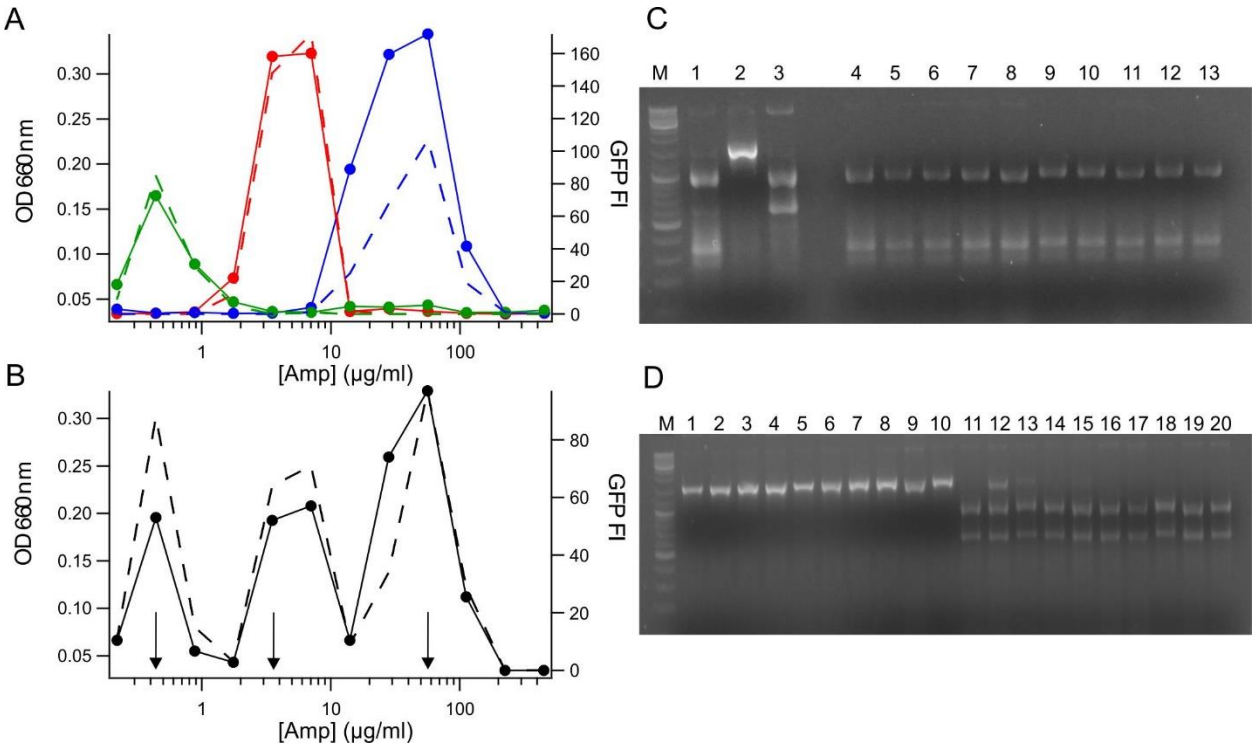


Figure 3-10: Separation of a mixed culture with increased genetic analysis.
The separation experiment shown in Figure 3-9 was repeated several times with additional genetic analysis. Graphs A and B show the OD_{660} (solid trace) and GFP fluorescence (dashed trace) for pure cultures of cells with a GFP bandpass plasmid and a hitchhiker plasmid expressing WinZip B1/B1 (weak interaction, green), WinZip A2/A2 (moderate interaction, red), or WinZip A2/B1 (strong interaction, blue), and a mixed culture containing equal proportions of all three strains (black). The mixed culture showed growth in three distinct peaks. Based on the activity of the pure culture of each strain, three wells were selected from the mixed culture that were expected to contain pure cultures (arrows). Streak plates were made (10 from each well) to yield single colonies for genetic analysis. Colonies were genotyped using colony PCR to amplify the hitchhiker plasmid, followed by a digestion targeting the WinZip coding sequences. The resulting banding pattern was analyzed by gel electrophoresis (C, D). The markers are shown in lanes M. In the gel shown in (C), lanes 1-3 are controls, showing the banding pattern from purified plasmid DNA of WinZip B1/B1, WinZip A2/A2 and WinZip A2/B1 respectively. Lanes 4–13 show the analysis of ten colonies isolated from the low Amp well. In the gel shown in (D), lanes 1–10 show the analysis of ten colonies isolated from the medium Amp well, and lanes 11–20 show the analysis of ten colonies isolated from the high Amp well. Lanes 12 and 13 show an incomplete digest. All 30 analysed colonies showed the expected genotype. WinZip B1/B1 was recovered in the low Amp well, WinZip A2/A2 was recovered in the medium Amp well, and WinZip A2/B1 was recovered in the high Amp well.

We further analyzed the relationship between cellular Amp resistance levels and the interaction strength of the expressed protein pairs using a second series of protein–protein interactions for which extensive *in vitro* data are also available.[148,149,158] We constructed a series of hitchhiker vectors containing various lengths of the bZIP domain of cJun and cFos (Table 3-2).

Table 3-2: cJun and cFos series of hitchhiker constructs.

Hitchhiker pair*	**Notes**	**T_m (°C)**	**Source**
cJun/cJun	Truncated leucine zippers	24	Ref. 148
cJun(bZIP)/cJun(bZIP)	Full-length bZIP sequences	30	Ref. 149
cFos/cJun	Truncated leucine zippers	16	Ref. 148
cFos(bZIP)/cJun(bZIP)	Full-length bZIP sequences	50	Ref. 149
cFos(ZIP)/cJun(bZIP)	Full-length cJun bZIP sequence, cFos leucine zipper	53	Ref. 149

*The first protein in the pair is fused to ssTorA, and the second protein is fused to β-lactamase

These proteins dimerize via a coiled-coil leucine zipper region and bind to AP-1 sites on DNA via a positively charged basic region. cJun can both homodimerize and heterodimerize with cFos, whereas cFos cannot form stable homodimers.[159] These constructs were tested in the hitchhiker-bandpass assay (Figure 3-11). The Amp resistance levels observed, correlated well with published *in vitro* data. The truncated cJun leucine zipper homodimer and truncated cFos/cJun leucine zipper heterodimer are known to interact weakly.[148] Hitchhiker-bandpass strains based on these interactions grew in low concentrations of Amp. The strain with a hitchhiker interaction based on the full-length cJun homodimer exhibited activity similar to the strain with the truncated cJun interaction, potentially because the increase in interaction strength gained by extending the length of the cJun leucine zipper is not strong enough to offset the electrostatic repulsion of the basic domains in the full-length version. The strain containing the full-length cJun/cFos hitchhiker pair grew in much higher concentrations of Amp, consistent with the strong interaction reported in *in vitro* measurements.[149] The hitchhiker plasmid with the full-length cFos/cJun interaction was modified by removing the cFos basic domain from the full-length bZIP construct. The hitchhiker-bandpass strain with this interaction grew in even higher concentrations of Amp, indicating an increase in affinity. Removing the basic domain from one protein in a bZIP pair has been shown previously to increase the affinity of the complex, because this eliminates the electrostatic repulsion between the two positively charged basic domains.[149]

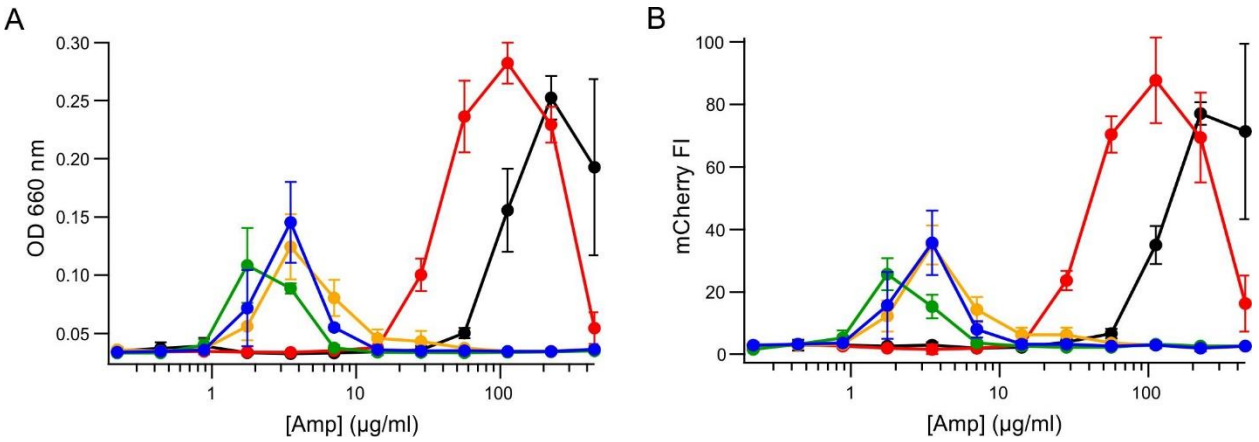


Figure 3-11: Hitchhiker-bandpass cultures with interactions based on the AP-1 transcription factor.

Components of the AP-1 transcription factor were tested to compare known protein–protein interaction strength with observed cellular Amp resistance levels in the hitchhiker-bandpass assay. Cells contained a bandpass plasmid with an mCherry marker, and a hitchhiker plasmid with one of the following protein pairs: truncated cJun/cJun (blue), cJun(bZIP)/cJun(bZIP) (yellow), truncated cFos/cJun (green), cFos(bZIP)/cJun(bZIP) (red) or cFos(ZIP)/cJun(bZIP) (black). The optical density measurements (A) mirrored the fluorescence measurements from the mCherry bandpass plasmid (B). The graph shows the mean of the optical density measurements (n = 3), and error bars reflect the SD.

3.3.6 Co-expressing an inhibitor protein to modulate hitchhiker interaction strength

Although the hitchhiker circuit provides a good assessment of protein–protein interaction strength, other factors are expected to influence the cellular Amp resistance observed for a given pair. These factors include variations in the effective concentration of hitchhiker pairs, the propensity for a pair to form correct ssTorA/β-lactamase heterodimers (which is decreased in hitchhiker pairs with homodimerizing interaction partners), and details of how a particular pair interacts with the Tat transport machinery. The Tat export pathway has a quality control mechanism that inhibits the transport of improperly folded proteins.[160] While this feature can be beneficial, adding selection pressure for well-folded proteins,[145,160] it is possible that transport efficiency may be altered by the stability of a protein–protein complex (*i.e.*, its "foldedness" at a particular temperature) in addition to the affinities of its components. To avoid any influence this may have on the selection for protein–protein interactions, we explored the option of incorporating an inhibitor of a hitchhiker protein–protein interaction to allow the binding strength of different inhibitors to be compared with respect to a single hitchhiker export pair.

Inhibitor constructs were cloned into a third compatible plasmid (Figure 3-1) and were designed to contain a C-terminal fusion to BFP[151] to monitor inhibitor expression levels and increase *in vivo* stability. When a strong inhibitor is co-expressed in the circuit, it is expected that it would bind to the hitchhiking protein(s) and prevent successful transport of β-lactamase by outcompeting the hitchhiker pair dimerization, restricting cell growth to lower concentrations of Amp. The responsiveness of the hitchhiker-bandpass assay to inhibitor strength was investigated by building and testing a series of cFos inhibitors. Previously, it was shown that potent inhibitors for bZIP proteins could be designed by appending an acidic extension to a leucine zipper, creating favorable electrostatic interactions with the basic domain of the native bZIP protein.[149,150,161] We created a series of cFos-based inhibitors containing an acidic extension of increasing length (Table 3-3).

Table 3-3: Fos-based inhibitor series.

Inhibitor	**Notes**	**T_m with cJun (bZIP) (°C)**	**Source**
BFP	Negative control	n/a	
A0-Fos-BFP	Fos leucine zipper	53.4	Ref. 149
A2-Fos-BFP	Two heptad repeats of the acidic extension	n.d.	
A4-Fos-BFP	Four heptad repeats of the acidic extension	72.1	Ref. 149

Co-expression of the inhibitors with the cFos(ZIP)/cJun(bZIP) hitchhiker pair shifted growth to lower concentrations of Amp in comparison to a non-inhibiting BFP negative control (Figure 3-12). The rank order of the decrease in Amp resistance correlated with the binding strength of the inhibitors.

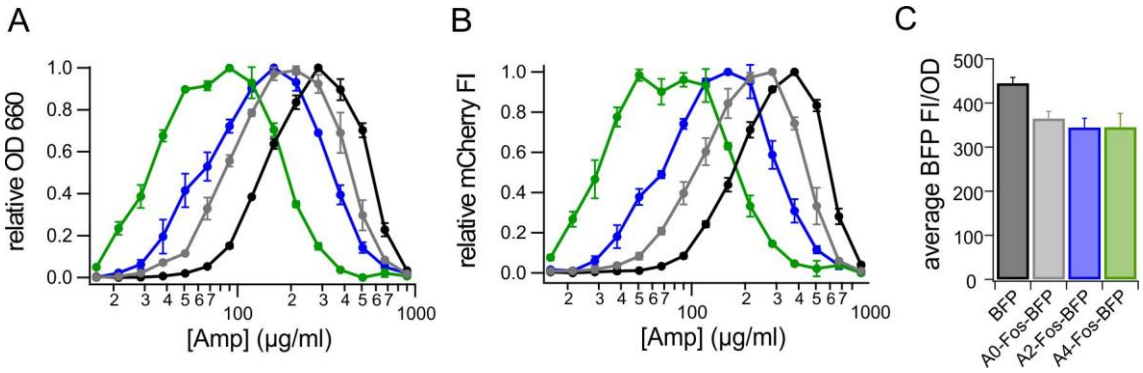


Figure 3-12: Assessment of inhibitor strength using the hitchhiker-bandpass assay.
The cFos(ZIP)/cJun(bZIP) hitchhiker pair was co-expressed with a cFos-based inhibitor with an acidic extension of increasing length: BFP negative control (black), A0-Fos-BFP (gray), A2-Fos-BFP (blue) and A4-Fos-BFP (green). Growth (OD_{660}) (A), expression of the fluorescent mCherry reporter (B) and inhibitor expression levels (BFP fluorescence) (C) are shown. OD and mCherry fluorescence values show the mean (n = 3), and error bars reflect the SD. BFP fluorescence values are shown as the mean of the BFP fluorescence normalized by the OD, using data from wells that had growth (defined as an OD within 75% of the maximum OD of the culture).

To confirm the ability of the hitchhiker-bandpass system to respond to inhibitors of different strength, we repeated these experiments in another protein–protein interaction system, using CREB, a bZIP transcription factor that is analogous to cJun and cFos but functions as a homodimer. CREB hitchhiker pairs were constructed containing the CREB bZIP with an N-terminal ssTorA fusion, and the CREB leucine zipper fused to the β-lactamase. We created a series of inhibitors based on A4-CREB,[25,150] a construct containing the CREB leucine zipper with a four-heptad repeat acidic extension, by making point mutations designed to weaken its binding (Table 3-4).

Table 3-4: CREB-based inhibitor series.

Inhibitor	**Notes**	**T_m with CREB(ZIP) (ºC)***
BFP	negative control	n/a
A4-CREB-(2x Gly *f*)-BFP	zipper with mutations in *f* positions (N318G, E325G)	31
A4-CREB-(2x Gly *d*)-BFP	zipper with mutations in *d* positions (L316G, L323G)	-2
A4-CREB-BFP	Positive control, strong inhibitor	34

* Prediction by bCIPA[148,158]

Coiled-coil sequences form helices with a periodicity of seven residues. The residues within this heptad repeat are designated *a* - *g* to describe their position within the coil. Residues in the *d* positions are located along the coiled-coil interface and are typically involved in hydrophobic interactions.[162,163] We mutated two *d* position residues in the A4-CREB-BFP leucine zipper from Leu to Gly (residues L316G, L323G in the murine CREB sequence), a change that is expected to dramatically decrease stability (Table 3-4).[162,163] In addition, we made a version of A4-CREB-BFP with two *f* position residues in the leucine zipper mutated to Gly (residues N318G, E325G in the murine CREB sequence). The *f* position is surface exposed, and Gly substitution at these sites is expected to have less of an effect on coiled-coil stability (Table 3-4).[162] The activity of these inhibitors in the hitchhiker-bandpass assay (Figure 3-13) showed that the *d* position mutations drastically decreased the strength of the inhibitor, while the *f* position mutations had no detectable effect compared to the non-mutated A4-CREB-BFP inhibitor.

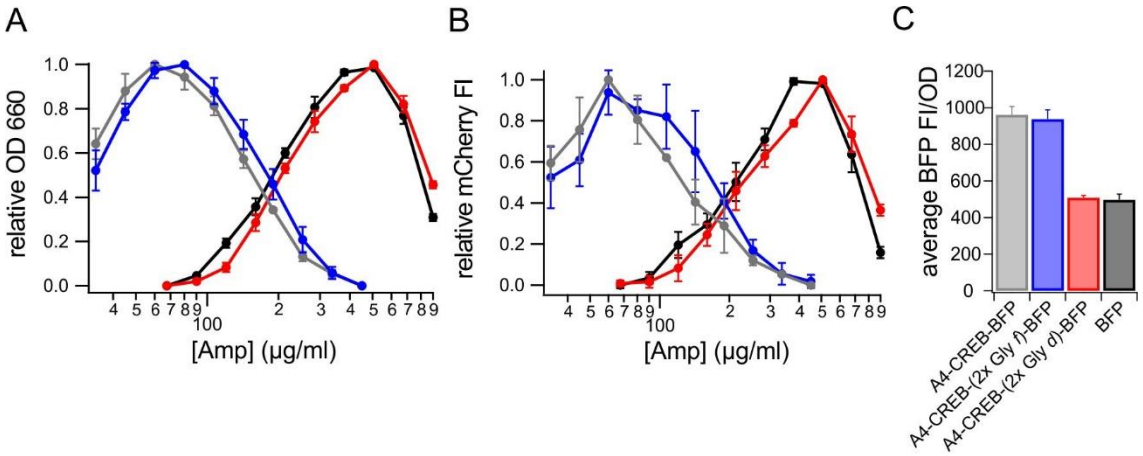


Figure 3-13: Assessment inhibitor strength using a second set of interaction partners.
The CREB hitchhiker pair was co-expressed with a CREB-based inhibitor: BFP negative control (black), A4-CREB-BFP (gray), A4-CREB-BFP with two *f* positions mutated to Gly (blue) and A4-CREB-BFP with two *d* positions mutated to Gly (red). Growth (OD_{660}) (A), expression of the fluorescent mCherry reporter (B) and inhibitor expression levels (BFP fluorescence) (C) are shown. OD and mCherry fluorescence values show the mean (n = 3), and error bars reflect the SD. BFP fluorescence values are shown as the mean of the BFP fluorescence normalized by the OD, using data from wells that had growth (defined as an OD within 75% of the maximum OD of the culture).

To confirm that the change in Amp resistance of the cells was due to a specific interaction between the expressed inhibitor and the hitchhiker pair, we tested the system with a non-specific inhibitor. To do so, we co-expressed the strong inhibitor A4-CREB-BFP with the cFos(ZIP)/cJun(bZIP) hitchhiker pair. This had no detectable effect on the Amp resistance of the cells in comparison to cells co-expressing a BFP negative control with the cFos(ZIP)/cJun(bZIP) hitchhiker pair (Figure 3-14).

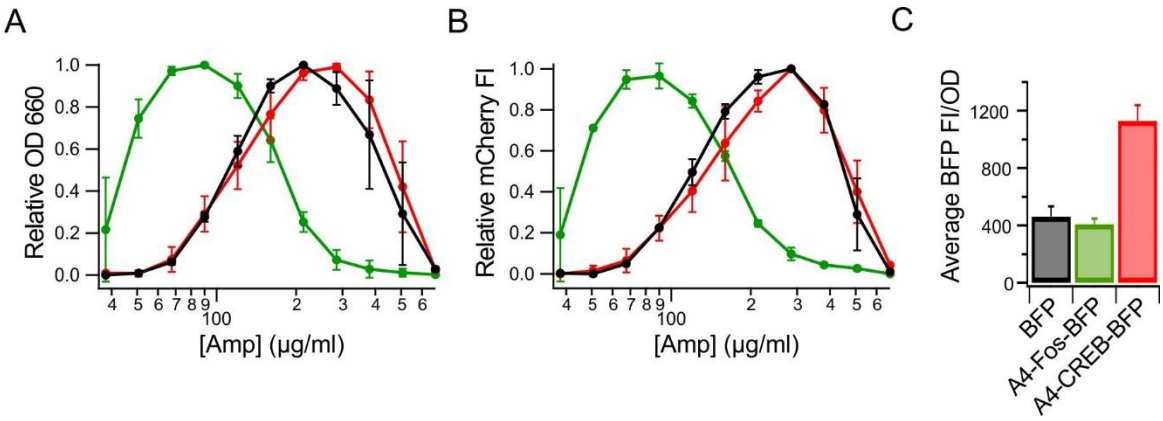


Figure 3-14: Testing the hitchhiker-bandpass assay with a non-specific inhibitor.
Using the hitchhiker pair cFos(ZIP)/cJun(bZIP) and the mCherry bandpass plasmid, A4-Fos-BFP, A4-CREB-BFP, and BFP were co-expressed to test inhibitor specificity. Cultures expressing the strong inhibitor (A4-Fos-BFP, green) grew in low concentrations of Amp, characteristic of weakly interacting hitchhiker pairs (*i.e.* strong inhibition of hitchhiker complex formation by the inhibitor). Cultures expressing the non-specific inhibitor (A4-CREB-BFP, red) or the non-binding control (BFP, black) grew in the same Amp concentrations, indicating hitchhiker complex formation was not inhibited. Growth (OD_{660}) (A), expression of the fluorescent mCherry reporter (B) and inhibitor expression levels (BFP fluorescence) (C) are shown. Data shows the mean values (n = 3); error bars show the SD. BFP fluorescence values, indicating inhibitor expression levels, are shown as the mean of the BFP fluorescence normalized by the OD, using data from wells that had growth (defined as an OD_{660} within 75% of the maximum OD of the culture).

3.4 Summary

Protein–protein interactions *in vivo* are tuned through evolutionary pressure to optimize function; this tuning may produce weak-, intermediate- or high-affinity interactions. Currently available selection systems, however, focus on the development of interactions with high affinity, recovering all interactions above a set threshold. Using well-defined test cases of known affinity, it was shown that the hitchhiker-bandpass circuit allows selection pressure to be tuned to recover protein–protein interactions with affinities falling within a desired range. The ability to select for weaker or intermediate, or switchable protein–protein interaction affinity *in vivo* is expected to be useful for biotechnological applications requiring tuned interactions.

4 Evaluation of small molecule modulators of protein-protein interactions in the hitchhiker-bandpass circuit

4.1 Introduction

The ability to select for protein-protein interactions of a desired affinity using the hitchhiker-bandpass circuit could be useful for the development of small molecule modulators of protein-protein interactions. The genetic circuit would offer flexibility, enabling evaluation of compounds designed to either inhibit or stabilize protein-protein interactions, and a high degree of robustness to false positive results, since the read-out is always a positive signal (*i.e.* β-lactamase activity that is increased or decreased). To this end, hitchhiker constructs were developed for a well-characterized protein-protein interaction, and the interaction was challenged in the hitchhiker-bandpass circuit with an array of commercially available small molecule inhibitors, testing this idea as a proof-of-principle study. Unfortunately, *E. coli* were found to be impermeable to all small molecule inhibitors tested. What follows is a summary outlining the project objective, encountered difficulties, attempted solutions and overall conclusion.

The model protein-protein interaction chosen for this study was selected from the B cell lymphoma 2 (Bcl-2) gene family. Members of this family have been found to play a critical role in numerous types of cancer;[164] as such, these proteins are well-characterized and numerous commercially available inhibitors have been developed to target their interactions.

4.1.1 B cell lymphoma 2 gene family

Members of the Bcl-2 gene family regulate the intrinsic apoptosis pathway by balancing intracellular pro- and anti-apoptotic signals.[164,165] This family is named after Bcl-2, the first discovered apoptotic regulator, which was found to be overexpressed in follicular lymphomas.[166,167] Unlike classical growth-promoting oncogenes, Bcl-2 enables cells to survive under conditions normally leading to apoptosis, effectively allowing the cells to persist until other oncogenes become active.[168] In healthy cells, pro- and anti-apoptotic Bcl-2 family members interact to regulate programmed cell death; it is now well-established that evasion of apoptosis by cancer cells frequently occurs through the up-regulation of anti-apoptotic Bcl-2 family members, such as Bcl-2.[164]

Bcl-2 is a small, all α-helical protein with a hydrophobic groove that mediates protein-protein interactions.[169] Bcl-2 binds and sequesters the Bcl-2 homology-3 (BH-3) domain of pro-apoptotic effectors, such as Bcl-2-associated X protein (BAX), and pro-apoptotic BH3-only proteins, such as Bcl-2-associated agonist of cell death (BAD).[170] The crystal structure of Bcl-2 is shown in Figure 4-1AB as a ribbon diagram and as a surface, bound to the BH3 domain of BAX.

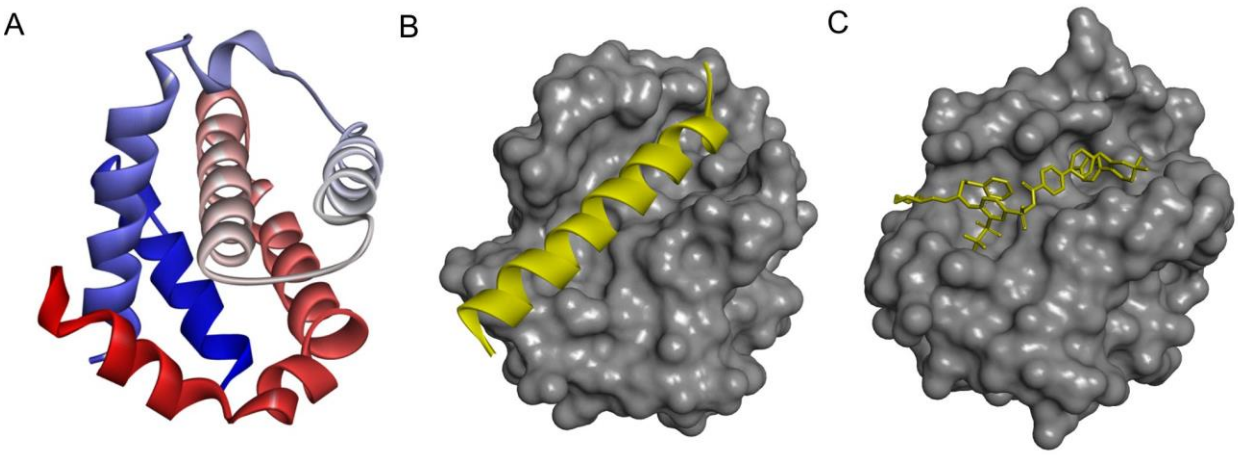


Figure 4-1: Crystal structures of Bcl-2.
(A) Bcl-2 is shown as a ribbon diagram (2XA0,[169] coloured from the N- to C-terminus blue to red) and as a surface, bound to (B) the BH3 domain of BAX (2XA0[169]) or (C) the inhibitor navitoclax (4LVT[170]). The models were visualized using BIOVIA Discovery Studio 2019.

4.1.2 BH3-mimetics to antagonize Bcl-2 proteins

The development of inhibitors targeting specific anti-apoptotic Bcl-2 family proteins, to restore the natural process of programmed cell death, holds compelling therapeutic potential for cancer treatment.[165] Over the last two decades, numerous small, nonpeptidic organic compounds have been developed that bind to the BH-3 binding groove of anti-apoptotic Bcl-2 family members, blocking their ability to sequester pro-apoptotic signals (Figure 4-1C). Many of these inhibitors have progressed through various stages of clinical trials,[170–172] and the Bcl-2 inhibitor venetoclax has been granted approval by the U.S. Food and Drug Administration (FDA) for several treatment regimens for specific types of leukemia,[165,173] demonstrating the efficacy of this approach.

To test the ability of the hitchhiker-bandpass circuit to detect inhibition of protein-protein interactions by small molecule inhibitors, a hitchhiker construct was developed based on the interaction of Bcl-2 with the BH3 domain BAD, and the interaction was challenged using several commercially available Bcl-2 inhibitors.

4.2 Materials and Methods

4.2.1 Bacterial strains and media

All DNA manipulations were performed in *E. coli* XL1-Blue (Agilent Technologies). Hitchhiker-bandpass assays were performed in *E. coli* SNO301[139,140] as discussed in Chapter 3, or in *E. coli* SNO301 *ΔtolC*, as discussed in the main text. *E. coli* SNO301 has resistance to Strep.[92]

LB broth, LB agar, and DYT broth were used for cell growth. Unless otherwise indicated, antibiotics were used at the following concentrations: Amp 100 µg/mL; Spec 100 µg/mL; Strep 100 µg/mL; Cm 25 µg/mL; Kan 50 µg/mL; and Tet 20 µg/mL.

4.2.2 Plasmid construction

The Bcl-2/BAD hitchhiker interaction was constructed in the pACYCDuet1*tac* hitchhiker vector based on the design outlined in Chapter 3 using Gibson assembly.[113] Oligonucleotides were ordered from Sigma-Aldrich. Purification of DNA fragments and plasmid DNA was performed using a DNA clean-up kit and a plasmid isolation kit respectively, both from Macherey-Nagel. The genes for Bcl-2 and Bcl-X_L were acquired from AddGene (plasmids 8768 and 8749 respectively).[174,175] The coding sequence for Bcl-2 was modified to contain a loop from the homologue Bcl-X_L, as described by Petros *et al.*[176] in order to promote solubility, as discussed in the main text. The Bcl-2 gene was inserted into the first expression site of the pACYCDuet1*tac* hitchhiker vector as a fusion to the signal sequence ssTorA. The coding sequence of the BAD peptide was ordered as oligonucleotide fragments, annealed and inserted into the second expression site in the hitchhiker vector as an N-terminal fusion to the β-lactamase gene. The plasmid sequence was confirmed via sequencing by GATC Biotech/Eurofins. The plasmid map for the Bcl-2/BAD hitchhiker interaction is shown in Figure 4-2. The mCherry bandpass plasmid (pTS1-mCherry), described in Chapter 3, was used for all hitchhiker-bandpass experiments.

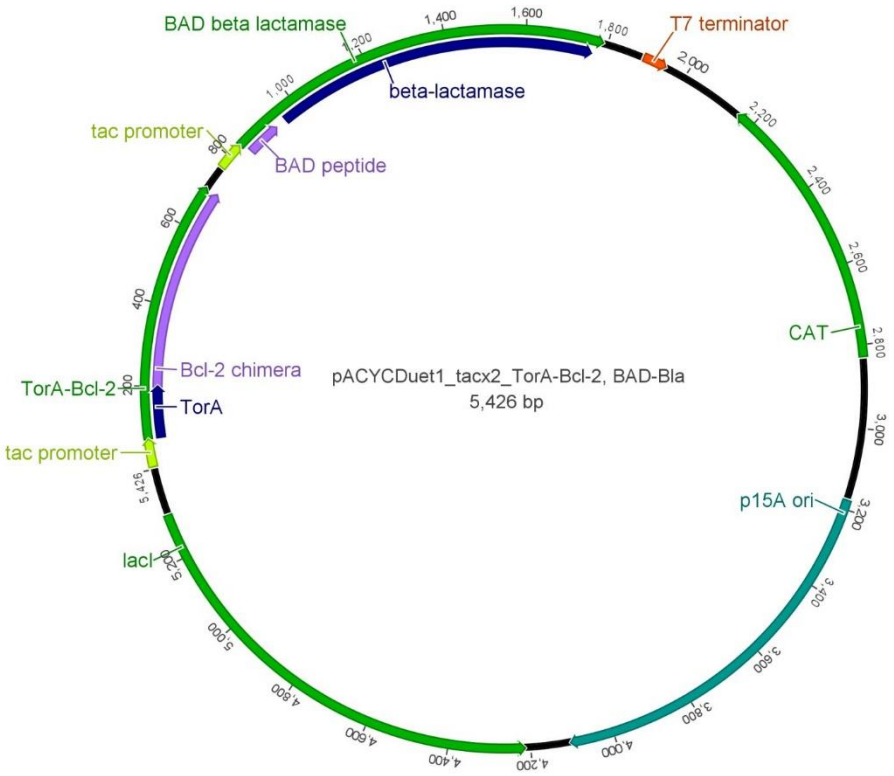


Figure 4-2: A hitchhiker plasmid for the Bcl-2/BAD interaction.
The hitchhiker plasmid for the Bcl-2/BAD interaction was constructed using a pACYCDuet1 plasmid backbone with *tac* promoters for expression *E. coli* SNO301. Bcl-2 was expressed as a fusion to ssTorA, and BAD was expressed as a fusion to β-lactamase. The plasmid contains a *p15A* origin of replication, a Cm resistance gene (*cat*, encoding chloramphenicol acetyltransferase, CAT) and a gene for the lac repressor (*lacI*) to decrease uninduced expression from the *tac* promoters. The plasmid map was created using Geneious v5.0.[152]

4.2.3 Genomic modification

The deletion of the gene for the TolC membrane transporter was performed using the phage λ Red recombinase method.[177] The λ Red plasmids, pSIJ8 and pKD4, were obtained from AddGene.[177] The plasmid pSIJ8 encodes the λ Red genes and flippase (Flp) recombinase gene (AddGene: 68122). The plasmid pKD4 encodes a template to design a PCR fragment for genomic integration and gene disruption (AddGene: 45605).

First, a DNA fragment was amplified that was designed to integrate into the genome of *E. coli* at the site of the *tolC* gene. The completed fragment contained a Kan resistance gene, flanked by

Flp recognition target (*FRT*) sites, flanked by 50 base pairs of sequence homologous to the targeted *tolC* gene. Primers were used with 5′ phosphorothioate bonds to protect the linear oligonucleotide from degradation upon subsequent electroporation (see below). The homologous sequence used for genomic integration at the *tolC* site was described by Baba *et al.*[178] The primer sequences are shown in Table 4-1. The resulting PCR fragment was purified using a PCR clean-up kit from Macherey-Nagel.

Table 4-1: Primers used for generation of the *tolC* deletion oligonucleotide.

Name	**Sequence (5′ - 3′)**
dTolC-Keio_fwd	A*A*T*AATTTTACAGTTTGATCGCGCTAAATACTGCTTCACCAC AAGGAATGGTGTAGGCTGGAGCTGCTTC
dTolC-Keio_rev	C*A*G*ACGGGGCCGAAGCCCCGTCGTCGTCATCAGTTACGGAA AGGGTTATGGGTCCATATGAATATCCTCCTTAGTTC

The asterisks (*) represent phosphorothioate bonds. The sequence shown in black is homologous with the *tolC* gene, and the underlined sequence in grey is homologous with the PCR template.

Next, electrocompetent cells expressing the λ Red phage were prepared. A culture of *E. coli* SNO301 transformed with pSIJ8 was used to make electrocompetent cells following a standard protocol, with two protocol modifications: 1) all incubation steps were performed at 30°C because pSIJ8 has a temperature-sensitive origin of replication, and 2) during the preparation of the log phase culture, 20 mM of arabinose was added to induce expression of the λ Red genes once the culture reached an OD_{600} of ~ 0.1.

The freshly made electrocompetent cells were electroporated with the prepared PCR fragment for the *tolC* deletion (with incubation of the outgrowth culture at 30°C). The cells were plated on an LB agar plate containing Amp for plasmid maintenance and Kan to select for genomic integration of the electroporated fragment, and were incubated overnight at 30°C.

The recovered colonies were genotyped using colony PCR to confirm the genomic integration of the electroporated fragment. There was a high rate of false positive colonies. A strain showing successful integration was selected and grown overnight at 30°C in 5 mL DYT with Kan and Amp to remove satellite colonies. Next, to remove the integrated Kan antibiotic cassette, the overnight culture was pelleted, resuspended in 5 mL DYT with Amp and rhamnose (50 mM) to induce expression of Flp recombinase and grown at 30°C for 5 h. The culture was then plated on

an LB agar plate with Amp. After overnight growth, individual colonies were screened for loss of the Kan resistance gene via replica plating and colony PCR. A successful colony was selected, and the cells were cured of the plasmid pSIJ8 by growing a culture at 37°C overnight (with Strep for strain selection, but without Amp or Kan pressure). The new strain is referred to as *E. coli* SNO301 *ΔtolC*.

4.2.4 Hitchhiker-bandpass assay

The hitchhiker-bandpass assay was performed as described in Chapter 3. Briefly, *E. coli* SNO301 (or SNO301 Δ*tolC*) were co-transformed to obtain cells with the Bcl-2/BAD hitchhiker plasmid and the mCherry bandpass plasmid. Hitchhiker-bandpass starter cultures were prepared by diluting an overnight culture to an OD_{660} of 0.5 in DYT with 15% glycerol and stored in aliquots at -80°C. The assay was performed in round-bottom 96-well plates prepared with DYT supplemented with plasmid maintenance antibiotics (Spec and Cm), an inducer (IPTG 0.5 mM) and bandpass selective pressure (Tet as well as an array of Amp concentrations). When applicable, small molecule inhibitors were added as described in the main text. Unless otherwise indicated, starter cultures were inoculated to a starting OD_{660} of 0.003. The prepared 96-well plates were incubated overnight at 32°C with 450 rpm. After incubation, the cells were pelleted via centrifugation (1500 rcf, 15 min) and the supernatant was removed. Cells were resuspended in 100 µL of 1× PBS and transferred to flat-bottom 96-well plates for evaluation in an Infinite M1000 Pro plate reader (Tecan) to measure OD_{660} and mCherry fluorescence (ex. 587 nm, em. 610 nm).

4.3 Results and Discussion

4.3.1 Development and analysis of Bcl-2/BAD hitchhiker pairs

A hitchhiker construct for the human Bcl-2/BAD interaction was developed based on the interaction studied by Petros *et al.*,[176] so that the reported *in vitro* affinity data could be used to approximate the binding affinity of the hitchhiker pairs. Petros *et al.* replaced a putative unstructured loop in Bcl-2 with a shortened loop from the homologous protein Bcl-X_L to circumvent previously reported poor solubility of wild-type Bcl-2.[176] This structural change was reported not to affect the affinity of Bcl-2 for its BH3 ligands and improved the behavior of the protein *in vitro*.[176] The reported solubility-optimized chimeric sequence (containing residues 1-

34 of Bcl-2, 35-50 of Bcl-X_L and 92-207 of Bcl-2) was prepared using the genes for Bcl-2 (isoform 2) and Bcl-X_L, which were acquired from AddGene (plasmids 8768 and 8749 respectively).[174,175] The resulting modified Bcl-2 sequence was inserted into the pACYCDuet1*tac* plasmid backbone with an N-terminal fusion to the signal sequence ssTorA. The BAD peptide sequence was ordered as oligonucleotides, annealed and inserted into the second expression site in pACYCDuet1*tac* with a C-terminal fusion to a β-lactamase gene. The amino acid sequences for the described hitchhiker constructs are shown in Figure 4-3.

ssTorA-Bcl-2

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAGSAHAGRTGYDNREIVMK
YIHYKLSQRGYEWDAGDDVEENRTEAPEGTESEVVHLTLRQAGDDFSRRYRRDFAEMS
SQLHLTPFTARGRFATVVEELFRDGVNWGRIVAFFEFGGVMCVESVNREMSPLVDNIAL
WMTEYLNRHLHTWIQDNGGWDAFVELYGPSMR*

BAD-β-lactamase

MASNLWAAQRYGRELRRMSDEFVDSFKKGAPSGSSGGSGGSSHPETLVKVKDAEDQL
GARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQN
DLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHV
TRLDRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVA
GPLLRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQI
AEIGASLIKHWGGHHHHH*

Figure 4-3: Amino acid sequences for the Bcl-2/BAD hitchhiker pair.
Bcl-2 (isoform 2) was designed with an N-terminal fusion to ssTorA (green). The modified sequence for Bcl-2 is shown in dark blue. The residues from Bcl-X_L are underlined (see text). The BAD peptide (red) was designed as an N-terminal fusion to β-lactamase (light blue). Capping sequences were used to stabilize the fold of the short BAD peptide (pink). A short flexible linker (grey) was used in the fusion. The sequence contains a C-terminal His tag (yellow).

The Bcl-2/BAD hitchhiker interaction was tested in the hitchhiker-bandpass assay. A strain of *E. coli* SNO301 co-transformed with the Bcl-2/BAD hitchhiker plasmid and a bandpass plasmid containing an mCherry fluorescent marker (pTS1-mCherry, see Chapter 3), was inoculated to a starting OD_{660} of 0.001 into prepared media with plasmid maintenance antibiotics (Spec and Cm), an inducer (IPTG 500 μM) and bandpass selective pressure (Tet and an array of Amp concentrations). After overnight incubation (32°C, 450 rpm), maximal growth and mCherry fluorescence was observed with Amp selective pressure of 500 μg/mL, indicating a strong

interaction of the hitchhiker pairs (Figure 4-4). This is consistent with the observation of Petros *et al.,* who reported the 25-residue BAD peptide to interact with the Bcl-2 chimera with an affinity of 8 nM.[176]

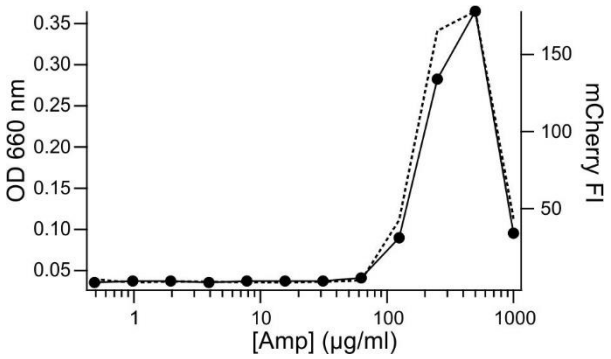


Figure 4-4: Hitchhiker-bandpass experiment with Bcl-2/BAD interaction.
A hitchhiker-bandpass experiment using the Bcl-2/BAD hitchhiker pair showed bandpass activity characteristic of strongly interacting protein pairs. Cells were inoculated into prepared media with plasmid maintenance antibiotics, an inducer and bandpass selective pressure (Tet and an array of Amp concentrations as annotated above). The graph shows optical density (···) and fluorescence (—) measurements.

4.3.2 Testing small molecule inhibitors

A set of five Bcl-2 inhibitors were purchase from ApexBio to test with the Bcl-2/BAD hitchhiker interaction. The inhibitors are summarized in Table 4-2. The chemical structures of the inhibitors are shown in the Appendix.

Table 4-2: Bcl-2 inhibitors tested in the Bcl-2/BAD hitchhiker-bandpass circuit.

Compound	**Molecular weight (g/mol)**	**Affinity with Bcl-2**	**Reference**
navitoclax (ABT-263)	974.61	< 1 nM	Ref. 171
TW-37	573.70	0.29 µM	Ref. 179
(+)-apogossypol	462.53	micromolar	Ref. 180
HA14-1	409.23	micromolar	Ref. 181
obatoclax mesylate	413.50	0.22 µM	Ref. 172

The Bcl-2/BAD hitchhiker interaction was challenged with each of the inhibitors in the hitchhiker-bandpass assay. Inhibition of Bcl-2/BAD hitchhiker complex formation would be observed as a decrease in cellular Amp resistance, as was demonstrated in Chapter 3 with peptide inhibitors. A schematic of the genetic circuit is shown in Figure 4-5.

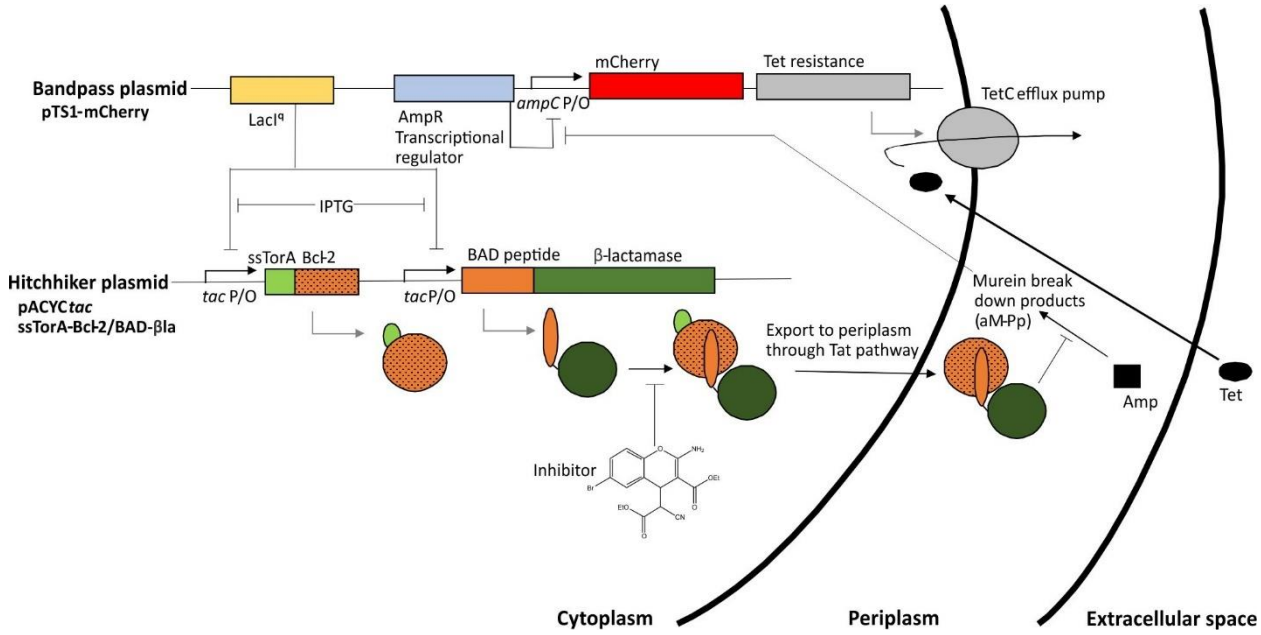


Figure 4-5: Schematic of the hitchhiker-bandpass circuit with small molecule inhibitors.
The hitchhiker-bandpass circuit, described in Chapter 3, was modified to contain hitchhiker constructs encoding the Bcl-2/BAD interaction (hitchhiker plasmid). Hitchhiker complex formation was challenged with small molecule inhibitors binding to Bcl-2. Binding of an inhibitor displaces the BAD peptide, inhibiting hitchhiker complex formation; this results in a decrease in cellular Amp resistance.

To test the Bcl-2 inhibitors, the hitchhiker-bandpass assay was performed as described above, with supplementation of each Bcl-2 inhibitor. Figure 4-6 shows the results from testing the Bcl-2/BAD interaction with navitoclax and TW-37 (with the tested inhibitor concentrations indicated in the figure legend). The results from testing obatoclax mesylate (10 μM), HA14-1 (10 μM) and apogossypol (1 μM) were comparable (data not shown). None of the tested inhibitors affected the cellular Amp resistance levels.

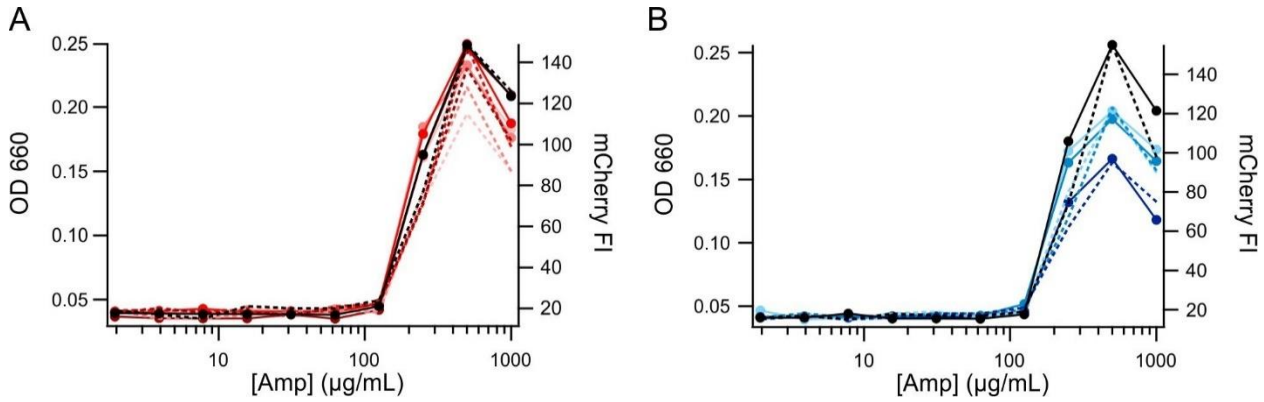


Figure 4-6: Testing small molecule inhibitors with the Bcl-2/BAD interaction.
Bcl-2/BAD hitchhiker-bandpass cultures were supplemented with inhibitors to test for *in vivo* inhibition. (A) Cultures were supplemented with navitoclax to yield a final concentration of 0.1 nM, 1 nM, 10 nM, or 100 nM (coloured from lightest to darkest shades of red). (B) Cultures were supplemented with TW-37 to yield a final concentration of 0.03 µM, 0.3 µM or 3 µM (coloured from lightest to darkest shades of blue). A negative control (1% DMSO) is shown in black in both (A) and (B). Inhibitors were dissolved in DMSO (1% final concentration). Optical density (—) and fluorescence measurements (···) from the bandpass plasmid are shown.

The most plausible reason for the observed lack of inhibition is that the inhibitors were not in the cytoplasm, either due to a lack of cell penetration or active efflux (see below). Several attempts were made to promote cytoplasmic accumulation of the inhibitors: the inhibitors were tested at elevated concentrations (HA14-1 at 40 µM, apogossypol at 10 µM, TW-37 at 200 µM, navitoclax at 200 µM); the overnight cultures for the bandpass starter stocks (see Methods section) were primed with inhibitors (10 µM navitoclax or 10 µM TW-37); and the amount of DMSO used in the assay was increased up to 5%. None of the attempted solutions showed any effect.

4.3.3 Modifying the permeability of *E. coli* to small molecules

The Bcl-2 inhibitors can penetrate mammalian cells; however, the permeability of the inhibitors to prokaryotic cells cannot be inferred due to structural differences of the cell wall. Indeed, the permeability barrier of prokaryotic cells is an unsolved problem that presents a major challenge in other areas of research, most notably in the development of novel antibiotics,[182] but also in the application of bacterial bioreporters[183] and cell-based mutagen screening.[184]

4.3.3.1 Permeability of Gram-negative bacteria

Gram-negative bacteria, such as *E. coli*, have a complex, multilayered cell envelope that functions as a highly sophisticated selectively permeable barrier.[185,186] The cell envelope consists of an inner membrane (IM) and an outer membrane (OM), framing an intervening layer of peptidoglycan in the periplasmic space.[185] The IM is a phospholipid bilayer; it restricts the passage of hydrophilic compounds due to the hydrophobic nature of the membrane interior.[185] The OM is an asymmetrical bilayer composed of phospholipids in the inner leaflet and lipopolysaccharides (LPS) in the outer leaflet.[186] The LPS content significantly slows diffusion of lipophilic solutes due to strong lateral interactions between LPS molecules.[186] A certain degree of permeability, however, is permitted by channel-forming proteins in the OM, called porins, that function as nonspecific diffusion channels.[186] Porins allow the passage of low-molecular-weight hydrophilic molecules (up to approximately 600 kDa); however, the channels show solute charge preferences and can be inactivated to regulate cell permeability.[186,187]

In addition to the permeability barrier created by the cell envelope, *E. coli* expresses numerous multidrug efflux pumps which expel adverse compounds that manage to make their way into the cytoplasm.[188] The efflux pumps function as modular, tripartite systems composed of an OM channel, an active transporter and a periplasmic fusion protein.[186,189] In *E. coli*, a single OM channel, TolC, is used for all efflux pumps.[190] The complex AcrAB-TolC has been shown to play a dominant role in drug efflux, expelling a multitude of structurally diverse compounds from the cell.[186,189,191] As such, inhibition of TolC is viewed as a potential strategy to increase the susceptibility of antibiotic-resistant bacteria[189,191] and it has been proposed that bacteria lacking multidrug pumps could be useful for drug discovery.[182,192] Strains with a *tolC* deletion have been created previously,[88,191,193,194] showing that this knockout is not lethal.

Improving the cytoplasmic concentration of the Bcl-2 inhibitors could be achieved either by permeabilizing the *E. coli* OM or by inhibiting endogenous multidrug efflux pumps. Permeabilizing the OM, however, has been shown to allow leakage of periplasmic enzymes, such as β-lactamase.[195] Therefore, it was decided to delete the OM channel TolC to inhibit endogenous multidrug efflux pumps.

4.3.3.2 Deletion of the TolC efflux channel

The deletion of the TolC efflux channel was achieved using the λ Red protocol[177,178] for genomic knockouts, as is detailed in the Methods section. The resulting strain is referred to as *E. coli* SNO301 *ΔtolC*. As reported previously, the deletion of *tolC* affected cellular antibiotic resistance.[191] New antibiotic working concentrations were determined experimentally for the modified bandpass strain. Spec and Cm were used at half their standard concentrations (50 μg/mL and 12.5 μg/mL respectively) and Tet was used at 1 μg/mL. Contrary to previous reports,[191] the antibiotic resistance of Amp was not found to be affected by the knockout.

4.3.3.3 Hitchhiker-bandpass assays using *E. coli* SNO301 *ΔtolC*

The efflux channel knockout strain was tested to determine if removal of the TolC efflux channel enabled detection of the activity of the small molecule inhibitors in the hitchhiker-bandpass assay. *E. coli* SNO301 *ΔtolC,* co-transformed with the mCherry bandpass plasmid and the Bcl-2/BAD hitchhiker plasmid, was inoculated to a starting OD_{660} of 0.005 in prepared media containing plasmid maintenance antibiotics (Spec 50 μg/mL and Cm 12.5 μg/mL), an inducer (IPTG 500 μM) and bandpass selection pressure (Tet 1 μg/mL and an array of Amp concentrations ranging from 750 - 1000 μg/mL). Five replicates were prepared to test each inhibitor (HA14-1, apogossypol, obatoclax mesylate, navitoclax and TW-37). Inhibitors were dissolved in DMSO and added a final concentration of 10 μM (with a final DMSO concentration of 4%). A sixth replicate was performed with a 4% DMSO blank. The cells were incubated overnight at 32°C with 450 rpm. The following day, maximal growth for all cultures (including the negative control) was observed in media containing 250 – 750 μg/mL of Amp, indicating that none of the compounds inhibited the Bcl-2/BAD interaction (Figure 4-7). The culture supplemented with obatoclax mesylate did not grow, suggesting the compound was toxic.

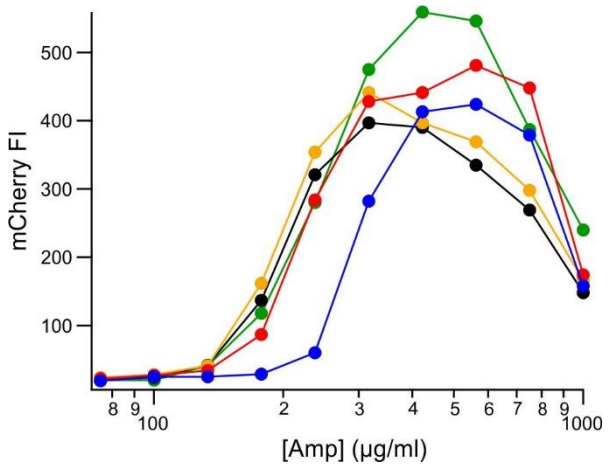


Figure 4-7: Testing for inhibition using *E. coli* SNO30l *ΔtolC.*
The efflux channel knockout strain *E. coli* SNO301 *ΔtolC* did not enable detection of Bcl-2 inhibitor activity in the hitchhiker-bandpass circuit. Shown are hitchhiker-bandpass cultures (mCherry bandpass plasmid, Bcl-2/BAD hitchhiker plasmid) with one of the following Bcl-2 inhibitors: apogossypol (green), HA14-1 (yellow), navitoclax (red) or TW-37 (blue). A culture with a DMSO blank is shown in black. The culture with obatoclax mesylate did not grow (not shown). All inhibitors were added to a final concentration of 10 µM with a 4% final DMSO concentration. The apparent increase in interaction strength shown by the culture with TW-37 (blue) is likely an artifact of the relatively mild Amp gradient.

4.4 Conclusions

Due to the permeability barrier of *E. coli,* it is not feasible to screen small molecule modulators of protein-protein interactions using the hitchhiker-bandpass circuit. While the Bcl-2/BAD protein-protein interaction behaved well in the hitchhiker circuit (Figure 4-4), no inhibition of the protein interaction was seen when the cultures were supplemented with known Bcl-2 inhibitors.

The lack of observed inhibition by the Bcl-2 inhibitors was hypothesized to be due to either a lack of penetration of the inhibitors into the *E. coli* cells, or due to expulsion of the inhibitors by endogenous multidrug efflux pumps. It may be possible to break the permeability barrier in *E. coli* by controlled hyperporination of the OM,[196] or by using OM permeabilizing agents such as polymyxin nonapeptide.[197] Permeabilizing the OM, however, would likely affect cellular Amp resistance and has been reported to result in leaking of periplasmic enzymes,[195] both of which would interfere with the bandpass circuit. Permeabilizing the OM, therefore, would likely be problematic. Instead, the endogenous efflux pumps in the bandpass strain *E. coli* SNO301 were inhibited by removing the OM channel TolC, by creating a genetic knockout. This did not solve

the problem; the new strain also did not show inhibition of the Bcl-2/BAD interaction by the Bcl-2 inhibitors in the hitchhiker-bandpass assay. This indicates that the lack of observed activity of the Bcl-2 inhibitors was not due to the expulsion of the inhibitors but rather due to their lack of permeability.

While controlled permeabilization of the OM could be attempted, it seems unlikely that a solution could be found that would enable unbiased entry of all compounds. Such a permeability bias would result in an unacceptably high number of false negatives. It is concluded that the idea of screening small molecules in the hitchhiker-bandpass circuit is not viable.

5 Towards the development of photo-controlled inhibitors of AP-1 and CREB transcription factors

5.1 Introduction

The spatiotemporal control afforded by photo-switchable protein activity provides unparalleled experimental precision, allowing complex molecular processes to be probed on a timescale relevant to their activity. Discussed here are on-going efforts towards the development of photo-controlled inhibitors of two mammalian transcription factors, AP-1 and CREB, using a directed evolution approach facilitated by the hitchhiker-bandpass circuit.

The hitchhiker-bandpass circuit, described in detail in Chapter 3, enables selection for strong or for weak protein-protein interactions; this makes it possible to evaluate switchable inhibitors in high-throughput, enabling a directed evolution approach. Recent examples of photo-switchable protein development using directed evolution have shown this method to be very successful.[15,16,33,40,47,48] Rational design strategies, while powerful, cannot definitively predict the effect of changes to a protein sequence, leading to labor-intensive cycles of protein design and analysis. The large structural changes associated with light-induced signal transmission make the rational design of photo-switchable proteins in particular a formidable task. Indeed, the rational design of photo-switchable proteins has often led to unexpected results.[18,25,48,198,199] The application of a directed evolution approach, using the newly developed hitchhiker-bandpass circuit, is hoped to facilitate the development and optimization of photo-switchable AP-1 and CREB inhibitors.

5.1.1 AP-1 and CREB

The transcription factors AP-1 and CREB orchestrate gene expression to translate extracellular stimuli into alterations in cellular processes. AP-1 activity regulates numerous processes including cellular proliferation, differentiation and apoptosis.[200] Not surprisingly, malfunctioning AP-1 activity has been observed in various types of cancer, making this transcription factor an important therapeutic target.[201,202] CREB is best known for its central role in the mechanisms underlying learning, memory and emotion.[203–205] The CREB signaling pathway is therefore viewed as a potential therapeutic target for neurodegenerative diseases and psychological

disorders, and it has even been suggested that modulating CREB activity could be used to improve normal cognitive performance.[206] Although the significance of AP-1 and CREB activity has been clearly established, the development of therapeutics targeting these transcription factors is hampered by an incomplete understanding of their activities on a molecular level.

Structurally, AP-1 and CREB seem optimized for extensive regulatory flexibility. These transcription factors bind to DNA as obligate dimers composed of homo- and heterodimerizing monomers from several subfamilies.[159,207] Monomers dimerize with a range of affinities,[119] and each monomer can interact with a unique array of regulatory proteins that tune recognition of the DNA-binding site, modulate activity levels and initiate downstream processes.[201,208,209] This combinatorial design, coupled with other regulatory mechanisms,[207] provides the means to precisely tune transcriptional events. This highly complex system leads to confounding observations, such as the discovery that monomers in the AP-1 complex can interact synergistically or antagonistically depending on the context.[202]

High levels of spatiotemporal complexity have been observed through experimental modulation of AP-1 and CREB activity in tissue culture and model organisms. This complexity is exemplified by CREB activity in the brain. Wallace *et al.* demonstrated that elevating CREB activity in rats before or after training in a model system for depression resulted in opposite phenotypes.[210] Likewise, it was shown that increasing the expression of CREB prior to fear conditioning increased the strength of a long-term fear memory,[211] and that such memories are preferentially encoded in neurons overexpressing CREB compared to untreated neighbouring neurons within the amygdala.[205,212]

The understanding of AP-1 and CREB activity has been improved using experiments based on elevating expression levels or reducing activity through knockouts; however, the spatial resolution and timescale of such experiments is insufficient to fully capture their complex patterns of activity. The experimental precision required to more carefully probe the roles of these transcription factors *in vivo* may be achieved using photo-controlled inhibitors.

5.1.2 Selection of active and inactive inhibitors using the hitchhiker-bandpass circuit

The hitchhiker-bandpass circuit will be used to evaluate libraries of photo-switchable AP-1 and CREB inhibitors, as shown in Figure 5-1. The photo-switchable inhibitor constructs will be co-expressed with hitchhiker pairs based on AP-1 or CREB; selection for weakly interacting hitchhiker pairs or for strongly interacting hitchhiker pairs will recover active and inactive inhibitors respectively. The use of a selection method enables larger libraries to be evaluated compared to screening methods, because cells expressing library variants that do not meet the selection criteria are eliminated. In addition, using an *in vivo*-based method for library evaluation imposes selective pressure for proteins that are soluble, stabile and specific, as variants are analyzed within the host proteome.

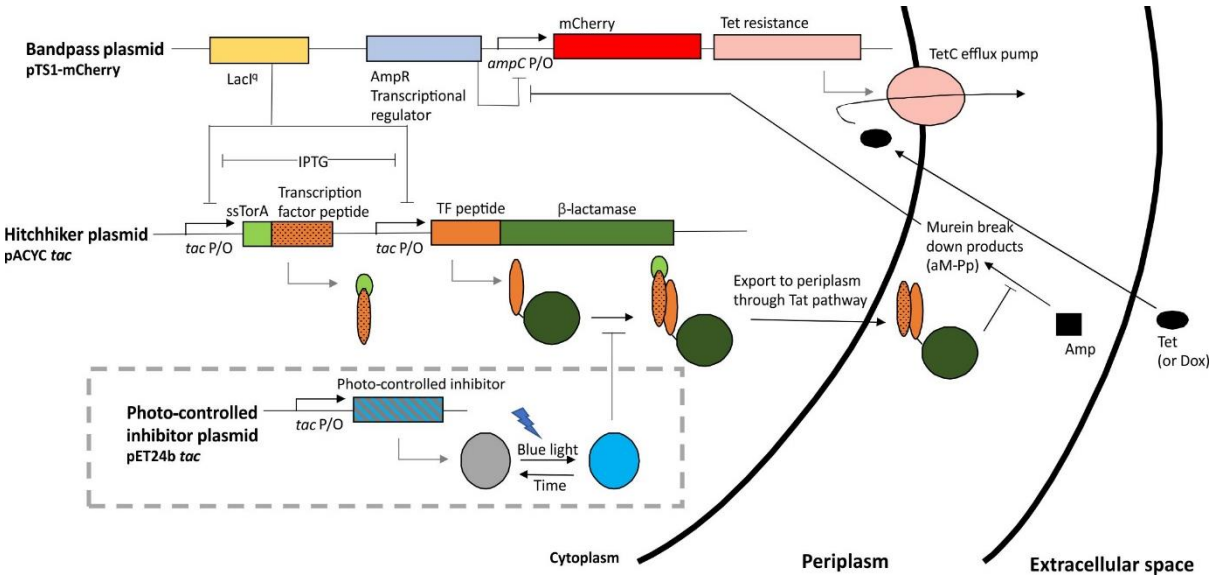


Figure 5-1: Schematic of the hitchhiker-bandpass genetic circuit designed to select for on- and off-states of the photo-controlled inhibitors of AP-1 and CREB.
The hitchhiker-bandpass circuit creates a bandpass for protein-protein interactions, as described in Chapter 3. The libraries of the photo-controlled inhibitors will be incorporated into the genetic circuit (grey dashed box). When a photo-controlled inhibitor is in its light-state (blue circle), it should interact with its target bZIP domain (either AP-1 or CREB) and thus interfere with hitchhiker complex formation, ultimately resulting in a decrease in cellular Amp resistance levels. When a photo-controlled inhibitor is in its dark-state (grey circle), it should not bind its target bZIP domain. Rounds of selections will be performed for active inhibitors under blue light using low Amp selection pressure and Tet (or Dox, as discussed later), and for inactive inhibitors in the darkness using high Amp selection pressure.

5.2 Materials and Methods

5.2.1 Bacterial strains and media

E. coli XL1-Blue (Agilent Technologies) was used for all DNA manipulations and to test for protein expression. *E. coli* SNO301[139,140] was used for hitchhiker-bandpass experiments, as discussed in Chapter 3. LB agar and DYT media were used for cell growth. Unless otherwise indicated, antibiotics were used at the following concentrations: Amp 100 μg/mL; Spec 100 μg/mL; Kan 50 μg/mL; Cm 25 μg/mL; and Tet 20 μg/mL.

5.2.2 Plasmid construction

Plasmids were constructed using Gibson assembly[113] and SLiCE.[142] Purification of DNA fragments (*e.g.* from agarose gels and after enzymatic reactions) was performed using a DNA clean-up kit from Macherey-Nagel. Plasmid purifications from *E. coli* were performed using a plasmid isolation kit from Macherey-Nagel. All plasmids were confirmed via sequencing (GATC Biotech/Eurofins). The photo-controlled inhibitor constructs were assembled in a modified pET vector (originally from Novagen) containing a *tac* promoter, enabling expression in *E. coli* SNO301 and XL1-Blue. The hitchhiker plasmids for the AP-1 and CREB interactions and bandpass plasmid were reported in Chapter 3. The plasmids discussed in this chapter are listed in Table 5-1.

Table 5-1: Plasmids used to develop photo-controlled inhibitors of AP-1 and CREB.

Plasmid name	**Description**
pTS1-mCherry	Bandpass plasmid (see Chapter 3)
pACYCDuet-*tac*-ssTorA-CREB(bZIP), CREB(ZIP)-βla	Hitchhiker plasmid for CREB interaction
pACYCDuet-*tac*-ssTorA-cFos(ZIP), cJun(bZIP)-βla	Hitchhiker plasmid for AP-1 interaction
pET24b-*tac*-BFP	Inhibitor negative control
pET24b-*tac*-cLOV-BFP	Inhibitor negative control
pET24b-*tac*-A4-CREB-BFP	Strong inhibitor control for the CREB hitchhiker interaction
pET24b-*tac*-BFP-cLOV-A4-Fos-His-v5	Preliminary photo-controlled AP-1 inhibitor design, template used to create libraries
pET24b-*tac*-BFP-cLOV-A4-CREB-His-v5	Preliminary photo-controlled CREB inhibitor design, template used to create libraries

5.2.3 Small-scale expression evaluation

Expression of the synthetic photo-controlled inhibitor constructs was confirmed via small-scale test expressions. Constructed plasmids were transformed into chemically competent *E. coli* XL1-Blue and plated on LB agar plates with Kan for plasmid selection. After overnight incubation at 37°C, a single colony was used to inoculate 15 mL of DYT media with Kan. The culture was incubated overnight at 37°C with 200 rpm. The following day, the overnight culture was used to inoculate 15 mL of fresh DYT with Kan to an OD_{600} of 0.1. The culture was incubated at 37°C with 200 rpm, and the OD_{600} was monitored. At an OD_{600} of 0.5 - 0.6, protein expression was induced with IPTG (1 mM) and the temperature was dropped to 30°C. Expression was allowed to proceed overnight, and culture samples (with a volume equivalating 1 mL of 1 OD_{600} culture) were taken 5 h post induction and the following morning. Expression was evaluated via SDS PAGE, western blotting and/or fluorescence measurements as appropriate.

5.2.4 SDS PAGE and western blotting

Cell pellets (1 mL of 1 OD_{600} culture) were boiled in 100 µL of Roti-Load-1 loading dye (Roth) for 15 min, and 20 µL was loaded on a gel. SDS PAGE gels were cast with a gradient from 7.5 – 15% acrylamide, run in a Tris-Glycine buffer and stained with Coomassie. When western blotting, the unstained SDS PAGE gel was transferred to PVDF membrane using a semi-dry western blotting apparatus. Blots were probed using mouse peroxidase-labelled monoclonal anti-poly-His antibodies (Sigma), detected using SignalFire™ ECL Reagent (Cell Signaling Technology) and imaged using a Fusion SL (Vilber Lourmat) imaging system.

5.2.5 Fluorescence spectroscopy

Cell samples (1 mL of 1 OD_{600} culture) were pelleted via centrifugation at 5000 rpm (2340 rcf) for 5 min. The pellets were resuspended in 1 mL of 1× PBS. The samples were measured in a flat bottom, clear 96-well plate (100 µL culture volume) in an Infinite M1000Pro plate reader (Tecan). OD_{600} and blue fluorescence (ex. 401 nm, em. 457 nm) were measured. Fluorescence measurements were normalized by cell density, and a negative control sample (cells without a plasmid) was included to account for intrinsic cell fluorescence.

5.2.6 Library generation

Libraries of the photo-controlled inhibitors were generated using PCR and Gibson assembly, as described in the main text. After incubation, the Gibson assembly reaction mixtures were diluted 3-fold in water and electroporated into freshly-made electrocompetent *E. coli* XL1-Blue cells. Cell dilutions of 1:10 and 1:100 were plated on LB agar plates (with Kan for plasmid selection) to quantify the electroporation efficiency. The remainder of the library was plated on large LB agar trays with Kan. Following incubation at 37°C, the colonies on the LB agar plates were quantified and the colonies on the LB agar trays were pooled. Each library was stored in cells as glycerol stocks, as well as in the form of purified plasmid DNA, recovered from the pooled colonies. The library sequence variation was examined using sequencing, as discussed in the main text.

5.2.7 Hitchhiker-bandpass assay

The hitchhiker-bandpass assay was performed as described in Chapter 3. Briefly, *E. coli* SNO301 was transformed to obtain strains with a hitchhiker plasmid, a bandpass plasmid and a photo-controlled inhibitor plasmid (or control). In this work, only the mCherry bandpass plasmid (*i.e.* pTS1-mCherry) was used to avoid unwanted excitation of the fluorescent protein when using blue light irradiation. Hitchhiker-bandpass starter cultures were prepared by diluting an overnight culture to a standardized OD_{660}. The assay was performed in round-bottom 96-well plates prepared with DYT supplemented with plasmid maintenance antibiotics (Kan, Spec and Cm), an inducer (IPTG 1 mM) and bandpass selective pressure (Tet or Dox (*vide infra*), as well as an array of Amp concentrations). Starter cultures were inoculated to a calculated starting OD_{660} and the 96-well plates were incubated overnight at 32°C with 450 rpm with the lid on. Irradiation was performed using an array of blue LEDs (450 nm) with an intensity of approximately 250 μW/cm^2. After incubation, the cells were pelleted via centrifugation at 4000 rpm (3724 rcf) for 15 min and the supernatant was removed. Cells were resuspended in 100 μL of 1× PBS and transferred to a flat-bottom 96-well plate for evaluation in an Infinite M1000Pro plate reader (Tecan). OD_{660} as well as the fluorescence of BFP (ex. 402 nm, em. 457 nm) and mCherry (ex. 587 nm, em. 610 nm) were measured.

5.3 Results and Discussion

5.3.1 Inhibitors targeting AP-1 and CREB

AP-1 and CREB are structurally related transcription factors, characterized by containing a bZIP domain that mediates dimerization and DNA binding. Due to the high structural homology of the bZIP domains of AP-1 and CREB, photo-controlled inhibitors for both transcription factors can be developed in parallel.

Inhibitors for AP-1 and CREB were designed by appending a negatively charged acidic extension to a leucine zipper from each AP-1 and CREB, based on the work by members of the Vinson research group (Figure 5-2).[149,150,161] In this design, the leucine zipper region conveys binding specificity while the negatively charged acidic extension creates favourable electrostatic interactions with the native basic domain. The acidic leucine zipper is referred to as an aZIP. The dominant negative (DN) versions of AP-1 and CREB are referred to as DN-AP1 and DN-CREB respectively.

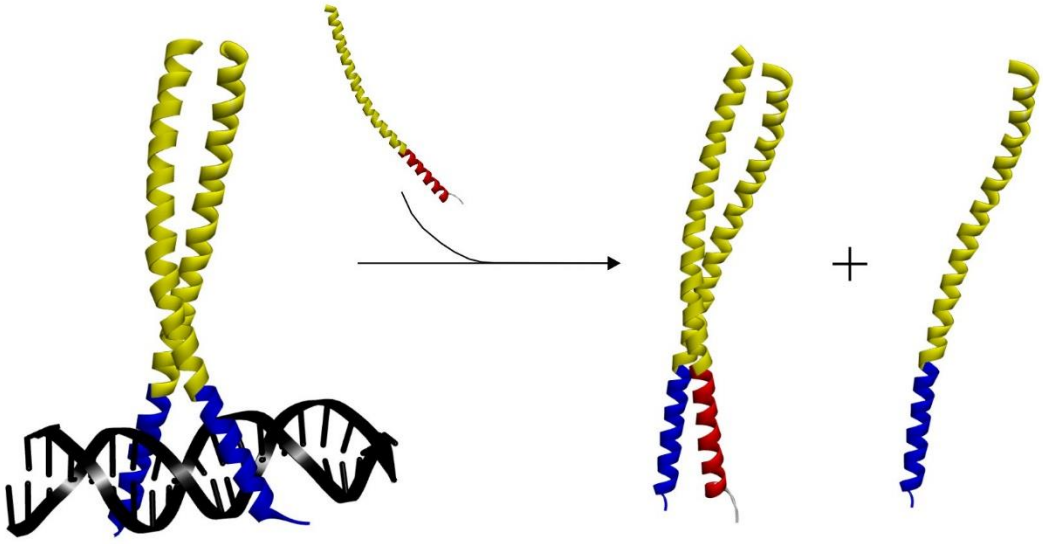


Figure 5-2: Schematic of inhibitor design for bZIP domains.
Strong inhibitors of bZIP proteins can be developed by fusing a leucine zipper with an acidic extension, in place of the native basic domain. The leucine zipper is shown in yellow, the basic domain in blue and the acidic extension in red. The inhibitor design is modelled using the experimentally determined structures of cFos (PDB code 1FOS[213]) and KIF21A, a long regulatory coiled-coil (PDB code 5D3A [unpublished]).

5.3.2 Photoactive protein scaffolds

As discussed in Chapter 1, there are numerous strategies that can be employed to confer photo-responsiveness on a target effector. The ability to evaluate photo-switchable inhibitor activity in high-throughput using the hitchhiker-bandpass assay makes it feasible to design and test numerous libraries based on different photoactive scaffolds and fusion designs. At the outset, this work will focus on different fusion strategies using three photoactive scaffolds: *As*LOV2, *Rs*LOV and PYP.

*As*LOV2 undergoes a reversible photocycle involving large protein structural rearrangements. The photocycle is triggered by the absorption of a photon of blue light by an FMN chromophore packed in the protein core, which leads to the formation of a transient cysteinyl-FMN adduct with a proximal cysteine residue.[12,214] The covalent bond formation causes a rearrangement of the hydrogen bonding network in the protein,[215] propagating the signal to the terminal structural motifs and ultimately causing the displacement and unfolding of an N-terminal α-helix (A′α) and a C-terminal α-helix (Jα) from the protein core (Figure 5-3).[8,216,217] The flavin-cysteinyl adduct decays spontaneously and the protein structure rearranges to the dark-state, completing the photocycle with a half-life of ~ 80 s, depending on the sequence length used for the LOV domain.[217] The large photoinduced structural changes in the *As*LOV2 domain have been successfully used on numerous occasions to create optogenetic tools.[3,22,23,25,103] These design strategies, in addition to reported photocycle-tuning point mutations,[216,218,219] provide extensive data for engineering optogenetic tools.

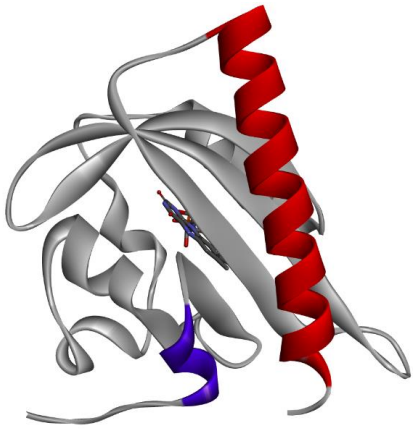


Figure 5-3: Crystal structure of LOV.
The structure of *As*LOV2 (PDB code 2V0U[216]) is shown in the dark-state. The core of LOV is shown in grey with the A′α helix highlighted in blue (T407 – I411) and the Jα helix highlighted in red (R521 – K544). The boundaries of the A′α and Jα helices are defined according to Zayner *et al.*[217]

The LOV domain from *R. sphaeroides, Rs*LOV, is a particularly interesting photoactive scaffold because it has been shown to undergo light-induced dimer dissociation,[20] a process which could be used to cage an effector through dark-state circularization. The dimerization is mediated by a C-terminal helix-loop-helix extension that forms a four-helical bundle in the dark-state (Figure 5-4).[20] The helix-loop-helix dimerization motif is flanked by a short N-terminal helical extension that likely also undergoes light-state structural rearrangments.[20] *Rs*LOV shares photochemistry common to other LOV domains, with a reversible photocycle mediated by the formation of a transient cysteinyl-FMN adduct triggered by blue light irradiation.[20] The light-state has an extended lifetime of approximately 40 minutes.[20]

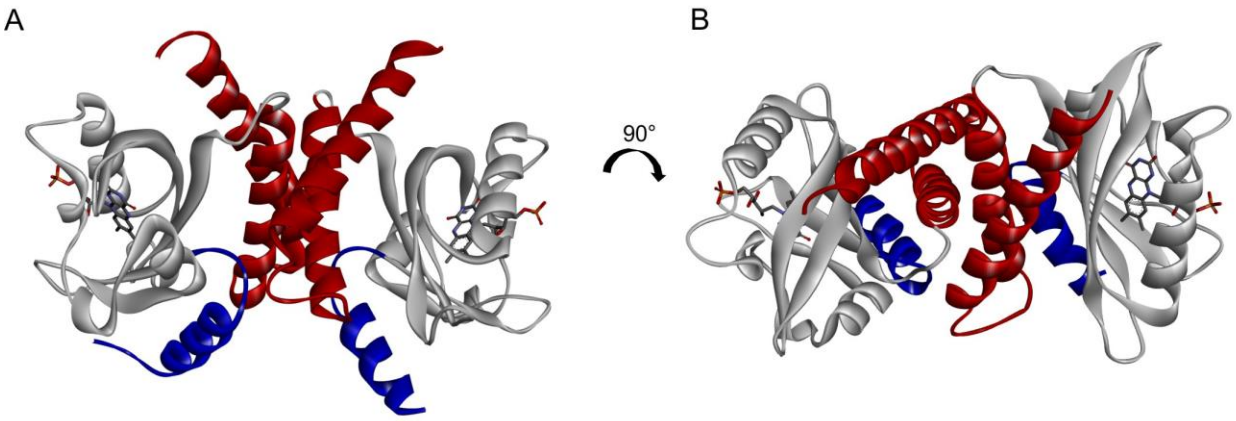


Figure 5-4: *Rs*LOV dark-state structure.
The crystal structure of a dimer of dark-state *Rs*LOV (PDB code 4HJ4[20]) is shown with a side view (A) and a top view (B). The C-terminal helix-loop-helix motif, composed of the Jα and Kα helices, are shown in red, mediating the dimer interface. The N-terminal helical extension is shown in blue.

PYP from *H. halophila* undergoes a reversible photocycle triggered by absorption of a photon of blue light by a covalently bound *p*-coumaric acid-based chromophore.[13,220] This causes the chromophore to undergo a *trans-cis* isomerization, which leads to global protein conformational changes, producing a light-state structure with characteristics of a molten globule (Figure 5-5).[10] PYP relaxes thermally to the dark-state with a half-life of ~ 1 s.[221] A variety of mutagenesis studies[221–225] have identified residues that tune the absorbance spectrum and alter signaling kinetics, providing useful data for engineering PYP-based photo-switchable tools. Most light-switchable constructs developed using PYP have been designed to cage effectors by creating chimeric fusions with the N-terminus of PYP.[25,95,96,198] Recently, a circularly permutated version of PYP was developed[14] that has been shown to be a useful scaffold for engineering switchable proteins.[16–18]

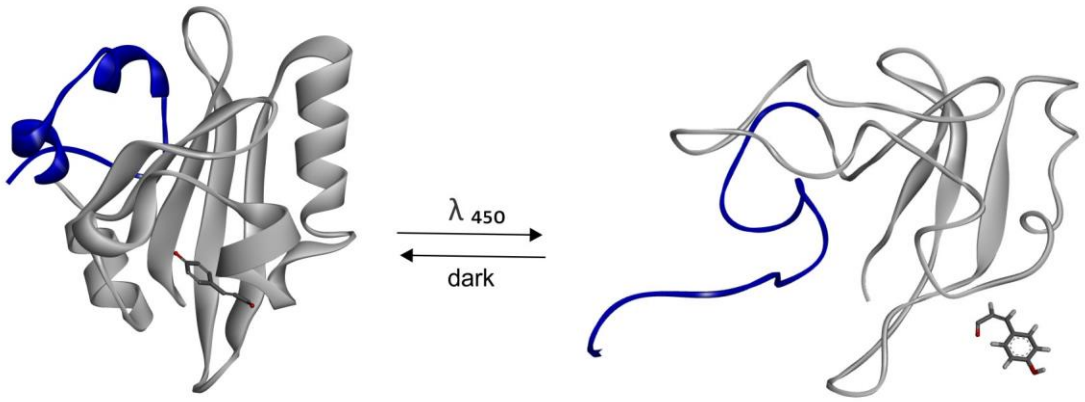


Figure 5-5: Dark-state and light-state structures of PYP.
The structure of PYP is shown in the dark-state (left, PDB code 1NWZ[226]) and as a snapshot of the light-state ensemble (right, PDB code 2KX6[10]). The N-terminal cap (residues 1-25), frequently used in engineering PYP-based optogenetic tools, is shown in blue.

In the following section, the design and production of libraries based on a caging strategy using an *As*LOV2 scaffold is discussed. Following this, several other protein design strategies, using PYP or *Rs*LOV, are considered for future development.

5.3.3 Photo-switchable inhibitors designed using an *As*LOV2 scaffold

As a first attempt, *As*LOV2 was selected as a photoactive scaffold to develop libraries of photo-controlled AP-1 and CREB inhibitors. Protein fusions were designed with the intention that in the dark-state, the inhibitor would pack on the core of LOV and be unavailable for binding with its target, while in the light-state, the increased structural flexibility of the photoactive scaffold would allow the inhibitor to bind its target bZIP monomer.

First, DN-AP1 and DN-CREB were each fused to the *As*LOV2 Jα helix with a region of sequence overlap, based on a strategy used by Lungu *et al.*[22] The *As*LOV2 Jα helix contains rudimentary sequence homology with the inhibitor acidic extension sequence, as shown in the sequence alignment in Figure 5-6A. Based on this alignment, the acidic extension was embedded in the Jα helix with the intension of creating a bi-functional chimeric sequence.

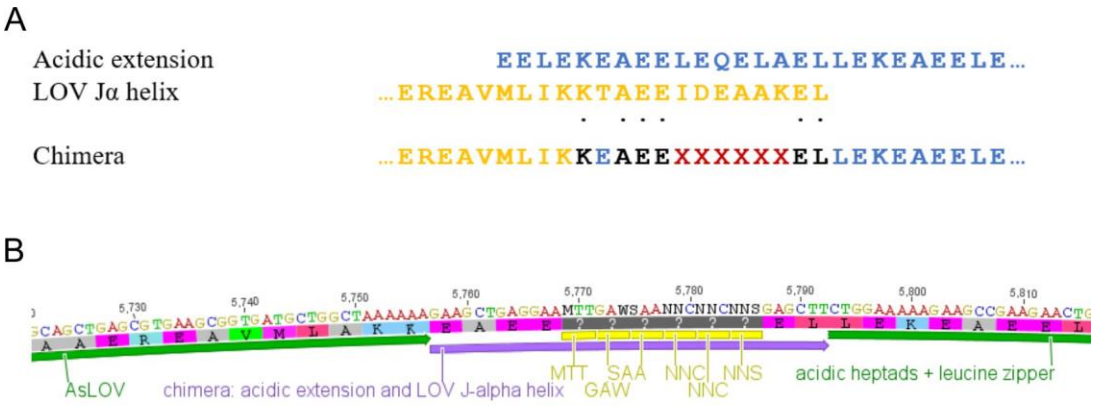


Figure 5-6: Sequence alignment for chimera design.
(A) A sequence alignment of the acidic extension and *As*LOV2 Jα helix is shown; the resulting designed chimeric sequence is shown below. Residues that are identical in the acidic extension and Jα helix are marked with a dot. In the designed sequence, residues are coloured according to the sequence of origin; 6 residues, marked with red x's, will be targeted for mutagenesis. (B) The chimeric overlap sequence is shown in detail with the sites targeted for mutagenesis annotated. Table 5-2 details the degenerate codons used. The figure was created with Geneious version 5.0.[152]

Within the chimeric sequence, six residues were chosen for randomization using a degenerate codon strategy, shown in detail in Figure 5-6B. At the first three sites targeted for randomization, degenerate codons were chosen such that the identity of the encoded residue was varied equally between the residue in the Jα helix sequence and the residue in the acidic extension sequence. For example, the mutagenic codon MTT was used at the first site, where M represents an equal mixture A or C, resulting in a mixture of CTT (encoding Leu) and ATT (encoding Ile) codons at this site. The remaining three sites were randomized more freely, using NNC or NNS degenerate codons, where N represents an equal mixture of all four nucleotides and S represents an equal mixture of G or T. The codons used in randomization, and the resulting encoded residues, are outlined in Table 5-2. Note that the library could potentially encode a stop codon in the sixth position.

Table 5-2: Degenerate codons used to randomize the chimeric overlap sequence.

Codon	Translation	Notes
MTT, where M = A/C	Ile/Leu	
GAW, where W = A/T	Glu/Asp	
SAA, where S = G/T	Gln/Glu	
NNC, where N = A/C/T/G	all expect Trp/Gln/Met/Glu	
NNS	all possible residues	includes 1 stop codon

The effectors DN-AP1 and DN-CREB are considerably longer than other peptides that have been caged in the LOV Jα helix.[22,23,227] In order to avoid constitutive binding activity due to dark-state availability of the inhibitors, a circular permutation in the *As*LOV2 scaffold was introduced based on a recently described circularly permutated *As*LOV2 (cLOV) scaffold.[15] cLOV was developed by creating new N- and C-termini by breaking the protein backbone in a loop between two β-sheets (Hβ and Iβ), such that K503 became the new N-terminus.[15] The original N- and C-termini were fused together with DN-AP1 or DN-CREB. Tethering both the N- and C-termini of the inhibitors to cLOV is expected to increase dark-state steric hindrance and may therefore yield tighter photo-control over the inhibitor activity. The cLOV scaffold also takes advantage of the photoinduced structural changes in the *As*LOV2 N-terminal A′α helix,[216] contributing to the light-state structural flexibility.

A hypothetical model of the designed cLOV-inhibitor constructs is shown in Figure 5-7. The sequence is modelled on the experimentally determined dark-state crystal structure of LOV[216] using the sequence for cLOV-DN-AP1. cLOV-DN-CREB would be expected to have a similar structure. The structures of the A′α and Jα helices in the light-state are hypothetical.

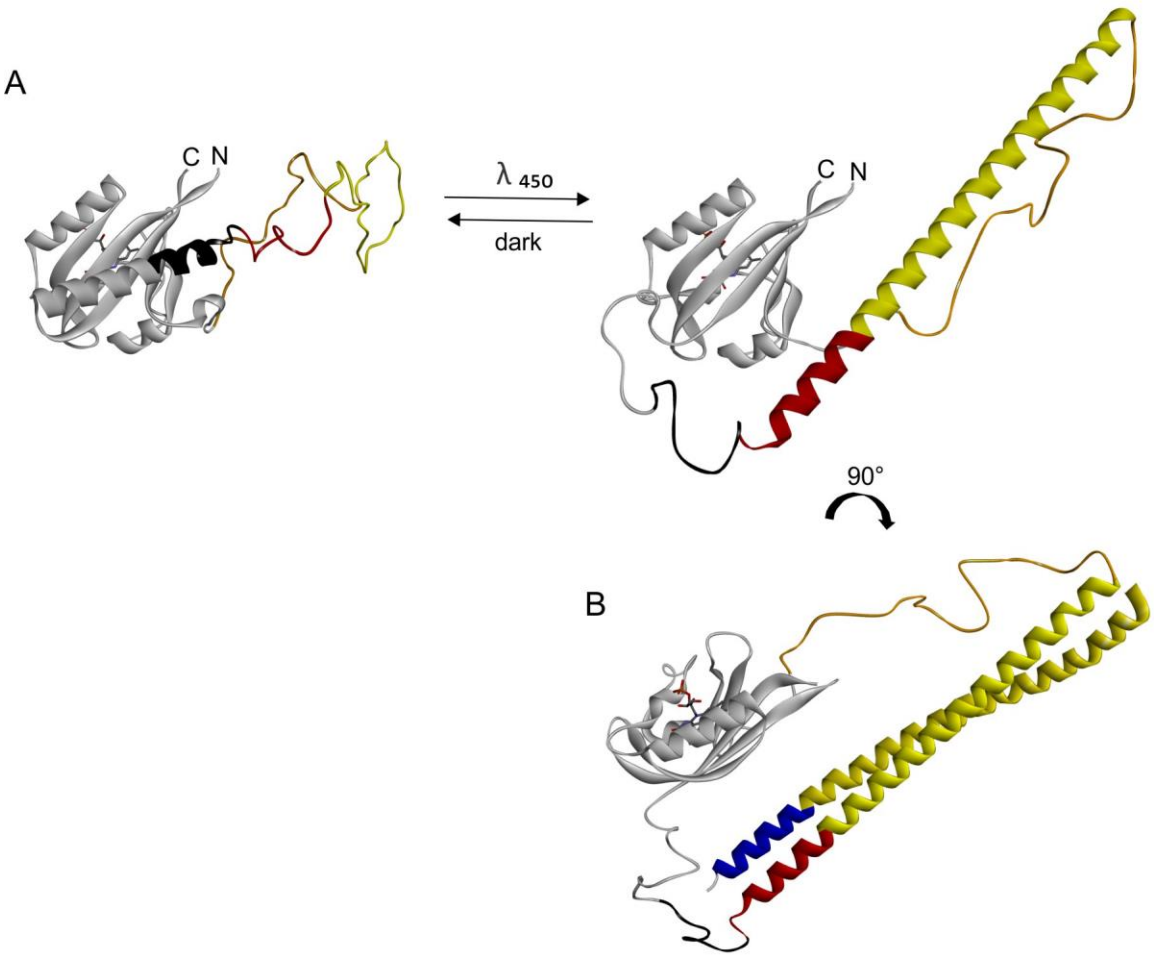


Figure 5-7: Hypothetical structural model of the designed photo-controlled inhibitors.
A hypothetical structural model of cLOV-DN-AP1 is shown, created on the dark-state crystal structure of *As*LOV2 (PDB code 2V0U[216]). (A) The dark-state (left) and light-state (right) of cLOV-DN-AP1 are shown. The model shows the LOV domain (grey), chimeric region (black), acidic extension (red), leucine zipper (yellow), and a linker (orange). (B) A rotated view of the light-state model is shown interacting with a bZIP monomer. The bZIP structures are coloured as in Figure 5-2 (leucine zipper in yellow, acidic extension in red and basic domain in blue). The structure of the LOV Jα and A′α helices in the light-state conformation are hypothetical. The models were created with BIOVIA Discovery Studio 2019.

A flexible linker was added C-terminal to the inserted inhibitor. The length of this linker sequence is expected to be a critical component of the design; it must be long enough to ensure sufficient structural flexibility to allow cLOV to assume its native fold, yet it must sterically restrict dark-state activity of the inhibitor. The length of this linker was therefore targeted for randomization using a series of primers designed to truncate its length.

Libraries were designed using a PCR-based strategy to vary the chimeric sequence and linker length (Figure 5-8). A segment of the designed photo-controlled inhibitor coding sequence was amplified with a forward primer containing degenerate codons and a set of reverse primers designed to bind at staggered sites along the linker. The reverse primers were designed to create linkers of 0, 4, 8, 12, 16, 22 or 30 amino acids in length. Each primer additionally contained sequence homologous with the plasmid backbone to enable insertion of the library fragments using Gibson assembly.[113] The plasmid backbone is a modified pET24b plasmid with a *tac* promoter, chosen for its compatibility with the other plasmids used in the hitchhiker-bandpass circuit. The primers used in library construction are listed in the Appendix.

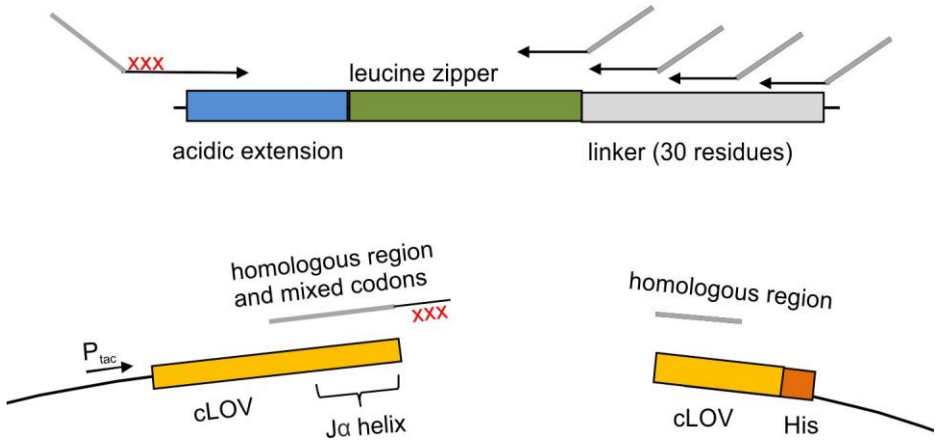


Figure 5-8: Schematic of the PCR-based library generation.
The library was created using one forward primer with mixed nucleotide content (red x's) and seven reverse primers, designed to truncate the length of the linker. For simplicity, only four reverse primers are shown. All primers contained homologous sequence on the 5′ end for recombination with the parent plasmid (thick grey lines). The library fragments were amplified from a truncated template composed of the coding sequence for the acidic extension (blue), leucine zipper (either cFos or CREB, green) and linker (grey). The template did not contain the sequence targeted for randomization by the forward primer (to avoid bias in primer annealing) or the sequence used in homologous recombination (to avoid annealing of the 5′ end of the reverse primers). The homologous sequences used in the primers are mapped on the plasmid backbone to illustrate the locations of the homology (thick grey lines). The plasmid backbone shows the *tac* promoter (P_{tac}), cLOV sequence (yellow) and a His tag (orange).

The forward primer with mixed nucleotide content encodes 43,008 possible protein sequence variants. Combined with the seven reverse primers, each library encodes 301,056 different protein variants in total.

5.3.3.1 Verifying protein expression

Several versions of the cLOV-inhibitor constructs were designed and tested for expression. To detect expression, constructs were fused to a blue fluorescent protein variant, mTagBFP2[151] (referred to as BFP for simplicity). Expression of the constructs was monitored using fluorescence measurements, SDS PAGE and western blotting. Expression was only detected with constructs containing an N-terminal BFP fusion; no expression was detected with a C-terminal fusion or without BFP, suggesting BFP additionally acts as a solubility tag.

The sequences used in the most recent designs for cLOV-DN-AP1 and cLOV-DN-CREB are shown in the Appendix, as well as data confirming their expression in *E. coli*. The sequence for the acidic extension originates from the work of Vinson and colleagues[149] and sequences for the leucine zippers from AP-1 and CREB originate from the human cFos bZIP domain and the human CREB bZIP domain respectively. After verifying expression, the designed constructs were used as templates to generate libraries.

5.3.3.2 Library generation

To generate the libraries, seven PCR reactions were performed in parallel for each cLOV-DN-AP1 and cLOV-DN-CREB using a prepared template with the appropriate leucine zipper sequence. Individual reactions were performed with each reverse primer, rather than pooling the reverse primers, to avoid enrichment of shorter fragments. A representative agarose gel with the resulting amplified inserts for cLOV-DN-AP1 is shown in Figure 5-9.

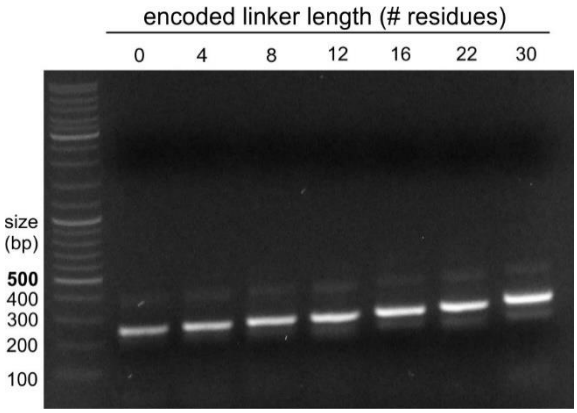


Figure 5-9: Representative agarose gel showing the insert library for cLOV-DN-AP1.
The library inserts were prepared in seven parallel PCR reactions and the products of the PCR reactions were run on a 1% agarose gel in 1× TAE. The gel shows inserts with incrementally increasing linker lengths. The ladder (GeneRuler 1 kb Plus DNA ladder) is labelled on the left.

To evaluate the library construction protocol, an aliquot of each library fragment was inserted into the prepared plasmid backbone via Gibson assembly and electroporated into *E. coli* XL1-Blue. The recovered colonies from each electroporation were pooled and used to inoculate an overnight culture, creating one mixed culture per electroporation. After overnight growth, the cultures were used to purify plasmid DNA for sequencing (performed by GATC/Eurofins). The sequencing chromatograms confirmed that the linker fragments assembled as expected (data not shown) and that the chimeric overlap region contained sequence variation, consistent with the mutagenic strategy (Figure 5-10). Note that in the sites with mixed nucleotide content, there was an enrichment G and C, likely because a G-C base pair contains three hydrogen bonds, while an A-T base pair contains only two hydrogen bonds. This creates a bias in the amino acid content of the resulting library. This is especially evident at the first site targeted for randomization, where the nucleotide content was varied between C and A.

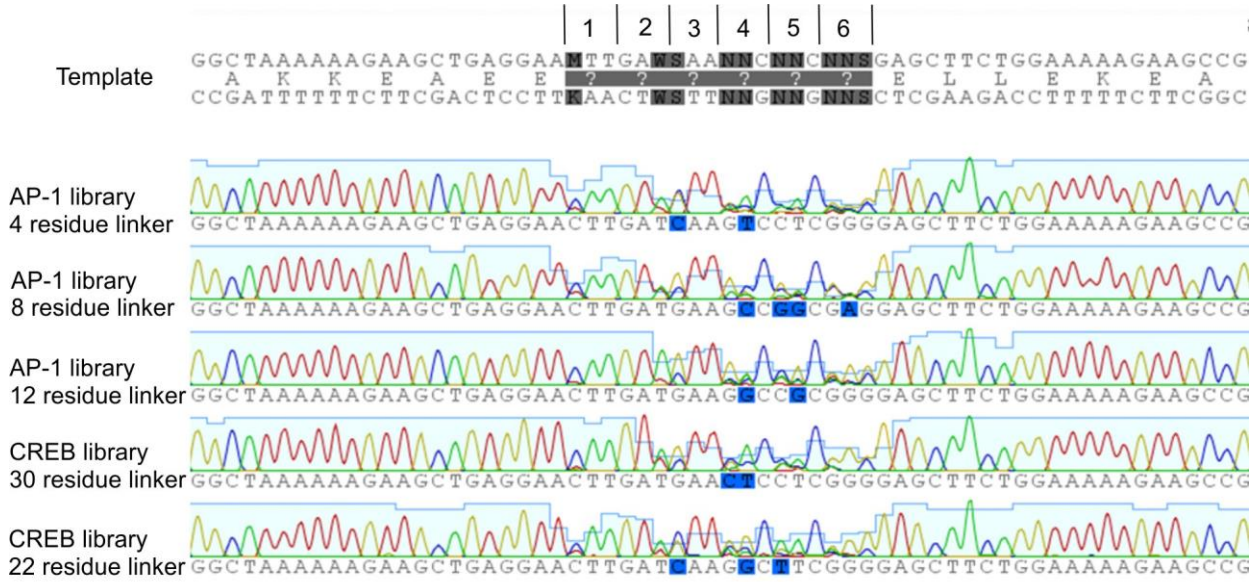


Figure 5-10: Sequencing chromatogram of the chimeric sequence.
The chromatogram shows mixed base content at all sites targeted for mutagenesis. The top sequence shows the nucleotide (sense and antisense) and protein sequences of the template. The six codons targeted for randomization are labelled. The chromatogram traces show the sequencing results from selected Gibson assembly reactions (one linker length per reaction). The light blue bar indicates the quality of the sequencing signal at each base. Nucleotides highlighted in blue were marked by the software as having low certainty. The chromatograms were visualized using Geneious v5.0.[152]

The remaining Gibson assembly reaction mixtures from each library were pooled, electroporated into *E. coli* XL1-Blue and plated on large agar trays with antibiotic pressure for plasmid selection. Quantification of the libraries showed approximately 1-fold coverage of the theoretical library size for each the AP-1-based library and the CREB-based library. Ideally, library coverage should be at least 10-fold; however, even with 10-fold coverage there is a very high probability that some theoretical variants will not be physically represented in the library.[228] Better coverage could be obtained by increasing the number of cell aliquots electroporated with the Gibson assembly products. Despite this poor coverage, the library was stored for future evaluation.

5.3.4 Additional design strategies

In addition to the library described above, other libraries will be designed using different randomization strategies as well as alternative photoactive scaffolds. Several designs are briefly discussed below.

The *Rs*LOV domain could enable dark-state protein caging via circularization. This strategy will be explored by fusing each DN-AP1 and DN-CREB between two *Rs*LOV monomers, with the aim of caging the effectors in the dark-state and uncaging the effectors through light-induced dissociation of the *Rs*LOV domains. It has been shown on numerous occasions that linker design plays a critical role in the development of switchable proteins.[47,66,199,229] To sample a range of different linkers, libraries will be created that vary the length and sequence of linkers connecting the effectors to the *Rs*LOV monomers at both the N- and C-terminal fusions.

Protein engineering using circular permutations has been shown to be an effective strategy to couple the function of unrelated proteins. Circularly permutated scaffolds of *As*LOV2[15] and PYP[14] have been recently introduced; however, the sequence space accessible through randomized circular permutations has not yet been fully explored for these proteins. Random circular permutation libraries have been successfully used to create allosteric protein switches.[61,67] It has been shown that the short sequence duplications and insertions additionally sampled in these libraries can yield sequences with switching behaviour.[67] In an interesting parallel, a cPYP variant with an inadvertent 9-residue duplication in a β-sheet was found to display switching through light-state β-strand slippage.[230] Randomized circular permutation libraries of PYP and *As*LOV2 will each be developed with DN-AP1 and DN-CREB using the method described by Kanwar and colleagues, as discussed in Chapter 1 (see Figure 1-3B).[62] Briefly, each of the photoactive scaffold genes will be duplicated in tandem using an effector sequence as an intervening linker. The resulting sequences will be used as templates with an array of primers to amplify libraries of gene fragments; complementary 5′ and 3′ truncations of the photoactive scaffold genes will yield circularly permutated sequences.

Recently, it has been shown that the length of the A′α and Jα helices in the *As*LOV2 domain affects the biophysical characteristics of LOV.[217] For example, the length of the A′α helix has been reported to both alter the photocycle kinetics and to modulate the extent of the structural

response in the Jα helix.[217] This suggests that altering these sequences could be an effective strategy for photo-caging. To explore this option, a library will be designed using cLOV with DN-inhibitor insertions, by incrementally varying the N- and C-terminal fusions of inhibitors with the cLOV scaffold.

The ability to select for switchable inhibitor activity in high-throughput using the hitchhiker-bandpass circuit enables the exploration of a variety of protein designs.

5.3.5 Optimization of the hitchhiker-bandpass circuit selection procedure under blue light irradiation

Prior to use in selections, the hitchhiker-bandpass assay was tested with blue light irradiation to detect and control for any effects caused by the light. There are two major areas of concern regarding the irradiation: the stability of Tet under blue light, and the effect of the light on the growth of *E. coli*.

5.3.5.1 Stability of the negative selection agent

Tet is yellow in colour, indicating it absorbs light in the blue region; the blue light irradiation could therefore lead to its degradation. Tet was substituted with Dox, an analogue of Tet with increased blue light stability. Testing Dox in the hitchhiker-bandpass circuit showed that Dox can be successfully substituted for Tet to impose negative selective pressure for β-lactamase activity. An array of Dox concentrations were tested to determine the optimal concentration for selection (Figure 5-11). Dox was found to be more potent than Tet. Bandpass activity was observed beginning at 2 μg/mL Dox and at ≥ 5 μg/mL, the selective pressure became inhibitory even to cells expected to express the Tet resistance gene. At concentrations of 3-4 μg/mL, Dox yielded appropriate negative selective pressure for β-lactamase activity. In comparison, Tet was used at a concentration of 20 μg/mL.

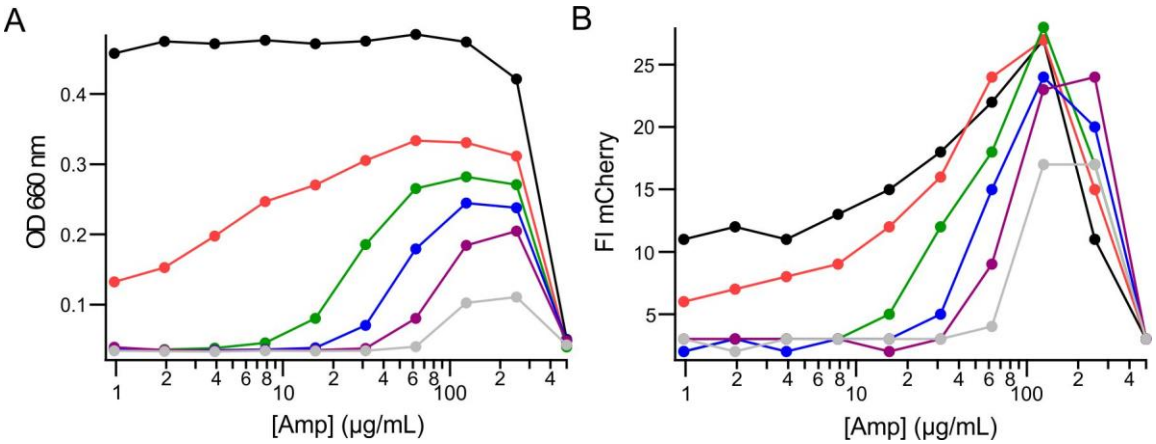


Figure 5-11: Determining the working concentration of Dox.
A range of Dox concentrations were tested in the hitchhiker-bandpass circuit to determine the optimal working concentration. A hitchhiker-bandpass experiment was performed with cells containing an mCherry bandpass plasmid, a plasmid encoding the CREB hitchhiker pair, and a plasmid encoding BFP (as a placeholder for future inhibitor constructs). The circuit was tested using Dox at the following concentrations: 0 (black), 1 (red), 2 (green), 3 (blue), 4 (purple) and 5 (grey) µg/mL. The media additionally contained plasmid maintenance antibiotics (Spec, Cm and Kan), an inducer (IPTG 1 mM) and Amp at the annotated concentrations. The graphs show (A) cell growth and (B) mCherry fluorescence from the bandpass plasmid.

5.3.5.2 Use of Dox with blue light irradiation

Testing the hitchhiker-bandpass circuit under blue light irradiation revealed that Dox is also sensitive to blue light. A hitchhiker-bandpass experiment was performed with an mCherry bandpass plasmid, a hitchhiker plasmid for CREB and a plasmid encoding either a strong inhibitor (A4-CREB-BFP) or a non-binding control (cLOV-BFP). Cells were inoculated into prepared media containing plasmid maintenance antibiotics, 1 mM IPTG, 4 µg/mL Dox, and an array of Amp concentrations. Two identical 96-well plates were prepared in parallel; one plate was incubated with constant blue light irradiation (450 nm, 250 µW/cm^2) and the second plate was incubated in the dark. After incubation (17 h, 32°C, 450 rpm), it was apparent that the irradiation affected the performance of the circuit (Figure 5-12). The optical density measurements (Figure 5-12A,C) showed that cultures incubated under blue light irradiation were able to grow in low concentrations of Amp that were inhibitory to identical cultures incubated in the dark. The observed survival of cells in low concentrations of Amp was attributed to light-induced Dox degradation. The mCherry fluorescence measurements (Figure 5-12B,D) showed bandpass activity as expected in all cultures; the *ampC* promoter was only induced in a range of

Amp concentrations. Importantly, the mCherry fluorescence measurements showed identical maxima in the light and dark cultures. This indicates the Amp concentration was unaffected by the blue light incubation, which is a critical requirement for light/dark library evaluation. However, to be able to reliably perform selections, the photoinduced degradation of Dox needs to be prevented. The loss of the Dox selection pressure would produce false positives for active inhibitors by allowing cell growth in low concentrations of Amp.

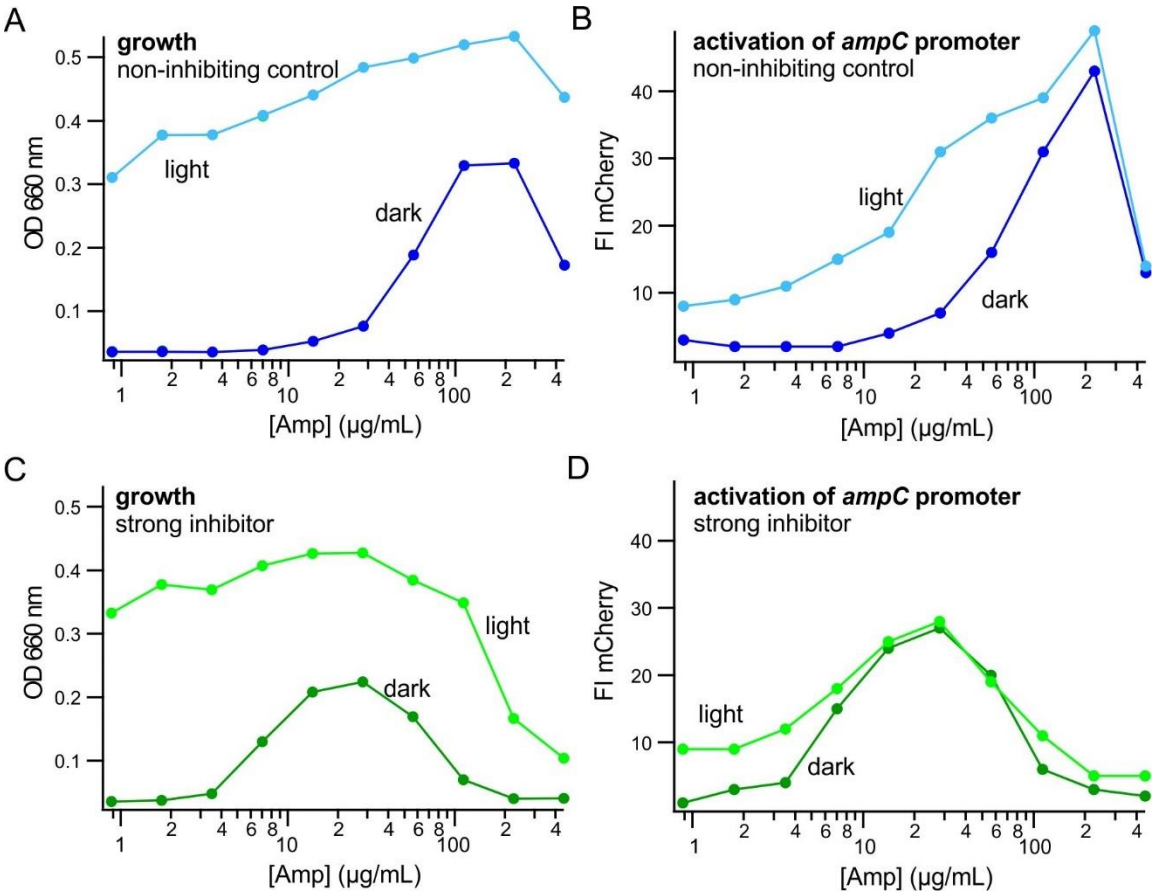


Figure 5-12: Assessment of the effects of blue light irradiation on the hitchhiker-bandpass assay using Dox selective pressure.
Hitchhiker-bandpass experiments were performed with cells containing an mCherry bandpass plasmid, a CREB hitchhiker plasmid and a plasmid encoding either a non-binding control, cLOV-BFP (A and B, blue shades) or a strong inhibitor, A4-CREB-BFP (C and D, green shades). The darker shades show cultures incubated in the dark; the lighter shades show cultures incubated with blue light irradiation. Growth (OD_{660}) is shown in (A) and (C), while activation of the *ampC* promoter, measured by expression of the fluorescent marker mCherry, is shown in (B) and (D). Media contained plasmid maintenance antibiotics (Spec, Cm and Kan), an inducer (IPTG 1 mM), and Dox (4 µg/mL) in addition to the annotated Amp concentrations.

It seems possible to overcome the sensitivity of Dox to blue light irradiation. The light-sensitivity of Dox was evaluated by incubating solutions of Dox in conditions replicating the light and dark hitchhiker-bandpass conditions, and analyzing the solutions using absorbance spectroscopy after overnight incubation (Figure 5-13). Indeed, the solution of Dox incubated with blue light irradiation at 32 °C showed degradation. The solution of Dox incubated in the dark at 32 °C was identical to the negative control, a solution incubated in the dark at 4°C. The spectra show that Dox absorbs very minimally in the blue light region used for irradiation, indicating degradation might be avoided by using high-pass filters to eliminate short-wavelength light.

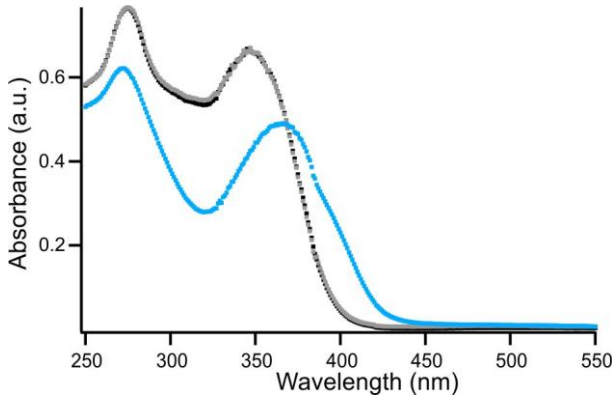


Figure 5-13: Absorbance spectroscopy of Dox.
Absorbance spectroscopy of Dox showed light-induced degradation. Solutions of Dox (25 µg/mL) were incubated with blue light irradiation (450 nm at ~ 250 µW/cm²) at 32°C (blue trace), in the dark at 32°C (black trace) or in the dark at 4°C (grey trace).

5.3.5.3 Effect of blue light irradiation on *E. coli* growth

Blue light can have a significant effect on *E. coli* growth and metabolism.[110,111] Certain intracellular small molecules, such as NADH and flavins, absorb light in the blue wavelength range (400 – 500 nm), which can lead to the generation of reactive oxygen species (ROS).[110] ROS can react with and cause damage to macromolecules such as proteins, lipids and DNA.[110]

The photodegradation of Dox in the hitchhiker-bandpass assay masked any effects of blue light irradiation on the growth of *E. coli*. Effects from the blue light irradiation will need to be carefully monitored by comparing growth rates of cultures under blue light irradiation with those of cultures incubated in dark conditions.

5.4 Conclusions and future directions

The ability to analyze switchable inhibitor activity in high-throughput makes it possible to explore a wide range of protein designs via directed evolution. Photo-controlled AP-1 and CREB inhibitor constructs were developed by fusing AP-1 and CREB inhibitors in a circularly permutated *As*LOV2 scaffold. Libraries were developed by randomizing a set of residues connecting the Jα helix with the N-terminus of the inhibitors, as well as by varying the length of a linker connecting the C-terminus of the inhibitors with the cLOV scaffold. In addition to these libraries, several other designs were discussed, including libraries using an *Rs*LOV scaffold, randomized circular permutations using PYP and *As*LOV2 scaffolds, and libraries designed to vary the length of the *As*LOV2 terminal motifs. The resulting libraries will be evaluated by using the hitchhiker-bandpass circuit to select for active and for inactive inhibitors in light and dark conditions.

In order for the hitchhiker-bandpass assay to be used for selections under blue light irradiation, optimization of the irradiation conditions is required. Current irradiation conditions lead to degradation of Dox, the antibiotic used to impose negative selective pressure for β-lactamase activity. It seems likely that this degradation can be avoided by using high-pass filters to eliminate short-wavelength light. In addition, it may be beneficial to decrease light intensity or to use irradiation pulses rather than continuous irradiation. Controls will be required to confirm that the irradiation used is sufficient to initiate the photocycle of the photoactive scaffolds.

The ability to select specifically for either the on- or the off-state of the photo-switchable inhibitors enables the evaluation of large libraries of protein variants. The hitchhiker-bandpass circuit is expected to be a useful platform for the development of the switchable inhibitors of protein-protein interactions.

6 Detection of *p*-coumaric acid incorporation into photoactive yellow protein variants *in vivo*

Note: This chapter was adapted with permission from K.E. Brechun, D. Zhen, A. Jaikaran, V. Borisenko, M. Kumauchi, W.D. Hoff, K.M. Arndt, G.A. Woolley. Detection of incorporation of *p*-coumaric acid into photoactive yellow protein variants *in vivo.* 2019. *ACS Biochemistry.* 58(23): 2682-2694.

Author contributions: KEB performed the experiments, analyzed data and wrote the manuscript. DZ performed the apo/holo protein standard curve. DZ and AJ helped optimize the *in vivo* assays. VB performed preliminary wavelength search experiments. GAW derived the quench model. KMA and GAW supervised the project and wrote the manuscript.

The article may be accessed via the following link:
https://pubs.acs.org/doi/pdf/10.1021/acs.biochem.9b00279

6.1 Introduction

Photoactive yellow protein (PYP) is a flagship member of the Per-Arnt-Sim (PAS) domain superfamily and has been the subject of a wide range of studies aimed at understanding how signaling operates at a molecular level in this large family of regulators.[215,221] In PYP, signaling is initiated by absorption of a photon by the covalently attached *p*-coumaric acid chromophore, triggering a *trans-cis* isomerization event that leads to global changes in the protein structure. A variety of mutagenesis studies[222–225] and a systematic Ala scan of PYP[221] have explored the importance of specific residues for protein production, spectral tuning and signaling kinetics.

More recently, PYP has been adopted as a switching element in several fusion proteins designed as optogenetic tools.[16–18,25,95,96,198,230–233] Its small size and robustness make it relatively easy to link to effector domains resulting in soluble, well-behaved constructs. The relatively large conformational changes that PYP undergoes during signaling make it likely to affect the structure and function of fused target proteins.

In its native host, PYP acquires the *p*-coumaric acid chromophore via an intrinsic metabolic pathway (Figure 6-1). It has been proposed that holo-PYP is formed via reaction of apo protein with the CoA thioester of *p*-coumaric acid, which is produced by *p*-coumaroyl-CoA ligase (4CL, also called pCL).[234–237] The metabolite *p*-coumaric acid can be produced from the amino acid

tyrosine through the action of tyrosine ammonia lyase (TAL).[238] The genes encoding both 4CL and TAL are found near the PYP gene in certain organisms.[234–236,239] These enzymes are also present in plants,[240,241] implying holo-PYP may be produced if heterologously expressed there. However, *E. coli*, yeast and mammalian cells do not appear to produce *p*-coumaric acid, so the formation of holo-PYP requires the addition of activated chromophore.[16] For optogenetic applications of PYP, biosynthetic production of holo protein is desirable since it obviates the need to continually add exogenous activated chromophore. In addition, when testing new optogenetic designs, it is useful to have a means for confirming reconstitution of the PYP domain in the target cell or tissue in a manner that is independent of the intended function of the optogenetic tool.[25]

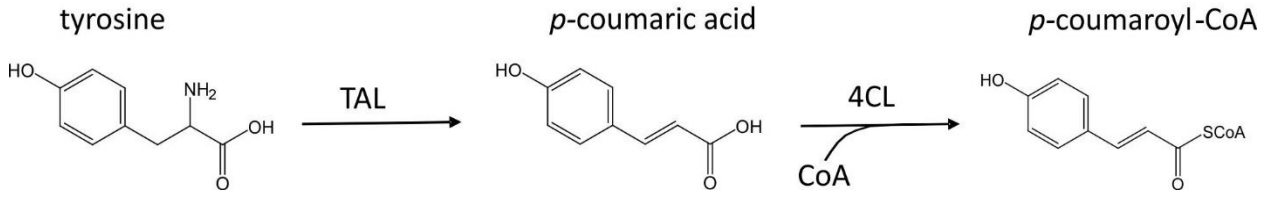


Figure 6-1: Metabolic pathway producing *p*-coumaroyl-CoA.
Production of *p*-coumaroyl-CoA from tyrosine via the activity of tyrosine ammonia lyase (TAL) and *p*-coumaroyl-CoA ligase (4CL).

In addition to their usefulness for enabling biosynthetic production of holo-PYP, heterologous expression of the enzymes TAL and 4CL has been explored for use in other metabolic engineering contexts. 4CL catalyzes a key step in the plant phenylpropanoid pathway, a metabolic route which diversifies to produce numerous commercially valuable secondary metabolites, including flavonoids, curcuminoids, structural polymers, and stilbenes.[242] Numerous homologues of TAL and 4CL have been analyzed[243,244] and selected versions have been engineered to modify substrate specificity,[245] remove feedback inhibition,[246] or to channel substrate shuttling.[247] Other metabolic engineering strategies, such as optimizing culture growth conditions,[248] balancing enzyme expression levels,[249] and controlling the carbon flux,[250] have been used to improve yields. While most metabolic engineering studies involving TAL and 4CL have used bacteria or yeast as hosts, the genes for TAL and 4CL have also been introduced into

mammalian cells in order to produce resveratrol, a stilbene with numerous promising pharmaceutical properties.[247]

Kyndt *et al.* reported that co-expression of *H. halophila* PYP with 4CL and TAL from *Rhodobacter capsulatus* leads to reconstitution of holo-PYP in *E. coli*,[235] a bacterial host that does not natively express these enzymes. Production of holo-PYP was detected by the yellow colour of the holo protein, and the reconstituted protein was purified and analyzed spectroscopically.[235] Holo protein production through co-expression of TAL and 4CL in *E. coli* has also been reported for several PYP homologues, namely *R. capsulatus* PYP[234] as well as the PYP domains of Ppd from *Thermochromatium tepidum*[251] and Ppr from *Rhodospirillum centenum*.[252,253] However, to our knowledge, biosynthetic reconstitution has not been explored for PYP mutants or fusion constructs intended for use as optogenetic tools, nor have factors affecting the efficiency and extent of *in vivo* reconstitution been addressed. In the case of PYP from *H. halophila* or *T. tepidum,* biosynthetic production of holo protein was incomplete, as judged by absorbance spectra and mass spectrometric analysis of purified protein.[235,251] These factors are critical for use of PYP-based optogenetic tools.

We wished to develop a reliable way of assessing functional reconstitution of the PYP domain *in vivo* in real time. Assessing absorbance ratios (dark-state PYP at 446 nm versus total protein at 280 nm) or using mass spectrometry are not straightforward ways of quantifying holo protein production since these methods require purification of the protein, and it is unclear to what extent apo protein is lost during purification. Here, we present PYP-mTagBFP-fusions as sensors for monitoring PYP chromophore uptake and photocycle characteristics. For simplicity, we refer to the blue fluorescent protein variant as BFP. We first characterize a wild-type PYP-BFP-fusion in detail and analyze PYP reconstitution *in vitro.* We then use the BFP-fusion *in vivo* to analyze the reconstitution of PYP and PYP variants with enzymatically produced chromophore, and in the process, optimize a metabolic pathway producing the activated chromophore, *p*-coumaroyl-CoA, in *E. coli.* We then show that this system supports the reconstitution of various PYP-based optogenetic constructs in *E. coli*.

6.2 Materials and Methods

6.2.1 Bacterial strains

E. coli XL1-Blue and DH5α were used for cloning and plasmid propagation. *E. coli* BL21(DE3) was used for protein expression. The genotype of each strain is shown in the following table.

Table 6-1: List of strains.

Strain	**Genotype**
E. coli XL1-Blue	*endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'(::Tn10 proAB$^+$ lacI^q Δ[lacZ]M15)* hsdR17*(r_K^- m_K^+)*
E. coli DH5α	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ$^-$
E. coli BL21(DE3)	F$^-$ *ompT gal dcm lon hsdS$_B$*($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$(λ^S)

6.2.2 Plasmid design and construction

Plasmids were constructed using classical restriction enzyme-based cloning, Gibson assembly[113] and SLiCE.[142] A list of the plasmids used is included in the Appendix.

The gene sequences of the PYP constructs were encoded in the vector pQE80L (Qiagen), designed with a fusion to mTagBFP on the 3' end via a short linker (encoding 9 residues). As mentioned above, we refer to this fluorescent protein as BFP. The BFP gene contained two point mutations (C222S and I174A) introduced into the sequence via QuickChange reactions. The C222S mutation changed a surface-exposed Cys to Ser. It was found that when activated chromophore was added in excess, this Cys became modified by *p*-coumaric acid in addition to the Cys in the active site of PYP (unpublished observation). This side reaction was avoided by the C222S mutation. The I174A mutation was made based on the theory of Subach *et al.* that this mutation stabilizes the BFP chromophore and slows a process that leads to the destruction of the BFP chromophore (*i.e.* yielding the variant mTagBFP2).[151] The PYP genes used were wtPYP from *H. halophila* (also known as PYP(A), [NCBI accession number P16113.3]), as well as the *H. halophila* PYP homologue, PYP(B) (NCBI accession number WP_011814122.1). The optogenetic PYP-based constructs GCN4-Δ25PYP(v2 K143F E35L), opto-DN-CREB, cPYP and cPYP M121E were cloned into pQE80L so that they would be expressed with BFP fused to the

N-terminus via a flexible linker, as discussed in the main text. The Ala point mutations in PYP were introduced via QuickChange reactions.

The genes for TAL and 4CL were co-expressed from the vector pRSFDuet^{TM}-1 (Novagen) in the first and second multiple cloning sites respectively. The genes for *Rhodobacter sphaeroides* TAL (NCBI accession number WP_011339422.1), *A. thaliana* 4CL 1 (NCBI accession number U18675) and *Idiomarina loihiensis* 4CL (NCBI accession number WP_011235612.1) were synthesized by BioBasic (Toronto, Canada).

The BFP-fusion plasmid (pQE80L-PYP-BFP) and chromophore biosynthesis plasmid (pRSFDuet-v2 *Rs*TAL *At*4CL) are available from AddGene.

6.2.3 Protein expression for purification

Bacterial expression and reconstitution of holo-PYP containing constructs with chemically synthesized *p*-coumaroyl-*S*-thiophenyl ester was adapted from the work of Devanathan *et al*.[232,254] Apo-PYP-BFP was expressed and purified using the same protocol, excluding the addition of activated PYP chromophore. Activated PYP chromophore, S-phenyl E-3-(4-hydroxyphenyl)prop-2-enethioate, was synthesized according to the method of Changenet-Barret *et al.*[232,255]

DNA was transformed into chemically competent *E. coli* BL21(DE3) cells and plated onto LB agar containing 100 µg/mL Amp. A single colony from the plate was used to inoculate 25 mL of LB broth supplemented with Amp (100 µg/mL) and grown overnight with shaking at 37°C. The following day, 1 L of LB supplemented with Amp (100 µg/mL) was inoculated with the 25 mL overnight culture. Cells were grown at 37°C until an optical density at OD_{600} of 0.5 - 0.6 was reached, then induced with IPTG to a final concentration of 1 mM. The temperature was lowered to 25°C and the cells were grown for a further 1.5 h, after which, 25 mg of thiophenyl activated PYP chromophore dissolved in 1 mL of ethanol was added to the medium. The cells were grown for a further 2.5 h then collected via centrifugation at 4000 rpm (3724 rcf). The resulting cell pellet was frozen at -20°C until purification.

6.2.4 Protein purification

The cell pellet was resuspended in lysis buffer containing 50 mM sodium phosphate (pH 8.0) with 300 mM NaCl, 5 mM $MgCl_2$ and freshly added dithiothreitol (DTT, 1 mM final concentration). Cells were disrupted through sonication and the cell debris was removed through centrifugation at 12,000 rpm (10976 rcf) for 45 min. The supernatant was filtered with a 0.45 μm filter and loaded onto a Ni-NTA column equilibrated with lysis buffer. The resin was washed sequentially with 5 column volumes (CV) of lysis buffer, 2 CV of high salt buffer (lysis buffer with 2 M NaCl), 2 CV of low imidazole buffer (lysis buffer with 10 mM imidazole), 2 CV of high imidazole buffer (lysis buffer with 25 mM imidazole) and then the protein was eluted with 2 CV of lysis buffer containing 200 mM imidazole. The protein was then dialyzed against 1× TAE (40 mM Tris-acetate pH 7.5, 1 mM EDTA) with 100 mM NaCl at 4°C, or for *in vitro* reconstitution experiments, the eluted apo-PYP-BFP was immediately buffer exchanged into 1× PBS (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl) at 4°C using a centrifugal spin column with a 10 kDa molecular weight cut-off (MWCO) filter. Protein mass was confirmed using ESI-MS.

6.2.5 UV-visible absorbance spectrophotometry

Absorbance spectra were recorded using a PE Lambda 35 or 25 spectrometer or using a diode array UV-visible spectrophotometer (Ocean Optics Inc., USB4000), with a temperature-controlled cuvette holder (Quantum Northwest, Inc.). Samples were measured in a 1.0 cm path length quartz cuvette. Concentrations were determined using the following parameters: ε_{399} of 50.6 x 10^3 $M^{-1}cm^{-1}$ for BFP I174A C222S,[151] ε_{446} of 46 x 10^3 $M^{-1}cm^{-1}$ for dark-state PYP,[95] and ε_{333} of 21.0 x 10^3 $M^{-1}cm^{-1}$ for *p*-coumaroyl-CoA.[256] Irradiation was achieved using a Luxeon III Star Royal Blue (455 nm) LED operating at 700 mA (~50 mW/cm^2). The thermal relaxation rate of PYP was determined by fitting the absorbance at 446 nm to a mono-exponential function.

6.2.6 *In vitro* chromophore reconstitution

The reconstitution of purified apo-PYP-BFP was monitored via UV-visible spectrophotometry using *p*-coumaroyl-CoA, purchased from PlantMetaChem (Gießen, Germany) and chemically synthesized *p*-coumaroyl-*S*-thiophenyl ester. For reconstitution with *p*-coumaroyl-CoA, the absorbance spectrum of apo-PYP-BFP (7 μM) was monitored with the addition of 10-fold molar

equivalents of *p*-coumaroyl-CoA in PBS at 20°C over a period of 12 h, with scans taken every 10 min. For reconstitution with thiophenyl activated chromophore, the absorbance spectrum of apo-PYP-BFP (7 μM) was monitored with the addition of 3-fold molar equivalents of *p*-coumaroyl-*S*-thiophenyl ester in PBS over a period of 70 min, with continuous scans. The age of the purified protein (hours or days post-sonication of the expression culture) was noted.

6.2.7 Protein refolding

Purified apo-PYP-BFP in PBS was denatured by incubating the protein in 2 M guanidine hydrochloride (GnHCl) for 2 h at room temperature. The protein was then refolded by exchanging the buffer to PBS solutions with sequentially lower concentrations of GnHCl (1.5 M, 500 mM, 250 mM, 100 mM, 50 mM, 1 mM, followed by two washes with PBS without GnHCl) using a 10 kDa MWCO centrifugal spin column at 4°C over a period of several hours.

6.2.8 *In vivo* chromophore reconstitution

Chemically competent *E. coli* BL21(DE3) were co-transformed with the chromophore biosynthesis plasmid (pRSFDuet-v2 TAL 4CL) and a PYP-BFP plasmid (pQE80L with BFP fused to a PYP mutant, homologue or optogenetic construct), and plated on LB agar plates containing Amp (100 μg/mL) and Kan (50 μg/mL). For expression, 3 colonies were picked and each inoculated into a 50 mL flask containing 15 mL of DYT media with Amp, Kan and 1% glucose. The 3 biological replicate cultures were grown overnight at 37°C with 200 rpm. For the log phase expression protocol, each overnight culture was inoculated into 15 mL of fresh DYT media with Amp (100 μg/mL) and Kan (50 μg/mL) to a starting OD_{600} of 0.1. The cultures were incubated at 37°C until an OD_{600} of 0.5 - 0.6 was reached. The cultures were then induced with IPTG to a final concentration of 1 mM and incubated at 25°C for expression. For the stationary phase expression protocol, the overnight cultures were pelleted at 4000 rpm (3724 rcf) for 10 min, resuspended in 15 mL of fresh DYT with Kan (50 μg/mL), Amp (100 μg/mL) and IPTG (1 mM), and returned to the empty 50 mL flasks. The cultures were incubated at 25°C with 200 rpm. When indicated, cultures were reconstituted by adding *p*-coumaroyl thiophenol ester freshly dissolved in ethanol (final concentration of 30 μg/mL) and the cultures were incubated at 25°C for an additional 1 h prior to taking samples.

Throughout the experiment, timepoint samples were taken to monitor holo-PYP-BFP content. Timepoint samples were prepared by taking a volume of culture media equivalent to 1 mL of 1.5 OD_{600} culture. The samples were pelleted at 5000 rpm (2340 rcf) for 5 min and resuspended in 1 mL of ice-cold, filtered PBS. The samples were immediately incubated on ice for 15 min to stop further protein synthesis and maturation. The culture timepoint samples were measured in triplicate in a flat-bottom, clear 96-well plate (100 µL culture volume per well) in a plate reader (TECAN, Infinite M1000Pro). Prior to the measurement, samples were allowed to settle and dark-adapt for 20 min. Optical density was measured at 600 nm. Fluorescence was measured from the bottom of the plate with excitation at 420 nm and emission at 460 nm (5 nm bandwidth each) with a kinetic interval totaling 60 – 90 s (measurements every 0.2 s).

6.2.9 Fitting of time-dependent fluorescence data

Raw BFP fluorescence time-course data were fit to Eq. 1 (*vide infra*) using the global fit function in Igor Pro. We assumed that the fluorescence observed in all cases was dominated by BFP fluorescence since PYP fluorescence is much less intense. The fit was constrained such that the fluorescence of apo-PYP-BFP, was greater than the fluorescence of *cis* holo-PYP-BFP, which in turn was greater than *trans* holo-PYP-BFP, *i.e.* $Fl_A \geq Fl_C > Fl_T$ since the absorbance spectrum of *cis* holo-PYP shows less overlap than *trans* holo-PYP with BFP emission. The ratio of BFP fluorescence intensity of *trans* holo-PYP-BFP versus apo-PYP-BFP was determined using purified protein. The ratio of Fl_A to Fl_T in the fitted *in vivo* data ranges from ~2.6 – 2.8 (see Appendix XI). The k_f value for all *in vivo* reconstitution experiments would be expected to be similar, since the irradiation intensity was constant for all experiments, the cell density was set to a standard OD and the quantum yield of PYP isomerization is not expected to vary significantly between samples. The k_f value in the fit data sets ranges from 0.048 – 0.053 (see Appendix XI). To fit the BFP-GCN4-Δ25PYP dataset, the fitting was calibrated to the alternative BFP-PYP-fusion design, as discussed in the results and discussion. The fit parameters for these data sets are listed in Table 6-3 and Appendix XI. In all cases, protein concentrations were restricted to positive values.

6.2.10 SDS PAGE and western blotting

Cell pellets (1 mL of 1 OD_{600} culture) or purified protein samples were boiled in 100 µL of Roti®-Load 1 loading dye (Roth) for 15 min, and 20 µL was loaded on a gel. SDS PAGE gels were cast with a gradient from 7.5 – 15% acrylamide, run in a Tris-Glycine buffer and stained with Coomassie Blue. When western blotting, gels were run with 5 µL of a 1/10 dilution of the prepared samples. The unstained SDS-PAGE gels were transferred to PVDF membranes using a semi-dry western blotting apparatus. Blots were probed using peroxidase-labeled mouse monoclonal anti-poly-His antibodies (Sigma), detected using SignalFire™ ECL Reagent (Cell Signaling Technology) and imaged using a Fusion SL (Vilber Lourmat) imaging system.

6.3 Results and Discussion

6.3.1 Protein design

While absorbance measurements can detect functional holo-PYP, these measurements require high protein concentrations and *in vivo* measurements suffer from severe scattering artifacts. Intrinsic fluorescence of the *p*-coumaric acid chromophore in PYP can be detected[257] but this is weak (Figure 6-2).

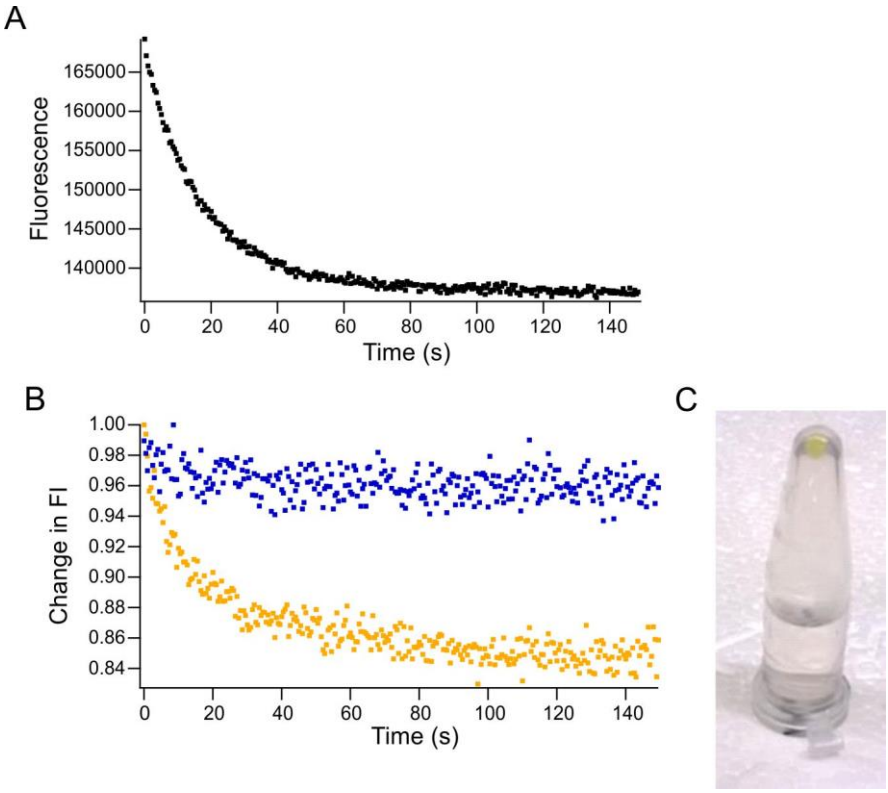


Figure 6-2: Monitoring PYP photo-switching using PYP fluorescence.
PYP photo-switching can be monitored by the loss of dark-state (*trans*) fluorescence; however, the signal is very weak. (A) Exciting a sample of 50 μM holo-PYP at 446 nm and monitoring fluorescence emission at 495 nm over time shows a loss of PYP fluorescence as the protein isomerizes to the *cis* state. (B) *In vivo* measurements in *E. coli* show a poor signal-to-noise ratio. A culture with fully holo-PYP (obtained by supplementing the culture with *p*-coumaroyl-thiophenyl ester) showed a weak curve (yellow). A culture with partially holo-PYP (obtained through enzymatic chromophore synthesis) had a barely detectable curve (blue), although the cell pellet was visibly yellow (C).

Moreover, to detect the extent of reconstitution, one also needs to detect levels of apo protein *in vivo*. We therefore opted to design a fusion protein that linked a blue fluorescent protein to PYP. Blue fluorescence provides a relative measure of total (apo plus holo) protein levels. In addition, we expected that blue fluorescence intensity could be modulated by the isomeric state of the PYP chromophore (*vide infra*).

We fused a BFP variant optimized for brightness and stability (specifically mTagBFP2-C222S,[151] referred to as BFP) to PYP via a flexible linker intended to minimize any effect of BFP on the PYP conformational changes required for the photocycle, but to keep the proteins

close enough in space so that quenching of BFP fluorescence by the *trans* form of PYP could occur via FRET. Figure 6-3A shows a model of the construct based on the crystal structures of the two components. Figure 6-3B shows the excitation and emission spectra of BFP together with the absorbance spectra of the *trans* (dark-adapted) and *cis* (irradiated) forms of PYP. Since the absorbance spectrum of the *trans* form of PYP has a greater degree of spectral overlap (J =1.8 x 10^{12} $nm^4/(M*cm)$) than the *cis* form (J = 2.5 x 10^{10} $nm^4/(M*cm)$) with BFP emission, we expected the *trans* form would act as a more efficient quencher of BFP fluorescence than the *cis* form.

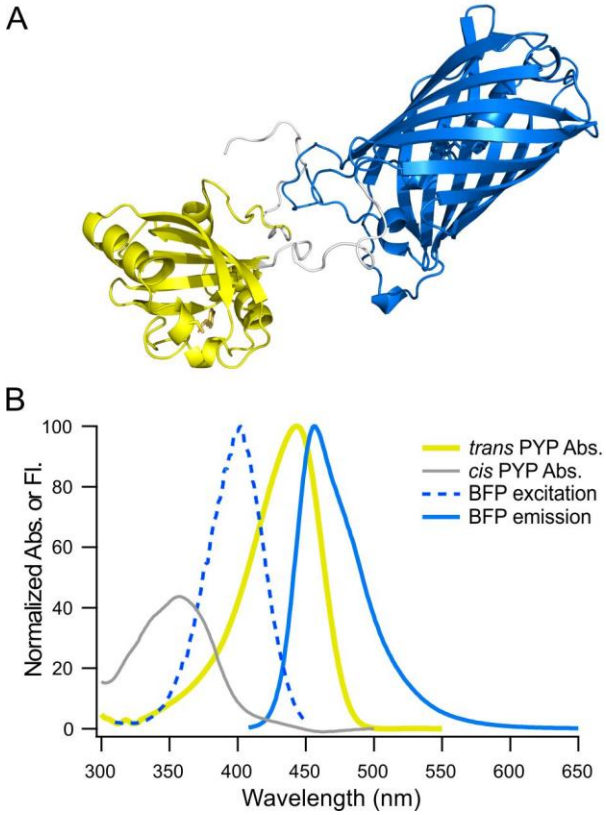


Figure 6-3: Structural model of PYP-BFP and spectral overlap.
(A) PYP was fused to the N-terminus of BFP via a flexible linker. The model shown was created using experimentally determined structures of PYP (1NWZ)[226] and mTagBFP (3M24),[151] and visualized using PyMol. PYP is shown in yellow, the linker and N-terminal His tag in grey and BFP in blue. (B) The absorption and emission spectra of BFP are shown together with the absorbance spectra of the *trans* (dark-adapted) and *cis* (blue-light irradiated) forms of Δ25PYP. PYP with an N-terminal deletion was used for this experiment because the light-state has an extended lifetime, facilitating absorbance measurements.[258,259] Despite this, the irradiated spectrum showed 11% dark-state protein content. The spectrum was adjusted to remove the contribution from dark-state protein. The spectra were normalized such that the maxima of the BFP spectra and dark-state PYP were equal.

6.3.2 Characterization of PYP-BFP

The designed PYP-BFP construct (see the Appendix for the sequence) was expressed in *E. coli*, fully reconstituted using synthetic activated chromophore, and purified. The protein mass and purity were confirmed by ESI-MS (Appendix VIII) and SDS PAGE (Figure 6-4).

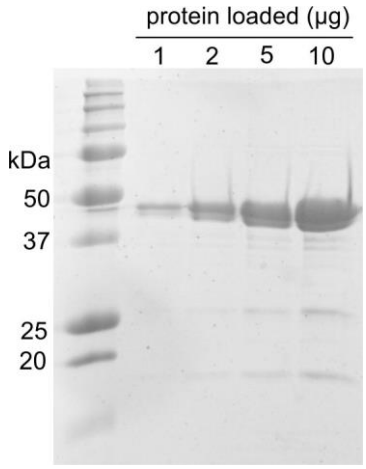


Figure 6-4: Mass and purity of PYP-BFP.
An SDS PAGE is shown with purified holo-PYP-BFP loaded with increasing protein concentrations (1, 2, 5, 10 µg protein from left to right). Cleavage of the BFP domain, as observed by Subach *et al.*[151] would be expected to yield 24 and 19 kDa fragments.

UV-visible absorbance spectra of the dark-adapted and light-adapted states of holo-PYP-BFP are shown in Figure 6-5, with wild-type PYP provided for comparison. Thermal *cis*-to-*trans* relaxation of PYP-BFP was 10-fold slower than wild-type PYP, with half-lives of 5.4 s and 0.54 s respectively (Figure 6-5C,D). This change in photo-cycle kinetics may be due to transient interactions of the BFP domain with PYP in the light-state. The light-state structure of PYP has been shown to be highly hydrophobic[260] and to have characteristics of a molten globule,[261] which could lead to non-specific protein interactions. In addition, it has been observed previously that modifications to PYP (such as the addition of a His tag[10]) can have an impact on the thermal relaxation rate.

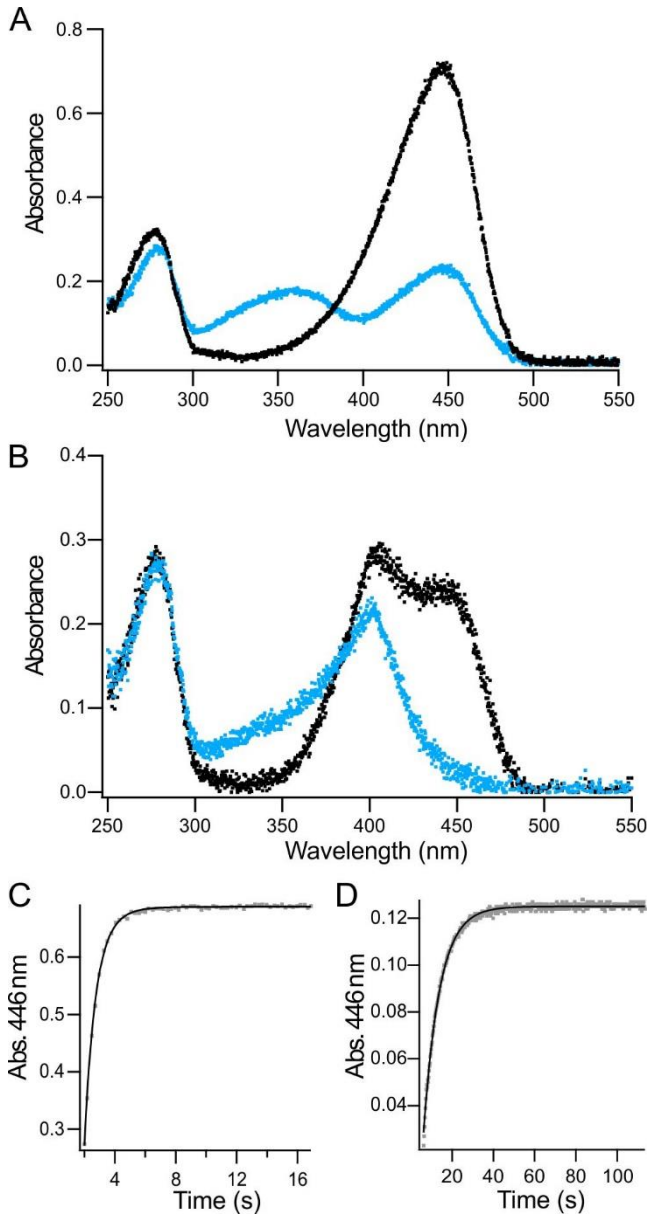


Figure 6-5: Biophysical characterization of PYP-BFP.
The UV-visible absorbance spectra of the dark-adapted (black) and light-adapted states (blue) of wild-type PYP (A) and wtPYP-BFP (B) are shown. PYP absorbs at 446 nm in the dark-state and 350 nm in the light-state. BFP absorbs at 399 nm. The light spectra were obtained by simultaneously irradiating samples with 450 nm light (measured on a diode array spectrophotometer at 20°C). The light spectrum of wild-type PYP is not in the fully light-adapted state due to its rapid thermal relaxation. The thermal relaxation of wild-type PYP and PYP-BFP are shown in (C) and (D) respectively. Samples were irradiated with 450 nm light and the recovery of absorbance at 446 nm (dark-state PYP) was monitored over time (at 20°C). The absorbance at 446 nm was fit to a mono-exponential curve, yielding a half-life of 0.54 s and 5.4 s for PYP and PYP-BFP respectively.

We next investigated the effect of the PYP photo-state on the observed BFP fluorescence intensity. Exciting samples of dark-adapted holo-PYP-BFP with certain wavelengths of light produced a time-dependent increase in BFP fluorescence emission, leading to a plateau (Figure 6-6A). We attribute the time-dependent increase in BFP fluorescence to decreased quenching as the *trans* form of PYP is converted to the *cis* form. The plateau represents a steady-state degree of quenching produced by a mixture of *trans*/*cis* PYP that depends on the light intensity used for the measurement, the wavelength, and the thermal relaxation rate of *cis*-PYP to *trans*-PYP. We experimentally determined the optimal wavelengths to maximize the change in observed fluorescence intensity by testing different combinations of excitation and emission wavelengths, and monitoring BFP fluorescence intensity over time (Figure 6-6B). The largest change in BFP fluorescence intensity was obtained by irradiating samples at 420 nm and monitoring emission at 460 nm. Both PYP and BFP absorb strongly at 420 nm, leading to fluorescence and/or photo-isomerization. Apo-PYP-BFP showed no significant changes in BFP fluorescence intensity at any of the wavelengths tested (Figure 6-6B).

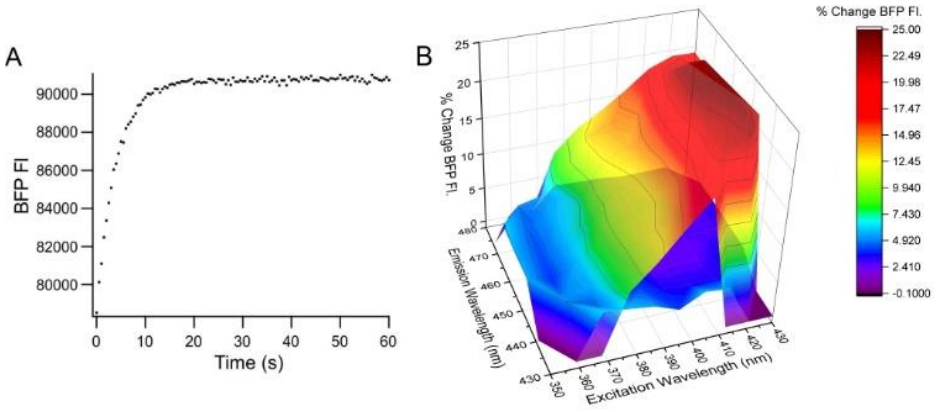


Figure 6-6: The effect of PYP photo-state on PYP-BFP fluorescence.
(A) Irradiation of dark-state holo-PYP-BFP produces a time-dependent increase in BFP fluorescence leading to a plateau. (B) The optimal wavelength to maximize the effect of PYP photo-state on the observed BFP fluorescence was determined using an *E. coli* culture with holo-PYP-BFP. BFP fluorescence was monitored for 60 s for each wavelength combination. The percent change in BFP fluorescence intensity for each wavelength combination is shown as a heat map (upper surface). The lower surface shows the lack of response of apo-BFP-PYP.

The time-dependent changes BFP fluorescence intensity can be modeled using Eq. 1.

$$Fl_{obs} = [apo] * Fl_{apo} + [holo] * \left(Fl_C + \left((Fl_T - Fl_C) * \left(\frac{k_b}{(k_b+k_f)} + \left(1 - \frac{k_b}{(k_b+k_f)} \right) e^{-(k_f+k_b)t} \right) \right) \right) \text{(Eq. 1)}$$

Where [apo] and [holo] are the total concentration of apo-PYP-BFP and holo-PYP-BFP respectively, Fl_{apo} , Fl_C , and Fl_T are the fluorescence intensities of apo-PYP-BFP, *cis* holo-PYP-BFP and *trans* holo-PYP-BFP respectively, k_f is the rate constant of the forward reaction (photoisomerization to *cis*) and k_b is the rate constant of the reverse reaction (thermal relaxation to *trans*). A detailed derivation of Eq. 1 can be found in the Appendix. Fitting the time-dependent changes in BFP fluorescence intensity to Eq. 1, in principle, allows determination of the fraction of apo-PYP-BFP and holo-PYP-BFP in a sample, as well as PYP photocycle kinetic parameters (k_b, k_f).

To confirm that fitting fluorescence data to Eq. 1 can be used to calculate the fractions of apo and holo protein, we purified apo- and holo-PYP-BFP, mixed the proteins in different proportions, and measured the resulting time-dependent changes in BFP fluorescence. The fitting procedure was able to correctly determine the percent holo content of each sample (Figure 6-7A). We next confirmed that the fitting procedure can be used to calculate rate constants in the PYP photocycle. We manipulated the light intensity used to excite dark-state holo-PYP-BFP to vary the forward rate constant (k_f), resulting in an increased rate at which the steady-state was reached (Figure 6-7B). The relaxation rate of holo-PYP-BFP obtained from the fitting (2.6 s) was in good agreement with the thermal relaxation rate calculated from direct measurement *in vitro* (5.4 s, Figure 6-5D).

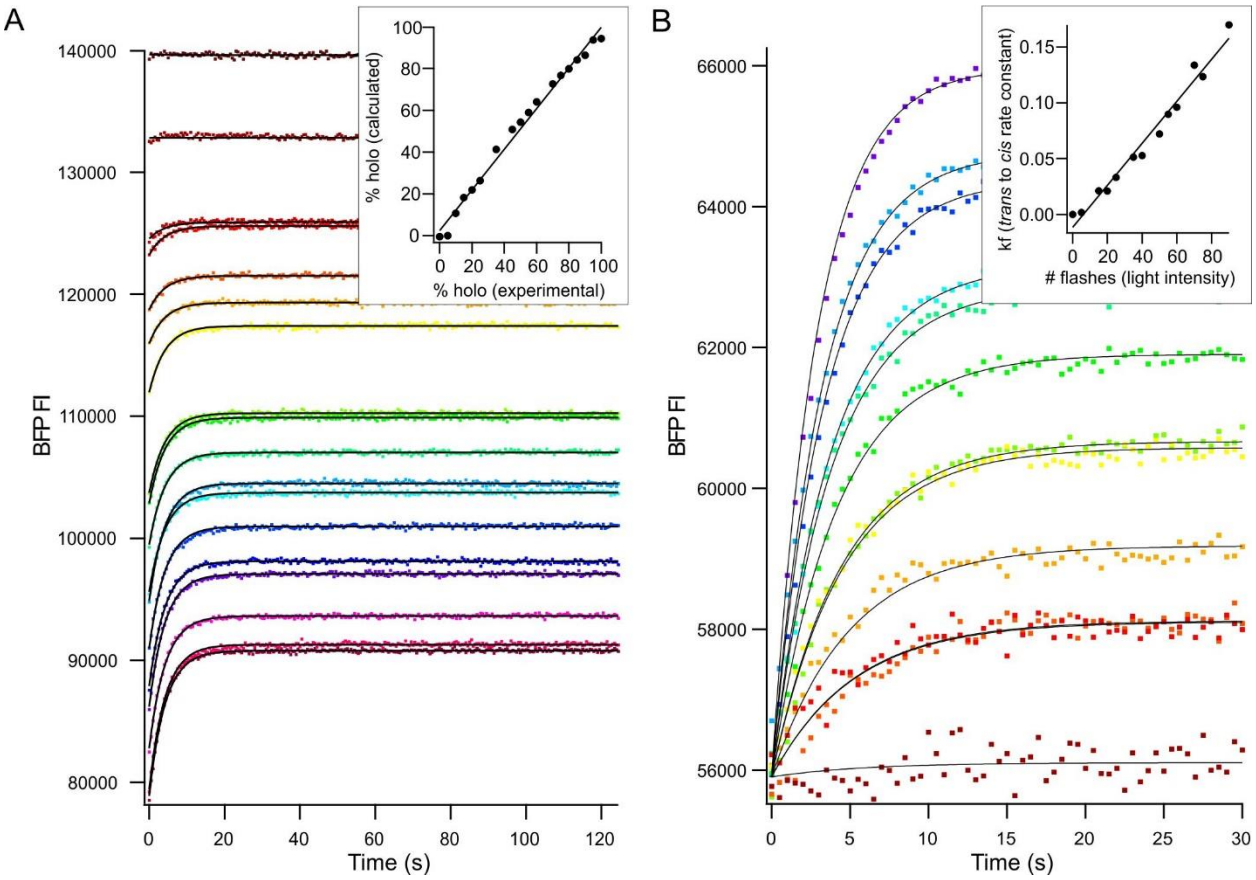


Figure 6-7: Determination of apo- and holo-PYP-BFP fractions, and k_f.
(A) Purified apo- and holo-PYP-BFP were mixed in different ratios, and BFP fluorescence was monitored at 460 nm over time with excitation at 420 nm. The traces are coloured as a rainbow from the lowest (dark red, highest BFP fluorescence) to highest (purple, lowest BFP fluorescence) percent holo. There is close agreement between the % holo protein calculated from fitting to Eq. 1 and the actual (experimental fraction) (insert). (B) The rate of chromophore isomerization (k_f) determined by fitting to Eq.1 is linearly dependent on the intensity of the incoming light as expected.

This analysis established that the BFP-PYP sensor is functional, and that it provides a measure of the thermal relaxation rate of the PYP photocycle and the percent holo protein. This sensor is therefore suitable for monitoring reconstitution of PYP.

6.3.3 Reconstitution of PYP-BFP *in vitro*

Normally, apo-PYP has been reconstituted *in vitro* using *p*-coumaroyl thiophenyl ester or *p*-coumaroyl anhydride to obtain holo-PYP.[220,254,262] Since the expected form of activated *p*-

coumaric acid *in vivo* is *p*-coumaroyl-CoA,[235] we wished to confirm that this compound was in fact competent to produce holo-PYP from apo-PYP in the absence of any other factors.

In previous *in vitro* reconstitution experiments with *p*-coumaroyl thiophenyl ester or *p*-coumaroyl anhydride, it has been noted that apo-PYP becomes unreactive over time,[222,262] and that the reactivity of purified apo-PYP cannot be revived using reducing agents,[254] indicating the change is not due to the formation of disulfide dimers of apo protein. We also observed that reactivity of apo-PYP-BFP decreased with the age of the purified protein (data not shown). We therefore performed *in vitro* reconstitution experiments immediately after purification (beginning 8 h after cell lysis). *P*-coumaroyl thiophenyl ester reacted rapidly with apo-PYP-BFP (Figure 6-8), producing >95% holo protein after approximately 20 min with 3-fold molar equivalents of chromophore at 20°C.

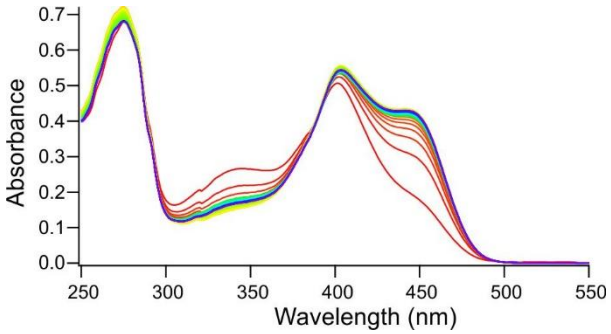


Figure 6-8: Reconstitution of apo-PYP-BFP with *p*-coumaroyl thiophenyl ester.
Purified apo-PYP-BFP was reconstituted immediately after purification (8 h post cell lysis) using 3-fold molar equivalents of thiophenyl activated *p*-coumaric acid at 20°C. The absorbance spectra show an increase at 446 nm (dark-state holo-PYP formation) and a decrease at 343 nm (loss of thiophenyl activated *p*-coumaric acid). The reconstitution was judged to have proceeded to approximately 95% completion (based on absorbance ratios).

The reaction with *p*-coumaroyl-CoA was slower and less efficient. Using a 10-fold molar excess of *p*-coumaroyl-CoA, reconstitution of apo-PYP-BFP proceeded to approximately 80% (based on absorbance ratios) after 6 h (Figure 6-9).

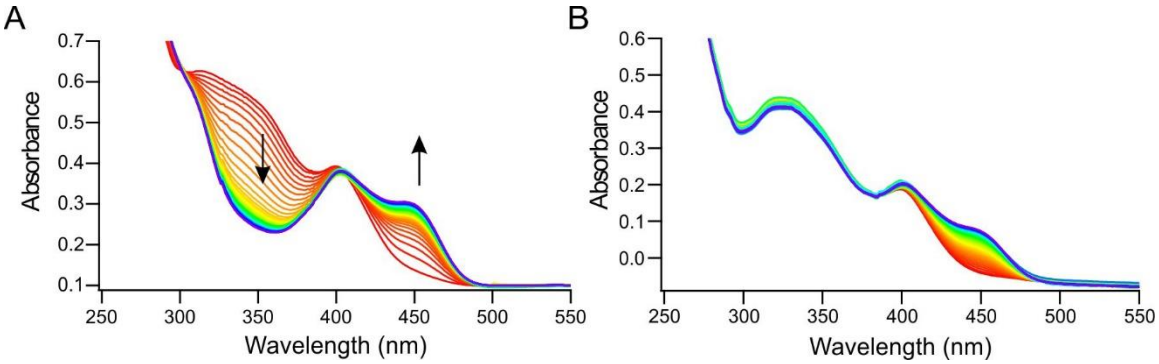


Figure 6-9: Reconstitution of apo-PYP-BFP with *p*-coumaroyl-CoA.
(A) Freshly purified apo-PYP-BFP was reconstituted using 10-fold molar excess of *p*-coumaroyl-CoA at 20°C. Absorbance scans were taken every 10 min over a period of 12 h. Development of dark-state holo-PYP-BFP can be seen by the increase in absorbance at 446 nm. Loss of *p*-coumaroyl-CoA can be seen by the decrease in absorbance at 333 nm. (B) A sample of apo-PYP-BFP was denatured and refolded, and the reconstitution was repeated as described previously. In this case, the loss of *p*-coumaroyl-CoA was not observed.

We confirmed that the holo-PYP-BFP obtained through *in vitro* reconstitution with *p*-coumaroyl-CoA had the correct mass (see Appendix IX) and was photoactive (Figure 6-10).

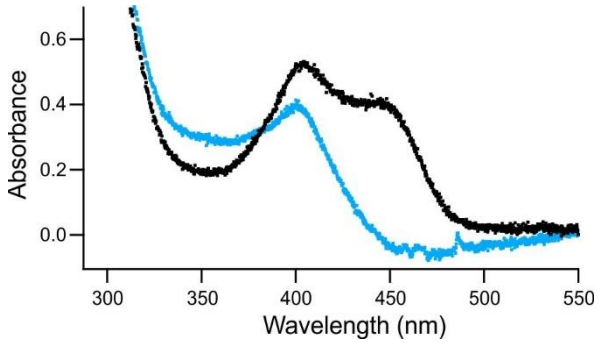


Figure 6-10: Analysis of reconstituted PYP-BFP.
Apo-PYP-BFP reconstituted with *p*-coumaroyl-CoA was photoactive. The light-adapted (blue) and dark-adapted (black) spectra of reconstituted PYP-BFP were recorded using a diode array. For the light spectrum, the protein was irradiated in the cuvette with a 450 nm LED.

Interestingly, in addition to reconstitution, we observed that *p*-coumaroyl-CoA (Abs. max at 333 nm) was rapidly degraded when mixed with apo-PYP-BFP. The observed loss of absorbance at 333 nm is consistent with spectral changes observed during heat-induced hydrolysis of *p*-coumaroyl-CoA in PBS (Figure 6-11A,B). In contrast, *p*-coumaroyl-CoA was stable in the presence of an unrelated protein (bovine serum album, BSA) (Figure 6-11C).

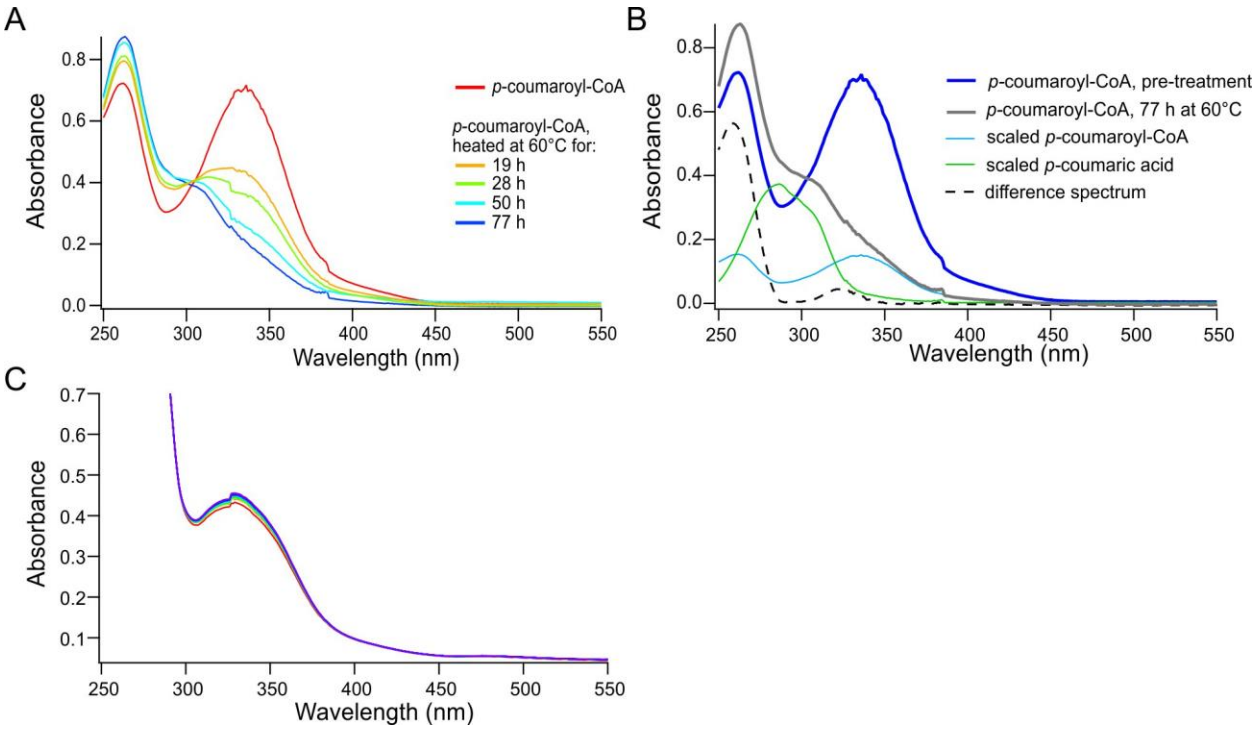


Figure 6-11: Heat-induced hydrolysis of *p*-coumaroyl-CoA and stability of *p*-coumaroyl-CoA in the presence of non-specific protein.
(A) A solution of *p*-coumaroyl-CoA in PBS was heated at 60°C to promote hydrolysis of the thioester. Heating resulted in a loss of absorbance at 333 nm and an increase at 290 nm and 260 nm. (B) Spectra of *p*-coumaroyl-CoA in PBS are shown prior to heating (blue) and post-treatment (77 h at 60°C, grey). The post-treatment spectrum is expected to be composed of liberated CoA and *p*-coumaric acid, as well as residual *p*-coumaroyl-CoA. Spectra of *p*-coumaroyl-CoA (light blue) and *p*-coumaric acid (green) in PBS are shown with the absorbance scaled to the post-treatment spectrum. Subtraction of these scaled spectra from the post-treatment spectrum, yielded a difference spectrum (dashed black trace) with strong absorbance at 258 nm, which is consistent with the published spectrum of CoA.[263] The small peak at 325 nm in the difference spectrum could indicate the presence of an additional product. (C) *P*-coumaroyl-CoA (50 µM) was found to be stable in the presence of BSA (50 µM). The sample was monitored in PBS over a period of 6 h with scans taken at 5 min intervals at 20°C.

Further, we found that after denaturing and refolding an older preparation of apo-PYP-BFP (9 days post purification), degradation of *p*-coumaroyl-CoA did not occur (Figure 6-9B), and the reactivity of the protein was slightly revived, compare to untreated apo-PYP-BFP samples of equal age (data not shown). We propose that apo-PYP may adopt additional folded state(s) that bind *p*-coumaroyl-CoA but catalyze *p*-coumaroyl-CoA hydrolysis rather than thioesterification. This competing hydrolysis reaction of the activated chromophore complicates the determination

of the reconstitution rate of apo-PYP-BFP with *p*-coumaroyl-CoA, however, it appears this reaction is significantly slower than the reaction with *p*-coumaroyl-thiophenyl ester, as expected based on their intrinsic reactivity.[264] The slow reconstitution rate observed with *p*-coumaroyl-CoA may suggest that, in the native system, chromophore reconstitution is facilitated by a chromophore-loading enzyme, such as a CoA transferase, as was speculated by Imamoto *et al.*[220] prior to the discovery of the role of TAL and 4CL, or perhaps via an interaction between 4CL and PYP. However, our results clearly show that *p*-coumaroyl-CoA alone is competent to yield holo-PYP.

6.3.4 Reconstitution of PYP-BFP *in vivo*

Next, we co-expressed PYP-BFP with the enzymes TAL and 4CL for chromophore biosynthesis and developed a protocol to monitor reconstitution *in vivo.* The metabolic pathway to produce *p*-coumaroyl-CoA is sensitive to imbalances in enzymatic activity (Figure 6-1); TAL is subject to feedback inhibition even at low levels of *p*-coumaric acid,[249] and a build-up of *p*-coumaric acid has been reported to be toxic to cell growth.[239,245] We rationalized that high 4CL activity would therefore be important for maintaining healthy growth rates. We expressed TAL and 4CL from the dual expression vector pRSFDuet^{TM}-1 (Novagen), which contains a pRSF1030 origin of replication (over 100 copies per cell [https://bionumbers.hms.harvard.edu]). PYP-BFP was expressed from the plasmid pQE80L (Qiagen), which contains a lower copy number origin of replication (ColE1, estimated 20-30 copies [Qiagen]). The expression of all three genes was controlled by IPTG-responsive promoters, allowing simultaneous induction. During expression, the production and reconstitution of PYP-BFP was monitored by taking culture samples at defined intervals and measuring time-dependent changes in BFP fluorescence. To prepare culture samples for fluorescence analysis, samples were diluted to a standard OD_{600}, pelleted, resuspended in an optically neutral buffer (PBS) and then cooled on ice for 15 min to stop protein production and maturation. Samples were transferred to 96-well plates, and were allowed to dark-adapt and equilibrate for 20 min. Each culture sample was transferred into three wells to allow the reproducibility of the measurements to be monitored with technical replicates. Fluorescence was measured by exciting dark-adapted samples at 420 nm and monitoring BFP emission at 460 nm over time. The raw BFP fluorescence data were fit to Eq. 1 to determine the fractions of apo and holo protein, as well as the photocycle kinetics as described above. The

fitted parameters for all data sets are reported in the Appendix. We observed that protein maturation and reconstitution continued in samples stored at 4°C (Figure 6-12), so measurements were performed with freshly prepared samples. It should be noted that, while the maturation of BFP is rapid ($t_{1/2}$ = 12 min at 37°C[151]), PYP with an immature BFP domain is not detected; measurements thus represent a 'snapshot' of the progress of the reconstitution.

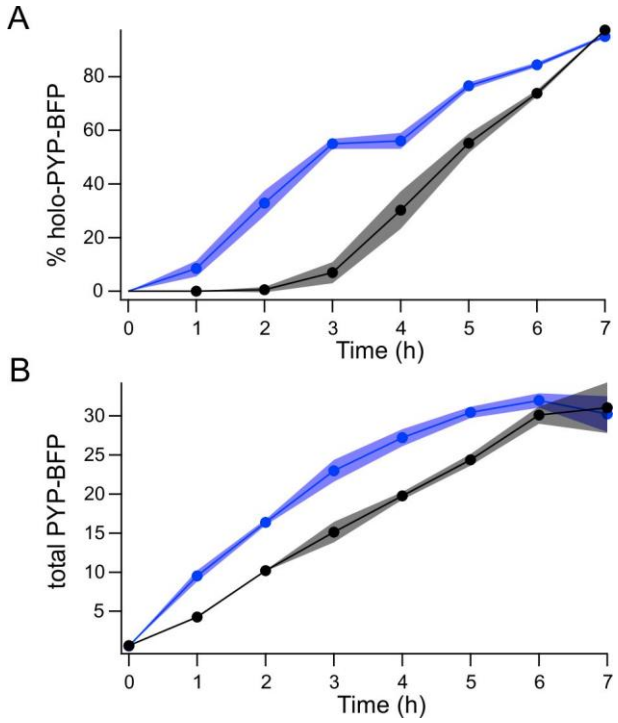


Figure 6-12: Maturation of holo-PYP-BFP at 4°C.
PYP-BFP was co-expressed with *Rs*TAL and *At*4CL, and the reconstitution was monitored by taking samples hourly. After the 6 h time point was taken, *p*-coumaroyl-thiophenyl ester was added to a final concentration of 30 µg/mL to complete reconstitution. The samples were measured immediately after preparation (black) and then once again after incubating overnight at 4°C (blue). After overnight incubation, the holo protein content and BFP fluorescence increased. The increase in BFP fluorescence was attributed to continued maturation of the BFP chromophore. The holo-PYP-BFP content (A) and total PYP-BFP content (B) are shown. The graphs show 3 biological replicates with standard deviation error envelopes.

We first attempted reconstitution in *E. coli* using TAL from *R. sphaeroides* (*Rs*TAL) and 4CL from *I. loihiensis* (*Il*4CL), organisms that both produce PYP homologues. For expression, *E. coli* was co-transformed with one plasmid encoding the chromophore biosynthesis enzymes and another plasmid encoding PYP-BFP. An isolated colony was picked and grown overnight in rich

media (DYT) with plasmid maintenance antibiotics and 1% glucose to repress expression of all three genes. The overnight culture was used the following day to inoculate 15 mL of fresh DYT (with plasmid maintenance antibiotics) to an OD_{600} of 0.1, and the culture was grown at 37°C until an OD_{600} of 0.5 was reached. The culture was then induced with 1 mM IPTG and incubated at 25°C for expression. Samples were taken periodically to check for active holo protein. Weak holo-PYP-BFP activity was detected at 48 h post-induction and did not increase with further incubation (Figure 6-13A). Fitting the fluorescence data showed that total PYP-BFP levels dropped after 30 h, likely due to the extended incubation of the cells in the same media.

Since holo-PYP-BFP was first observed in stationary phase cultures, we decided to induce the cells in stationary phase rather than in log phase. For this, the overnight culture was grown as before, then the following day, the entire culture was pelleted and resuspended in the same volume of fresh DYT containing 1 mM IPTG (and plasmid maintenance antibiotics). Using this protocol, holo-PYP-BFP was detected 11 h post-induction and reached approximately 85% reconstitution 24 h post-induction (Figure 6-13B). The total PYP-BFP content in the cultures increased throughout the expression.

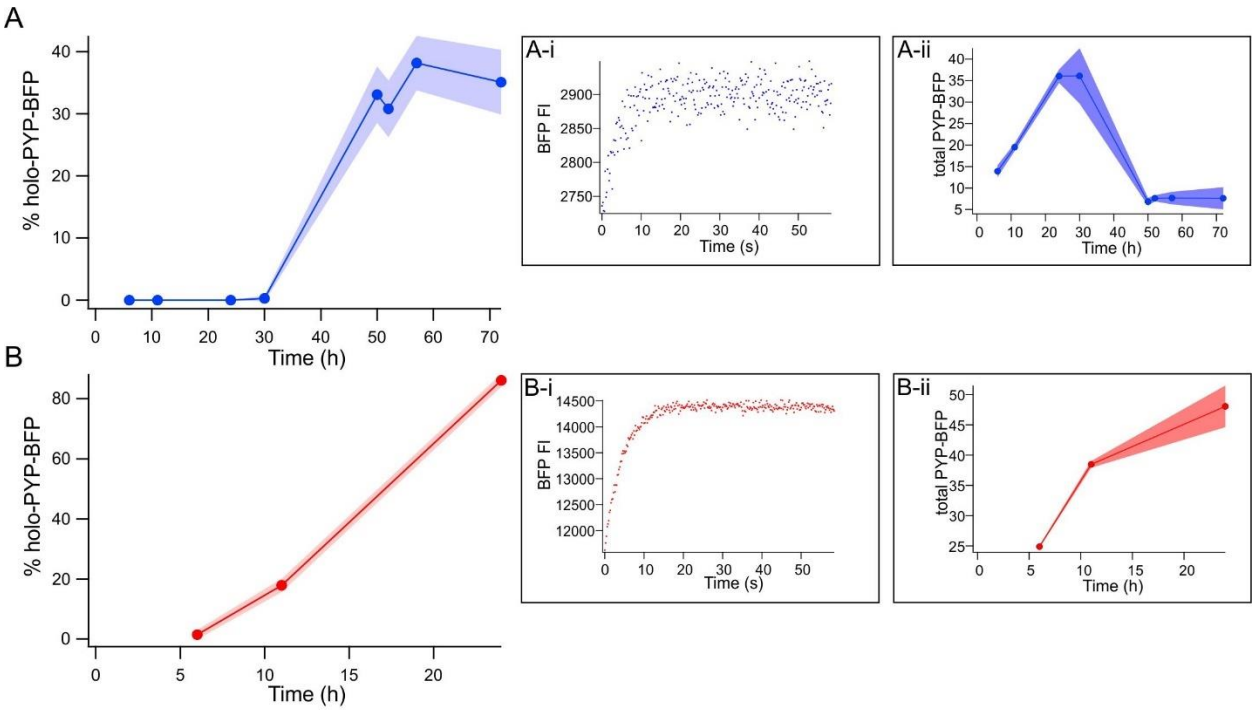


Figure 6-13: Comparison of log phase and stationary phase expression protocols.
E. coli was co-transformed with a plasmid encoding chromophore biosynthesis enzymes (*Rs*TAL and *Il*4CL) and a plasmid encoding PYP-BFP. Six isolated colonies were picked from the transformation plate; three colonies were expressed using the log phase protocol (A), and three colonies were expressed using the stationary phase protocol (B). Average holo-PYP-BFP content, determined by fitting the BFP fluorescence signal (n = 3, SD error envelopes) is shown for log phase (A) and stationary phase (B) expressions. The raw BFP fluorescence signal was significantly weaker in the log phase cultures compared to the stationary phase cultures. Raw fluorescence data are shown from one log phase culture at 72 h (A-i) and one stationary phase culture at 24 h (B-i). The signal from the log phase culture was comparable at earlier timepoints. Both cultures show increasing levels of PYP-BFP between 6 – 24 h (log phase cultures [A-ii] and stationary phase cultures [B-ii]). Incubation of the log phase culture was continued beyond 24 h to continue monitoring reconstitution (first observed at ~ 48 h). The log phase culture showed a sharp drop in protein PYP-BFP protein content after 30 h of incubation, possibly due to entry of the cells into death phase from extended incubation in the same media. The averages of three biological replicates are shown for each fitting with SD envelopes.

We attempted to further optimize the pathway, focusing our efforts on improving the expression of 4CL, since the pathway is expected to be sensitive to a build-up of *p*-coumaric acid. We codon optimized the *Il*4CL coding sequence for *E. coli* expression and added an N-terminal SUMO tag for solubility. Additionally, we reverted a point mutation in the *Il*4CL coding sequence of

unknown origin (R193S). The new version of the enzyme is referred to as syn-*Il*4CL. PYP-BFP was co-expressed with *Rs*TAL and syn-*Il*4CL using the stationary phase expression protocol. Holo-PYP-BFP was detected 8 h after induction of gene expression and reached approximately 35% reconstitution after 13 h of expression (Figure 6-14A), showing a slight improvement compared to reconstitution using the unaltered *Il*4CL.

We next exchanged syn-*Il*4CL with *A. thaliana* 4CL paralogue 1 (*At*4CL), an enzyme from an organism that does not produce a PYP homologue. Using *Rs*TAL and *At*4CL, holo-PYP-BFP could be detected 3 h post-induction and approached completion 10 h post-induction (Figure 6-14A).

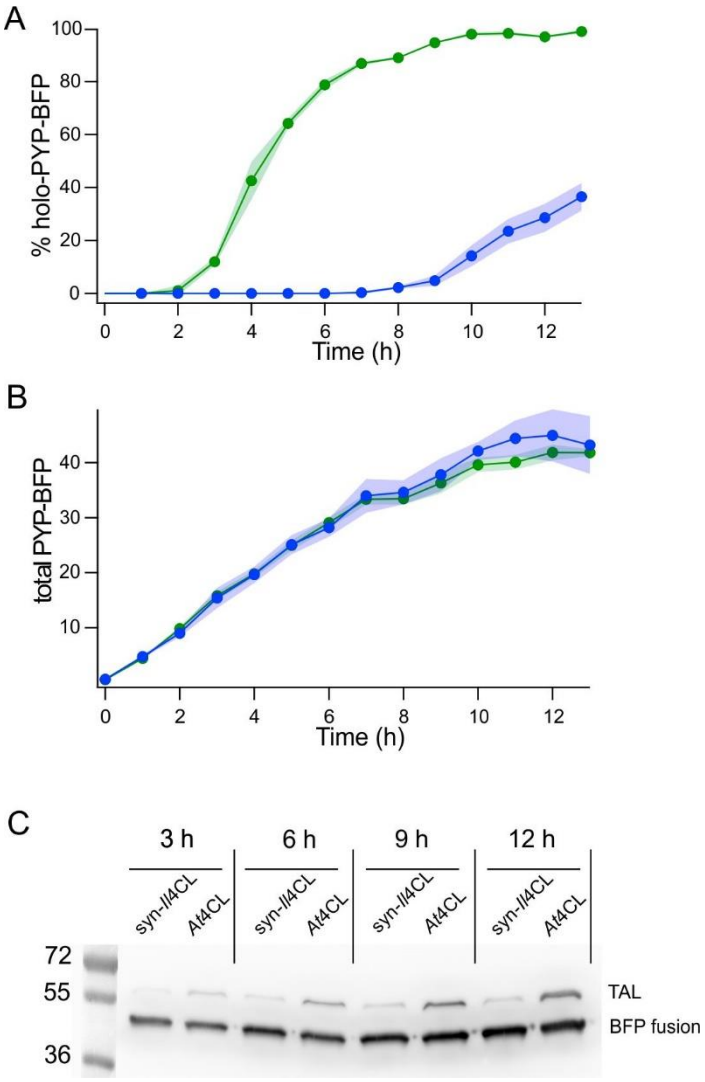


Figure 6-14: *In vivo* reconstitution of PYP-BFP using different 4CL variants.
PYP-BFP was reconstituted through co-expression with *Rs*TAL and either syn-*Il*4CL (blue) or *At*4CL (green). (A) Holo protein content was detected in the cultures with *At*4CL several hours earlier than in the cultures with syn-*Il*4CL. (B) The overall PYP-BFP protein content was similar in both cultures. This was confirmed with a western blot (C). The averages of three biological replicates for each genotype are shown. Envelopes show the SD. The western blot was probed with anti-His antibodies, detecting a His tag on *Rs*TAL (59 kDa) and PYP-BFP (43 kDa). The 4CL variants do not have a His tag and were not detected.

The raw fluorescence data from the reconstitution using *At*4CL is shown in Figure 6-15. Interestingly, the increased rate of PYP reconstitution observed using *At*4CL indicates that PYP does not need to interact with 4CL in order to accept the chromophore since *Arabidopsis* 4CL is unlikely to interact with *H. halophila* PYP.

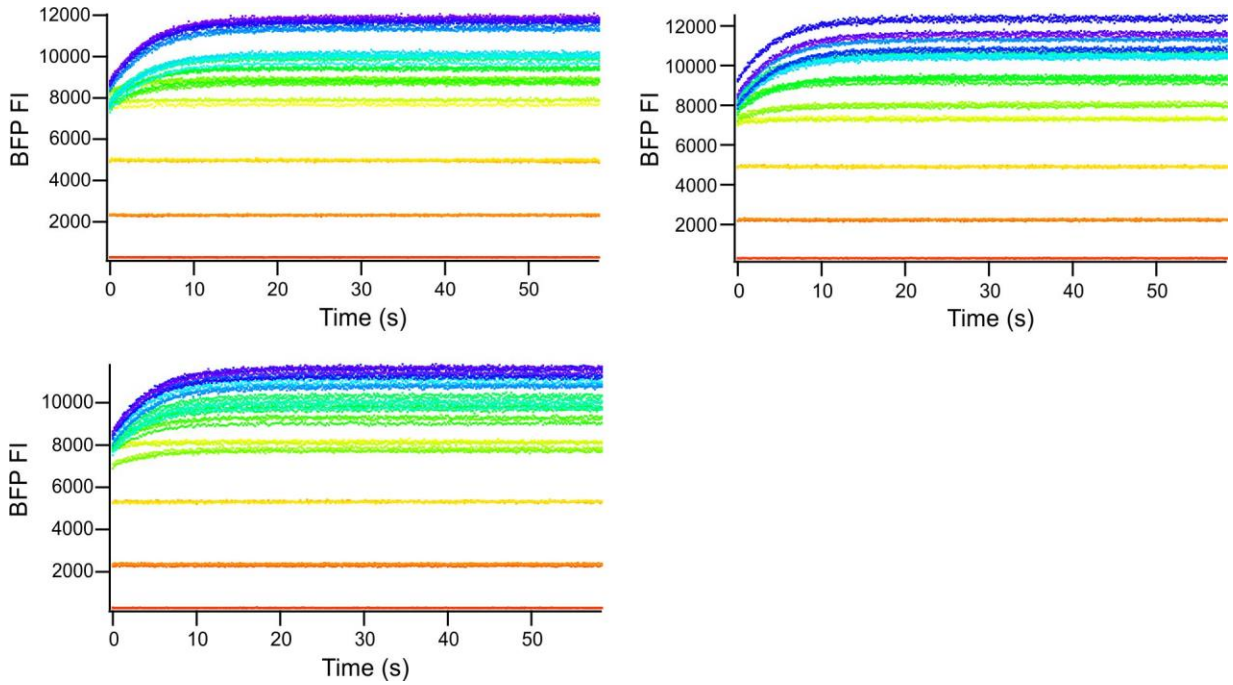


Figure 6-15: Raw data from an *in vivo* PYP-BFP reconstitution using *Rs*TAL and *At*4CL. The raw fluorescence data is shown for three cultures of PYP-BFP reconstituted via co-expression with *Rs*TAL and *At*4CL. Time points were taken hourly for 13 h (coloured in a rainbow spectrum with t = 0 red and t = 13 h purple). Each time point was measured with three technical replicates (included in the graphs above) to monitor the quality of the measurement. During the expression, BFP fluorescence levels increased as PYP-BFP was expressed and a curve began to develop as holo protein was formed. The steady-state fluorescence of holo-PYP-BFP shows quenching of BFP fluorescence due to the fraction of *trans* holo-PYP-BFP in the steady-state.

To further test the results of the fitting, the PYP-BFP reconstitution was repeated via co-expression of *Rs*TAL and *At*4CL, and a set of cultures was supplemented with *p*-coumaroyl-thiophenyl ester throughout the expression to immediately reconstitute all available apo protein (Figure 6-16). This experiment confirmed that there is apo protein available for reconstitution at the early time points during the expression.

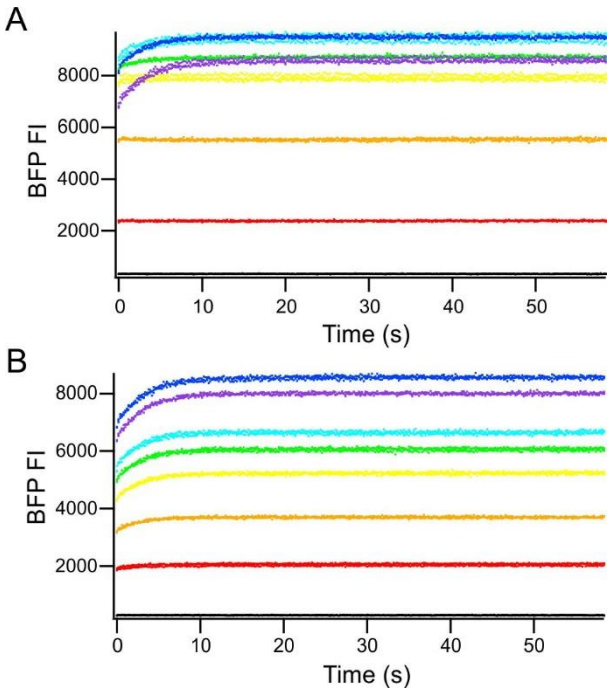


Figure 6-16: Comparing raw fluorescence data from cultures reliant on enzymatic chromophore synthesis versus cultures supplemented with additional activated chromophore.
Raw fluorescence data of cultures expressing PYP-BFP with *Rs*TAL and *At*4CL. The culture in (A) was reliant on enzymatic synthesis of chromophore for holo protein production, up until the final time point; after 6 h of expression, *p*-coumaroyl-thiophenyl ester was added to a final concentration of 35 µg/mL to complete reconstitution. The culture in (B) was supplemented with additional *p*-coumaroyl-thiophenyl ester throughout the expression (5 µg/mL per h). The graphs each show raw data from 1 culture, with 3 technical replicates (3 wells plated from the same culture) at the following timepoints: t = 0 (black), 1 h (red), 2 h (orange), 3 h (yellow), 4 h (green), 5 h (light blue), 6 h (dark blue), 7 h (purple). Note that the decrease in observed BFP fluorescence in the end timepoints in (A) is due to the development of holo-PYP-BFP; quenching occurs from the fraction of *trans* holo-PYP-BFP in the steady-state

Overall, the modifications described significantly improved the reconstitution, starting from a protocol yielding partial reconstitution that required two days of expression, to a protocol yielding complete reconstitution after 10 h of expression. It might be possible to further increase the rate of holo protein production by independently inducing the chromophore biosynthesis enzymes, to allow a supply of *p*-coumaroyl-CoA to be created prior to the induction of PYP-BFP (if intracellular *p*-coumaroyl-CoA concentrations are stable). The simultaneous induction of TAL, 4CL and PYP-BFP, however, is advantageous because it enables holo protein content to be

monitored as *p*-coumaroyl-CoA is being produced, allowing for direct comparisons of the pathway modifications.

6.3.5 Analysis of PYP variants *in vivo*

To evaluate the ability to detect changes in PYP photocycle kinetics *in vivo* using this BFP-fusion, we introduced point mutations in the PYP domain known to affect the PYP thermal relaxation rate. Point mutations known to accelerate (S117A), to slow (V105A), or to have no effect (S8A and T101A) on the PYP thermal relaxation rate were introduced.[221]

We performed *in vivo* reconstitution with each of the mutants by co-expressing *Rs*TAL and *At*4CL using the stationary phase expression protocol. The resulting BFP fluorescence signal (Figure 6-17) was fit to analyse the reconstitution.

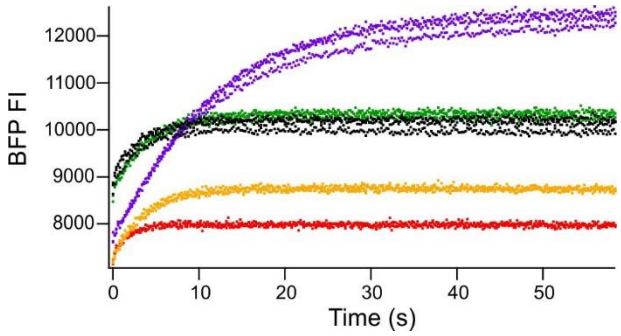


Figure 6-17: Raw data from cultures expressing PYP-BFP point mutant variants.
Raw fluorescence data of cultures co-expressing *Rs*TAL, *At*4CL and wild-type PYP-BFP (black) or PYP-BFP with one of the following point mutations: S8A (green), V105A (purple), T101A (yellow), or S117A (red). PYP-BFP S117A reaches the light-induced steady-state more quickly and has a greater proportion of the *trans* isomer in the steady-state, resulting in a stronger observed quench of the BFP fluorescence. Conversely, PYP-BFP V105A takes longer to reach a steady-state due to the longer lifetime of the *cis* isomer, and the steady-state exhibits less of quench (*i.e.* more *cis* state isomer). The graph shows 3 technical replicates for 1 culture of each genotype after 6 h of expression.

The point mutants expressed and reconstituted comparably with wild-type PYP-BFP (Figure 6-18).

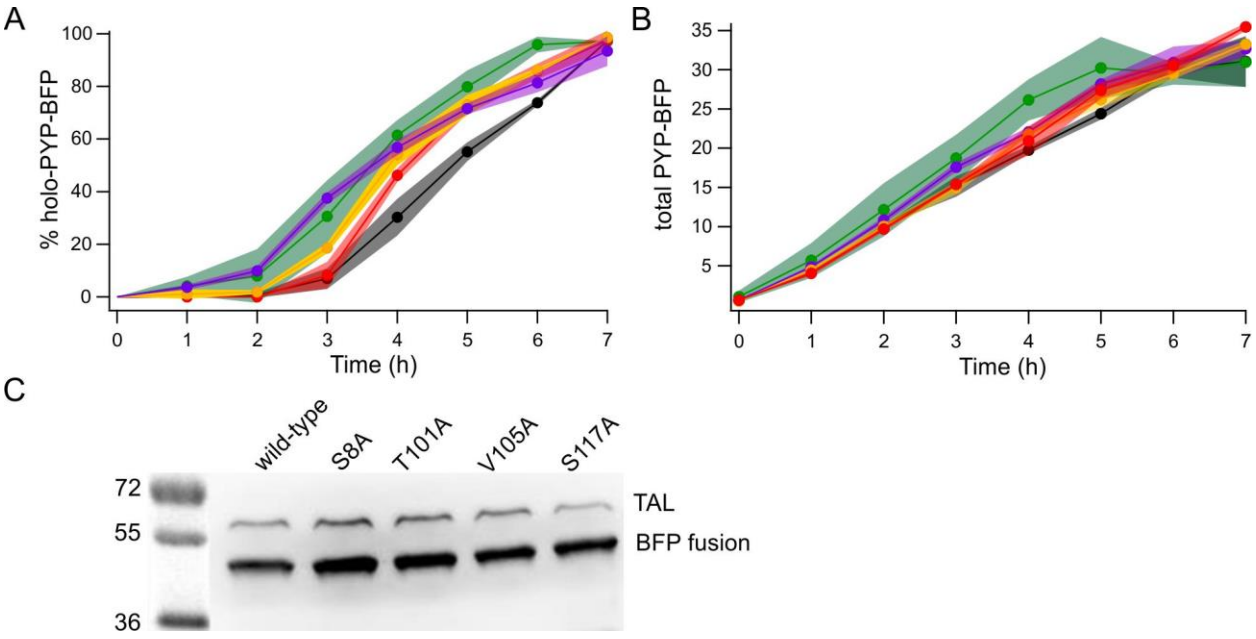


Figure 6-18: Reconstitution of cultures expressing a PYP-BFP point mutant variant.
In vivo monitoring of holo protein content (A) and total PYP-BFP (B) is shown for constructs with point mutations in the PYP domain of PYP-BFP (wild-type is shown in black, S8A in green, T101A in yellow, V105A in purple and S117A in red). The graphs show the average and SD of 3 biological replicates. (C) A western blot showed that the mutants expressed at slightly higher levels than the wild-type protein. The culture samples used in the western blot were produced in an independent expression and do not report directly on the samples shown in the fitting. The western blot was probed with anti-His antibodies, detecting a His tag on *Rs*TAL (59 kDa) and on each of the different PYP-BFP mutants (43 kDa). 4CL does not have a His tag and was not detected.

To evaluate the relaxation rate of each PYP-BFP mutant, the time-dependent BFP fluorescence intensity data from the wild-type protein and the four mutants were analyzed using a global fit, with the fluorescence terms (Fl_T, Fl_C, and Fl_{apo}) and k_f constrained among all data sets, and the terms k_b, [apo] and [holo] fit independently for each data set (Table 6-2). The fitting predicted a holo content of 75 - 90% for the samples. The calculated thermal relaxation rates for the PYP-BFP constructs with neutral mutations (S8A and T101A mutants) were comparable to that of the wild-type, while the V105A mutant had a slower relaxation rate, and the S117A mutant had a faster relaxation rate as expected. These data show that the effects of the point mutations on the PYP photocycle can be determined *in vivo*.

Table 6-2: Thermal relaxation rates of PYP point mutants.

Point mutant in PYP domain of PYP-BFP	**Expected effect on photocycle with respect to wild-type**	***In vitro* rate constant (Philip *et al.*)[221] (s^{-1})**	**Calculated *in vivo* rate constant (k_b, SD with n = 3 biological replicates) (s^{-1})**	**Calculated *in vivo* half-life of the light-state (s)**
Wild-type	n/a	1.1	0.30 (+/- 0.001)	2.3 (+/- 0.01)
S8A	Neutral	1.1	0.21 (+/- 0.01)	3.3 (+/- 0.2)
T101A	Neutral	1.1	0.23 (+/- 0.003)	3.0 (+/- 0.03)
V105A	Slower	0.1	0.041 (+/- 0.007)	17 (+/- 2.6)
S117A	Faster	2.0	0.68 (+/- 0.15)	1.1 (+/- 0.2)

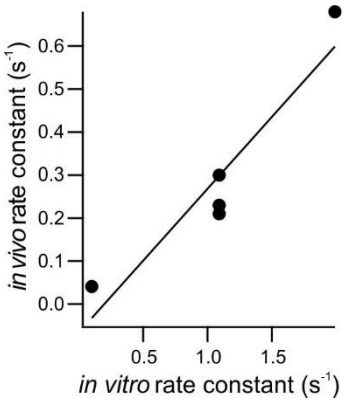


Figure 6-19: Correlation of rate constants determined *in vivo* with reported rate constants from *in vitro* analysis.
The rate constants determined *in vivo* via fitting BFP fluorescence time-course data (Eq. 1) correlate well with reported *in vitro* rate constants.[221]

6.3.6 Analysis of biosynthetic chromophore uptake in PYP homologues

Since the original discovery of PYP in *H. halophila,*[265] genome sequencing projects have revealed a large number of PYP homologues.[224,236] Within the PYP family, pairwise sequence identity can be as low as 19%, however the residues forming the chromophore pocket are highly conserved.[224,236] The genome of *H. halophila* encodes two PYP homologues, designated PYP(A) and PYP(B), with a sequence identity of 60% (Figure 6-20).

```
PYP(A)      MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGK        60
PYP(B)      MGTLIFGRQDLENRLAAMTPEEIDDLPFGVIQIDQHGRILLYNATEGAITGRDPEAMIGR        60
            *  : ** :*:** ** *   ::*.* **.**:* .*.** ***:** ******: :**:

PYP(A)      NFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWV        120
PYP(B)      DFFNDVAPCGHTEAFYGRFQEGVRHGDLNEIFDYTFDYRMAPTKVRVHMKRALSGDTYWI        120
            :**:***** .:  ***:*:***  *:** :*:*****:*:****:****:*****:**:

PYP(A)      FVKRV*----- 125
PYP(B)      FVKRISAPAA* 130
            ****:
```

Figure 6-20: Sequence alignment of two PYP homologues.
Sequence alignment of PYP(A) (also referred to as wild-type PYP) and PYP(B) from *H. halophila* using Clustal Omega.[266–268] Sequence identity is indicated by a star (*). Residues with strongly similar properties are marked by a colon (:) and residues with weakly similar properties are marked by a period (.).

PYP(A) from *H. halophila* is the most commonly studied PYP version, and is the version used in this work unless otherwise mentioned (it is also referred to as wild-type PYP). Compared to PYP(A), PYP(B) has a slightly blue-shifted absorbance maximum (443 nm) and a longer-lived light-state (k_b = 1.3 x 10^{-2} s, $t_{½}$ = 53 s).[269] We constructed a PYP(B)-BFP-fusion and co-expressed it with *Rs*TAL and *At*4CL using the stationary phase expression protocol. Reconstitution of PYP(B)-BFP occurred more slowly than wild-type PYP and the expression level was lower (Figure 6-21).

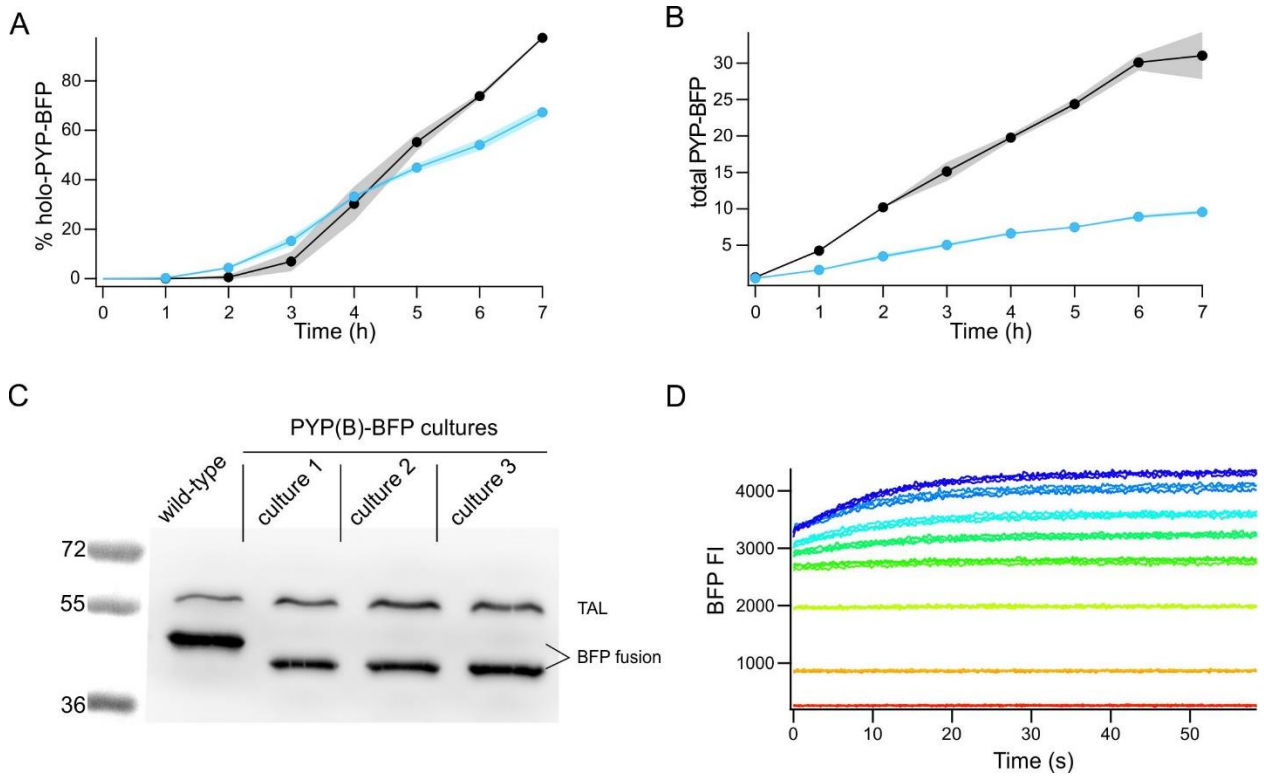


Figure 6-21: *In vivo* reconstitution of PYP homologues.
PYP-BFP versions of the two PYP homologues from *H. halophila* were reconstituted *in vivo* by co-expression with *Rs*TAL and *At*4CL. Culture samples were taken hourly. After 6 h of expression, *p*-coumaroyl-CoA (30 µg/mL) was added to the cultures; the final time point shows cultures supplemented with additional activated chromophore. Fitting BFP fluorescence data showed that PYP(B)-BFP (blue) reconstituted more slowly than the wild-type PYP-BFP (black) (A) and expressed at a lower level (B). The data shows the average values with SD envelopes for 3 biological replicates of each strain. (C) A western blot confirmed that the expression level of PYP(B)-BFP was lower than that of wild-type PYP-BFP. The blot was probed with anti-His antibodies, detecting a His tag on *Rs*TAL (59 kDa) and either wild-type PYP-BFP (43 kDa) or PYP(B)-BFP (44 kDa). Three cultures of PYP(B)-BFP are shown. PYP(B)-BFP runs faster on a gel than would be predicted by its molecular weight. The coding sequence was fully sequenced and did not contain any truncations or other mutations. The culture samples used in the western blot were produced in an independent expression and do not report directly on the samples shown in the fitting. (D) The raw fluorescence data for the reconstitution of PYP(B)-BFP showed lower BFP fluorescence levels compared to wild-type PYP-BFP (see Figure 6-16A). The graph shows the raw data from 1 culture of PYP(B)-BFP with 3 technical replicate measurements per timepoint (hourly for 7 h, with time point measurements coloured in a rainbow spectrum starting from t = 0 in red).

6.3.7 *In vivo* reconstitution of PYP-based optogenetic tools

Protein engineering efforts to develop PYP-based optogenetic tools have focused on co-opting the photo-induced structural changes in the N-terminus of PYP to affect the function of a fused target protein in a light-dependent manner. The N-terminal region (residues 1-28) consists of two α-helices that pack against the PAS core.[270] In the light-state, these helices unfold, and the chromophore is exposed to the solvent.[271] Photo-controlled constructs have been successfully developed by redesigning the N-terminal region to create a chimeric fusion with a target protein.[25,95,96] Ui *et al.* have also reported a fusion of the pore-forming protein hemolysin with the C-terminus of PYP.[233] In addition, a circularly permutated PYP (cPYP) variant was developed as an alternative scaffold for engineering photo-controlled proteins.[14] This construct was designed such that a target peptide could be inserted between the native PYP N- and C-termini, to maximize the effect of the N-terminal structural changes.[17] Recently, phage display methods have been used to develop interaction partners for cPYP that selectively bind the light-adapted or dark-adapted states.[16]

In vivo reconstitution of PYP-based optogenetic tools using biosynthesized chromophore has not yet been demonstrated. Previously, reconstitution of PYP-based optogenetic tools has been achieved by adding chemically synthesized activated *p*-coumaric acid derivatives, such as *p*-coumaric acid anhydride,[233] *p*-coumaroyl thiophenyl ester[25,95,96] or fluorogenic thioesters[272] to cells in culture. PYP-based optogenetic tools can contain substantial changes to the PYP structure and these changes may impact the ability to produce holo protein from the less reactive *p*-coumaroyl-CoA *in vivo*. We therefore tested the reconstitution of several PYP-based optogenetic tools using the optimized chromophore biosynthesis pathway in *E. coli* described above.

We constructed BFP-fusion versions of two optogenetic constructs that contained changes to the N-terminal part of PYP: GCN4-Δ25PYP(v2 K143F E35L), a light-responsive version of the yeast transcription factor GCN4,[96,231] and opto-DN-CREB, a photo-controlled inhibitor of the mammalian transcription factor CREB.[25] We also constructed a BFP fusion of the cPYP scaffold.

We found that C-terminal BFP fusions of GCN4-Δ25PYP and opto-DN-CREB expressed poorly, so constructs were re-designed with the BFP fused to the N-terminus of the PYP-based construct instead. The N-terminal BFP-fusion contained a new linker that was longer than the original version (30 residues versus 9 residues), which would be expected to decrease the energy transfer efficiency between the two domains. To analyze the effect of the altered fusion, an N-terminal fusion of BFP to wild-type PYP was also prepared, allowing the activity of the new construct to be compared to the original PYP-BFP design (Figure 6-22). As expected, the N-terminal BFP-fusion construct showed a weaker quench of BFP fluorescence.

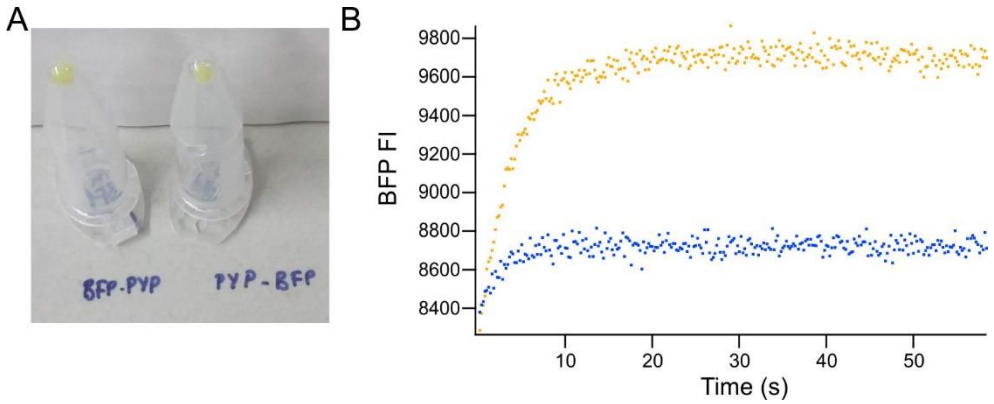


Figure 6-22: Comparison of the activity of two BFP-fusion designs.
The BFP-fusions consist of wild-type PYP with either an N-terminal or a C-terminal fusion to BFP (BFP-PYP or PYP-BFP respectively). BFP-PYP has a longer linker than PYP-BFP. These two constructs were reconstituted *in vivo* by co-expression with *Rs*TAL and *At*4CL. Reconstitution of both constructs appeared to occur at a comparable rate based on the development of a yellow pellet. (A) After 5 h of expression, both BFP-PYP (left) and PYP-BFP (right) had bright yellow pellets. (B) The culture of PYP-BFP (yellow) showed a much larger quench in BFP fluorescence intensity compared to the culture of BFP-PYP (blue). The raw fluorescence data for three technical replicates of each culture is shown.

Changing the fusion between BFP and wild-type PYP altered the energy transfer between the two domains. In order to quantify this difference, BFP fluorescence data from both BFP-fusion designs were fit simultaneously. Fitting the BFP fluorescence time-course data from both designs together enabled the parameters affected by the altered linker to be determined (Figure 6-23, Table 6-3), calibrating the fitting for use with the alternative BFP-fusion design. The rate constant of the reverse reaction (k_b, relaxation to dark-state PYP) should be the same in both constructs. Likewise, the BFP fluorescence intensity of BFP fusion with an apo-PYP domain

(Fl_{apo}) should not be affected by the altered fusion design and the fluorescence intensity for the BFP-fusion with a *cis* holo-PYP domain (Fl_C) should only be minorly affected. These parameters were linked among all datasets for the fitting. The parameters Fl_T, k_f, [apo] and [holo] were fit independently for each dataset. Overall these data show that the BFP-fusion strategy can be used to detect reconstitution and photo-switching activity of PYP constructs in a manner that is robust with respect to changes in the linker and fusion design.

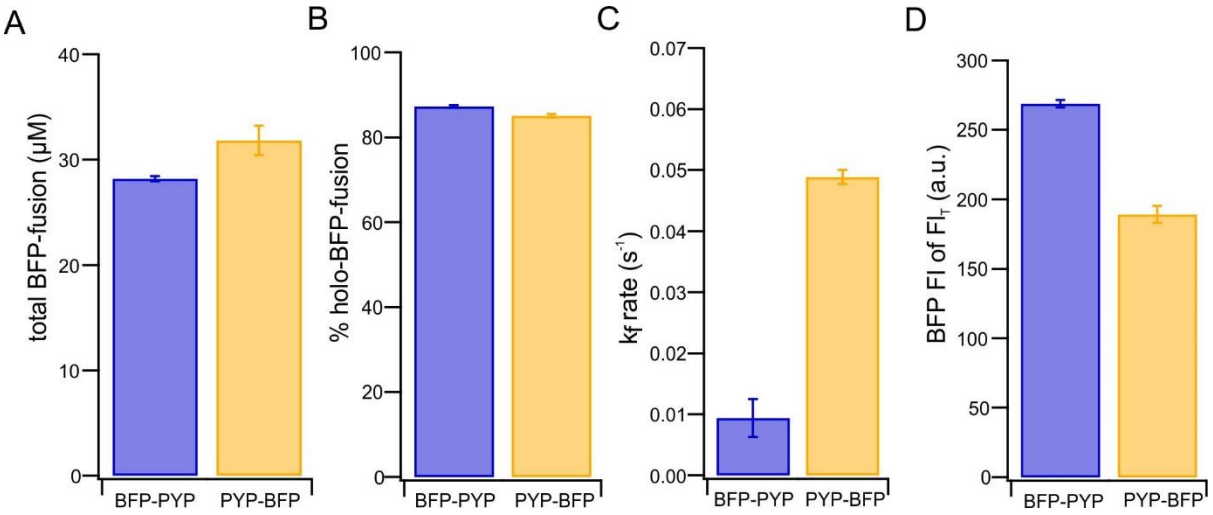


Figure 6-23: Fitted parameters from the two BFP-fusion designs.
The fitting showed that the total BFP-fusion concentration (A) and the percent holo BFP-fusion (B) was roughly the same in both BFP-fusion designs. In BFP-PYP, the rate of the forwards reaction (k_f, C) was decreased, and the BFP fluorescence with *trans* holo-PYP (Fl_T, D) was increased, compared to PYP-BFP. The graphs show the average values with SD error bars (n = 3) for cultures of BFP-PYP (blue) and PYP-BFP (yellow).

Table 6-3: Summary of the fitted parameters of the PYP-BFP and BFP-PYP data.

Parameter	**Both BFP fusions**	
Fl_A (arbitrary units)	536	
Fl_C (arbitrary units)	527	
k_b s^{-1}	0.244	
	PYP-BFP	**BFP-PYP**
Fl_T (arbitrary units)	189	269
k_f s^{-1}	0.049	0.0094
[apo] µM	4.8	3.6
[holo] µM	27.1	24.6

We then monitored chromophore uptake in BFP-GCN4-Δ25PYP by co-expressing the BFP-fusion with *Rs*TAL and *At*4CL, and measuring PYP-induced changes in BFP fluorescence at regular intervals during the expression (Figure 6-24). BFP-GCN4-Δ25PYP reconstituted more slowly than wild-type PYP-BFP. After 11 h of expression, the culture had approximately 80% holo protein. After this point, *p*-coumaroyl-thiophenyl ester was added to the cultures (30 µg/mL) and after an additional hour of incubation, photo-switching was measured, showing that holo protein content had increased to approximately 94%.

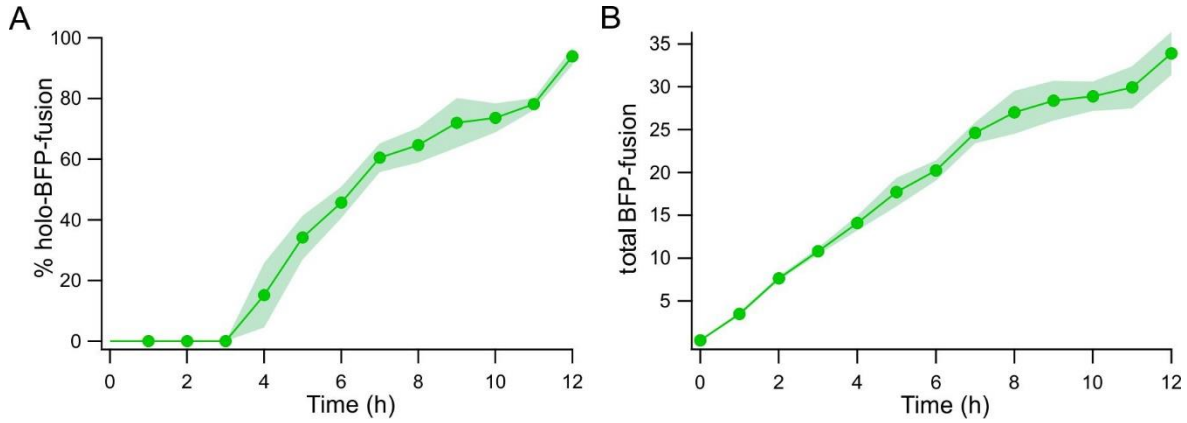


Figure 6-24: Reconstitution an engineered PYP with a Δ25PYP scaffold: GCN4-Δ25PYP. A BFP-fusion with GCN4-Δ25PYP was reconstituted *in vivo* by co-expressing *At*4CL and *Rs*TAL. (A) BFP-GCN4-Δ25PYP showed holo protein content starting at 4 h and slowly increased to approximately 80% reconstitution at 11 h post induction. Cultures were then supplemented with *p*-coumaroyl-thiophenyl ester (30 µg/mL), resulting in approximately 94% reconstitution at 12 h. (B) Total BFP-GCN4-Δ25PYP protein content increased steadily. The graphs show the results of three biological replicates (SD error envelopes).

We next reconstituted a BFP-fusion version of opto-DN-CREB by co-expressing *Rs*TAL and *At*4CL. The light-state of opto-DN-CREB has a long lifetime ($t_{1/2}$ = 20 min),[25] and as a result, the protein would require an extended period of irradiation for a photo-stationary state to develop. Therefore, to evaluate reconstitution of this construct, we instead irradiated samples with 450 nm light to isomerize the PYP-based domain to the light-state, and then measured BFP fluorescence intensity at programmed time points as opto-DN-CREB returned to the dark-state. Photo-switching was observed as a loss of BFP fluorescence over time. Cultures of BFP-opto-DN-CREB were found to be photoactive, indicating reconstitution with enzymatically produced chromophore had occurred (Figure 6-25); however, the reproducibility of the relaxation rate was low. This could be due to variable non-specific binding of opto-DN-CREB to cellular DNA, which was noted previously to shorten the lifetime of the light-state.[25]

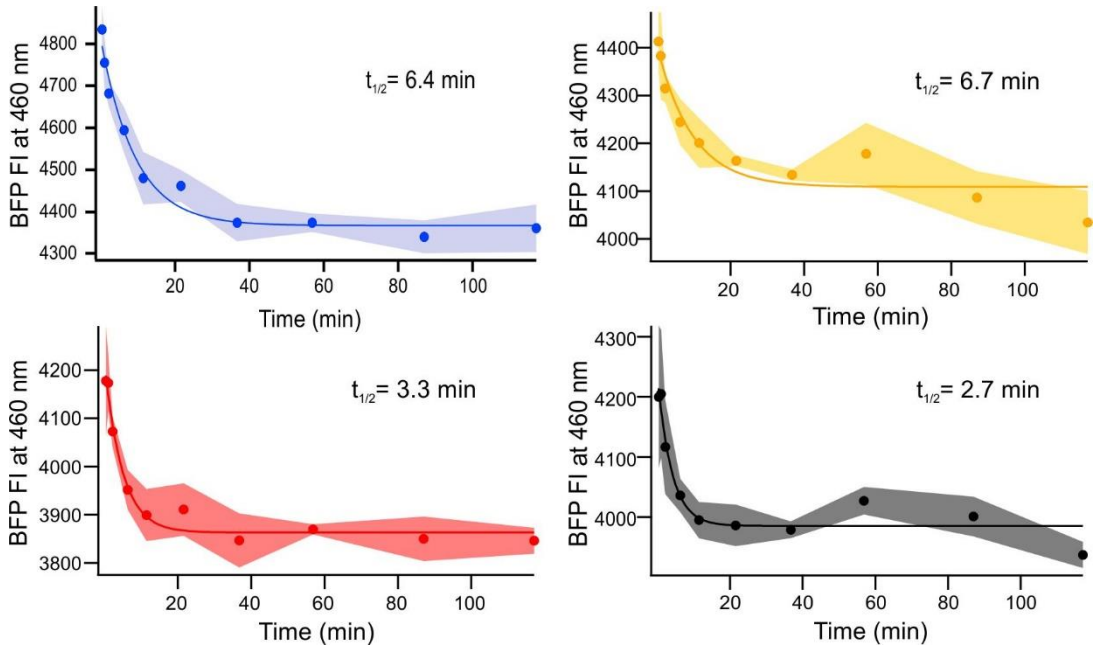


Figure 6-25: Reconstitution an engineered PYP with a Δ25PYP scaffold: opto-DN-CREB. A BFP-fusion with opto-DN-CREB was reconstituted *in vivo* by co-expressing *At*4CL and *Rs*TAL. Photo-switching of the PYP domain of BFP-opto-DN-CREB was monitored at 30°C by irradiating culture samples with 450 nm light to obtain light-state protein and subsequently measuring the BFP fluorescence intensity (ex. 401 nm, em. 460 nm) during the recovery of the dark-state (with shaking between measurements). Cells from an overnight reconstitution were used for analysis (with no synthetic chromophore added). Shown are 4 biological replicates with SD error envelops showing 3 technical replicates (see text).

The protocol used to examine the recovery of the dark-state of the PYP-based domain was performed with cells expressing only BFP. No change was seen in the BFP fluorescence intensity over time (Figure 6-26).

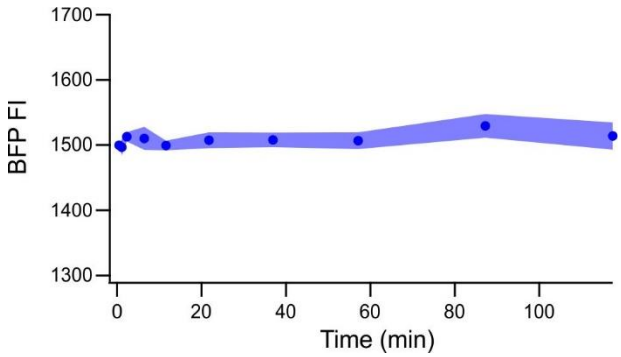


Figure 6-26: Monitoring BFP fluorescence intensity of cells with BFP alone.
Monitoring BFP fluorescence intensity of cells expressing BFP alone showed no change in intensity after irradiation. Samples were irradiated with blue light (450 nm), and BFP fluorescence (ex. 401 nm, em. 460 nm) was measured at programmed intervals over 2 h with shaking between measurements. Data shows measurements from 3 biological replicates with a SD envelope.

We next investigated biosynthetic reconstitution of the cPYP scaffold. The photocycle of cPYP is an order of magnitude faster than that of wild-type PYP ($t_{1/2}$ = 0.14 s and $t_{1/2}$ = 1.5 s respectively).[14] Due to the short lifetime of the light-state, it was not possible to evaluate holo protein content based on changes in BFP fluorescence intensity since the photo-stationary state was reached too quickly. We instead monitored the development of a yellow pellet in cells co-expressing BFP-cPYP, *Rs*TAL and *At*4CL (Figure 6-27A). The first signs of yellow could be seen 4 h after induction and continued to intensify. We then made a point mutation in BFP-cPYP known to increase the lifetime of the light-state (M121E, M100E in wild-type PYP numbering).[14] We analyzed biosynthetically reconstituted samples using the same method used for the opto-DN-CREB construct. The recovery of the dark-state fit to a mono-exponential curve with a half-life of 11.4 +/- 2.9 min (n = 4), which is similar to the previously reported relaxation rate for cPYP M121E ($t_{1/2}$ = 12 min, 20°C) measured *in vitro*[14] (Figure 6-27B).

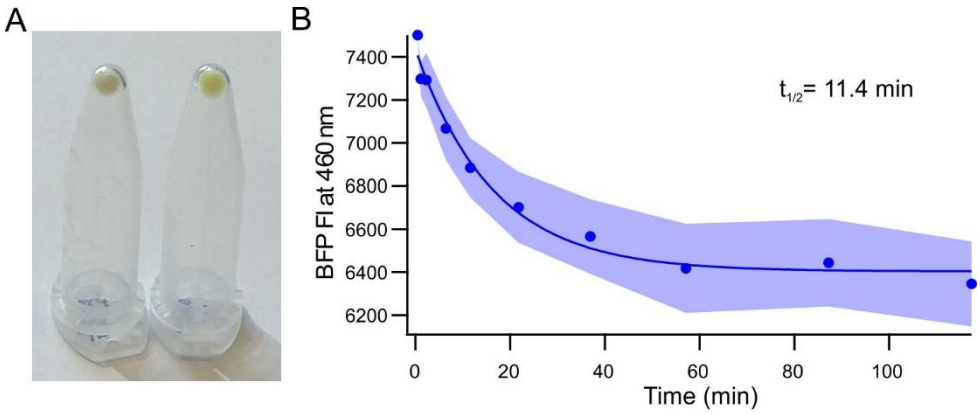


Figure 6-27: Enzymatic reconstitution of cPYP.
BFP-fusions with cPYP were reconstituted *in vivo* by co-expression of *At*4CL and *Rs*TAL. (A) Reconstitution of BFP-cPYP was monitored by the development of a yellow pellet. Left: cell pellet with expression repressed with glucose. Right: cell pellet after 7 h of expression. (B) Photo-switching of BFP-cPYP M121E was monitored by irradiating cell samples with blue light (450 nm) and monitoring the BFP fluorescence during the recovery of the PYP dark-state. Cells from an overnight reconstitution (16 h) were used for analysis. The graph shows the average fluorescence values of 4 cultures (SD error envelopes).

These data demonstrate for the first-time reconstitution of different PYP-based optogenetic tools with biosynthesized chromophore. The slower reconstitution rate observed with the GCN4-Δ25PYP BFP-fusion compared to wild-type PYP suggests that the ability to take up chromophore can be impacted significantly by the protein structural changes required to create optogenetic tools. In addition, when PYP-based optogenetic tools are deployed in different cell types, it is expected that biosynthetic production of chromophore will have to be re-optimized depending on the host chosen, as has been the case for other optogenetic tools that employ distinct chromophores.[273] Thus BFP-fusions of PYP and PYP analogues constitute a robust sensor highly useful for monitoring and optimizing PYP reconstitution and function in diverse settings.

7 Conclusions

7.1 Summary

Directed engineering strategies are highly effective tools for the development and optimization of synthetic proteins. Despite this, the majority of attempts to create novel photo-switchable proteins have followed a rational design approach, due in part to the difficulty of finding an appropriate method to evaluate switchable protein activity. This work describes the development of two novel genetic circuits that can be used to evaluate libraries of switchable proteins. In addition, a PYP-BFP-fusion construct was developed and used to analyze chromophore reconstitution in PYP, in order to facilitate the use of this photoactive protein scaffold to develop photo-switchable protein tools.

7.1.1 Development of genetic circuits for engineering switchable protein activity via directed evolution

In Chapter 2, a genetic circuit was discussed that functions as a screen for conditional DNA-binding activity. This circuit was designed by placing the DNA recognition sequence of the protein of interest in the promoter region of a gene encoding the fluorescent protein, tdTomato. As a result, DNA binding of the target protein, and the lack thereof, was reported based on the absence or presence of tdTomato fluorescence, respectively. Careful tuning of expression levels, culture conditions and incubation times was required to optimize the assay. Under optimized conditions, the circuit was shown to have a good dynamic range, based on the fluorescence obtained when using control constructs with constitutive DNA-binding and non-binding activities. Known photo-switchable DNA-binding proteins were shown to produce a detectable difference in cellular tdTomato fluorescence in the light versus the dark, indicating the sensitivity of the circuit was adequate for evaluating libraries of photo-switchable constructs. Indeed, screening a small library successfully recovered a known photo-switchable construct as well as a novel photo-switchable variant with a similar dynamic range.

The hitchhiker-bandpass circuit, discussed in Chapter 3, is to our knowledge the first example of a selection system able to directly select for protein-protein interactions of a desired affinity range. The circuit was developed by combining a circuit linking the affinity of a protein pair to β-lactamase activity,[130,131] with a bandpass circuit for β-lactamase activity.[92] This created a

bandpass circuit for protein-protein interactions, which enables selection for protein interactions of desired affinity using different concentrations of Amp. It was shown that the concentration of Amp permitting cell growth was linked to the interaction strength of an expressed protein pair. This allowed a mixed culture of strains expressing interacting protein pairs with different affinities to be separated into pure cultures based on protein interaction strength. These separations were demonstrated on LB agar plates and in liquid culture. The liquid culture separations were robust and complete, as was demonstrated using strains with different fluorescent markers. The compatibility of the circuit with liquid-culture format was important, because it enabled Amp concentrations to be more finely controlled, yielding more precise experimental control. A protein inhibitor was incorporated into the circuit and it was shown that the circuit could differentiate between inhibitors of various binding affinities. This demonstrated the potential for using the hitchhiker-bandpass circuit to develop inhibitors of protein-protein interactions with tuned, or switchable, binding affinity.

The tunable positive selection enabled by the hitchhiker-bandpass circuit is unique and expected to be a very useful addition to the array of circuits available for evaluating protein function. The use of a positive selection to select both for interacting proteins as well as for non-interacting proteins, is robust against false positive results. Negative selectable markers are known to suffer from a high rate of false positives due to the occurrence of inactivating mutations. In the hitchhiker-bandpass circuit, a false positive would arise through an unintended change in the activity level of β-lactamase. This could only occur through point mutations altering β-lactamase catalytic activity or the expression level of one or both of the hitchhiker pairs. These events could only arise from a limited number of mutations and would be much less likely to occur than a mutation leading to gene inactivation.

The robustness of the hitchhiker-bandpass circuit to false positives inspired the use of the circuit in other contexts; studies in Chapter 4 investigated the use of the hitchhiker-bandpass circuit for evaluating small molecule modulators of protein-protein interactions. A hitchhiker pair based on the interaction between Bcl-2 and BAD was constructed and shown to require high concentrations of Amp for growth in the hitchhiker-bandpass circuit, as would be expected based on reported *in vitro* data.[176] This interaction was then challenged with a series of commercially available small molecule Bcl-2 inhibitors. However, inhibition of the protein interaction was not

detected with any of the tested compounds. The lack of observed inhibition could have been due to the lack of compound permeation through the *E. coli* cell wall and/or due to active efflux of the compounds. Permeabilizing the *E. coli* OM would be expected to cause leakage of periplasmic contents, including β-lactamase. Instead, the OM channel TolC, was deleted to inhibit endogenous efflux pumps. The TolC knockout, however, did not affect the activity of the chemical inhibitors in the hitchhiker-bandpass circuit. Gram-negative bacteria are known to be highly resistant to permeation by small molecules.[186,197] It was concluded that the hitchhiker-bandpass circuit cannot be used to evaluate small molecule modulators of protein-protein interactions, because the permeability barrier presented by the *E. coli* cell wall would result in a high number of false negative results.

Chapter 5 discusses on-going efforts to develop photo-controlled inhibitors of the transcription factors CREB and AP-1 via directed evolution. The hitchhiker-bandpass circuit is expected to enable the selection of active and inactive inhibitors using low and high concentrations of Amp respectively. Chimeric fusions of peptide inhibitors were constructed with a circularly permutated *As*LOV2 scaffold, and these constructs were used to generate libraries. Alternative protein design strategies are discussed, such as using other photoactive protein scaffolds (*e.g.* *Rs*LOV or PYP), and alternative library strategies are considered, including libraries creating new circular permutations or varying linkers and fusions. In addition, the use of the hitchhiker-bandpass assay under blue light irradiation was discussed. The negative selection agent for β-lactamase activity, Tet, was substituted with Dox, an analogue with increased light stability. Dox was unfortunately found to be unstable under constant blue light irradiation, indicating that irradiation conditions will require adjustment to avoid Dox degradation.

7.1.2 Analysis of chromophore reconstitution in PYP

Chapter 6 presents a synthetic protein, consisting of a fusion between PYP and BFP, that was designed to analyze chromophore reconstitution in PYP. By fusing PYP to BFP, the photo-state of the PYP domain can be monitored by changes in BFP fluorescence levels. This construct was characterized, and it was shown that time-dependent changes in BFP fluorescence intensity could be fit to determine PYP photocycle kinetics and the fraction of apo and holo protein in a sample. Next, PYP-BFP was used to analyze PYP reconstitution *in vitro.* Although it was postulated that

p-coumaroyl-CoA is the activated species responsible for reconstitution of PYP,[234–237] it has never been shown that these species directly react with one another, potentially due to the instability of purified apo-PYP. This work clarifies this outstanding question in the literature, by showing that *p*-coumaroyl-CoA alone is competent to yield holo-PYP. Next, TAL and 4CL, enzymes catalyzing the production of *p*-coumaroyl-CoA from tyrosine, were co-expressed with PYP-BFP, and holo protein production was monitored *in vivo*. Chromophore biosynthesis was optimized by modifying the expression protocol and improving 4CL activity by using different gene homologues. The ability to monitor PYP photocycle kinetics *in vivo* was demonstrated, using PYP-BFP constructs containing point mutations in the PYP domain known to affect the photocycle. Finally, the BFP-fusion design was used to monitor chromophore reconstitution in optogenetic constructs containing engineered variants of PYP. This is the first example of biosynthetic reconstitution of PYP-based optogenetic tools and is an important step towards the use of these tools *in vivo.*

7.2 Future directions

7.2.1 Future use of the fluorescence-based screening circuit for conditional DNA-binding

The fluorescence-based screening system discussed in Chapter 2 could be used to develop other switchable DNA-binding proteins. The tdTomato reporter plasmid contains an AP-1 DNA binding site, which is also recognized by other DNA-binding proteins.[213] Alternatively, the AP-1 binding site could be exchanged with the binding site of other target proteins. Optimization of the circuit (*i.e.* protein induction levels, expression times) would be required for use with different proteins and, in particular, if the DNA recognition site is exchanged.

While the FACS-based screening approach enables this assay to be performed in high-throughput, this genetic circuit could be also converted to a dual selection. Exchanging the fluorescent marker gene with a dual selectable marker, such as TetA, would allow for selection of both DNA-binding activity and the lack thereof. Using the example of the TetA selectable marker, growth in the presence of Tet would impose positive selection for operon expression (non-binding state), while growth in the presence of $NiCl_2$ would impose negative selection for operon expression (binding-state). Such a system could potentially enable liquid culture

selections, allowing for competitive enrichment of rare library variants with a growth advantage. Negative selectable markers, however, have a high rate of inactivation due to the selection pressure for loss-of-function mutations. The rate of inactivating mutations would need to be evaluated to determine if this would obstruct circuit performance.

7.2.2 Future applications of the hitchhiker-bandpass circuit

The hitchhiker-bandpass circuit offers a unique approach to develop protein interactions with tuned or switchable binding affinity. The hitchhiker circuit has been shown to function with a wide variety of protein interactions,[130,131,133,134,153] indicating the hitchhiker-bandpass circuit should be highly generalizable. The ability to directly select for protein interactions with weak or moderate affinity is novel, and could be useful for applications requiring protein interactions with tuned affinity, such as in affinity-based drug delivery systems.[274] The ability to co-express a third interacting protein in the circuit increases the experimental flexibility, enabling the use of tripartite protein interactions or protein inhibitors. This circuit should be widely applicable for generating switchable protein-protein interactions, whether in response to light or other stimuli.

7.2.3 Future use of the PYP-BFP sensor

The PYP-BFP sensor is expected to facilitate use of PYP-based optogenetic tools by enabling photo-switching and reconstitution of the PYP domain to be monitored *in vivo*. We observed reconstitution of GCN4-Δ25PYP was delayed compared to wild-type PYP, indicating the ability to accept chromophore imposes a constraint on the design of PYP-based photo-switchable tools. This should be investigated in greater detail to gain an understanding of how structural changes affect chromophore reconstitution.

The optimized biosynthetic pathway for reconstitution of PYP should prove useful for future work with PYP, enabling this photoreceptor to be used as a fully genetically encoded element in diverse hosts. The chromophore biosynthesis pathway is anticipated to require host-specific optimization in order to balance the enzymatic activity of TAL and 4CL. The PYP-BFP sensor will facilitate this, by enabling *in vivo* chromophore production to be monitored.

Bibliography

1. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230–233 (2015).

2. Paonessa, F. *et al.* Regulation of neural gene transcription by optogenetic inhibition of the RE1-silencing transcription factor. *Proc. Natl. Acad. Sci.* **113**, E91–E100 (2016).

3. Wu, Y. I. *et al.* A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature* **461**, 104–108 (2009).

4. 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).

5. 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).

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

7. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16**, 378–384 (2005).

8. Harper, S. M., Neil, L. C. & Gardner, K. H. Structural basis of a phototropin light switch. *Science.* **301**, 1541–1545 (2003).

9. Huala, E. *et al.* Arabidopsis NPH1: A protein kinase with a putative redox-sensing domain. *Science.* **278**, 2120–2123 (1997).

10. Ramachandran, P. L. *et al.* The short-lived signaling state of the photoactive yellow protein photoreceptor revealed by combined structural probes. *J. Am. Chem. Soc.* **133**, 9395–9404 (2011).

11. Meyer, T. E., Yakali, E., Cusanovich, M. A. & Tollin, G. Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin. *Biochemistry* **26**, 418–423 (1987).

12. Salomon, M., Christie, J. M., Knieb, E., Lempert, U. & Briggs, W. R. Photochemical and mutational analysis of the FMN-binding domains of the plant blue light receptor, phototropin. *Biochem. Biophys. Res. Commun.* **39**, 9401–9410 (2000).

13. Hoff, W. D. *et al.* Thiol ester-linked p-coumaric acid as a new photoactive prosthetic group in a protein with rhodopsin-like photochemistry. *Biochemistry* **33**, 13959–13962 (1994).

14. Kumar, A., Burns, D. C., Al-Abdul-Wahid, M. S. & Woolley, G. A. A circularly permuted photoactive yellow protein as a scaffold for photoswitch design. *Biochemistry* **52**, 3320–3331 (2013).

15. Lu, H. *et al.* A yeast system for discovering optogenetic inhibitors of eukaryotic translation initiation. *ACS Synth. Biol.* **8**, 744–757 (2019).

16. Reis, J. M. *et al.* Discovering selective binders for photoswitchable proteins using phage display. *ACS Synth. Biol.* **7**, 2355–2364 (2018).

17. Kumar, A., Ali, A. M. & Woolley, G. A. Photo-control of DNA binding by an engrailed homeodomain-photoactive yellow protein hybrid. *Photochem. Photobiol. Sci.* **14**, 1729–1736 (2015).

18. Reis, J. M., Burns, D. & Woolley, G. A. Optical control of protein-protein interactions via blue light induced domain swapping. *Biochemistry* (2014). doi:10.1021/bi500622x

19. 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).

20. Conrad, K. S., Bilwes, A. M. & Crane, B. R. Light-induced subunit dissociation by a light-oxygen-voltage domain photoreceptor from Rhodobacter sphaeroides. *Biochemistry* **52**, 378–391 (2013).

21. Möglich, A. & Moffat, K. Engineered photoreceptors as novel optogenetic tools. *Photochem. Photobiol. Sci.* **9**, 1286–1300 (2010).

22. Lungu, O. I. *et al.* Designing photoswitchable peptides using the AsLOV2 domain. *Chem. Biol.* **487**, 109–113 (2012).

23. 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 Synth. Biol.* **3**, 788–795 (2014).

24. Cunniff, B., Mckenzie, A. J., Heintz, N. H. & Alan, K. H. AMPK activity regulates trafficking of mitochondria to the leading edge during cell migration and matrix invasion. *Mol. Biol. Cell* mbc-E16 (2016).

25. Ali, A. M. *et al.* Optogenetic inhibitor of the transcription factor CREB. *Chem. Biol.* **22**, 1531–1539 (2015).

26. Renicke, C., Schuster, D., Usherenko, S., Essen, L. O. & Taxis, C. A LOV2 domain-based optogenetic tool to control protein degradation and cellular function. *Chem. Biol.* **20**, 619–626 (2013).

27. Usherenko, S. *et al.* Photo-sensitive degron variants for tuning protein stability by light. *BMC Syst. Biol.* **8**, 128 (2014).

28. Bonger, K. M., Rakhit, R., Payumo, A. Y., Chen, J. K. & Wandless, T. J. General method for regulating protein stability with light. *ACS Chem. Biol.* **9**, 111–115 (2014).

29. Müller, K., Zurbriggen, M. D. & Weber, W. An optogenetic upgrade for the Tet-OFF system. *Biotechnol. Bioeng.* **112**, 1483–1487 (2015).

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

31. Shcherbakova, D. M., Shemetov, A. A., Kaberniuk, A. A. & Verkhusha, V. V. Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools.

Annu. Rev. Biochem. **84**, 519–550 (2015).

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

33. Guntas, G. *et al.* Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proc. Natl. Acad. Sci.* **112**, 112–117 (2015).

34. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat. Commun.* **6**, 6256 (2015).

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

36. Pathak, G. P., Strickland, D., Vrana, J. D. & Tucker, C. L. Benchmarking of optical dimerizer systems. *ACS Synth. Biol.* **3**, 832–838 (2014).

37. O'Neill, P. R. & Gautam, N. Subcellular optogenetic inhibition of G proteins generates signaling gradients and cell migration. *Mol. Biol. Cell* **25**, 2305–2314 (2014).

38. Jost, A. P. T. & Weiner, O. D. Probing yeast polarity with acute, reversible, optogenetic inhibition of protein function. *ACS Synth. Biol.* **4**, 1077–1085 (2015).

39. Yang, X., Jost, A. P.-T., Weiner, O. D. & Tang, C. A light-inducible organelle-targeting system for dynamically activating and inactivating signaling in budding yeast. *Mol. Biol. Cell* **24**, 2419–2430 (2013).

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

41. Beyer, H. M. *et al.* Red light-regulated reversible nuclear localization of proteins in mammalian cells and zebrafish. *ACS Synth. Biol.* **4**, 951–958 (2015).

42. Yumerefendi, H. *et al.* Light-induced nuclear export reveals rapid dynamics of epigenetic modifications. *Nat. Chem. Biol.* **12**, 2–6 (2016).

43. Niopek, D., Wehler, P., Roensch, J., Eils, R. & Di Ventura, B. Optogenetic control of nuclear protein export. *Nat. Commun.* **7**, 10624 (2016).

44. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.* **5**, 4404 (2014).

45. Lee, S. *et al.* Reversible protein inactivation by optogenetic trapping in cells. *Nat. Methods* **11**, 633–636 (2014).

46. Taslimi, A. *et al.* An optimized optogenetic clustering tool for probing protein interaction and function. *Nat. Commun.* **5**, 4925 (2014).

47. Ohlendorf, R., Schumacher, C. H., Richter, F. & Möglich, A. Library-aided probing of linker determinants in hybrid photoreceptors. *ACS Synth. Biol.* **5**, 1117–1126 (2016).

48. Richter, F. *et al.* Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Res.* **44**, 10003–10014 (2016).

49. Tee, K. L. & Tuck, S. W. Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology Advances* **31**, 1707–1721 (2013).

50. Smith, J. M. Natural selection and the concept of a protein space. *Nature* **225**, 563–564 (1970).

51. Reetz, M. T. Controlling the enantioselectivity of enzymes by directed evolution: Practical and theoretical ramifications. *Proc. Natl. Acad. Sci.* **101**, 5716–5722 (2004).

52. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res.* **2**, 28–33 (1992).

53. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258–261 (1998).

54. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).

55. Stemmer, W. P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci.* **91**, 10747–10751 (1994).

56. Müller, K. M. *et al.* Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. *Nucleic Acids Res.* **33**, e117 (2005).

57. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).

58. Fellouse, F. A., Wiesmann, C. & Sidhu, S. S. Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci.* **101**, 12467–12472 (2004).

59. Ribeiro, L. F., Amarelle, V., Ribeiro, L. F. C. & Guazzaroni, M.-E. Converting a periplasmic binding protein into a synthetic biosensing switch through domain insertion. *Biomed Res. Int.* **2019**, 4798793 (2019).

60. Lee, J. *et al.* Surface sites for engineering allosteric control in proteins. *Science.* **322**, 438–442 (2008).

61. Guntas, G., Mitchell, S. & Ostermeier, M. A molecular switch created by in vitro recombination of nonhomologous genes. *Chem. Biol.* **11**, 1483–1487 (2004).

62. Kanwar, M., Wright, R. C., Date, A., Tullman, J. & Ostermeier, M. Protein switch engineering by domain insertion. *Methods Enzymol.* **523**, 369–388 (2013).

63. Guntas, G. & Ostermeier, M. Creation of an allosteric enzyme by domain insertion. *J. Mol. Biol.* **336**, 263–273 (2004).

64. Tullman, J., Guntas, G., Dumont, M. & Ostermeier, M. Protein switches identified from diverse insertion libraries created using S1 nuclease digestion of supercoiled-form plasmid DNA. *Biotechnol. Bioeng.* **108**, 2535–2543 (2011).

65. Jones, D. D. Triplet nucleotide removal at random positions in a target gene: The

tolerance of TEM-1 β-lactamase to an amino acid deletion. *Nucleic Acids Res.* **33**, e80 (2005).

66. Edwards, W. R., Busse, K., Allemann, R. K. & Jones, D. D. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Res.* **36**, e78 (2008).

67. Guntas, G., Mansell, T. J., Kim, J. R. & Ostermeier, M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc. Natl. Acad. Sci.* **102**, 11224–11229 (2005).

68. Jäckel, C., Kast, P. & Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **37**, 153–173 (2008).

69. Smith, G. P. & Petrenko, V. A. Phage display. *Chem. Rev.* 391–410 (1997). doi:10.1021/cr960065d

70. Lipovsek, D. & Plückthun, A. In vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods* **290**, 51–67 (2004).

71. Leemhuis, H., Kelly, R. M. & Dijkhuizen, L. Directed evolution of enzymes: Library screening strategies. *IUBMB Life* **61**, 222–228 (2009).

72. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246 (1989).

73. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623–627 (2000).

74. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci.* **95**, 5752–5756 (1998).

75. Licitra, E. J. & Liu, J. O. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc. Natl. Acad. Sci.* **93**, 12817–12821 (2002).

76. SenGupta, D. J. *et al.* A three-hybrid system to detect RNA-protein interactions in vivo. *Proc. Natl. Acad. Sci.* **93**, 8496–8501 (1996).

77. Horswill, A. R., Savinov, S. N. & Benkovic, S. J. A systematic method for identifying small-molecule modulators of protein-protein interactions. *Proc. Natl. Acad. Sci.* **101**, 15591–15596 (2004).

78. Park, S. H. & Raines, R. T. Genetic selection for dissociative inhibitors of designated protein-protein interactions. *Nat. Biotechnol.* **18**, 847–851 (2000).

79. Ried, J. L. & Collmer, A. An nptl-sacB-sacR cartridge for constructing directed, unmarked mutations in Gram-negative bacteria by marker exchange-eviction mutagenesis. *Gene* **57**, 239–246 (1987).

80. Gay, P., Le Coq, D., Steinmetz, M., Berkelman, T. & Kado, C. I. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J. Bacteriol.* **164**, 918–921 (1985).

81. Collins, C. H., Leadbetter, J. R. & Arnold, F. H. Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. *Nat. Biotechnol.* **24**, 708–712 (2006).

82. Yokobayashi, Y. & Arnold, F. H. A dual selection module for directed evolution of genetic circuits. *Nat. Comput.* **4**, 245–254 (2005).

83. Li, X., Thomason, L. C., Sawizke, J. A., Costantino, N. & Court, D. L. Positive and negative selection using tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. *Nucleic Acids Res.* **41**, e204 (2013).

84. Tominaga, M., Ike, K., Kawai-Noma, S., Saito, K. & Umeno, D. Rapid and liquid-based selection of genetic switches using nucleoside kinase fused with aminoglycoside phosphotransferase. *PLoS One* **10**, e0120243 (2015).

85. Chopra, I. & Roberts, M. Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232–260

(2001).

86. Muranaka, N., Sharma, V., Nomura, Y. & Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Res.* **37**, e39 (2009).

87. Nomura, Y. & Yokobayashi, Y. Dual selection of a genetic switch by a single selection marker. *BioSystems* **90**, 115–120 (2007).

88. Devito, J. A. Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of Escherichia coli. *Nucleic Acids Res.* **36**, e4 (2008).

89. Tashiro, Y., Fukutomi, H., Terakubo, K., Saito, K. & Umeno, D. A nucleoside kinase as a dual selector for genetic switches and circuits. *Nucleic Acids Res.* **39**, e12 (2011).

90. Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33**, e36 (2005).

91. Hoffmann, S. A., Kruse, S. M. & Arndt, K. M. Long-range transcriptional interference in E. coli used to construct a dual positive selection system for genetic switches. *Nucleic Acids Res.* gkw125 (2016). doi:10.1093/nar/gkw125

92. Sohka, T. *et al.* An externally tunable bacterial band-pass filter. *Proc. Natl. Acad. Sci.* **106**, 10135–10140 (2009).

93. Heins, R. A., Choi, J. H., Sohka, T. & Ostermeier, M. In vitro recombination of non-homologous genes can result in gene fusions that confer a switching phenotype to cells. *PLoS One* **6**, e27302 (2011).

94. Tullman, J., Nicholes, N., Dumont, M. R., Ribeiro, L. F. & Ostermeier, M. Enzymatic protein switches built from paralogous input domains. *Biotechnol. Bioeng.* **113**, 852–858 (2015).

95. Morgan, S.-A., Al-Abdul-Wahid, S. & Woolley, G. A. Structure-based design of a

photocontrolled DNA binding protein. *J. Mol. Biol.* **399**, 94–112 (2010).

96. Fan, H. Y. *et al.* Improving a designed photocontrolled DNA-binding protein. *Biochemistry* **50**, 1226–1237 (2011).

97. Ye, H., Baba, M. D., Peng, R. & Fussenegger, M. Device enhances blood-glucose homeostasis in mice. *Science.* **332**, 1565–1569 (2011).

98. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).

99. Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R. & Tabor, J. J. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods* **11**, 449–455 (2014).

100. Motta-Mena, L. B. *et al.* An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat. Chem. Biol.* **10**, 196–202 (2014).

101. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. A light-switchable gene promoter system. *Nat. Biotechnol.* **20**, 1041–1044 (2002).

102. Polstein, L. R. & Gersbach, C. A. Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. *J. Am. Chem. Soc.* **134**, 16480–16483 (2012).

103. Strickland, D., Moffat, K. & Sosnick, T. R. Light-activated DNA binding in a designed allosteric protein. *Proc. Natl. Acad. Sci.* **105**, 10709–10714 (2008).

104. Carey, M. F., Peterson, C. L. & Smale, S. T. Confirming the functional importance of a protein – DNA interaction. *Cold Spring Harb. Protoc.* 733–758 (2012). doi:10.1101/pdb.top070060

105. Carey, M. F., Peterson, C. L. & Smale, S. T. Experimental strategies for the identification of DNA-binding proteins. *Cold Spring Harb. Protoc.* 18–34 (2012). doi:10.1101/pdb.top067470

106. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).

107. Calvin, N. M. & Hayakawa, P. C. High-efficiency transformation of bacterial cells by electroporation. *J. Bacteriol.* **170**, 2796–2801 (1988).

108. Lutz, R., Lozinski, T., Ellinger, T. & Bujard, H. Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Res.* **29**, 3873–3881 (2001).

109. Bagh, S., Mandal, M. & McMillen, D. R. Minimal genetic device with multiple tunable functions. *Phys. Rev. E* **82**, 021911 (2010).

110. Abana, C. M. *et al.* Characterization of blue light irradiation effects on pathogenic and nonpathogenic Escherichia coli. *Microbiol. Open* **6**, e466 (2017).

111. Adler, H. I. & Hardigree, A. A. Postirradiation growth, division, and recovery in bacteria. *Radiat. Res.* **25**, 92–102 (1965).

112. Aoust, J. Y. D., Giroux, J., Barran, L. R., Schneider, H. & Martin, W. G. Some effects of visible light on Escherichia coli. *J. Bacteriol.* **120**, 799–804 (1974).

113. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

114. Tonikian, R., Zhang, Y., Boone, C. & Sidhu, S. S. Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat. Protoc.* **2**, 1368–1386 (2007).

115. Nooren, I. M. A. & Thornton, J. M. Diversity of protein-protein interactions. *EMBO J.* **22**, 3486–3492 (2003).

116. Norris, A. & Boeke, J. D. Silent information regulator 3: The Goldilocks of the silencing complex. *Genes Dev.* **24**, 115–122 (2010).

117. Slansky, J. E. & Jordan, K. R. The Goldilocks model for TCR - Too much attraction might not be best for vaccine design. *PLoS Biol.* **8**, e1000482 (2010).

118. Lever, M. *et al.* Architecture of a minimal signaling pathway explains the T-cell response to a 1 million-fold variation in antigen affinity and dose. *Proc. Natl. Acad. Sci.* **114**, E267–E267 (2017).

119. Rodriguez-Martinez, J. A., Reinke Aaron W., Bhimsaria Devesh, Keating, A. E. & Ansari, A. Z. Combinatorial bZIP dimers define complex DNA-binding specificity landscapes. *Elife* 1–29 (2017). doi:10.7554/eLife.19272

120. Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G. & Orengo, C. Transient protein-protein interactions: structural, functional, and network properties. *Structure* **18**, 1233–1243 (2010).

121. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2011).

122. Lage, K. Protein – protein interactions and genetic diseases: The interactome. *Biochim. Biophys. Acta* **1842**, 1971–1980 (2014).

123. Ryan, D. P. & Matthews, J. M. Protein – protein interactions in human disease. *Curr. Opin. Struct. Biol.* **15**, 441–446 (2005).

124. Thaxton, J. E. & Li, Z. To affinity and beyond: Harnessing the T cell receptor for cancer immunotherapy. *Hum. Vaccines Immunother.* **10**, 3313–3321 (2014).

125. Pakulska, M. M., Miersch, S. & Shoichet, M. S. Designer protein delivery: From natural to engineered affinity-controlled release systems. *Science.* **351**, (2016).

126. Estojak, J., Brent, R. & Golemis, E. A. Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol.* **15**, 5820–5829 (1995).

127. Younger, D., Berger, S., Baker, D. & Klavins, E. High-throughput characterization of protein – protein interactions by reprogramming yeast mating. *Proc. Natl. Acad. Sci.* **114**,

12166–12171 (2017).

128. Vanantwerp, J. J. & Wittrup, K. D. Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol. Prog.* **16**, 31–37 (2000).

129. Reich, L. L., Dutta, S. & Keating, A. E. SORTCERY — A high-throughput method to affinity rank peptide ligands. *J. Mol. Biol.* **427**, 2135–2150 (2015).

130. Strauch, E. & Georgiou, G. A bacterial two-hybrid system based on the twin-arginine transporter pathway of E. coli. *Protein Sci.* **16**, 1001–1008 (2007).

131. Waraho, D. & DeLisa, M. P. Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proc. Natl. Acad. Sci.* **106**, 3692–3697 (2009).

132. Berks, B. C. The twin-arginine protein translocation pathway. *Annu. Rev. Biochem.* **84**, 843–864 (2015).

133. Patel, R., Smith, S. M. & Robinson, C. Protein transport by the bacterial Tat pathway. *Biochim. Biophys. Acta* **1843**, 1620–1628 (2014).

134. Rodrigue, A. *et al.* Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway. *J. Biol. Chem.* **274**, 13223–13228 (1999).

135. Zeng, X. & Lin, J. Beta-lactamase induction and cell wall metabolism in Gram-negative bacteria. *Front. Microbiol.* **4**, 1–9 (2013).

136. Dietz, H., Pfeifle, D. & Wiedemann, B. The signal molecule for β-lactamase induction in Enterobacter cloacae is the anhydromuramyl-pentapeptide. *Antimicrob. Agents Chemother.* **41**, 2113–2120 (1997).

137. Lindberg, F., Westman, L. & Normark, S. Regulatory components in Citrobacter freundii ampC beta-lactamase induction. *Proc. Natl. Acad. Sci.* **82**, 4620–4624 (1985).

138. Vadlamani, G. *et al.* The β-lactamase gene regulator AmpR is a tetramer that recognizes and binds the D-Ala-D-Ala motif of its repressor UDP-N-acetylmuramic acid (MurNAc)-

pentapeptide. *J. Biol. Chem.* **290**, 2630–2643 (2015).

139. Lindberg, F., Lindquist, S. & Normark, S. Inactivation of the ampD gene causes semiconstitutive overproduction of the inducible Citrobacter freundii beta-lactamase. *J. Bacteriol.* **169**, 1923–1928 (1987).

140. Normark, S. & Burman, L. G. Resistance of Escherichia coli to penicillins: fine-structure mapping and dominance of chromosomal beta-lactamase mutations. *J. Bacteriol.* **132**, 1–7 (1977).

141. Valtonen, S. J., Kurittu, J. S. & Karp, M. T. A luminescent Escherichia coli biosensor for the high throughput detection of β-lactams. *J. Biomol. Screen.* **7**, 127–134 (2002).

142. Zhang, Y., Werling, U. & Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* **40**, e55 (2012).

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

144. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).

145. Speck, J. *et al.* TAT hitchhiker selection expanded to folding helpers, multimeric interactions and combinations with protein fragment complementation. *Protein Eng. Des. Sel.* **26**, 225–242 (2012).

146. Arndt, K. M. *et al.* A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. *J. Mol. Biol.* **295**, 627–639 (2000).

147. Arndt, K. M., Pelletier, J. N., Müller, K. M., Plückthun, A. & Alber, T. Comparison of in vivo selection and rational design of heterodimeric coiled coils. *Structure* **10**, 1235–1248 (2002).

148. Mason, J. M., Schmitz, M. A., Müller, K. M. & Arndt, K. M. Semirational design of Jun-

Fos coiled coils with increased affinity: Universal implications for leucine zipper prediction and design. *Proc. Natl. Acad. Sci.* **103**, 8989–8994 (2006).

149. Olive, M. *et al.* A dominant negative to activation protein-1 (AP1) that abolishes DNA binding and inhibits oncogenesis. *J. Biol. Chem.* **272**, 18586–18594 (1997).

150. Ahn, S. *et al.* A dominant-negative inhibitor of CREB reveals that it is a general mediator of stimulus-dependent transcription of c-Fos. *Mol. Cell. Biol.* **18**, 967–977 (1998).

151. Subach, O. M., Cranfill, P. J., Davidson, M. W. & Verkhusha, V. V. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS One* **6**, e28674 (2011).

152. Drummond, A. *et al.* Geneious v5.0. (2010).

153. Waraho-Zhmayev, D., Meksiriporn, B., Portnoff, A. D., DeLisa, M. P. & Bradbury, A. Optimizing recombinant antibodies for intracellular function using hitchhiker-mediated survival selection. *Protein Eng. Des. Sel.* **27**, 351–358 (2014).

154. Kaplan, J. B., Reinke, A. W. & Keating, A. E. Increasing the affinity of selective bZIP binding peptides through surface residue redesign. *Protein Sci.* **23**, 940–953 (2014).

155. Mason, J. M., Müller, K. M. & Arndt, K. M. Considerations in the design and optimization of coiled coil structures. in *Protein Engineering Protocols. Methods in Molecular Biology, vol 352* (eds. Arndt, K. M. & Müller, K. M.) 35–70 (Humana Press, 2007). doi:10.1385/1-59745-187-8:35

156. Sohka, T., Heins, R. A. & Ostermeier, M. Morphogen-defined patterning of Escherichia coli enabled by an externally tunable band-pass filter. *J. Biol. Eng.* **3**, 10 (2009).

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

158. Hagemann, U. B., Mason, J. M., Müller, K. M. & Arndt, K. M. Selectional and mutational scope of peptides sequestering the Jun-Fos coiled-coil domain. *J. Mol. Biol.* **381**, 73–88 (2008).

159. Hess, J., Angel, P. & Schorpp-Kistner, M. AP-1 subunits: quarrel and harmony among siblings. *J. Cell Sci.* **117**, 5965–5973 (2004).

160. Fisher, A. C., Kim, W. & DeLisa, M. P. Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway. *Protein Sci.* **15**, 449–458 (2006).

161. Krylov, D., Olive, M. & Vinson, C. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. *EMBO J.* **14**, 5329–5337 (1995).

162. Mason, J. M. & Arndt, K. M. Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem* **5**, 170–176 (2004).

163. Moitra, J., Krylov, D. & Vinson, C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry* **36**, 12567–12573 (1997).

164. Schenk, R. L., Strasser, A. & Dewson, G. BCL-2: Long and winding path from discovery to therapeutic target. *Biochem. Biophys. Res. Commun.* **482**, 459–469 (2017).

165. Ashkenazi, A., Fairbrother, W. J., Leverson, J. D. & Souers, A. J. From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors. *Nat. Rev. Drug Discov.* **16**, 273–284 (2017).

166. Tsujimoto, Y., Finger, L. R., Yunis, J., Nowell, P. C. & Croce, C. C. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science.* **226**, 1097–1099 (1984).

167. Graninger, W. B., Seto, M., Boutain, B., Goldman, P. & Korsmeyer, S. J. Expression of Bcl-2 and Bcl-2-Ig fusion transcripts in normal and neoplastic cells. *J. Clin. Invest.* **80**, 1512–1515 (1987).

168. Vaux, D. L., Cory, S. & Adams, T. M. Bcl-2 promotes the survival of haemotopoeitic cells and cooperates with c-myc to immortalize pre-B cells. *Nature* **258**, 1955–1957. (1988).

169. Ku, B., Liang, C., Jung, J. U. & Oh, B.-H. Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Res.* **21**, 627–641 (2011).

170. Souers, A. J. *et al.* ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **19**, 202–208 (2013).

171. Tse, C. *et al.* ABT-263: A potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res.* **68**, 3421–3428 (2008).

172. Nguyen, M. *et al.* Small molecule obatoclax (GX15-070) antagonizes MCL-1 and overcomes MCL-1-mediated resistance to apoptosis. *Proc. Natl. Acad. Sci.* **104**, 19512–19517 (2007).

173. FDA approves venetoclax in combination for AML in adults. *U.S. Food & Drug Administration* (2018). Available at: https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm626499.htm. (Accessed: 2nd March 2019)

174. Yamamoto, K., Ichijo, H. & Korsmeyer, S. J. BCL-2 is phosphorylated and inactivated by an ASK1/Jun N-terminal protein kinase pathway normally activated at G2 /M. *Mol. Cell. Biol.* **19**, 8469–8478 (2015).

175. Chao, D. T. *et al.* Bcl-XL and Bcl-2 repress a common pathway of cell death. *J. Exp. Med.* **182**, 821–828 (1995).

176. Petros, A. M. *et al.* Solution structure of the antiapoptotic protein Bcl-2. *Proc. Natl. Acad. Sci.* **98**, 3012–3017 (2001).

177. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97**, 6640–5 (2000).

178. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **4474**, (2006).

179. Mohammad, R. M. *et al.* Preclinical studies of TW-37, a new nonpeptidic small-molecule inhibitor of Bcl-2, in diffuse large cell lymphoma xenograft model reveal drug action on both Bcl-2 and Mcl-1. *Clin. Cancer Res.* **13**, 2226–2235 (2007).

180. Becattini, B. *et al.* Rational design and real time, in-cell detection of the proapoptotic activity of a novel compound targeting Bcl-XL. *Chem. Biol.* **11**, 389–395 (2004).

181. Han, X. *et al.* Structure-based discovery of an organic compound that binds Bcl-2 protein and induces apoptosis of tumor cells. *Proc. Natl. Acad. Sci.* **97**, 7124–7129 (2002).

182. Lewis, K. Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.* **12**, 371–387 (2013).

183. Mabekou, S. S., Lee, S. C., Dinh, T. H., Won, K. & Mitchell, R. J. Enhanced sensitivity and responses to viologens from a whole-cell bacterial bioreporter treated with branched polyethyleneimines. *J. Appl. Microbiol.* **123**, 1478–1487 (2017).

184. Ames, B. N., Lee, F. D. & Durston, W. E. An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc. Natl. Acad. Sci.* **70**, 782–786 (1973).

185. Madigan, M. & Martinko, J. *Brock, Biology of Microorganisms*. (Pearson Prentice Hall, 2006).

186. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (1996).

187. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim. Biophys. Acta - Biomembr.* **1794**, 808–816 (2009).

188. Masi, M., Réfregiers, M., Pos, K. M. & Pagès, J. M. Mechanisms of envelope permeability and antibiotic influx and efflux in Gram-negative bacteria. *Nat. Microbiol.* **2**, (2017).

189. Koronakis, V., Eswaran, J. & Hughes, C. Structure and function of TolC: The bacterial exit duct for proteins and drugs. *Annu. Rev. Biochem.* **73**, 467–489 (2004).

190. Zgurskaya, H. I., Rybenkov, V. V., Krishnamoorthy, G. & Leus, I. V. Trans-envelope multidrug efflux pumps of Gram-negative bacteria and their synergism with the outer membrane barrier. *Res. Microbiol.* **169**, 351–356 (2018).

191. Sulavik, M. C. *et al.* Antibiotic susceptibility profiles of Escherichia coli strains lacking multidrug efflux pump genes. *Antimicrob. Agents Chemother.* **45**, 1126–1136 (2001).

192. Hsieh, P.-C., Siegel, S. A., Rogers, B., Davis, D. & Lewis, K. Bacteria lacking a multidrug pump: A sensitive tool for drug discovery. *Proc. Natl. Acad. Sci.* **95**, 6602–6606 (1998).

193. Coldham, N. G., Webber, M., Woodward, M. J. & Piddock, L. J. V. A 96-well plate fluorescence assay for assessment of cellular permeability and active efflux in Salmonella enterica serovar Typhimurium and Escherichia coli. *J. Antimicrob. Chemother.* **65**, 1655–1663 (2010).

194. Dhamdhere, G. & Zgurskaya, H. I. Metabolic shutdown in Escherichia coli cells lacking the outer membrane channel TolC. *Mol. Microbiol.* **77**, 743–754 (2010).

195. Dixon, R. A. & Chopra, I. Leakage of periplasmic proteins from Escherichia coli mediated by polymyxin B nonapeptide. *Antimicrob. Agents Chemother.* **29**, 781–788 (1986).

196. Krishnamoorthy, G. *et al.* Breaking the permeability barrier of Escherichia coli by controlled hyperporination of the outer membrane. *Antimicrob. Agents Chemother.* **60**, 7372–7381 (2016).

197. Vaara, M. Agents that increase the permeability of the outer membrane. *Microbiol. Rev.* **56**, 395–411 (1992).

198. Morgan, S.-A. & Woolley, G. A. A photoswitchable DNA-binding protein based on a truncated GCN4-photoactive yellow protein chimera. *Photochem. Photobiol. Sci.* **9**, 1320–1326 (2010).

199. Cosentino, C. *et al.* Engineering of a light-gated potassium channel. *Science.* **348**, 707–710 (2015).

200. Shaulian, E. & Karin, M. AP-1 as a regulator of cell life and death. *Nat. Cell Biol.* **4**, E131-6 (2002).

201. Tewari, D. *et al.* Targeting activator protein 1 signaling pathway by bioactive natural agents: Possible therapeutic strategy for cancer prevention and intervention. *Pharmacol. Res.* **128**, 366–375 (2018).

202. Shaulian, E. AP-1 - The Jun proteins: Oncogenes or tumor suppressors in disguise? *Cell. Signal.* **22**, 894–899 (2010).

203. Carlezon, W. A., Duman, R. S. & Nestler, E. J. The many faces of CREB. *Trends Neurosci.* **28**, 436–445 (2005).

204. Yin J.C. & Tully T. CREB and the formation of long-term memory. *Curr. Opin. Neurobiol.* **6**, 264–268 (1996).

205. Han, J.-H. *et al.* Selective erasure of a fear memory. *Science.* **323**, 1492–1496 (2009).

206. Barco, A., Pittenger, C. & Kandel, E. R. CREB, memory enhancement and the treatment of memory disorders: promises, pitfalls and prospects. *Expert Opin. Ther. Targets* **7**, 101–114 (2003).

207. Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nat. Rev. Mol. Cell Biol.* **2**, 599–609 (2001).

208. Saura, C. A. & Cardinaux, J. R. Emerging roles of CREB-regulated transcription coactivators in brain physiology and pathology. *Trends Neurosci.* **40**, 720–733 (2017).

209. Fonseca, G. J. *et al.* Diverse motif ensembles specify non-redundant DNA binding activities of AP-1 family members in macrophages. *Nat. Commun.* **10**, (2019).

210. Wallace, T. L., Stellitano, K. E., Neve, R. L. & Duman, R. S. Effects of cyclic adenosine monophosphate response element binding protein overexpression in the basolateral amygdala on behavioral models of depression and anxiety. *Biol. Psychiatry* **56**, 151–160 (2004).

211. Nestler, E. J. *et al.* Long-term memory is facilitated by cAMP response element-binding protein overexpression in the amygdala. *J. Neurosci.* **21**, 2404–2412 (2001).

212. Han, J.-H. *et al.* Neuronal competition and selection during memory formation. *Science.* **316**, 457–460 (2007).

213. Glover, J. N. & Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373**, 257–261 (1995).

214. Swartz, T. E. *et al.* The photocycle of a flavin-binding domain of the blue light photoreceptor phototropin. *J. Biol. Chem.* **276**, 36493–36500 (2001).

215. Kottke, T., Xie, A., Larsen, D. S. & Hoff, W. D. Photoreceptors take charge: Emerging principles for light sensing. *Annu. Rev. Biophys.* **47**, 291–313 (2018).

216. 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).

217. Zayner, J. P., Mathes, T., Sosnick, T. R. & Kennis, J. T. M. Helical contributions mediate light-activated conformational change in the LOV2 domain of Avena sativa phototropin 1. *ACS Omega* **4**, 1238–1243 (2019).

218. Strickland, D. *et al.* Rationally improving LOV domain-based photoswitches. *Nat. Methods* **7**, 623–626 (2010).

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

220. Imamoto, Y., Ito, T., Kataoka, M. & Tokunaga, F. Reconstitution photoactive yellow protein from apoprotein and p-coumaric acid derivatives. *FEBS Lett.* **374**, 157–160 (1995).

221. Philip, A. F., Kumauchi, M. & Hoff, W. D. Robustness and evolvability in the functional anatomy of a PER-ARNT-SIM (PAS) domain. *Proc. Natl. Acad. Sci.* **107**, 17986–17991

(2010).

222. Genick, U. K. *et al.* Active site mutants implicate key residues for control of color and light cycle kinetics of photoactive yellow protein. *Biochemistry* **36**, 8–14 (1997).

223. Imamoto, Y. *et al.* Diverse roles of glycine residues conserved in photoactive yellow proteins. *Biophys. J.* **94**, 3620–3628 (2008).

224. Kumauchi, M., Hara, M. T., Stalcup, P., Xie, A. & Hoff, W. D. Identification of six new photoactive yellow proteins - diversity and structure-function relationships in a bacterial blue light photoreceptor. *Photochem. Photobiol.* **84**, 956–969 (2008).

225. Mihara, K., Hisatomi, O., Imamoto, Y., Kataoka, M. & Tokunaga, F. Functional expression and site-directed mutagenesis of photoactive yellow protein. *J. Biochem* **121**, 876–880 (1997).

226. Getzoff, E. D., Gutwin, K. N. & Genick, U. K. Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation. *Nat. Struct. Biol.* **10**, 663–668 (2003).

227. Mart, R. J., Meah, D. & Allemann, R. K. Photocontrolled exposure of pro-apoptotic peptide sequences in LOV proteins modulates Bcl-2 family interactions. *ChemBioChem* **17**, 698–701 (2016).

228. Denault, M. & Pelletier, J. N. Protein library design and screening. in *Protein Engineering Protocols. Methods in Molecular Biology, vol 352* (eds. Arndt, K. M. & Müller, K. M.) 127–154 (Humana Press, 2007).

229. Ribeiro, L. F. *et al.* Insertion of a xylanase in xylose binding protein results in a xylose-stimulated xylanase. *Biotechnol. Biofuels* **8**, 118 (2015).

230. Kumar, A., Nokhrin, S., Woloschuk, R. M. & Woolley, G. A. Duplication of a single strand in a β-sheet can produce a new switching function in a photosensory protein. *Biochemistry* **57**, 4093–4104 (2018).

231. Mazumder, M. *et al.* An Escherichia coli system for evolving improved light-controlled DNA-binding proteins. *Protein Eng. Des. Sel.* **28**, 293–302 (2015).

232. Reis, J. M. & Woolley, G. A. Photo control of protein function using photoactive yellow protein. **1408**, 79–92 (2016).

233. Ui, M. *et al.* Application of photoactive yellow protein as a photoresponsive module for controlling hemolytic activity of staphylococcal α-hemolysin. *Chem. Commun.* **48**, 4737–4739 (2012).

234. Kyndt, J. A. *et al.* Rhodobacter capsulatus photoactive yellow protein: genetic context, spectral and kinetics characterization, and mutagenesis. *Biochemistry* **43**, 1809–1820 (2004).

235. Kyndt, J. A. *et al.* Heterologous production of Halorhodospira halophila holo-photoactive yellow protein through tandem expression of the postulated biosynthetic genes. *Biochemistry* **42**, 965–970 (2003).

236. Meyer, T. E. *et al.* The growing family of photoactive yellow proteins and their presumed functional roles. *Photochem. Photobiol. Sci.* **11**, 1495–1514 (2012).

237. Kort, R. *et al.* The xanthopsins: a new family of eubacterial blue-light photoreceptors. *EMBO J.* **15**, 3209–3218 (1996).

238. Kyndt, J., Meyer, T. E., Cusanovich, M. & Van Beeumen, J. J. Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. *FEBS Lett.* **512**, 240–244 (2002).

239. Xue, Z., McCluskey, M., Cantera, K., Sariaslani, F. S. & Huang, L. Identification, characterization and functional expression of a tyrosine ammonia-lyase and its mutants from the photosynthetic bacterium Rhodobacter sphaeroides. *J. Ind. Microbiol. Biotechnol.* **34**, 599–604 (2007).

240. Li, Z. & Nair, S. K. Structural basis for specificity and flexibility in a plant 4-coumarate:CoA ligase. *Structure* **23**, 2032–2042 (2015).

241. Hamberger, B. & Hahlbrock, K. The 4-coumarate:CoA ligase gene family in Arabidopsis thaliana comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc. Natl. Acad. Sci.* **101**, 2209–2214 (2004).

242. Dixon, R. A., Lamb, C. J., Masoud, S., Sewalt, V. J. H. & Paiva, N. L. Metabolic engineering: Prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses - A review. *Gene* **179**, 61–71 (1996).

243. Jendresen, C. B. *et al.* Highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **81**, 4458–4476 (2015).

244. Wang, S., Zhang, S., Zhou, T., Zeng, J. & Zhan, J. Design and application of an in vivo reporter assay for phenylalanine ammonia-lyase. *Appl. Microbiol. Biotechnol.* **97**, 7877–7885 (2013).

245. Xiong, D. *et al.* Improving key enzyme activity in phenylpropanoid pathway with a designed biosensor. *Metab. Eng.* **40**, 115–123 (2017).

246. Alberstein, M., Eisenstein, M. & Abeliovich, H. Removing allosteric feedback inhibition of tomato 4-coumarate:CoA ligase by directed evolution. *Plant J.* **69**, 57–69 (2012).

247. Zhang, Y. *et al.* Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and mammalian cells. *J. Am. Chem. Soc.* **128**, 13030–13031 (2006).

248. Watts, K. T., Lee, P. C. & Schmidt-Dannert, C. Exploring recombinant flavonoid biosynthesis in metabolically engineered Escherichia coli. *Chembiochem* **5**, 500–507 (2004).

249. Santos, C. N. S., Koffas, M. & Stephanopoulos, G. Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng.* **13**, 392–400 (2011).

250. Juminaga, D. *et al.* Modular engineering of L-tyrosine production in Escherichia coli. *Appl. Environ. Microbiol.* **78**, 89–98 (2012).

251. Kyndt, J. A., Fitch, J. C., Meyer, T. E. & Cusanovich, M. A. Thermochromatium tepidum photoactive yellow protein/bacteriophytochrome/ diguanylate cyclase: Characterization of the PYP domain. *Biochemistry* **44**, 4755–4764 (2005).

252. Kyndt, J. A., Fitch, J. C., Meyer, T. E. & Cusanovich, M. A. The photoactivated PYP domain of Rhodospirillum centenum Ppr accelerates the recovery of the bacteriophytochrome domain after white light illumination. *Biochemistry* **46**, 8256–8262 (2007).

253. Kamikubo, H. *et al.* The photoreaction of the photoactive yellow protein domain in the light sensor histidine kinase Ppr is influenced by the C-terminal domains. *Photochem. Photobiol.* **84**, 895–902 (2008).

254. Devanathan, S. *et al.* Preparation and properties of a 3,4-dihydroxycinnamic acid chromophore variant of the photoactive yellow protein. *Arch. Biochem. Biophys.* **340**, 83–89 (1997).

255. Changenet-Barret, P. *et al.* Excited-state relaxation dynamics of a PYP chromophore model in solution: Influence of the thioester group. *Chem. Phys. Lett.* **365**, 285–291 (2002).

256. Stöckigt, J. & Zenk, M. H. Chemical syntheses and properties of hydroxycinnamoyl-coenzyme A derivatives. *Zeitschrift für Naturforsch.* **30 c**, 352–358 (1975).

257. Meyer, T. E., Tollin, G., Causgrove, T. P., Cheng, P. & Blankenshipt, R. E. Picosecond decay kinetics and quantum yield of fluorescence of the photoactive yellow protein from the halophilic purple phototrophic bacterium Ectothiorhodospira halophila. *Biophys. J.* **59**, 988–991 (1991).

258. Harigai, M. *et al.* Amino acids in the N-terminal region regulate the photocycle of photoactive yellow protein. *J. Biochem.* **130**, 51–56 (2012).

259. van der Horst, M. A., van Stokkum, I. H., Crielaard, W. & Hellingwerf, K. J. The role of the N-terminal domain of photoactive yellow protein in the transient partial unfolding

during signalling state formation. *FEBS Lett.* **497**, 26–30 (2001).

260. Hendriks, J. *et al.* Transient exposure of hydrophobic surface in the photoactive yellow protein monitored with Nile Red. *Biophys. J.* **82**, 1632–1643 (2002).

261. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39–46 (2001).

262. Kroon, A. R. *et al.* Spectral tuning, fluorescence, and photoactivity in hybrids of photoactive yellow protein, reconstituted with native or modified chromophores. *J. Biol. Chem.* **271**, 31949–31956 (1996).

263. Giljanović, J. & Prkić, A. Determination of Coenzyme A (CoASH) in the presence of different thiols by using flow-injection with a UV/vis spectrophotometric detector and potentiometric determination of CoASH using an iodide ISE. *Molecules* **15**, 100–113 (2010).

264. Bracher, P. J., Snyder, P. W., Bohall, B. R. & Whitesides, G. M. The relative rates of thiol-thioester exchange and hydrolysis for alkyl and aryl thioalkanoates in water. *Orig. Life Evol. Biosph.* **41**, 399–412 (2011).

265. Meyer, T. E. Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototropic bacterium Ectothiorhodospira halophila. *Biochim. Biophys. Acta* **806**, 175–183 (1985).

266. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **43**, W580–W584 (2015).

267. Remmert, M. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2011).

268. Uludag, M. *et al.* Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600 (2013).

269. van der Horst, M. A. *et al.* From primary photochemistry to biological function in the

blue-light photoreceptors PYP and AppA. *Photochem. Photobiol. Sci.* **4**, 688–693 (2005).

270. Borgstahl, G. E., Williams, D. R. & Getzoff, E. D. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* **34**, 6278–6287 (1995).

271. Rubinstenn, G. *et al.* Structural and dynamic changes of photoactive yellow protein during its photocycle in solution. *Nat. Struct. Biol.* **5**, 568–570 (1998).

272. Kamikawa, Y. *et al.* Design of a protein tag and fluorogenic probe with modular structure for live-cell imaging of intracellular proteins. *Chem. Sci.* **7**, 308–314 (2015).

273. Kyriakakis, P. *et al.* Biosynthesis of orthogonal molecules using ferredoxin and ferredoxin-NADP+ reductase systems enables genetically encoded PhyB optogenetics. *ACS Synth. Biol.* **7**, 706–717 (2018).

274. Vulic, K., Pakulska, M. M., Sonthalia, R., Ramachandran, A. & Shoichet, M. S. Mathematical model accurately predicts protein release from an affinity-based delivery system. *J. Control. Release* **197**, 69–77 (2015).

275. Levine, I. N. *Physical Chemistry*. (McGraw Hill, 1988).

8 Appendices

Appendix I. List of plasmids used in Chapter 3

Table 8-1: Plasmids used in Chapter 3.

Plasmid name	Description
Bandpass vectors:	
pTS1-GFP	From Sohka *et al.,*[92] contains GFP marker
pTS1-mCherry	mCherry marker
pTS1-Topaz	Topaz marker
Hitchhiker vectors:	
pACYCDuet-tac-ssTorA-WinZipA2, WinZipA2-βla	WinZipA2 homodimer
pACYCDuet-tac-ssTorA-WinZipB1, WinZipB1-βla	WinZipB1 homodimer
pACYCDuet-tac-ssTorA-WinZipA2, WinZipB1-βla	WinZipA2/WinZipB1 heterodimer
pACYCDuet-tac-ssTorA-cJun, cJun-βla	Jun homodimer leucine zipper, 4.5 heptad repeats
pACYCDuet-tac-ssTorA-cFos, cJun-βla	Fos/Jun heterodimer leucine zipper, 4.5 heptad repeats
pACYCDuet-tac-ssTorA-cJun(bZIP), cJun(bZIP)-βla	bZIP Jun homodimer
pACYCDuet-tac-ssTorA-cFos(bZIP), cJun(bZIP)-βla	bZIP Fos/ bZIP Jun heterodimer
pACYCDuet-tac-ssTorA-cFos(ZIP), cJun(bZIP)-βla	ZIP Fos/ bZIP Jun heterodimer
pACYCDuet-tac-ssTorA-CREB(bZIP), CREB(ZIP)-βla	bZIP CREB/ ZIP CREB heterodimer
Inhibitor vectors:	
pET24b-tac-A4-CREB-BFP	4 heptads of acidic extension on the CREB leucine zipper
pET24b-tac-A4-CREB-(2xGly *f*)-BFP	Gly mutations in *f* position (N318G, E325G) of the leucine zipper of A4-CREB-BFP
pET24b-tac-A4-CREB-(2xGly *d*)-BFP	Gly mutations in *d* position (L316G, L323G) of the leucine zipper of A4-CREB-BFP
pET24b-tac-A4-Fos-BFP	Fos leucine zipper with 4 heptad repeats of the acidic extension
pET24b-tac-A2-Fos-BFP	Fos leucine zipper with 2 heptad repeats of the acidic extension
pET24b-tac-A0-Fos-BFP	Fos leucine zipper
pET24b-tac-BFP	Inhibitor negative control

Appendix II. Proteins used in Chapter 3

WinZip hitchhiker peptides:

ssTorA-WinZipA2

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAGSVAQLRERVKTLRAQNYELESEVQRLREQVAQLEFHHHHH*

ssTorA-WinZipB1

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAGSVDELQAEVDQLQDENYALKTKVAQLRKKVEKLEFHHHHH*

WinZipA2-BLA

MVAQLRERVKTLRAQNYELESEVQRLREQVAQLGAPSGSSGGSGGSS (+ β-lactamase with His tag)

WinZipB1-BLA

MVDELQAEVDQLQDENYALKTKVAQLRKKVEKLGAPSGSSGGSGGSS (+ β-lactamase with His tag)

Jun/Fos hitchhiker peptides and inhibitors:

ssTorA-cJun

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAASIARLEEKVKTLKAQNYELASTANMLREQVAQLGAP*

ssTorA-cFos

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAASTDTLQAETDQLEDEKYALQTEIANLLKEKEKLGAP*

cJun-BLA

MASIARLEEKVKTLKAQNYELASTANMLREQVAQLGAPSGSSGGSGGSS (+ β-lactamase with His tag)

ssTorA-cFos(bZIP)

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAGSKVEQLSPEEEEKRRIRRERNKMAAAKCRNRRRELTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPDEFHHHHH*

TorA-cJun(bZIP)

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAGSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTFEFHHHHH*

ssTorA-cFos(ZIP)

MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAGSTDTLQAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPDEFHHHHH*

cJun(bZIP)-BLA

MSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTFGAPSGSSGGSGGSS (+ β-lactamase with His tag)

A0-Fos-BFP

MAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPDGAPSGSSGGSGGSS (+ BFP)

A2-Fos-BFP
MLEKEAEELEQELAELEAETDQLEDEKSALQTEIANLLKEKEKLEFILAAHRPACKIPDGAPSG
SSGGSGGSS (+ BFP)
A4-Fos-BFP
MLEQRAEELARENEELEKEAEELEQELAELEAETDQLEDEKSALQTEIANLLKEKEKLEFILAA
HRPACKIPDGAPSGSSGGSGGSS (+ BFP)

CREB hitchhiker peptides and inhibitors:
ssTorA-CREB(bZIP)
MNNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATAGSKREVRLMKNREAARESRRKKKEY
VKSLENRVAVLENQNKTLIEELKALKDLEFHHHHH*
CREB(ZIP)-BLA
MASVKSLENRVAVLENQNKTLIEELKALKDLGAPSGSSGGSGGSS (+ β-lactamase
with His tag)

A4-CREB(2x Gly f)-BFP
MLEQQLEELAQENEELEKEAEELEQELAELENRVAVLEGQNKTLIGELKALKDLYSHKSDAGAP
SGSSGGSGGSS (+ BFP)
A4-CREB(2x Gly d)-BFP
MLEQQLEELAQENEELEKEAEELEQELAELENRVAVGENQNKTGIEELKALKDLYSHKSDAGAP
SGSSGGSGGSS (+ BFP)
A4-CREB-BFP
MLEQQLEELAQENEELEKEAEELEQELAELENRVAVLENQNKTLIEELKALKDLYSHKSDAGAP
SGSSGGSGGSS (+ BFP)

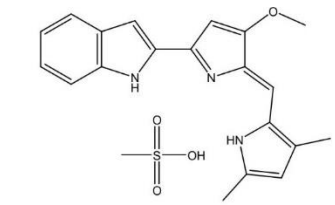
Fluorescent markers:
Topaz
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTF
GYGLQCFARYPDHMRQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSYQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK*
mCherry
MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWD
ILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKVK
LRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPV
QLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK*
GFP (GFPmut2)
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTF
AYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGI
DFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGP
VLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK*

mTagBFP2-C222S (as a C-terminal fusion to a peptide inhibitor)
SELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILATSFLYG
SKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFTSN
GPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANAKTTYRSKKPAKNLKMPGVY
YVDYRLERIKEANNETYVEQHEVAVARYSDLPSKLGHKLN*

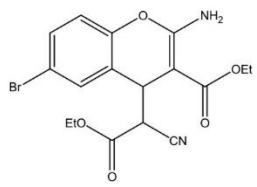
β-lactamase (as a C-terminal fusion with a hitchhiker peptide)
HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQE
QLGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLH
NMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQQLIDWMEADKVA
GPLLRSALPAGWFIADKSGAGERGSRGIIAALGPDGKPSRIVVIYTTGSQATMDERNRQIAEIG
ASLIKHWGGHHHHH*

Figure 8-1: Protein sequences used in Chapter 3.
The ssTorA sequence is shown in grey. In the bZIP proteins, the leucine zipper is shown in yellow, the basic domain is shown in green, and acidic extension is shown in blue. Mutations in the CREB inhibitors are shown in red. The CREB peptides have Cys to Ser mutations from previous work;[25] these are indicated with an underline. Stop codons are indicated with a star.

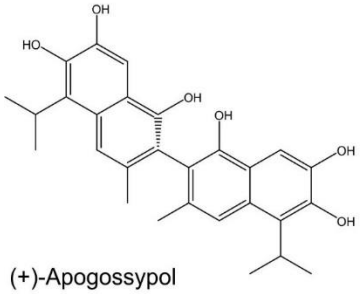
Appendix III. Bcl-2 inhibitors used in Chapter 4



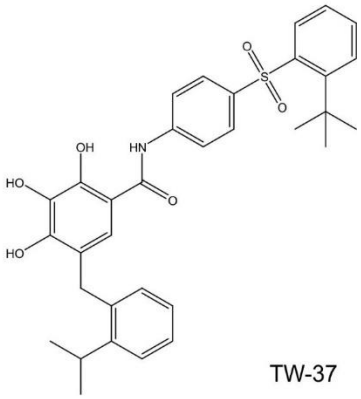
Obatoclax mesylate



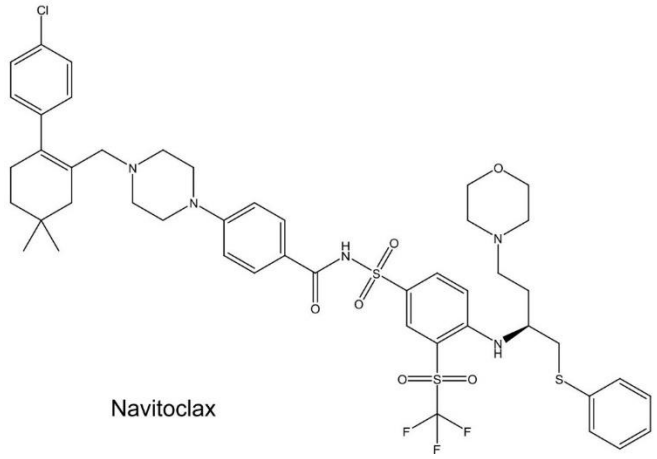
HA14-1



(+)-Apogossypol



TW-37



Navitoclax

Figure 8-2: Bcl-2 inhibitors used in Chapter 4.
The chemical structures of the Bcl-2 inhibitors used in Chapter 4 are shown above.

Appendix IV. Primers used in Chapter 5 for library construction

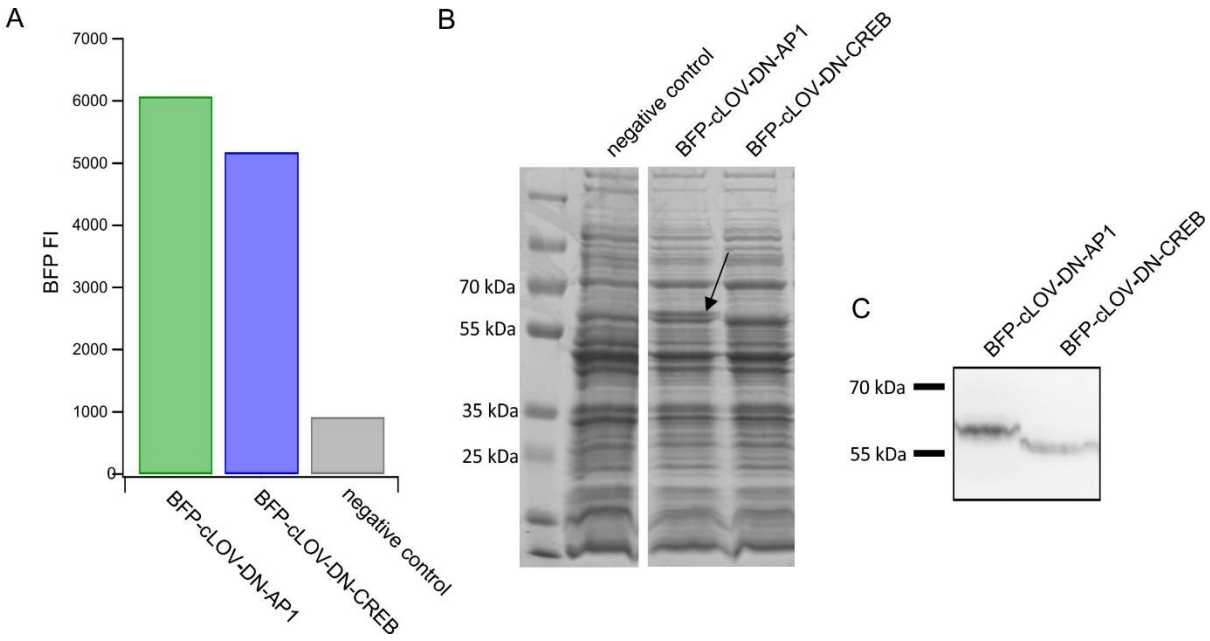
Table 8-2: Primers used in Chapter 5 for construction of the cLOV-DN-AP1 and cLOV-DN-CREB libraries.

Codons with mixed nucleotide content are underlined. Homologous sequence for recombination is shown in blue.

Primer name	**Sequence**	**Purpose**
BB-lib-fwd	CTGGCCACCACCCTGGAAC	Primer to amplify backbone to accept library inserts
BB-lib-rev	TTCCTCAGCTTCTTTTTTAGCCAGCA	Primer to amplify backbone to accept library inserts
A2-fwd (2)	CTGGAAAAAGAAGCCGAAGAA	Primer to amplify template fragment for library construction
template-rev	CTTCATGGCTGTACCAGC	Primer to amplify template fragment for library construction
Lib-cLOV-fwd	CTAAAAAAGAAGCTGAGGAAMTTGAWSAANNCNNCNNSGAGCTTCTGGAAAAAGAAGCC	Forwards primer for all library fragments
Link-0-CREB	CAGGGTGGTGGCCAGCTTCATAGCGTCGGATTTGTGAGAGTACAG	Library reverse primer for CREB without a C-terminal linker
Link-0-Fos	CAGGGTGGTGGCCAGCTTCATATCAGGGATCTTGCAGGCAGG	Library reverse primer for Fos without a C-terminal linker
Link-4-CREB	CAGGGTGGTGGCCAGCTTCATACCTGAACCGGTAGCGTCG	Library reverse primer for CREB yielding a 4 residue C-terminal linker
Link-4-Fos	CAGGGTGGTGGCCAGCTTCATACCTGAACCGGTATCAGGGATCT	Library reverse primer for Fos yielding a 4 residue C-terminal linker
Link-8	CAGGGTGGTGGCCAGCTTCATACCGCCAGAGCTACCTGAA	Library reverse primer (for CREB or Fos) yielding an 8 residue C-terminal linker
Link-12*	CGTTCCAGGGTGGTGGCCAGCGACGCGGTTGAACCGCC	Library reverse primer (for CREB or Fos) yielding a 12 residue C-terminal linker
Link-16	CAGGGTGGTGGCCAGCTTCATACCGCTACCTCCCGACGC	Library reverse primer (for CREB or Fos) yielding a 16 residue C-terminal linker
Link-22	CAGGGTGGTGGCCAGCTTCATAGAACTACCGCTCGAGCCA	Library reverse primer (for CREB or Fos) yielding a 22 residue C-terminal linker
Linker-GB-rev*	CGTTCCAGGGTGGTGGCCAGCTTCATGGCTGTACCAGCTGTACTACTTGCAGAACTACCGCTCGAGCCACC	Library reverse primer (for CREB or Fos) yielding a 30 residue C-terminal linker

* The homologous overlap sequence on these primers removes the first two residues on the second half of cLOV (MK). This was changed in subsequent primer designs.

Appendix V. Expression data and sequences used in Chapter 5



cLOV-DN-AP1 (BFP-cLOV-A4-Fos-His-v5)

MKGDVQYFIGVQLDGTEHVRDAAEREAVMLAKKEAEEIEQAAAELLEKEAEELEQELAELEAETDQLEDE
KSALQTEIANLLKEKEKLEFILAAHRPACKIPDTGSGSSGGSTASGGSGGSSGSSASSTAGTAMKLATTL
ERIEKNFIITDPRLPDNPIIFASDSFLQLTEYSREEILGRNCRFLQGPETDRATVRKIRDAIDNQTEVTV
QLINYTKSGKKFWNLFHLQPMRDQKLAAALEHHHHHH*

cLOV-DN-CREB (BFP-cLOV-A4-CREB-His-v5)

MKGDVQYFIGVQLDGTEHVRDAAEREAVMLAKKEAEEIEQAAAELLEKEAEELEQELAELENRVAVLENQ
NKTLIEELKALKDLYSHKSDATGSGSSGGSTASGGSGGSSGSSASSTAGTAMKLATTLERIEKNFIITDP
RLPDNPIIFASDSFLQLTEYSREEILGRNCRFLQGPETDRATVRKIRDAIDNQTEVTVQLINYTKSGKKF
WNLFHLQPMRDQKLAAALEHHHHHH*

Figure 8-3: Expression data and sequences of cLOV-DN-AP1 and cLOV-DN-CREB.
The fusion constructs of cLOV with DN-AP1 and DN-CREB expressed in *E. coli.* Expression was confirmed via (A) fluorescence measurements, (B) SDS PAGE and (C) western blotting using anti-His antibodies. The negative control is an *E. coli* culture without a plasmid. Note that expression of the CREB-based construct could not be seen via SDS PAGE; however, a band was visible in the western blot. The constructs are 57 kDa (BFP-cLOV-DN-AP1) and 55 kDa (BFP-cLOV-DN-CREB) in size. The protein sequences are shown below, showing cLOV (yellow), the acidic extension (blue), the leucine zipper of each Fos and CREB (green), a 30-residue linker (grey) and a His tag (orange). The chimeric overlap sequence is coloured according to residue origin, with homologous residues shown in black. Each construct contains an N-terminal fusion to mTagBFP, connected via a 23-residue linker (not shown). The Ser residue underlined in the CREB leucine zipper sequence was mutated from Cys in previous work.[25]

Appendix VI. List of plasmids used in Chapter 6

Table 8-3: Plasmids used in Chapter 6.

Name	**Description**
pQE80L-wtPYP	*H. halophila* PYP A (also referred to as wild-type)
pET24b-Δ25PYP	N-terminal deletion of wild-type PYP
pET24b-BFP	BFP only control
BFP fusion plasmids	
pQE80L-PYP-BFP	wild-type PYP-BFP fusion
pQE80L-PYP-BFP S8A	PYP-BFP fusion with photocycle neutral mutation in PYP
pQE80L-PYP-BFP T101A	PYP-BFP fusion with photocycle neutral mutation in PYP
pQE80L-PYP-BFP V105A	PYP-BFP fusion with mutation to slow PYP photocycle
pQE80L-PYP-BFP S117A	PYP-BFP fusion with mutation to accelerate PYP photocycle
pQE80L-PYP(B)-BFP	BFP fusion with *H. halophila* PYP B
pQE80L-BFP-cPYP	BFP fusion with cPYP
pQE80L-BFP-cPYP M121E	BFP fusion with cPYP, containing a point mutation to slow the cPYP photocycle
pQE80L-BFP-GCN4-Δ25PYP-v2	BFP fusion with GCN4-Δ25PYP(v2 K143F E35L)
pQE80L-BFP-PYP	BFP fusion with N-terminal BFP and longer linker
pQE80L-BFP-opto-DN-CREB	BFP fusion with opto-DN-CREB
Chromophore biosynthesis plasmids	
pRSFDuet-v2 *Rs*TAL *Il*4CL	*R. sphaeroides* TAL and *I. loihiensis* 4CL
pRSFDuet-v2 *Rs*TAL syn-*Il*4CL	*Rs*TAL and *Il*4CL, with *Il*4CL codon optimized for *E. coli* expression and tagged with SUMO
pRSFDuet-v2 *Rs*TAL *At*4CL	*Rs*TAL and 4CL from *A. thaliana*

Appendix VII. Proteins used in Chapter 6

PYP

MRGSHHHHHHGSDDDDKMEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEG
DITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKK
ALSGDSYWVFVKRV*

Δ25PYP

MGGSDDDDKLAFGAIQLDGDGNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFK
EGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRVKLAAALEHHHHHH*

PYP-BFP

MRGSHHHHHHGSDDDDKMEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEG
DITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKK
ALSGDSYWVFVKRVKLAAALESGSELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMR
IKVVEGGPLPFAFDILATSFLYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQ
DTSLQDGCLIYNVKIRGVNFTSNGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHL
IANAKTTYRSKKPAKNLKMPGVYYVDYRLERIKEANNETYVEQHEVAVARYSDLPSKLGHKLN*

PYP(B)-BFP

MRGSHHHHHHGSDDDDKMGTLIFGRQDLENRLAAMTPEEIDDLPFGVIQIDQHGRILLYNATEG
AITGRDPEAMIGRDFFNDVAPCGHTEAFYGRFQEGVRHGDLNEIFDYTFDYRMAPTKVRVHMKR
ALSGDTYWIFVKRISAPAAKLAAALESGSELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEG
TQTMRIKVVEGGPLPFAFDILATSFLYGSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGV
LTATQDTSLQDGCLIYNVKIRGVNFTSNGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLV
GGSHLIANAKTTYRSKKPAKNLKMPGVYYVDYRLERIKEANNETYVEQHEVAVARYSDLPSKLG
HKLN*

BFP-PYP

MSELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILATSFLY
GSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFTS
NGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANAKTTYRSKKPAKNLKMPGV
YYVDYRLERIKEANNETYVEQHEVAVARYSDLPSKLGHKLNGMDELYKSGLRSRAQASNEFGID
LSGLTLQMEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKQV
IGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVF
VKRVKLAAALE*

BFP-opto-DN-CREB

MSELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILATSFLY
GSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFTS
NGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANAKTTYRSKKPAKNLKMPGV
YYVDYRLERIKEANNETYVEQHEVAVARYSDLPSKLGHKLNGMDELYKSGLRSRAQASNEFGID
LSGLTLQMLEQQLEELAQENEELEKEAEELEQEVMELEFRVDNLEQQNLKLIEELLKLLLFGAI
QLDGDGNILQYNAAEGDITGRDPDQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLDTTFEYT
FDYQMTPTRVRVWMKKAYSGDSYWVYVKRS*

BFP-GCN4-d25PYP

MSELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILATSFLY
GSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFTS
NGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANAKTTYRSKKPAKNLKMPGV
YYVDYRLERIKEANNETYVEQHEVAVARYSDLPSKLGHKLNGMDELYKSGLRSRAQASNEFGID
LSGLTLQMKDPAALKRARNTEAARRSRARKLQRMKQLEDKVLELAFGNEHLEEELARLKKGQLD
RLAFGAIQLDGDGNILQYNQQEGDITGRDPSQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNL
QTMFEYTFDYQMTPTKVRVFMFKSKTGDSYWVFVTRVKLAAALE*

BFP-cPYP

MSELIKENMHMKLYMEGTVDNHHFKCTSEGEGKPYEGTQTMRIKVVEGGPLPFAFDILATSFLY
GSKTFINHTQGIPDFFKQSFPEGFTWERVTTYEDGGVLTATQDTSLQDGCLIYNVKIRGVNFTS
NGPVMQKKTLGWEAFTETLYPADGGLEGRNDMALKLVGGSHLIANAKTTYRSKKPAKNLKMPGV
YYVDYRLERIKEANNETYVEQHEVAVARYSDLPSKLGHKLNGMDELYKSGLRSRAQASNEFGID
LSGLTLQMKGDSYWVFVKRVGGSGGSGGMEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGD
GNILQYNAAEGDITGRDPKQVIGKNFFKDVAPCTDSPEFYGKFKEGVASGNLNTMFEYTFDYQM
TPTKVKVHMKKALSKL*

*Rs*TAL

MGSSHHHHHHSQDPKPMLAMSPPKPAVELDRHIDLDQAHAVASGGARIVLAPPARDRCRASEAR
LGAVIREARHVYGLTTGFGPLANRLISGENVRTLQANLVHHLASGVGPVLDWTTARAMVLARLV
SIAQGASGASEGTIARLIDLLNSELAPAVPSRGTVGASGDLTPLAHMVLCLQGRGDFLDRDGTR
LDGAEGLRRGRLQPLDLSHRDALALVNGTSAMTGIALVNAHACRHLGNWAVALTALLAECLRGR
TEAWAAALSDLRPHPGQKDAAARLRARVDGSARVVRHVIAERRLDAGDIGTEPEAGQDAYSLRC
APQVLGAGFDTLAWHDRVLTIELNAVTDNPVFPPDGSVPALHGGNFMGQHVALTSDALATAVTV
LAGLAERQIARLTDERLNRGLPPFLHRGPAGLNSGFMGAQVTATALLAEMRATGPASIHSISTN
AANQDVVSLGTIAARLCREKIDRWAEILAILALCLAQAAELRCGSGLDGVSPAGKKLVQALREQ
FPPLETDRPLGQEIAALATHLLQQSPVGAPAGRQACGRIMLKSNRK*

*Il*4CL

MTDSLQQHLIVTVIGDLIADELARMRPAESEYWKRRQWHEDDTLVAKNKSTKDNGEDDVVVDSL
ERLALAGRVVQFFHMGDSGVEDYLLRRNSLAEWAEVVLKSRQVHTQNLTVTTSGSTGQPKACEH
SWSALVEEVREFVRIFDNDYELSPVRIVALVPSHHIYGFLFTVLLPHLVDAPVLRGFKAYSHVR
NGGLRAGDAVVGFPELLTQLSSEMPPLPPGVLFISSAGPCPASTVHQLYAIGAARAVEIYGSSE
TAGMAYRSKPENNYRLLSRWRKNTENHQQLIDRQTKVIYEIPDNTQWHTEDEFQITGRVDKAVS
IRGINVFPAHIAKCLRQHPAVADATVRPMRSDEGYGLKAFIVLQENISETVTEQSVQTWLSDNL
CAAEIPERISFGEQLPINSMGKAQDWSIDNSPTGKPLN*

syn-*Il*4CL

MSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLR
FLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGSSGGMTDSLQQHLIVTVIGDLIADELARM
RPAESEYWKRRQWHEDDTLVAKNKSTKDNGEDDVVVDSLERLALAGRVVQFFHMGDSGVEDYLL
RRNSLAEWAEVVLKSRQVHTQNLTVTTSGSTGQPKACEHSWSALVEEVREFVRIFDNDYELSPV
RIVALVPSHHIYGFLFTVLLPHLVDAPVLRGFKAYSHVRNGGLRAGDAVVGFPELLTQLSSEMP
PLPPGVLFISSAGPCPASTVHQLYAIGAARAVEIYGSSETAGMAYRSKPENNYRLLSRWRKNTE
NHQQLIDRQTKVIYEIPDNTQWHTEDEFQITGRVDKAVSIRGINVFPAHIAKCLRQHPAVADAT
VRPMRSDEGYGLKAFIVLQENISETVTEQSVQTWLSDNLCAAEIPERISFGEQLPINSMGKAQD
WSIDNSPTGKPLN*

*At*4CL

MAPQEQAVSQVMEKQSNNNNSDVIFRSKLPDIYIPNHLSLHDYIFQNISEFATKPCLINGPTGH
VYTYSDVHVISRQIAANFHKLGVNQNDVVMLLLPNCPEFVLSFLAASFRGATATAANPFFTPAE
IAKQAKASNTKLIITEARYVDKIKPLQNDDGVVIVCIDDNESVPIPEGCLRFTELTQSTTEASE
VIDSVEISPDDVVALPYSSGTTGLPKGVMLTHKGLVTSVAQQVDGENPNLYFHSDDVILCVLPM
FHIYALNSIMLCGLRVGAAILIMPKFEINLLLELIQRCKVTVAPMVPPIVLAIAKSSETEKYDL
SSIRVVKSGAAPLGKELEDAVNAKFPNAKLGQGYGMTEAGPVLAMSLGFAKEPFPVKSGACGTV
VRNAEMKIVDPDTGDSLSRNQPGEICIRGHQIMKGYLNNPAATAETIDKDGWLHTGDIGLIDDD
DELFIVDRLKELIKYKGFQVAPAELEALLIGHPDITDVAVVAMKEEAAGEVPVAFVVKSKDSEL
SEDDVKQFVSKQVVFYKRINKVFFTESIPKAPSGKILRKDLRAKLANGL*

Figure 8-4: Protein sequences used in Chapter 6.
PYP (or the PYP-based domain) is shown in yellow, BFP is shown in blue, 4CL is shown in green and TAL is shown in red. Residues mutated (point mutations in cPYP and PYP) are indicated with an underline. Protein tags, His or SUMO, are shown in orange and light blue respectively. Linkers and vector sequences are shown in black.

Appendix VIII. ESI-MS of apo- and holo-PYP-BFP

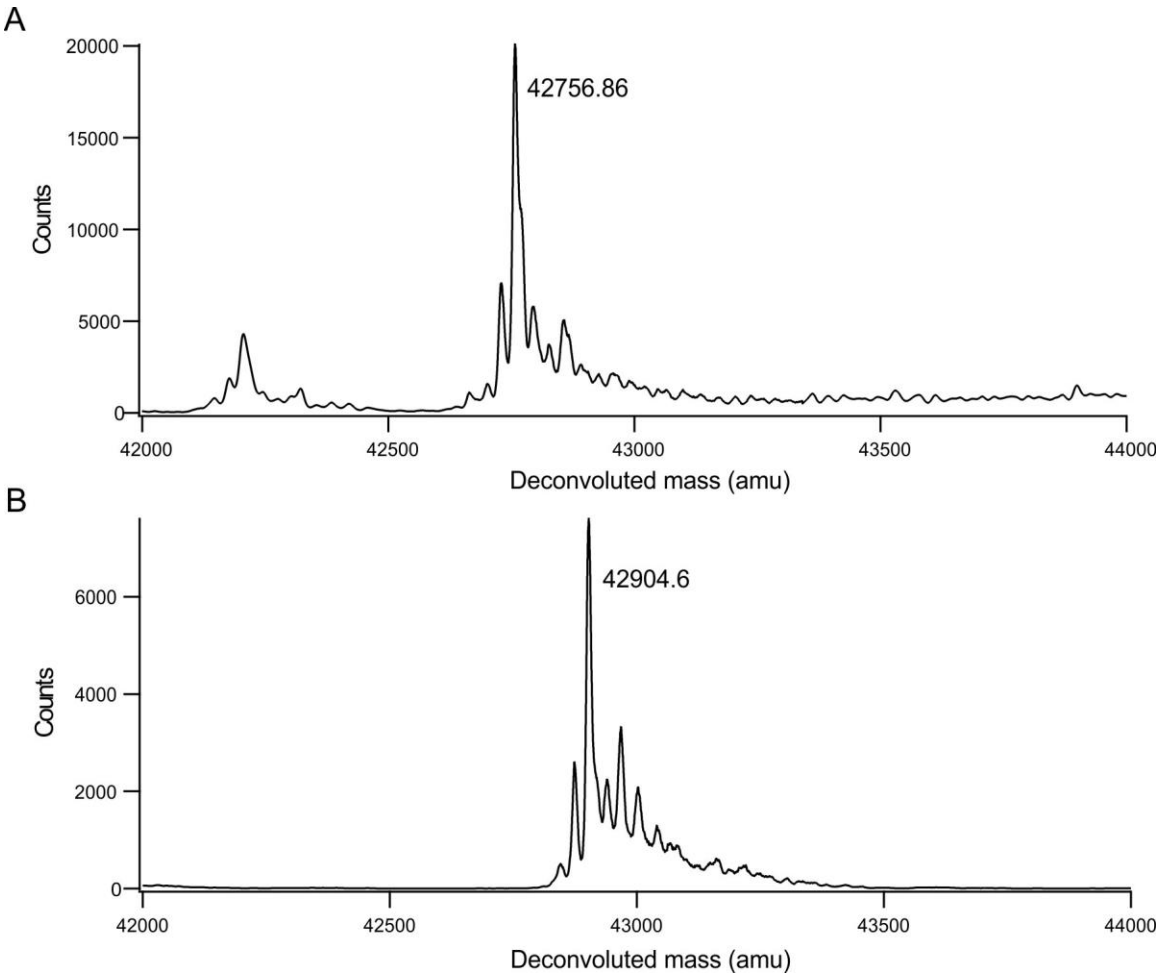


Figure 8-5: ESI-MS of apo- and holo-PYP-BFP.
The mass of purified apo- (A) and holo-PYP-BFP (B) was confirmed via ESI-MS. The expected mass of apo-PYP-BFP is 42759 Da and holo-PYP-BFP is 42905 Da

Appendix IX. ESI-MS of holo-PYP-BFP reconstituted *in vitro*

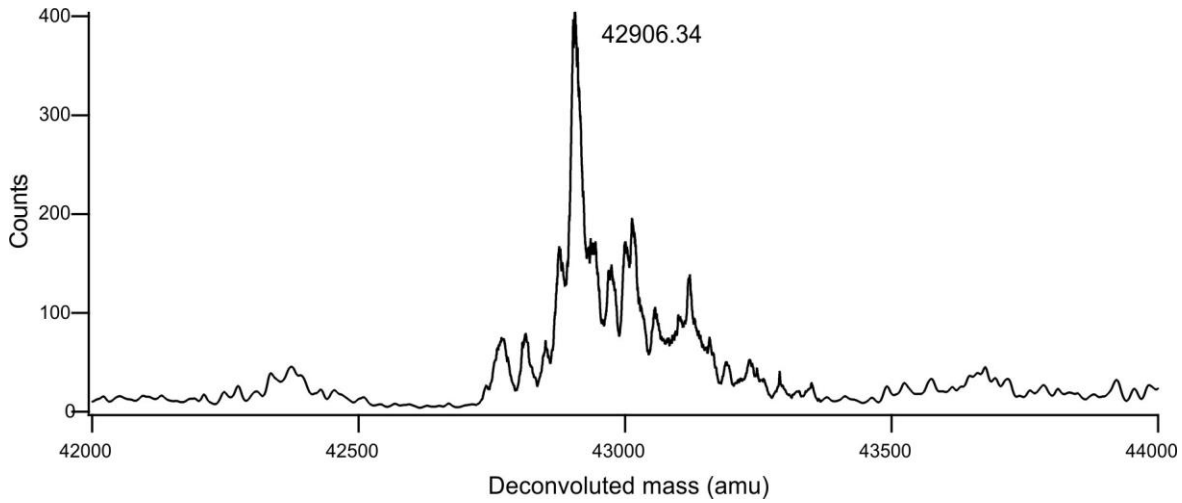
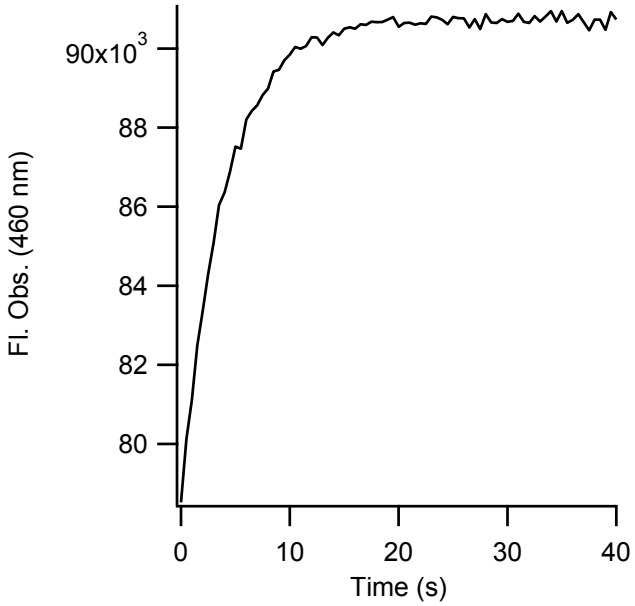


Figure 8-6: ESI-MS of holo-PYP-BFP reconstituted *in vitro.*
Apo-PYP-BFP reconstituted with *p*-coumaroyl-CoA had the correct mass, as confirmed by ESI-MS. The expected mass of holo-PYP-BFP is 42905 Da and the mass of apo-PYP-BFP is 42759.

Appendix X. Derivation of Eq. 1 from Chapter 6



The observed fluorescence depends on the concentration of each species (apo-PYP-BFP, *trans* holo-PYP-BFP, *cis* holo-PYP-BFP). The fluorescence from apo-PYP-BFP is time independent.

For holo-PYP-BFP the observed fluorescence Fl_{holo} depends on:

$$Fl_{holo} = [holo] * \left((fraction\_T) * Fl_T + (fraction\_C) * Fl_C\right) \qquad \text{(Eq. S1)}$$

where Fl_C , and Fl_T are the fluorescence intensities of *trans* holo-PYP-BFP and *cis* holo-PYP-BFP, and $(fraction\_T) + (fraction\_C) = 1$

After dark adaptation, $(fraction\_T) = 1$

The fluorescence time course observed when light (the 420 nm beam from the plate reader) is turned on, represents the system moving from all *trans* to a steady-state fraction *trans* $(fraction\_T_{ss})$ that depends on the rate constants k_f (*trans* to *cis* light-driven isomerization) and k_b (thermal back reaction *cis* to *trans*).

At steady state (ss): $[T]_{ss} * k_f = [C]_{ss} * k_b$ *or*

$$[holo] * (fraction\_T_{ss}) * k_f = [holo] * (fraction\_C_{ss}) * k_b$$

So, $(fraction\_T_{ss}) * k_f = \left(1 - fraction_{T_{ss}}\right) * k_b$

And, $fraction\_T_{ss} = \frac{k_b}{(k_b + k_f)}$ (Eq. S2)

The time course for approach to steady-state is given by (see Levine p. 522[275]):

$$fraction\_T = fraction\_T_{ss} + (fraction\_T_0 - fraction\_T_{ss})e^{-(k_f + k_b)t}$$ (Eq. S3)

Where $fraction\_T_0 = 1$; *i.e.* after dark-adaption, and *t* is time.

Substituting the expression above Eq. S2 for $fraction\_T_{ss}$:

$$fraction_T = \left(\frac{k_b}{(k_b + k_f)} + \left(1 - \frac{k_b}{(k_b + k_f)}\right)e^{-(k_f + k_b)t}\right)$$ (Eq. S4)

From Eq. S1 above:

$$Fl_{holo} = [holo] * \left((fraction\_T) * Fl_T + (1 - fraction\_T) * Fl_C\right)$$ (Eq. S5)

$$= [holo] * (Fl_C + (fraction\_T) * (Fl_T - Fl_C))$$ (Eq. S6)

Substituting Eq. S4 gives:

$$Fl_{holo} = [holo] * \left(Fl_C + \left((Fl_T - Fl_C) * \left(\frac{k_b}{(k_b + k_f)} + \left(1 - \frac{k_b}{(k_b + k_f)}\right)e^{-(k_f + k_b)t}\right)\right)\right)$$

Including apo-PYP-BFP gives:

$$Fl_{obs} = [apo] * Fl_{apo} + [holo] * \left(Fl_C + \left((Fl_T - Fl_C) * \left(\frac{k_b}{(k_b + k_f)} + \left(1 - \frac{k_b}{(k_b + k_f)} \right) e^{-(k_f + k_b)t} \right) \right) \right)$$

which is Eq. 1 in the Chapter 6.

Time dependent fluorescence data was fit to this equation using the Global Fitting protocol in Igor Pro. The values for k_b, k_f, Fl_T, Fl_C, and Fl_{apo} were linked in the fitting from data from genetically identical samples (*e.g.* all the timepoint measurements for a reconstitution of wild-type PYP-BFP), and only [holo] and [apo] varied.

Appendix XI. Fitted parameters from global fits in Chapter 6

Table 8-4: Fitted parameters from global fits from Chapter 6.

	wtPYP-BFP with *Rs*TAL, *Il*4CL (log or stationary phase expression) (Fig. 6-13)	**wtPYP-BFP with *Rs*TAL and syn-*Il*4CL or *At*4CL[a] (Fig. 6-14)**	**wtPYP-BFP with *Rs*TAL, *At*4CL[b] (Figs. 6-18, 6-21)**	**wtPYP-BFP with *Rs*TAL, *At*4CL (post 4°C incubation) (Fig. 6-12)**	**PYP(S8A)-BFP with *Rs*TAL, *At*4CL (Fig. 6-18)**	**PYP(T101A)-BFP with *Rs*TAL, *At*4CL (Fig. 6-18)**	**PYP(V105A)-BFP with *Rs*TAL, *At*4CL (Fig. 6-18)**	**PYP(S117A)-BFP with *Rs*TAL, *At*4CL (Fig. 6-18)**	**PYP(B)-BFP with *Rs*TAL, *At*4CL (Fig. 6-21)**	**Batch fitting of point mutants (Table 6-2)**	**BFP-GCN4-Δ25PYP with *Rs*TAL, *At*4CL (Fig. 6-24)[c]**
Fl_{apo}	531	552	552	559	521	557	527	533	550	537	568
Fl_{cis}	530	552	552	553	521	557	527	533	541	537	561
Fl_{trans}	196	200	202	202	200	206	195	197	239	199	233
k_b (s^{-1})	0.19	0.27	0.27	0.17	0.21	0.23	0.038	0.57	0.043	n/a	0.075
$t_{1/2}$ (s)	3.7	2.6	2.6	4.2	3.3	3.0	18	1.2	16	n/a	9.2
k_f (s^{-1})	0.048	0.048	0.048	0.053	0.048	0.048	0.048	0.049	0.048	0.049	0.010

Fluorescence intensity values are in arbitrary units. Each data set is derived from 3 biological replicates.

[a,b] The cultures of wtPYP-BFP co-expressed with *Rs*TAL and *At*4CL in the two marked columns are genetically identical. In (a) the cultures were monitored for reconstitution over 13 h, while in (b) the same experiment was repeated, but cultures were monitored for 7 h.

[c] Parameters for this fitting were determined as discussed in Figure 6-23, due to the alternative linker design of the BFP fusion.