ABSTRACT

GARDNER, LAURA KIMBERLY. Engineering of Prokaryotic and Eukaryotic Cells to Respond to Light. (Under the direction of Dr. Alexander Deiters).

Highly complex synthetic gene circuits have been engineered in living organisms to develop systems with new biological properties. In order to program synthetic gene circuits to perform their various functions, external control over their activity needs to be achieved. Any input for an engineered system must be accurate, precise and tunable to ensure stable output generation. Light serves as an excellent trigger to achieve precise control over synthetic systems as it can be regulated in wavelength, timing, intensity and location. Light-inducible systems are often generated through the installation of a light-cleavable photo-protecting (photocaging) group on a biologically active small molecule. The research herein describes various synthetic biological systems whose activity can be controlled by light through photocaged small molecules and other small molecule tools. Specifically we show the (1) control of gene expression in prokaryotic cells with photocaged inducer molecules, (2) control of protein function in pro- and eukaryotic cells through genetically encoded unnatural amino acids, (3) synthesis of small molecules for the control of miRNA function in eukaryotes, (4) design and synthesis of small molecules tools for use with an engineered enzyme for protein labeling and (5) the synthesis of a photocaged protein crosslinker to control enzyme function *in vivo*.

Engineering of Prokaryotic and Eukaryotic Cells to Respond to Light

by
Laura Kimberly Gardner

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Chemistry

Raleigh, North Carolina

2012

APPROVED BY:

_____	_____
Alexander Deiters, PhD Committee Chair	Reza Ghiladi, PhD
_____	
Robert Kelly, PhD	

DEDICATION

This work is dedicated to

My Mom and Dad who have always been supportive, loving and motivational

My brother who has always shown me to have multiple points of view

Fellow graduate students who have made this experience unforgettable

And to Steven Rogers, for being an excellent supporter, friend and teacher through my best and worst moments

BIOGRAPHY

The author, Laura Gardner, was born on September 22, 1987 to Peter and Mary Jo Gardner in Hickory, NC. She moved to Wake County in 1995 and graduated from Apex High School in 2005. After graduation, she began her undergraduate degree at North Carolina State University and worked in the lab of Dr. Reza Ghiladi to study the catalase and peroxidase activities of KatG mutants from *Mycobacterium tuberculosis.* Additionally, she completed the chemistry honors program, received an undergraduate scholarship for research and was an active member of the NCSU rowing team. After graduating magna cum laude from the chemistry department in 2009, she chose to continue her graduate studies at North Carolina State University.

Laura then began her graduate career under the direction of Dr. Alexander Deiters within the chemistry department and completed three years of research in organic synthesis and chemical biology. During graduate school, Laura completed a minor in biotechnology and received a GAANN molecular biotechnology fellowship. After completion of her master's degree, Laura plans to move to Kansas to live with her fiancé Steven Rogers, another graduate of the NCSU chemistry department and will pursue a career in chemical biology.

ACKNOWLEDGMENTS

I would like to acknowledge many people for helping me during my time in graduate school. In particular, I would like to acknowledge my advisor, Dr. Alex Deiters for being an excellent mentor and teacher during my graduate studies. I would also like to acknowledge Dr. Reza Ghiladi, my undergraduate advisor and graduate committee member, for introducing me to research and always giving fantastic advice and Dr. Robert Kelly, for developing such a great biotechnology program and support through the GAANN fellowship.

I would also like to thank all members of the Deiters Lab who have made lab a place I really enjoy. In particular I would like to thank Colleen, Kalyn, Jeane, Meryl, Yan, Qingyang, and Jessica for being excellent friends, providing some of the best bake goods I've ever eaten and being great dancing partners; be it downtown or in the lab. I would also like to thank Rajendra for all his knowledge, advice, great stories and Nepalese food; Dr. Hank Chou for helping me with all of my cloning questions; and Andrew for making rocks shows much more fun. Additionally I would like to thank members of the Ghiladi and Melander labs for always lending a helping hand, and Alex P. for assistance with pyrotechnics.

I would also like to mention all the help and love my family has given me throughout my graduate experience (including bringing me a birthday cake while I was studying for an exam). And very importantly, I would like to thank Steven Rogers who has always been able to help me see the light at the end of the tunnel.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

CHAPTER 1 – INTRODUCTION TO SYNTHETIC BIOLOGY

Biological systems are composed of a series of complex networks, connected by interactions of DNA, RNA, amino acids and small molecules. In recent years, much emphasis has been put on exploring these convoluted pathways and understanding the precise mechanisms on which they operate. As the wealth of information on genetic circuitry has grown, so has the ability to manipulate these systems to carry out new functions.

Proteins, oligonucleotides and small molecules are the building blocks for synthetic biology, a discipline that strives to apply electrical engineering principles to the design and understanding of biological processes for the creation of new systems with useful properties.[1-3] Synthetic biology requires the same bottom-up approach of electrical engineering to construct these genetic circuits (Figure 1.1).[4] The fundamental hardware (DNA, RNA, protein) is used to create switches and gates via reactions and interactions. These gates and switches are connected to form genetic circuits such as operons and feedback loops, which when assembled together create the complex pathways of a cell.

The first level of biological circuitry is the construction of molecular switches and gates. These components are essential pieces of molecular computing devices such as oscillators,[5] toggle-switches,[6] and feedback circuits.[7] Molecular switches and gates are created through protein-protein interactions, protein-DNA interactions and small molecule-protein interactions that generate an output signal. An input signal, often a small molecule, is required to initiate these interactions. Small molecules offer an excellent avenue for control

over biological processes such as gene expression as they can be easily manipulated synthetically and allow for temporal control over a system as one can control the timing of addition.

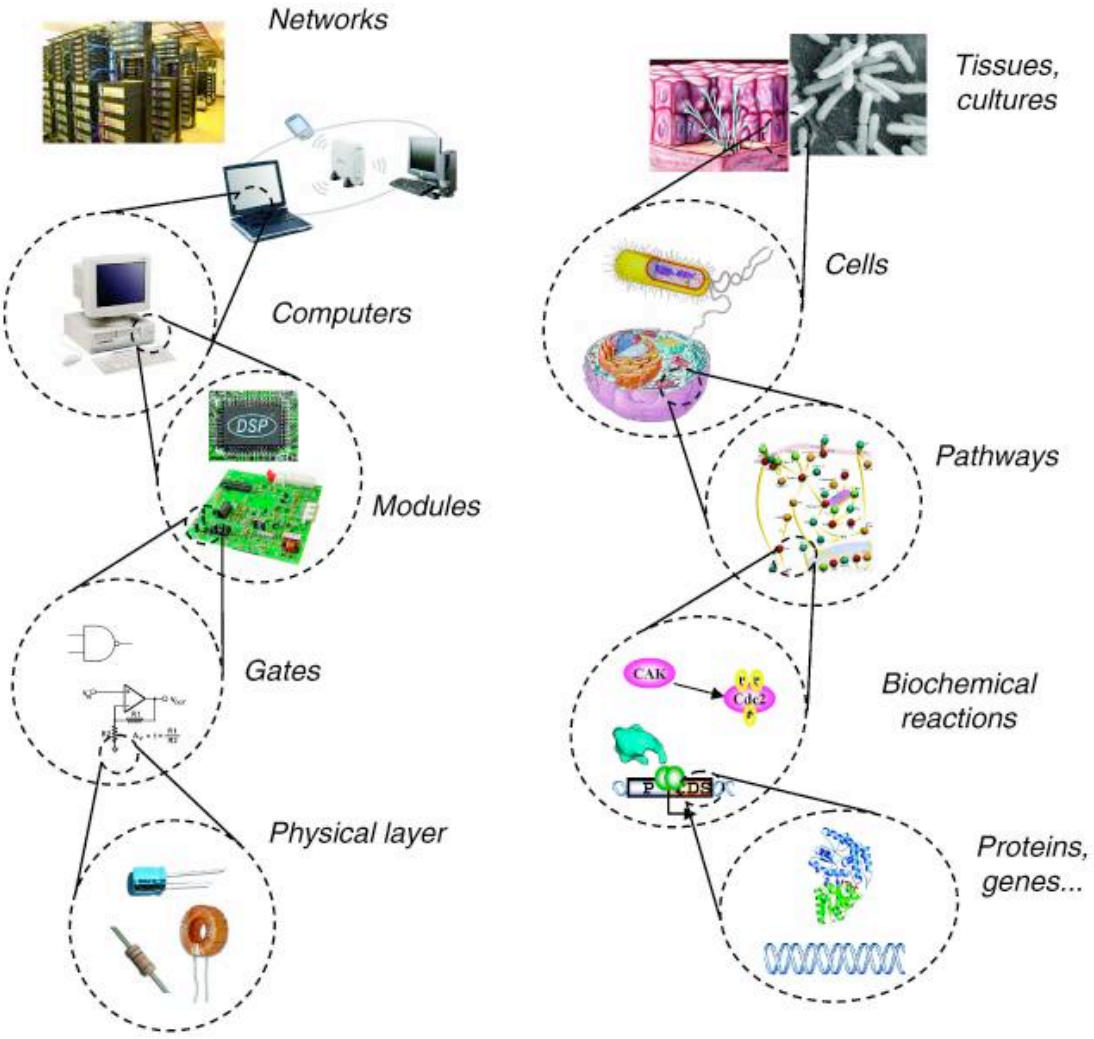


Figure 1.1. Computers and living organisms are constructed from a similar bottom-up approach. Adapted from *Mol Syst Bio.* **2006**, 2, 1.

Previously developed gene expression systems controlled by small molecules include antibiotic-inducible systems,[8] sugar-inducible systems[9] and biotin-mediated gene expression systems.[10] These methods allow for temporal control, but lack spatial control over gene expression. An alternative method to exhibit precise spatial control over gene expression is to use light as a trigger. To achieve photochemical control, a light-cleavable protecting group, termed "caging group", is installed on a biologically active molecule to render it temporarily inactive.[11] The caging group can then be removed from the molecule after irradiation with UV light; a process called "decaging" (Figure 1.2). Traditionally an *ortho*-nitrobenzyl group is the protecting group of choice as it can be applied to a variety of functional groups and is removed quickly after brief irradiation with UV light (365 nm).

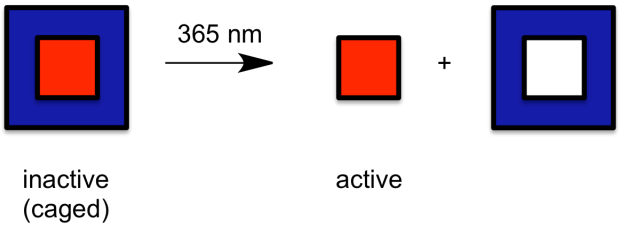


Figure 1.2. Decaging of a biologically inactive molecule with 365 nm UV light.

The first use of an *ortho*-nitrobenzyl (ONB) group on a biological molecule was photocaged ATP in 1978 by Kaplan et al.[12] The ONB group has since been applied to various molecules including nucleic acids, amino acids, metal ions and hormones.[13] This

ONB group can be placed on various functional groups including hydroxyls, carboxylates and amines, is easily synthesized, and can be modified to tune the absorbance maximum to longer wavelengths.[14]

The Deiters lab has a longstanding interest in using small molecules, in particular photocaged small molecules, as tools to modulate biological processes. In particular we are using small molecules and photocaged biomolecules to control gene expression and protein function in pro- and eukaryotic cells. In the presented research we aim to demonstrate:

1. Synthesis and application of photocaged small molecules to control gene expression in bacterial systems with high degrees of spatiotemporal control
2. Application of photocaged small molecules to construct a light-activated logic gate and a bandpass filter for further applications of light-based synthetic gene circuits
3. Incorporation of photocaged amino acids into bacterial proteins to control protein function and specific bioconjugation events with light
4. Evolution of amino-acyl tRNA synthetases in both pro- and eukaryotic cells to selectively incorporate unnatural amino acids into proteins *in vivo*
5. Synthesis of small molecules inhibitors of microRNA-21 to regulate miRNA activity in mammalian cells
6. Synthesis of a protein bioconjugatable dye for possible *in vivo* viscosity studies and protein labeling

7. Synthesis of various functional ligands for use with the HaloTag protein as chemical biology tools

Through the above objectives we are able to exhibit precise control over many biological pathways and regulation mechanisms using small molecules. We believe these methods provide novel ways to study biological systems and have wide application to the fields of chemical and synthetic biology.

CHAPTER 2 – PHOTOCHEMICAL CONTROL OF GENE EXPRESSION IN E. COLI

2.1 Introduction to light-activated gene expression

Small molecule inducible gene expression systems are a fundamental method of gene regulation in pro- and eukaryotic organism.[9] These gene switches are comprised of a repressor protein that binds to a promoter region upstream of the gene of interest inhibiting gene expression. Binding of a small molecule induces a conformational change and releases the repressor from the DNA thereby inducing gene expression. These switches exhibit high specificity for the small molecule and its cognate repressor protein, tight regulation between 'on' and 'off' states, and low basal levels of gene expression.[9] Small molecule activated gene switches have long been used in traditional molecular biology to control expression of recombinant proteins. These features, as well as the ease of manipulation of small organic molecules, make small molecule inducers targets for photochemical control of gene networks. Light-inducible small molecule systems are often generated through the installation of a light-cleavable photo-protecting (photocaging) group on the small molecule inducer of gene expression, rendering the small molecule inactive until the caging group is removed through irradiation.

The first instance of using a photocaged small molecule inducer to control a genetic circuit was the application of a photocaged estradiol for control of transcriptional activator activity in eukaryotic cells.[13] This light-inducible system demonstrated the modular nature of small molecule inducible systems, as the necessary components of transcriptional

regulation can be separated from their endogenous genes and recombined to create new synthetic circuits. Since then, various other photocaged small molecules have been developed to created light-activatable molecular switches including a caged ecdysone,[15] a caged IPTG for use with the *lac* operon,[16] caged toyocamycin for ribozyme-mediated gene expression,[17] caged doxycycline for use with the Tet-on system,[18] and caged rapamycin for light-mediated FKBP/FRB dimerization.[19-21]

Various photocaged analogues of doxycycline have been shown to control gene expression in eukaryotic systems, including transgenic mice.[22] Recently, Koh et. al, reported a new photocaged doxycycline, NvOC-Dox, to create photolithographic patterns of gene expression in mammalian cells.[23]

Discrete patterning was achieved in NIH 3T3 cells that grew into defined monolayers and contained a tetracycline-inducible GFP reporter (HRSp-GFP). Photolithography experiments were performed by treating cells harboring HRSp-GFP with **1**, followed by irradiation with UV light through a photomask applied to the bottom of a cell culture. Repeatable patterns of GFP expression with clearly defined edges were obtained using masks of different shapes (Figure 2.1). This system was also used to control cell adhesion of human embryonic kidney (HEK293T) cells to a mouse fibroblast (NIH 3T3) cellular monolayer in a spatially directed fashion. Light-inducible expression of ephrin A5, a membrane bound ligand that mediates cell-cell interactions by interacting with EphA7 receptors, allowed for photochemical control of cell attachment.

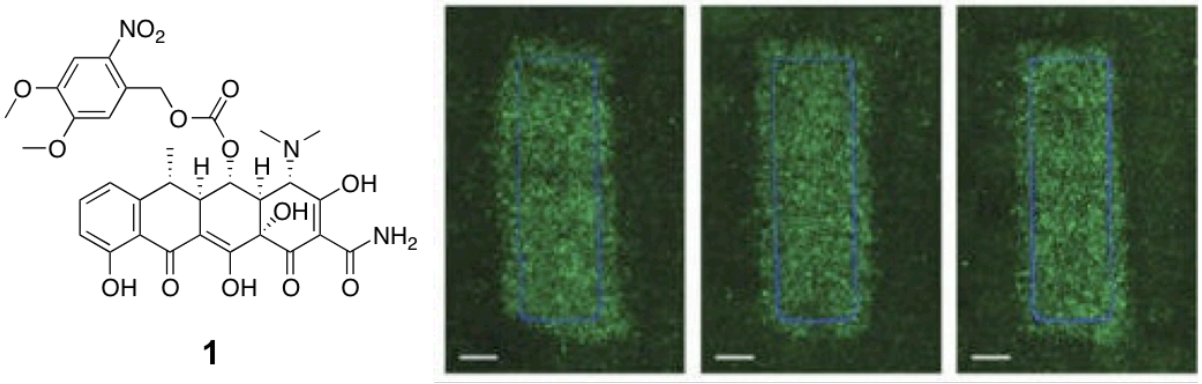


Figure 2.1. Photocaged doxycycline **1** was used to achieve spatial control over GFP expression. Adapted from *ACS Chem. Biol.* **2010**, *5*, 313.

Previously, the Deiters Lab developed a light-inducible gene-expression system for genes under control of the *lactose* operon in *E. coli*.[16] The *lactose* (*lac*) operon is commonly used in prokaryotes for IPTG inducible protein expression. In the absence of IPTG, the LacI repressor protein binds the *lac* operator *(lacO)* and prevents transcription of downstream genes. After IPTG addition, the small molecule binds LacI and causes the protein to undergo a conformational change, which allows for RNA polymerase (RNAP) to bind the *lacO* and initiate transcription (Figure 2.2).

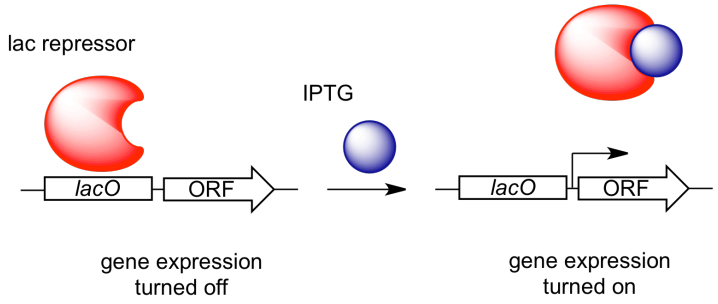


Figure 2.2. The IPTG (isopropyl-β-D-thio-galactoside) inducible *lac* operon.

A photocaged analogue of the inducer molecule, IPTG was synthesized to exhibit spatial and temporal control over the *lac* operon. After short irradiation with UV light, photocaged IPTG **2** is converted to IPTG **3** (Figure 2.3).

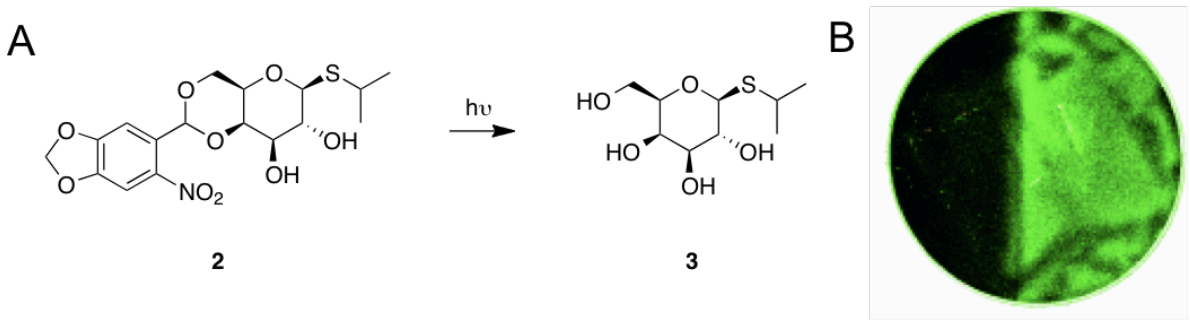


Figure 2.3. (A) Decaging of photocaged IPTG **2.** (B) Bacterial lithography of GFP expression; the left half of the plate was covered with a mask and the right half of the plate was exposed to UV light. Adapted from *Angew. Chem. Int. Ed.* **2007**, 460, 4290.

To demonstrate spatial control over IPTG-inducible expression, green fluorescent protein (GFP) was selected as a reporter. Bacterial cells transformed with pGFPuv were

plated onto agar plates pretreated with photocaged IPTG **2**. Half of the plate was covered with a mask, and the uncovered half was irradiated with UV light (365 nm, 25 W). The cells were then allowed to grow overnight before visualization. Distinct spatial control over GFP expression was observed (Figure 2.3).

This system introduced a new and facile method for controlling gene expression with light. As many genes used in recombinant protein expression are under control of the *Lac* operon, this methodology can be applied to a multitude of genes and proteins, and can be used to study elements involved in transcription. This system was demonstrated to work in *E. coli* but can potentially be applied to plant and mammalian cells that also use the *Lac* operon [24,25].

2.2 Control of gene expression with photocaged erythromycin

Many other inducible promoter systems are naturally present in bacteria to control phenomena such as quorum sensing [26] and antibiotic resistance.[27] Erythromycin, a 14-membered macrolactone, exhibits antibacterial activity by interfering with ribosome function and subsequently inhibiting protein synthesis.[28] Resistance to macrolide antibiotics has developed in various strains of bacteria.[28,29] The mechanisms of erythromycin resistance vary from ribosome modification, efflux of the drug, or enzymatic modification of the antibiotic itself.[27,28] One resistance mechanism found in *E. coli* is based on a gene cluster that encodes a 2'-phosphotransferase, Mph(A), that inactivates erythromycin by phosphorylating the hydroxyl group of the desosamine ring.[30,31] Studies of this gene

cluster revealed that *mphA*, the gene that encodes Mph(A), is regulated by the regulatory protein MphR(A) that binds upstream of its own promoter (PmphR) to prevent transcription of *mphA* and of itself.[32] When erythromycin is present, MphR(A) binds the antibiotic, undergoes a conformational change and releases the operator, thus activating transcription of the *mphA* gene and initiating erythromycin resistance (Figure 2.4). This negative feedback loop effectively allows *E. coli* to confer erythromycin resistance only when the antibiotic is present.

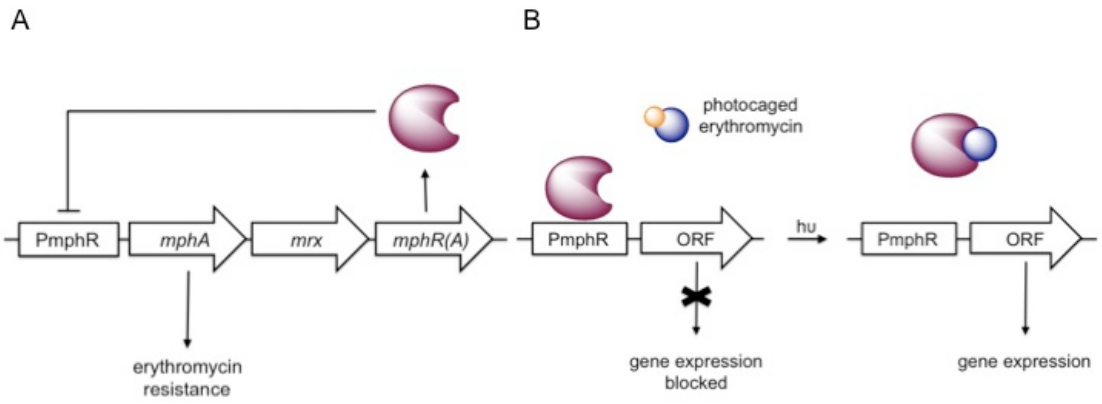


Figure 2.4. (A) Erythromycin resistance gene cassette found in some strains of *E. coli*. (B) Engineered MphR(A)/promoter system for conditional expression of reporter genes in the presence of erythromycin. Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

We generated a MphR(A)/promoter system (Figure 2.4) to enable conditional expression of three reporter genes (β-galactosidase, GFPuv and EGFP) in the presence of erythromycin.[31] We hypothesized that by creating a photocaged analogue of erythromycin

that is incapable of binding to MphR(A), we will be able to achieve spatial and temporal control over gene expression by photochemically regulating ligand binding. This system is an alternative method for photochemically controlled gene expression, and can be used in systems that contain a lac operon that is necessary for cellular processes.

The crystal structure of the MphR(A)/erythromycin complex (PDB 3FRQ) shows important hydrogen bonding interactions between multiple amino acid residues and the oxygen atoms of erythromycin (Figure 2.5).[31] We postulated that installation of a sterically demanding ONB photocaging group would disrupt the hydrogen bonding between the protein and the ligand and prevent binding of erythromycin to MphR(A). The ONB group could then be removed by irradiation with UV light (365 nm), restoring erythromycin binding to MphR(A) and activating gene expression.

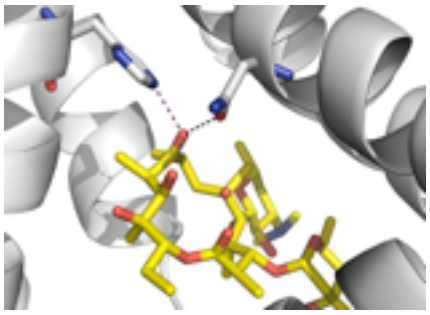
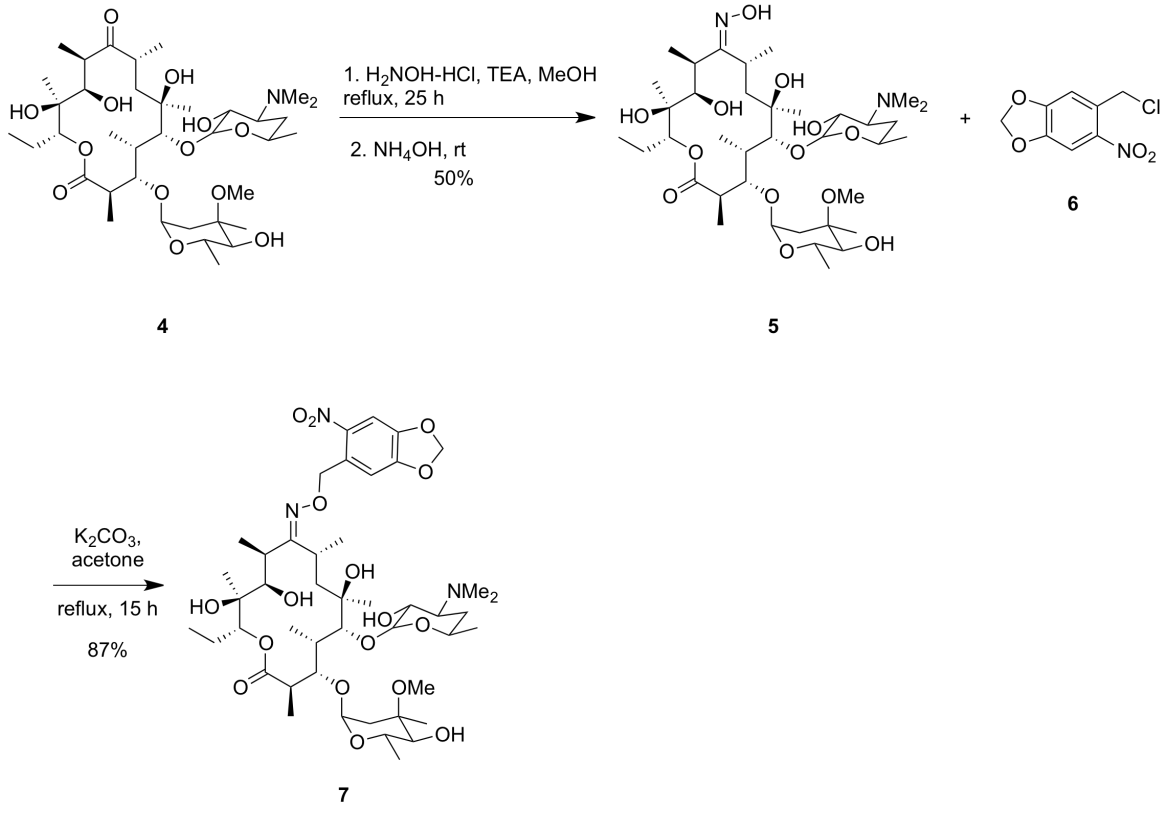


Figure 2.5. Crystal structure of the MphR(A) binding pocket with erythromycin (in yellow). Adapted from *Mol. BioSyst.* **2011**,*7*, 2554.

Yan Zou (Deiters Lab) synthesized the photocaged erythromycin **7** in two steps (Scheme 2.1) by reacting erythromycin **4** with hydroxylamine in the presence of triethylamine to obtain 9-oxime erythromycin **5**. The 9-oxime derivative **5** was then reacted with the ONB-chloride **6** in the presence of potassium carbonate and acetone to yield **7**.



Scheme 2.1. Synthesis of photocaged erythromycin **7**.

Our engineered erythromycin-inducible system (provided by the Cropp Lab) is encoded on two plasmids, pMLGFP and pJZ12. The constitutively expressed erythromycin

resistance gene, *mphA*, is encoded on pJZ12 to allow for constant resistance to the antibiotic. The erythromycin-inducible mphR(A) promoter and the gene of interest are encoded on pMLGFP (Figure 2.6).

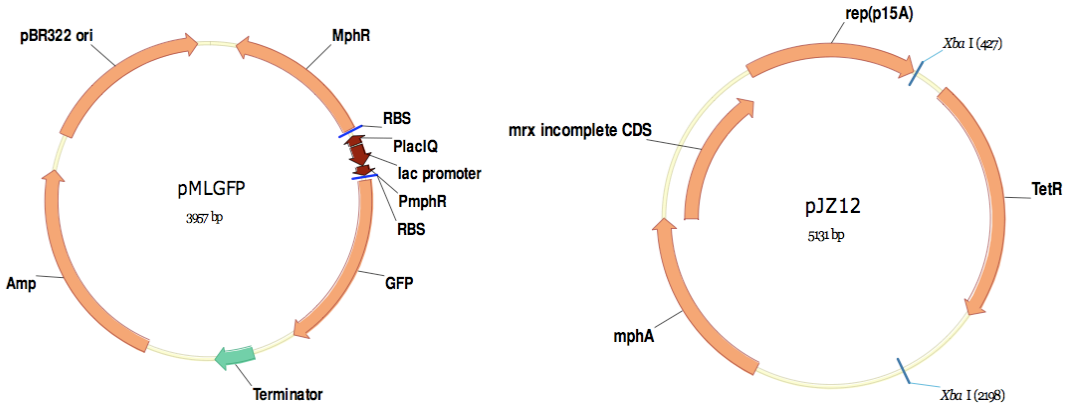


Figure 2.6. Two-plasmid system used in conjunction with photocaged erythromycin **7**. pMLGFP encodes the mphR(A) promoter and the reporter gene, while pJZ12 encodes the mphA phosphotransferase. Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

Since decaging of **7** with UV light yields **5** and not **4** we first confirmed that both had the same biological activity, as reported in literature.[33,34] Cells (Genehogs) containing pMLGFP and pJZ12, were grown to saturation and plated on plates pre-treated with **4** or **5** (10 uL, 1 mg/mL). Plates were then incubated overnight and visualized on a transilluminator. Comparable levels of fluorescence were observed on the two plates (Figure 3).

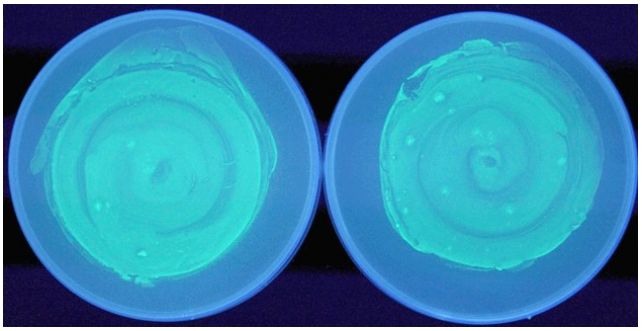


Figure 2.7. GFP expression in bacteria induced by erythromycin (**4**) (left) and the 9-oxime erythromycin (**5**) (right). Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

Decaging of photocaged erythromycin **7** (1 μM) was then performed in liquid cultures to determine the optimal amount of UV exposure. Cells containing pMLGFP and pJZ12 were grown to log phase (OD_{600} = 0.6) and induced with **5** or **7**. Cells were then transferred to a petri dish and irradiated on a transilluminator (365 nm, 25 W) for 5 minutes. After irradiation, cells were grown overnight, harvested by centrifugation, and resuspended in lysis buffer. Fluorescence was measured on a plate reader (Figure 2.8) (395/509 ex/em, Biotek Synergy 4 Microplate Reader).

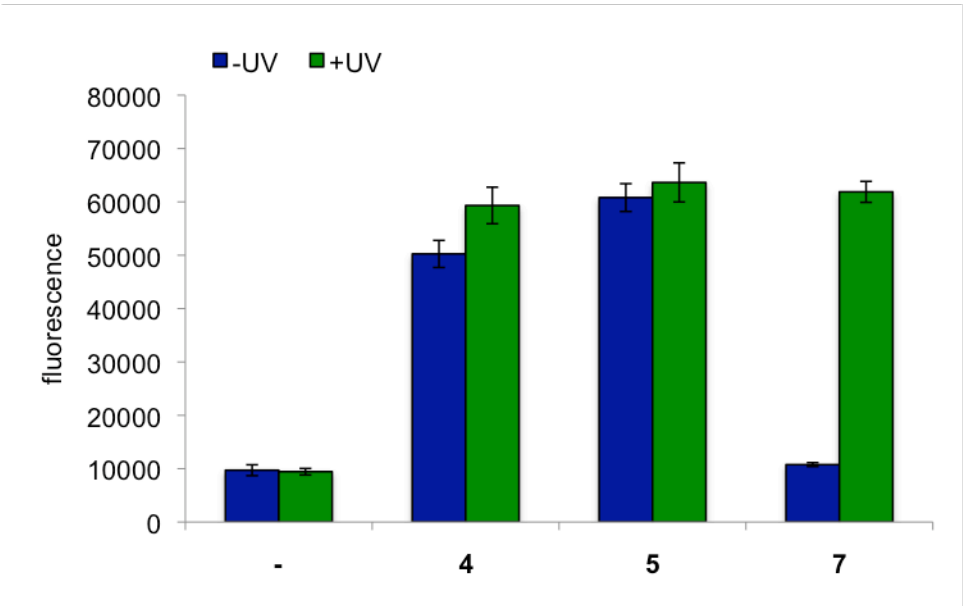


Figure 2.8. Decaging of photocaged erythromycin. In the absence of UV irradiation, **7** does not activate GFP expression. After 5 minutes UV irradiation, 96% of GFP expression is restored. Error bars represent standard deviations from three independent experiments. Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

To demonstrate that this system is applicable to multiple reporter genes, the *LacZ* reporter gene was utilized with the plasmid pBAD-mphR-P-lacZ (Figure 2.9).

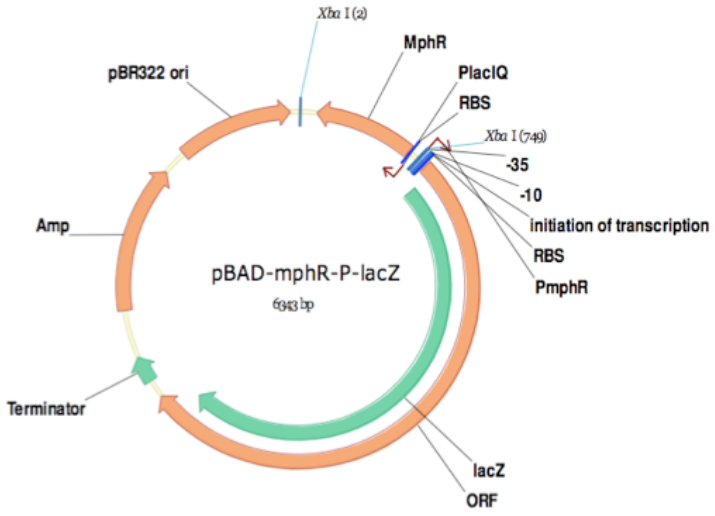


Figure 2.9. Erythromycin inducible *LacZ* reporter plasmid, provided by the Cropp Lab (VCU). Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

To perform a Miller assay,[35] cells containing the *LacZ* plasmid and pJZ12 were grown to log phase at 37 °C, and then induced with either **5** (1 μM) or **7** (1 μM) . Cells were then either irradiated for 5 minutes or kept in the dark, and then returned to 37 °C with shaking for 6 hours. After 6 hours, cells were lysed and reacted with *ortho*-nitrophenyl-β-galactoside for 10 minutes to quantify β-galactosidase expression.[35] Miller units were calculated according to the formula: $1000\times (OD_{420}-1.75\times OD_{550})]/(T\times V\times OD_{600})$. T = time in minutes, V= volume in mL (Figure 2.10).

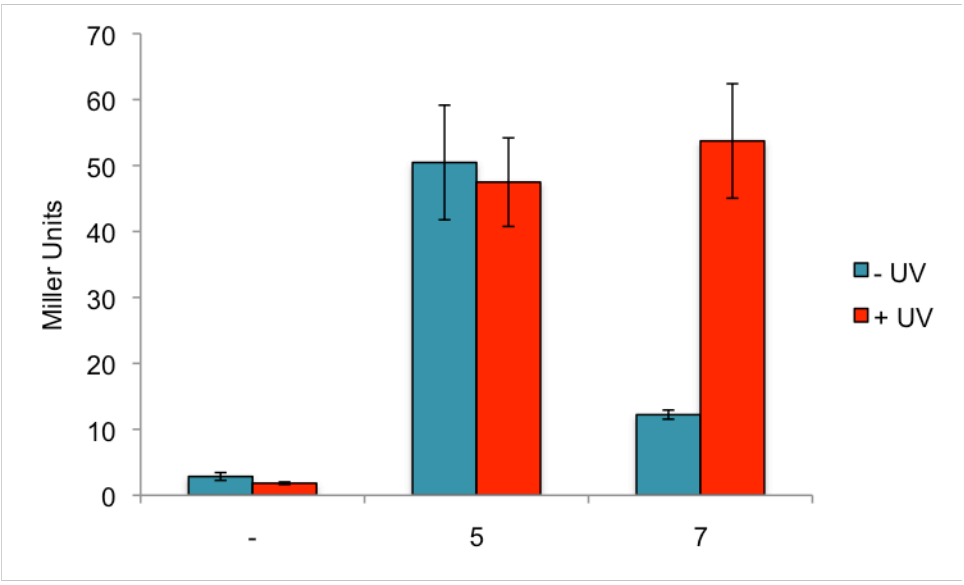


Figure 2.10. Miller assay to quantify β-galactosidase activity of *E. coli* induced with 9-Oxime **5** or photocaged erythromycin **7**; +/- UV irradiation. Photocaged erythromycin **7** induction + UV irradiation restored β-galactosidase activity to levels of the positive control. Error bars represent standard deviations from three independent experiments. Adapted from *Mol. BioSyst.*, **2011**,7, 2554.

Spatial control was then attempted using the GFPuv reporter. LB agar plates, containing ampicillin and tetracycline were pretreated with **7** and allowed to dry. Cells containing the pMLGFP plasmid were seeded onto the plate, half the plate was covered with an aluminum foil mask, and the uncovered half was then irradiated for 5 minutes at 365 nm. The plate was then incubated overnight at 37 °C before visualization. Initial attempts of spatial control on solid agar media showed a high background level of autofluorescence from the media when excited at 365 nm (Figure 2.11).

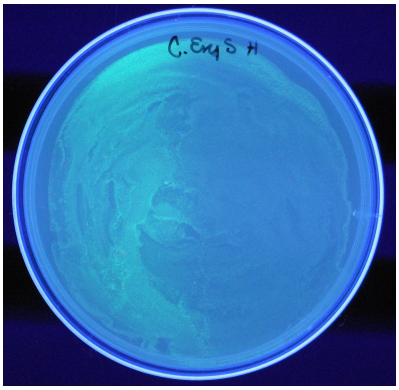


Figure 2.11. Spatial control of GFPuv expression using a photocaged erythromycin **7**. The left half of the plate was irradiated for 5 minutes at 365 nm, while the right half of the plate was covered with a mask. Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

In order to address the autofluorescence issue, the GFPuv gene was exchanged for eGFP, a mutant form of the green fluorescent protein, that exhibits increased fluorescence and a significantly red-shifted excitation wavelength compared to GFP.[36] The plasmid pMLGFP was digested with SpeI and PmeI and purified by gel electrophoresis. The *egfp* gene was amplified from pEGFP-N1 by PCR with Pfu Ultra DNA polymerase (Stratagene) to contain SpeI and PmeI restriction sites. Digested pMLGFP was then ligated with the PCR-amplified EGFP and then transformed into bacterial cells. Clones were sequenced to verify proper sequence and orientation of the EGFP gene. Positive clones were co-transformed into Top10 cells (genetically similar to DH10B) with pJZ12. We then repeated spatial control experiments with the eGFP reporter. Definitive spatial control was apparent (Figure 2.12).

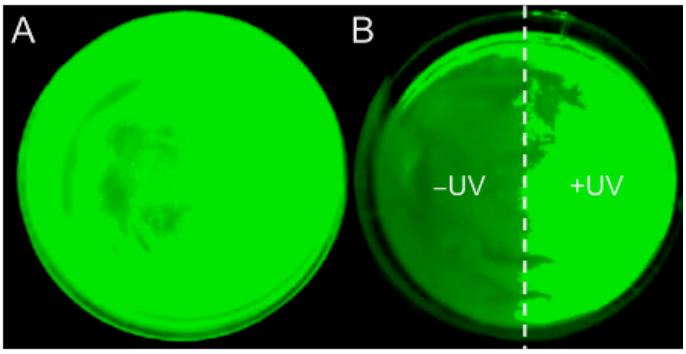


Figure 2.12. EGFP expression in a bacterial lawn exposed to (A) the 9-oxime erythromycin **5** and (B) the photocaged erythromycin **7** where half the plate was irradiated for 5 minutes at 365 nm. Adapted from *Mol. BioSyst.*, **2011**,7, 2554.

After demonstrating the MphR(A)/promoter in conjunction with **7** enables precise spatial and temporal control over gene expression, we then used this system to control bacterial signaling processes. Many of the modular components biological gene expression are analogous to engineered circuit devices. Genetic systems have previously been programmed to function as molecular computing devices, both *in vitro*[37,38] and *in vivo*.[39,40] In this context, logic gates are fundamental computational devices that generate outputs based on Boolean operations (AND, NOT, OR) between various inputs.[41] Recent developments in biological logic gates have paved the way for biological computing and engineering of synthetic biological circuits based on gene regulation.[1,37,42]

In comparison to other common inputs (e.g., small molecules, secondary metabolites, post-translational modifications, DNA, etc.), light allows for the programming of cellular circuitry with precise spatiotemporal control. As shown in Figure 2.13 A, the MphR/**7**-based AND gate uses irradiation with two different wavelengths (I_1 = 365 nm and I_2 = 483 nm) as

input signals and generates fluorescence emission at 509 nm as an output signal (O). The gate performs in agreement with its truth table (Figure 2.13), and thus sets the stage for the development of more complex photochemically controlled cellular circuits.

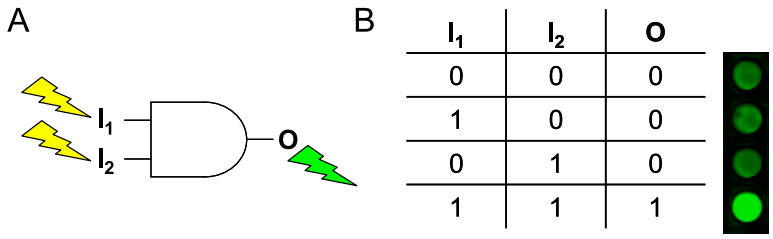


Figure 2.13. (A) Light-triggered bacterial AND gate and (B) truth table showing the resulting EGFP fluorescence imaged with a Typhoon scanner. Input I_1 = excitation at 365 nm, input I_2 = excitation at 483 nm, output O = emission at 509 nm. Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

Another common circuit in logic devices is a band-pass filter.[43] Band-pass filters are commonly used in signal processing and concentration-dependent band-pass filters have been applied to mammalian cells to create stable gene expression patterns.[44] We hypothesized that our light-activated gene expression system could exhibit band-pass like behavior based on the fact that the reporter gene activity responds to the irradiation time in a dose-dependent fashion until a maximum response is reached. Additionally, prolonged irradiation will not further increase the signal, but in turn will reduce it to basal levels due to the toxicity of the extended UV exposure. To test our hypothesis, **7** (1 μM) was added to cells containing pJZ12 and pMLEGFP in log phase and the cells were irradiated (365 nm, 25

W) for increasing periods of time, then allowed to express overnight. As expected, a band-pass shaped response was observed (Figure 2.14).

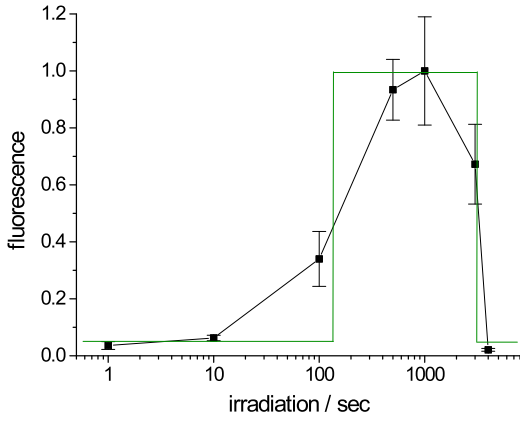


Figure 2.14. Irradiation-dependent band-pass filter of **7** based on EGFP expression. The band-pass filter is indicated in green. The fluorescence is normalized to the highest signal and the error bars represent standard deviations from three independent experiments. Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

To ensure this dramatic decrease in gene expression after prolonged irradiation was due to toxicity of UV irradiation, Top10 *E. coli* cells were grown to log phase, irradiated for varying amounts of time and their OD_{600} monitored over a 24 h period (Figure 2.15). A marked decrease in growth was observed after exposure to irradiation times of 1 h.

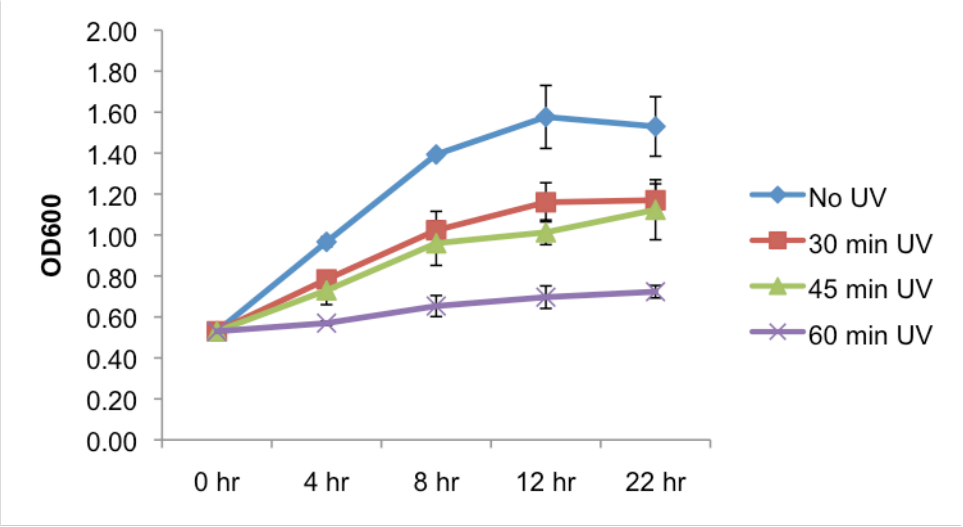


Figure 2.15. UV toxicity timecourse. Adapted from *Mol. BioSyst.*, **2011**,*7*, 2554.

In summary, we have demonstrated light-activatable gene expression in *E. coli* with an MphR(A)/promoter system that can be controlled with high spatial and temporal resolution. We have also applied this system to create a photo-activatable logic gate and bandpass filter, two building blocks of molecular computing devices. This technology is based on a photocaged erythromycin analogue **7** and provides a useful tool for the precise regulation of gene expression and thus the study of gene function. Notably, the MphR(A) regulatory system has previously been used in mammalian cells and mice [44,45] suggesting wide applicability of this system.

2.3 Photochemical control of DNA recombination in bacterial cells

DNA recombination is essential to many processes in pro- and eukaryotic cells such as the development of the immune system, antibiotic resistance, maintaining plasmid copy number and the translocation of genetic elements.[46] Site-specific recombinase proteins mediate these recombination events by catalyzing strand exchange between defined DNA target sites.[47] Two recombinase DNA target sites are bound by a monomer of the enzyme. A catalytic residue in the enzyme active site then acts as a nucleophile to cause single strand cleavage at a phosphodiester bond of the DNA backbone on each DNA molecule. This reaction generates a new 3′ phosphotyrosine bond and a 5′ OH, which allows for strand crossover to occur. The generated 5′ OH nucleophilic attacks the 3′ phosphotyrosine bond of the opposite DNA molecule, forming the Holliday Junction intermediate. Two more monomers of recombinase then bind the remaining two recognition sites and initiate strand cleavage. The newly recombined DNA is then generated after final crossover of strands (Figure 2.16).[48]

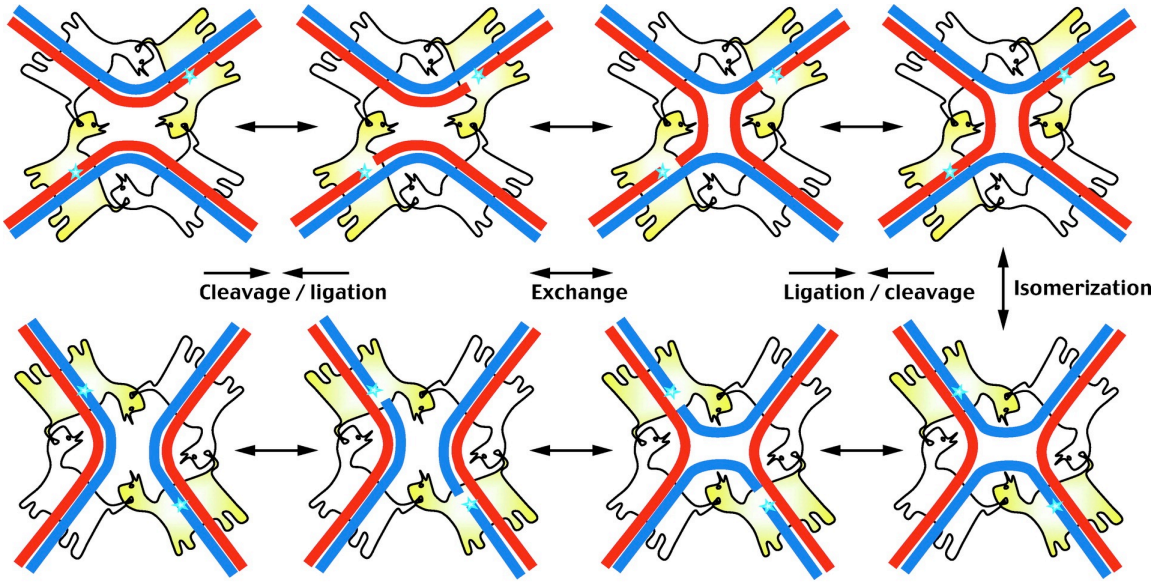


Figure 2.16. Site-specific recombination between double stranded DNA. Blue stars represent the catalytic center of the active recombinases (shown in yellow). Adapted from *Annu. Rev. Biochem.* **2006**, *75*, 567.

Depending on their active site residue, recombinases are classified as either tyrosine or serine recombinases.[48] Bacteriophage P1 Cre recombinase; a well-defined recombinase, initiates strand exchange between two 34-base pair *lox*P sites through a catalytic tyrosine residue.[46] Depending on the orientation of the two *lox*P sites, the DNA between them is either excised, integrated or inverted (Figure 2.17).[49]

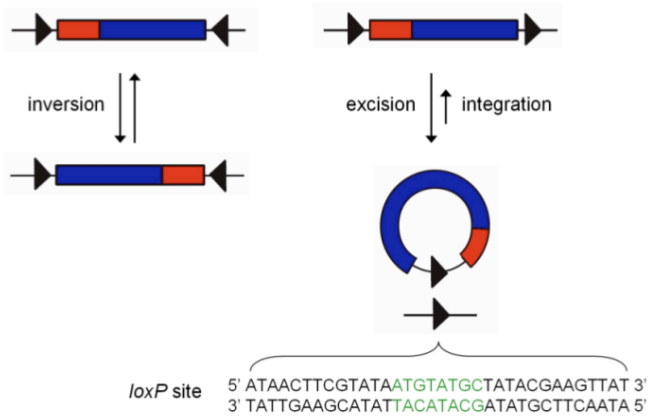


Figure 2.17. Cre DNA recombinase recognizes two, 34 base-pair *lox*P sites and either inverts, excises, or integrates the DNA between them; depending on their orientation. Adapted from *ACS Chem. Biol.* **2009**, *4*, 441.

Previously, several light-activated Cre enzymes have been developed, though photocaged hormones, a photocaged analogue of Cre, and a light inducible protein dimerization system.[50-52]

Cre has very low activity in *E. coli*, so to expand this technology to bacteria an alternative recombinase must be used. The *S. cerevisiae* Flippase recombinse (Flp), another member of the tyrosine family of recombinases, is analogous to Cre with the exception that it is functional in *E coli* as well as eukaryotic cells. Flp recognizes an 8 base-pair region flanked by two 13 base-pair sites, called *frt* sites, and excises/integrates or inverts the intervening DNA sequence. Construction of a photocaged version of Flp was also attempted in a similar manner to Cre, but protein could not be expressed (data not shown).

The reporter system to be used in conjunction with Flp recombinase, provided to us from the Voziyanov Lab (LA Tech), comprises of the pBAD33Flpe and pBU1 plasmids.[53]

The pBAD33Flpe plasmid contains a thermostable copy of the recombinase under control of the arabinose inducible pBAD promoter.[54] The pBUI plasmid encodes a *LacZ* gene that is flanked by two *frt* sites. In the presence of arabinose, Flpe is expressed, the *LacZ* gene is excised and no β-galactosidase activity is observed (Figure 2.18).

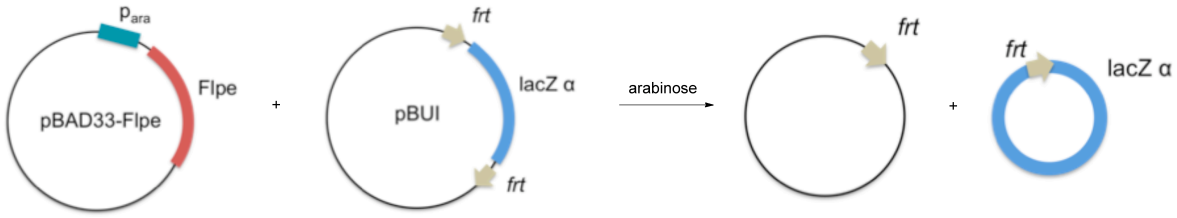


Figure 2.18. Plasmids for assaying Flpe expression in *E. coli*. Addition of arabinose induces expression of Flpe, that will then excise the *LacZ* gene from between two *frt* sites. Without an origin of replication, the *LacZ* gene is lost and no β-galactosidase is expressed.

In the absence of arabinose, Flpe is not expressed, the *LacZ* gene remains in the pBU1 plasmid, and β-galactosidase activity is observed. The expression of Flpe can be easily detected in *E. coli* by blue and white screening, which requires the addition of X-Gal (5-bromo-4-chloro-3-indolyl-*β*-D-galactopyranoside) to the growth media. X-Gal is a substrate for β-galactosidase that turns from colorless to blue when cleaved by the enzyme.

As we were unable to express photocaged Flpe, we then chose to photocage the inducer molecule of Flpe expression. We hypothesize that by synthesizing a photocaged analogue of L-arabinose; we will be able to exhibit spatiotemporal control over DNA recombination in *E. coli*, by regulating the expression of Flpe. This system is advantageous

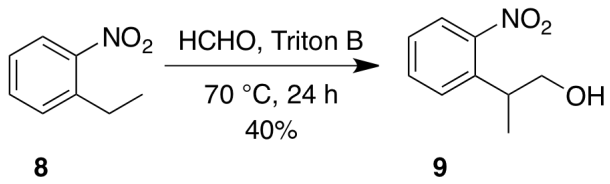
over the previously described light-activatable Cre as it will be functional in *E. coli* as well as eukaryotes such as *S. cerevisiae*.

A photocaged arabinose will not only be applicable to the Flpe system, but to any system in which a pBAD promoter regulates gene expression. Arabinose-inducible systems function similarly to IPTG-inducible systems. In the absence of arabinose, the homodimeric AraC protein (encoded by *araC*) is binds DNA to form a loop slightly upstream of the promoter.[55] This DNA loop prevents RNAP from transcribing genes downstream of pBAD . In the presence of arabinose, the AraC protein binds the sugar and changes its conformation to allow RNAP to bind the promoter and begin transcription.

2.3.1 Synthesis of a photocaged arabinose

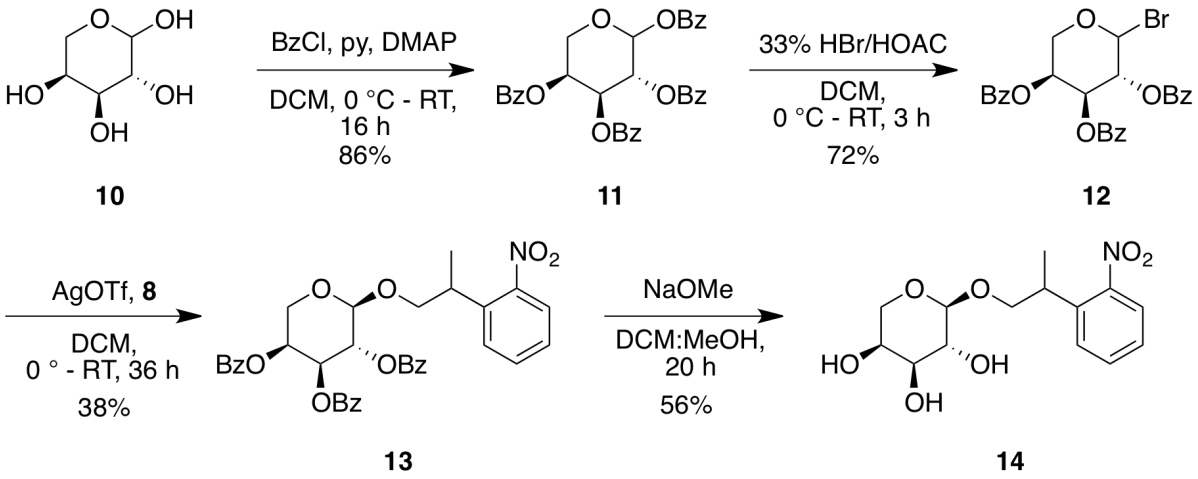
The 2-(2-nitrophenyl)propanol (NPP-OH) caging group was chosen as the photocaging group for arabinose over the traditional *o*-NB group for the following reasons: It has been reported that substitutions at the benzylic position lead to increases in photodeprotection rates,[56] and initial attempts to react arabinose with benzyl alcohol, 4-nitro benzyl alcohol and 2-nitro benzyl alcohol were unsuccessful.

NPP-OH was synthesized in one step according to literature procedure.[57] 2-Ethylnitrobenzene **8** was refluxed with paraformaldehyde and Triton-B (40% in methanol) for 24 hours to obtain **9** in 40% yield (Scheme 2.2).



Scheme 2.2. Synthesis of NPP-OH caging group.

To synthesize photocaged arabinose, L-(+)-arabinose **10** was reacted with benzoyl chloride, in the presence of pyridine and DMAP to yield 1,2,3,4-tetra-*O*-benzoyl-L-arabinopyranose (**11**), as a mixture of α and β isomers, in 86% yield. The tetrabenzoyl-protected sugar **11** was then brominated at the anomeric position with hydrogen bromide (33% w/v in acetic acid) to produce **12** in 72% yield. The brominated sugar **12** was then immediately reacted with the NPP-OH caging group and silver triflate in dichloromethane (DCM) overnight in the dark to yield the tribenzoyl photocaged L-arabinose **13** in 38% yield. Initial attempts to remove the benzoyl protecting groups from the sugar with ammonia in methanol were unsuccessful. Reaction of **13** with sodium methylate in methanol and DCM (3:1 MeOH:DCM) produced photocaged arabinose **14** in 56% yield (Scheme 7).



Scheme 2.3. Synthesis of photocaged arabinose **14**.

TLC confirmed decaging of compound **14** at a 1 mM concentration after 15 minutes of UV irradiation (handheld UV lamp, 365 nm).

2.3.2 In vivo control of DNA recombination with photocaged arabinose

Photocaged arabinose was then tested in *E. coli* for its ability to induce expression of genes under control of the arabinose-inducible pBAD promoter. Decaging was tested in Top10 cells transformed with pBADEGFP, which contains a copy of eGFP under control of pBAD. Cells were gown to log phase, and then induced with either arabinose **10** (100 μM) or photocaged arabinose **14** (100 μM), and irradiated for 15 minutes at 365 nm. Cells were then allowed to grow overnight before being harvested by centrifugation and lysed. Fluorescence was read on a plate reader (483/509 ex/em, Biotek Synergy 4 Microplate Reader) (Figure 2.19).

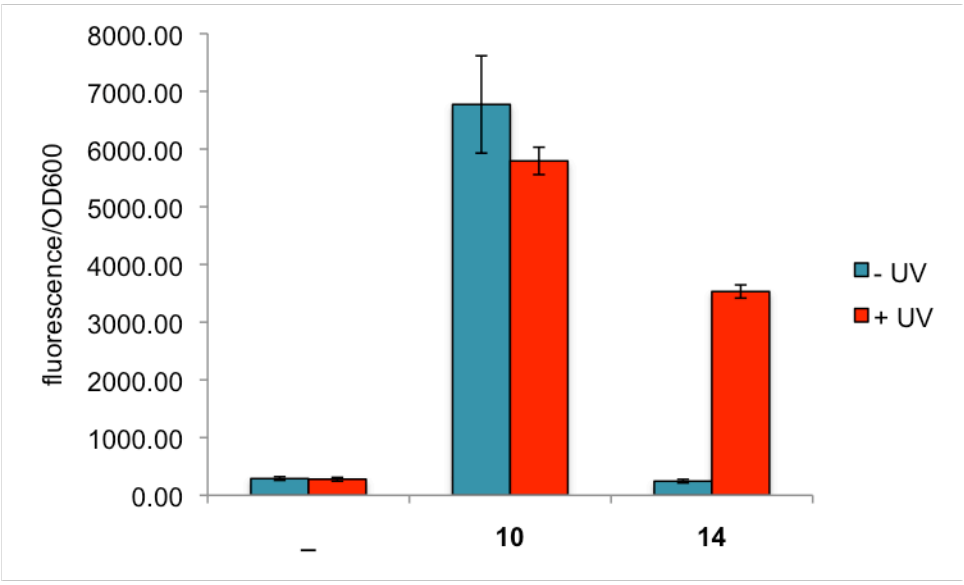


Figure 2.19. Decaging of photocaged arabinose **14** in liquid culture. Samples were irradiated for 15 minutes at 365 nm (25 W), followed by overnight protein expression. Fluorescence was measured by a Biotek Synergy 4 platereader (483/509 ex/em). Approximately 60% of fluorescence was restored after irradiation. Error bars represent standard deviations from three independent experiments.

Photocaged arabinose was then tested using the two-plasmid Flpe system.[53] pBAD33Flpe was transformed into competent cells harboring the plasmid pBUI and plated on LB agar plates containing the ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL) and pre-treated with either arabinose or photocaged arabinose (40 μL, 10 mM) and X-Gal (40 μL, 20% in DMF). One plate treated with photocaged arabinose was irradiated for 15 minutes. Plates were then incubated overnight at 37 °C. β-Galactosidase activity, and therefore Flpe expression, was measured by assessing the frequency of blue colonies. As expected, in the absence of any inducer molecule, β-galactosidase was expressed and colonies appeared blue on X-Gal media (Figure 2.20 A). In the presence of arabinose **10**,

active Flpe was produced and no β-galactosidase was observed due to an intracellular recombination event (Figure 2.20 B). In the presence of photocaged arabinose **14**, Flpe expression was blocked and β-galactosidase activity was inferred from blue colony color (Figure 2.20 C). However, irradiated plates pre-treated with photocaged arabinose **14** did not show complete activation of Flpe expression, and appeared predominantly blue on X-Gal media (Figure 2.20 D).

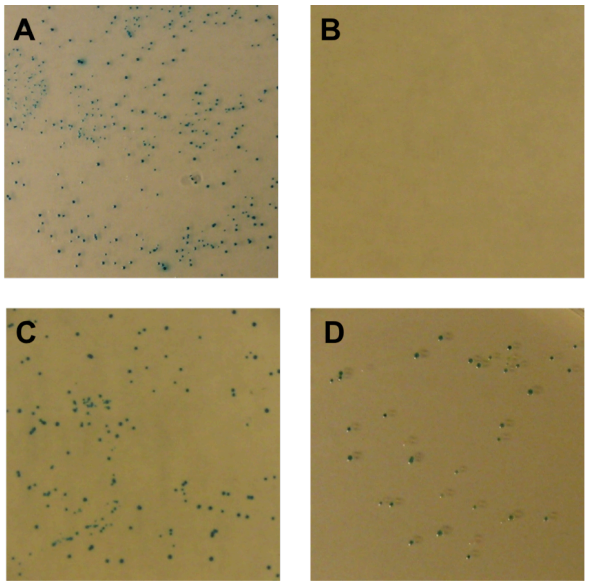


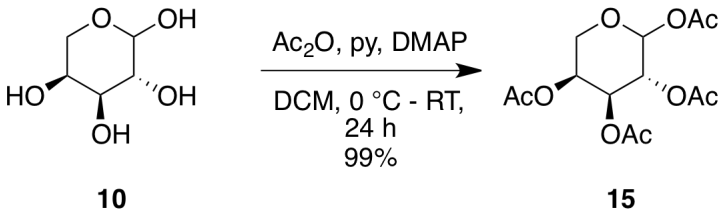
Figure 2.20. Flpe recombination activity assay. A) No compound, B) arabinose **10**, C) photocaged arabinose **14** -UV, D) photocaged arabinose **14** +UV (15 minutes).

Our initial results lead us to believe we are not achieving quantitative decaging of photocaged arabinose. Prolonged irradiations times of plates pre-treated with **14** for 20-30

minutes did not result in additional Flpe expression. After further investigation into the arabinose transport pathway, it was revealed that in bacterial cells arabinose undergoes an active transport pathway utilizing the AraE transport protein.[58,59] Thus, our photocaged arabinose is likely not recognized by the AraE protein and fails to enter the cell.

We then attempted to see if we could utilize a prodrug strategy to allow for passive transport of the photocaged sugar. Prodrugs are analogues of drugs molecules that are only activated *in vivo* and are used to alleviate solubility, membrane permeability or toxicity problems during drug delivery.[60] The most common form of chemical modification is the inclusion of esters functionalities to increase a drug's lipophilicity. Ester bonds are often used to mask nucleophilic or charged functional groups and are easily cleaved *in vivo* by endogenous esterase enzymes.[60,61] Acetylated versions of various sugars have shown to have increased bioactivity over their free hydroxyl counterparts.[62,63]

We synthesized a simple tetra-acetylated version of L-(+)-arabinose in one step through reaction of the sugar with pyridine and acetic anhydride (Scheme 2.4).



10 **15**

Scheme 2.4. Synthesis of 1,2,3,4-tetra-*O*-acetyl-L-arabinopyranose **15.**

The 1,2,3,4-tetra-*O*-acetyl-L-arabinopyranose (**15**) was then assayed for activation of gene expression in bacterial cells. *E. coli* containing pBADeGFP (Top10) were gown to log phase, and then induced with either arabinose **10** (100 μM) or **15** (100 μM) or uninduced. Cells were then allowed to grow overnight before being harvested by centrifugation and lysed. Fluorescence was read on a plate reader (483/509 ex/em, Biotek Synergy 4 Microplate Reader) (Figure 2.21).

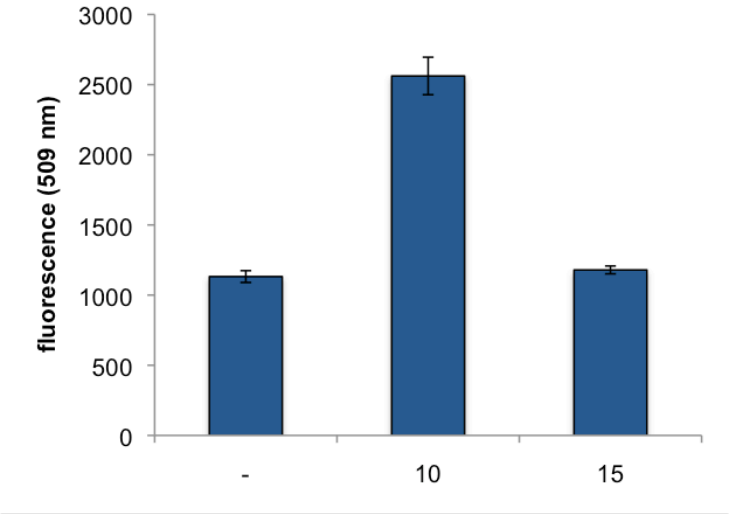


Figure 2.21. Arabinose controlled eGFP expression in bacterial cells.

No increase in gene expression was observed for cells induced with **15** when compared to the negative control, indicating the *E. coli* are either not transporting **15** into the cell or hydrolysis of the ester bonds was incomplete.

2.4 Summary and Outlook

We have successfully used small molecules as a means to achieve photochemical control over genetic circuits at the transcriptional level. A photocaged erythromycin was developed to effectively achieve spatial and temporal control over gene expression and as a component of light-activatable switches in bacterial cells. The photocaged erythromycin and MphR(A)/promoter system can be translated into the construction of larger synthetic systems.

We have also successfully synthesized a photocaged analogue of the sugar arabinose as an additional means to photochemically control gene expression in *E. coli* through the control recombinase activity. We found that the photocaged arabinose was not able to fully restore gene expression levels in p_{Bad} systems after irradiation, most likely due to the fact that arabinose is brought into *E. coli* through an active transport pathway. We believe the *araE* transport protein cannot recognize the photocaged arabinose and therefore cannot accumulate enough inducer molecule within the cell for gene activation. Our attempts to utilize a pro-drug strategy were also unsuccessful at activating gene expression. Either the acetyl groups did not sufficiently increase the cell permeability of the compound or they are not being quantitatively cleaved within the cell. We currently are exploring other methods to increase the cell permeability of our photocaged arabinose, which could be a useful tool for synthetic biology systems that already contain pBAD promoter switches.

2.5 Experimental Methods

Evaluation of erythromycin analogues. *E. coli* cells (GeneHogs DH10B) containing pJZ12 and pMLGFP were grown to saturation in LB broth containing ampicillin (50 μg/mL) and tetracycline (15 μg/mL). An aliquot (40 μL) was spread onto agar plates (10 cm) pre-treated with 40 μL of erythromycin or 9-oxime erythromycin (0.5 mg/mL in DMSO) and incubated overnight at 37 ºC, followed by visualization on a transilluminator (365 nm, 25 W).

Erythromycin decaging time course. *E. coli* (GeneHogs DH10B) containing pMLGFP and pJZ12 were grown at 37 °C overnight in LB media containing ampicillin (50 μg/mL) and tetracycline (15 μg/mL). Cells were diluted 50-fold into a 25 mL culture and grown to OD_{600} = 0.6 at 37 °C. Compound (either caged erythromycin **4** or 9-oxime erythromycin **5** in DMSO) was added to a final concentration of 1 μM. Cells (1 mL) treated with caged erythromycin were transferred to a 10 cm petri dish and irradiated at 365 nm for 30 sec, 1 min, 2 min, and 5 min. The cells were then transferred to a 96 well plate and grown for 12 hours at 37 °C, then pelleted by centrifugation and re-suspended in lysis buffer (100 μL, 50 mM Tris Base, 1 mM EDTA, 100 mM NaCl, 0.2% Triton X) and fluorescence was read on a plate reader (395/509 ex/em, Biotek Synergy 4 Microplate Reader).

β-Galactosidase (Miller) Assay. *E coli* (Top10, Invitrogen) containing pJZ12 and pBAD-mphR-P-lacZ were grown at 37 °C overnight in LB media containing ampicillin (50 μg/mL)

and tetracycline (15 μg/mL) . Cells were diluted 50-fold into a 25 mL culture and grown to OD_{600} = 0.6 at 37 °C. Compound (either caged erythromycin **7** or 9-oxime erythromycin **5** in DMSO) was added to a final concentration of 1 μM. Half of the cells (1 mL) treated with caged erythromycin and 9-oxime erythromycin were transferred to a 5 cm petri dish and irradiated at 365 nm for 5 minutes, and the remainder of the cells were not irradiated. The cells were then grown for 6 hours at 37 °C, and the OD_{600} was recorded. Cells (100 μL) were added to 900 μL of Z-buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 20 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethaonol), 25 μL of 0.1% SDS, 50 μL of $CHCl_3$ and vortexed for 1 minute. *O*-nitrophenyl-β-D-galactoside (200 μL, 4 mg/mL) substrate was added to each sample and incubated at 30 °C for 10 minutes. Na_2CO_3 (500 μL, 1 M) was added to stop the reaction. The absorption at 420 nm and 550 nm was recorded and Miller units were calculated using the following formula: $(1000 \times (Abs_{420} - 1.75 \times Abs_{550})/(\text{reaction time in minutes} \times \text{final volume in mL} \times OD_{600})$. Error bars represent standard deviations of three independent experiments.

Cloning of pMLEGFP. pMLGFP (3957 bp, 500 ng) was digested with SpeI and PmeI at 37 °C for 1 hour, and separated on a 1% argarose gel in TBE buffer at 80 V for 45 minutes. The digested plasmid backbone was excised from the gel and purified with the Qiagen Gel Extraction kit. EGFP was amplified from pEGFP-N1 by PCR with Pfu Ultra DNA polymerase (Stratagene) with forward (5'-GAACTAGTATGGTGAGCAAGGGCGAG-3') and reverse (5'-TTACTTGTACAGCTCGTCCATGCC-3', the 5' terminus was

phosphorylated) primers (IDT). The PCR product was verified by gel electrophoresis with a band apparent at ~800bp, and was purified, digested with SpeI and purified again on an ion-exchange column (Qiagen). Digested pMLGFP (50 ng) was then ligated with the PCR-amplified EGFP in 1:6, 1:3, and 1:0 ratios using T4 DNA ligase (NEB). Ligations were performed at 4 °C overnight, and then transformed into GC5 competent cells. Clones from the 1:6 ligation were sequenced to verify proper sequence and orientation of the EGFP gene. Positive clones were co-transformed into Top10 cells (genetically similar to DH10B) with pJZ12.

Bacterial logic gate. *E. coli* (Top10, Invitrogen) containing pMLEGFP and pJZ12 were at 37 °C overnight in LB media containing ampicillin (50 μg/mL) and tetracycline (15 μg/mL). Cells were diluted 50-fold into a 25 mL culture and grown to OD_{600} = 0.6 at 37 °C. **7** was added to a final concentration of 1 μM. Cells that were to receive I_1 were irradiated for 5 minutes (UVP High Performance Transilluminator, 365 nm, 25 W) in a petri dish (5 cm), and then were transferred to a 96 well plate, and grown overnight, as well as cells that were not irradiated. Cells were pelleted by centrifugation and resuspended in lysis buffer and were imaged on a Typhoon scanner.

Spatial control of EGFP expression. *E. coli* cells (Top10, Invitrogen) containing pJZ12 and pMLEGFP were grown to saturation in LB broth containing ampicillin (50 μg/mL) and tetracycline (15 μg/mL). An aliquot (10 μL) was spread onto agar plates (5 cm) pre-treated

with **7** or **5** (10 μL, 0.5 mg/mL in DMSO stock solution). Plates containing **7** were placed on a transilluminator (UVP High Performance Transilluminator, 365 nm, 25 W) and one half of the plate was covered with an aluminum foil mask. The other side of the plate was irradiated for 5 minutes at 365 nm. Plates were then incubated overnight at 37 °C, and then imaged for EGFP expression. (Typhoon FLA 7000, GE; FAM filter 473 nm/[Y520]).

Bandpass filter experiment. *E. coli* (Top10, Invitrogen) containing pMLEGFP and pJZ12 were at 37 °C overnight in LB media containing ampicillin (50 μg/mL) and tetracycline (15 μg/mL). Cells were diluted 50-fold into a 25 mL culture and grown to OD_{600} = 0.6 at 37 °C. **7** was added to a final concentration of 1 μM to 15 mL of cells. 1 mL aliquots of cells were then transferred to separate petri dishes (5 cm) and irradiated for 0 sec, 1 sec, 10 sec, 100 sec, 500 sec, 1000 sec, 3000 sec and 4000 sec on a transilluminator (UVP High Performance Transilluminator, 365 nm, 25W), transferred to a 96 well plate and allowed to grow overnight at 37 °C. Cells were pelleted by centrifugation and resuspended in lysis buffer. Fluorescence was read on a plate reader.

UV Toxicity Study. Top10 cells were grown in LB media to OD_{600} = 0.56 at 37 °C, and then 5 mL aliquots were irradiated for 30, 45, or 60 minutes (365 nm, 25 W), transferred to a 96 deep well block and allowed to grow at 37 °C with shaking. The optical density was recorded every 4 hours, for 12 hours, and after 22 hours incubation. Cells exhibited virtually no growth after being irradiated for 60 min.

Synthesis of 2-(2-nitrophenyl)propanol (9). Triton B (40% in MeOH, 4.66 mL, 11.18 mmol) and paraformaldehyde (1.12 g, 37.30 mmol) were added to 2-ethylnitrobenzene **8** (1.00 mL, 7.46 mmol). The reaction mixture was heated under reflux for 24 h and then concentrated *in vacuo*. The reaction mixture was neutralized with HCl (1 M, 10 mL), and then extracted with EtOAC (3 × 10 mL). The combined organic layers were washed with saturated aqueous $NaHCO_3$ (10 mL), brine (10 mL), dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel chromatography eluting with hexanes/EtOAc mixture (4:1) to give **9 (**550 mg, 40%) as a light yellow oil. The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[57]

Synthesis of 1,2,3,4-tetra-*O*-benzoyl-L-arabinopyranose (11). Triethylamine (3 mL, 21.52 mmol) was added at 0 ºC to a solution of L-(+)-arabinose **10** (500 mg, 3.33 mmol) and DMAP (0.5 mg, .004 mmol) in DCM (2 mL). Benzoyl chloride (2.3 mL, 19.95 mmol) was added to the reaction mixture at 0 ºC over 5 min. The reaction was stirred at r.t. overnight and then quenched with MeOH (1 mL) at 0 ºC. The solvent was then evaporated *in vacuo* and the reaction mixture was resuspended in EtOAc (15 mL). The solution was then washed with aqueous HCl (15 mL, 1M), saturated aqueous $NaHCO_3$ (15 mL), and brine (15 mL), and then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The product was purified by silica gel chromatography using hexanes/EtOAc mixture (9:1 → 4:1 → 1:1), yielding **11**

(1.47 g, 86%) as a light yellow solid. The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[64]

Synthesis of 2,3,4-tri-*O*-benzoyl-α-bromo-L-arabinopyranose (12). 1,2,3,4-tetra-*O*-benzoyl-L-arabinopyranose **11** (500 mg, 0.88 mmol) was dissolved in DCM (2.3 mL) and 200 μL of HBr (33% in AcOH) was added dropwise to the solution at 0 ºC. The reaction was stirred at 0 ºC for 2 h, and then diluted in DCM (15 mL), washed with ice water (10 mL), saturated aqueous $NaHCO_3$ (10 mL) and brine (10 mL) and dried over anhydrous Na_2SO_4. After filtration, the solvent was evaporated *in vacuo* to yield **12** as an off-white sticky solid (335 mg, 72% yield). The analytical data obtained matched reported literature results (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[64]

Synthesis of 2,3,4-tri-*O*-benzoyl photocaged arabinose (13). 2,3,4-tri-*O*-benzoyl-α-bromo-L-arabinopyranose **12** (100 mg, 0.19 mmol) was dissolved in DCM (500 μL) and cooled to 0 ºC. Silver trifluoromethanesulfonate (98 mg, 0.38 mmol) was added to the solution and the reaction mixture was stirred for 5 min at 0 ºC. 2-(2-nitrophenyl)propanol **9** (52 mg, 0.29 mmol) was added at 0 ºC, and the reaction mixture was stirred in the dark. After 36 h, the reaction was diluted in DCM (5 mL), washed with saturated aqueous $NaHCO_3$ (5 mL), brine (5 mL) and dried over anhydrous Na_2SO_4. The product was purified using silica gel chromatography eluting with hexanes/EtOAc (4:1) to give **13** as a white solid (45 mg, 38% yield). (R_f = 0.3). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 1.36 (dd, J^1 = 12.3 Hz, J^2 = 5.7, 3H),

3.58-3.78 (m, 2H), 3.86 (d, J = 5.4 Hz, 2H), 3.90-3.99 (m, 1H), 5.30 (dd, J^1 = 9.9 Hz, J^2 = 3.6 Hz, 1H), 5.65-5.69 (m, 2H), 5.82-5.90 (m, 1H), 7.19-7.60 (m, 15H), 7.82-7.90 (m, 2H), 7.97 (d, J = 6.9 Hz, 1H), 8.08 (d, J = 6.9 Hz, 1H). LRMS: *m/z* calcd for ($C_{35}H_{31}NO_{10}$) $[M+Na]^+$: 648.99; found: 648.20

Synthesis of photocaged arabinose (14). 2,3,4-tri-*O*-benzoyl photocaged arabinose **13** (45 mg, 0.07 mmol) was dissolved in DCM (400 μL). In a separate vial NaOMe (14 mg, 0.26 mmol) was dissolved in 100 μL MeOH, and the solution was added to the first solution and stirred at r.t. for 16 h. The reaction mixture was then neutralized with Dowex ion exchange resin (50WX8 50-100 mesh). The resin was filtered and washed with DCM and the combined washings were concentrated *in vacuo*. The product **14** was crystallized from ether as a white solid (13 mg, 56% yield). ^{1}H NMR (400 MHz, CD_3OD): δ = 1.36 (dd, J^1 = 7.5 Hz, J^2 = 4.5 Hz, 3H), 3.19 (m, 1H), 3.40 (dd, J^1 = 6.5 Hz, J^2 = 0.9 Hz, 2H), 3.55-3.70 (m, 3H), 3.78-3.80 (m, 2H), 4.71-4.74 (m, 1H), 7.38-7.42 (m, 1H), 7.62-7.64 (m, 2H), 7.71-7.72 (m, 1H). ^{13}C NMR (300 MHz, CD_3OD): δ = 16.9, 34.2, 63.0, 68.9, 69.4, 72.5, 73.2, 99.8, 123.5, 127.2, 128.3, 128.6, 132.4, 137.7. LRMS: *m/z* calcd for ($C_{14}H_{19}NO_7$) $[M+Na]^+$: 336.11; found: 336.10.

Decaging of photocaged arabinose *in vitro*. Solutions of photocaged arabinose **14** (1 mM) in hexanes were prepared for decaging experiments. A 1 mL aliquot was transferred to a quartz cuvette and irradiated for 5, 10 or 15 min with a handheld UV lamp (365 nm); taking

samples for TLC after each time point. Samples were then run on TLC (MeOH:DCM 1:9) and visualized with a UV lamp. Complete decaging was observed after 15 min, with ~50% decaging observed after 5 min.

Decaging of photocaged arabinose in bacterial cells, liquid culture. *E. coli* (Top10) containing pBAD-eGFP were grown at 37 °C overnight in LB media containing ampicillin (50 μg/mL). Cells were diluted 50-fold into a 25 mL culture and grown to OD_{600} = 0.6 at 37 °C. Compound (either **10** or **14** in DMSO) was added to a final concentration of 100 μM. Cells (4 mL) containing either arabinose, caged arabinose or no compound were transferred to a 10 cm petri dish and irradiated for 15 minutes (UVP High Performance Transilluminator, 365 nm, 25W). After irradiation the cells were grown in a 96-deep well block and incubated overnight at 37 °C, their OD_{600} read, then pelleted by centrifugation and re-suspended in lysis buffer (100 μL, 50 mM Tris Base, 1 mM EDTA, 100 mM NaCl, 0.2% Triton X) and fluorescence was read on a plate reader (483/509 ex/em, Biotek Synergy 4 Microplate Reader). The fluorescence was normalized to OD_{600}.

Decaging of Photocaged Arabinose in Bacterial Cells, on agar plates. *E. coli* (Top10) containing pBAD-eGFP were grown at 37 °C overnight in 3 mL LB media containing ampicillin (50 μg/mL). Cells (50 μL) were plated onto LB agar plates containing ampicillin (50 μg/mL) and pretreated with **10** or **14** (50 μL, 10 mM). One plated treated with photocaged arabinose was irradiated on a transilluminator for 5 min (UVP High Performance

Transilluminator, 365 nm, 25W). All plates were then incubated overnight and fluorescence was visualized on a Typhoon scanner (Typhoon FLA 7000, GE; FAM filter 473 nm/[Y520]).

Flpe Recombination Activity Assay in Bacterial Cells. Plasmids pBUI and pBAD33Flpe were transformed into Top10 cells and plated on media containing ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL). A single colony containing both plasmids was inoculated into 3 mL of LB containing ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL) and grown overnight at 37 °C. Agar plates containing ampicillin (50 μg/mL), chloramphenicol (34 μg/mL) and preteated with 40 μL X-Gal (20% in DMF) and either 10 μL of **10** or **14** (100 μM) were prepared. After the plates had been allowed to dry, one plate treated with photocaged arabinose was irradiated on a transilluminator for 15 min (UVP High Performance Transilluminator, 365 nm, 25W). After irradiation, 10 μL of a 1:1000 dilution of the overnight culture was plated on the plates and incubated overnight at 37 °C. Plates were then stored at 4 °C for 4-6 h to develop the blue color, and then photographed.

Synthesis of 1,2,3,4-tetra-*O*-acetyl-L-arabinopyranose (15). Pyridine (1.06 mL, 13.33 mmol) was added to a solution of L-(+)-arabinose **10** (800 mg, 5.33 mmol) and catalytic DMAP (0.5 mg, 0.004 mmol) in DCM (3 mL) at 0 ºC. Acetic anhydride (3.00 mL, 31.79 mmol) was added at 0 ºC over 5 min to the solution. The reaction mixture was stirred at r.t. overnight and then diluted in saturated aqueous $NaHCO_3$ (10 mL) and extracted with Et_2O (15 mL) for 15 min with stirring. The organic layer was then washed brine (15 mL), and then

dried over anhydrous Na_2SO_4, filtered and the solvent was evaporated *in vacuo* yielding **15** as white sticky solid (1.68 g, 99% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[65]

Activation of gene expression *via* an acetylated arabinose prodrug. *E. coli* (Top10) containing pBADeGFP were grown at 37 °C overnight in LB media containing ampicillin (50 μg/mL). Cells were diluted 50-fold into a 10 mL culture and grown to OD_{600} = 0.6 at 37 °C. Compound (either **10** or **15**) was added to a final concentration of 100 μM. The cells were grown in a 96-deep well block and incubated overnight at 37 °C, then pelleted by centrifugation and re-suspended in lysis buffer (100 μL, 50 mM Tris Base, 1 mM EDTA, 100 mM NaCl, 0.2% Triton X) and fluorescence was read on a plate reader (483/509 ex/em, Biotek Synergy 4 Microplate Reader).

CHAPTER 3 – UNNATURAL AMINO ACID MUTAGENESIS

3.1 Introduction to unnatural amino acid mutagenesis

Proteins are essential molecules in living cells. Examples of proteins are enzymes, antibodies, toxins and they are involved in signal transduction, transcription, cell motility and numerous other cellular functions.[66] Though the number and functions of proteins is vast, natural proteins are comprised of only the 20 canonical amino acids (and in some rare cases selenocysteine and pyrrolysine) that are encoded by 61 3-base codons. These 20 amino acids suffice for some proteins, but many proteins must undergo post-translational modifications such as methylation, phosphorylation, and glycosylation.[67] By expanding the genetic code of prokaryotes and eukaryotes with new amino acids we will gain great insight into protein structure and function as well as develop enhanced proteins with novel biological properties.[67,68]

Additionally, proteins represent excellent targets for the engineering of light-activated cellular networks as site-specific modification of active residues can be achieved through the *in vivo* incorporation of photocaged amino acids. Re-engineering of the genetic code to develop caged proteins has seen application in prokaryotes and eukaryotes as regulators of gene circuit.[67,69] Various photocaged amino acids are accessible for incorporation, which in turn makes a wide variety of protein gene-regulators accessible for modification. The light-regulation of enzymatic function was achieved through incorporation of photocaged analogues of tyrosine,[51] cysteine,[70] serine,[71] and lysine.[72]

Previous methods of introducing amino acids outside of the 20 natural amino acids (termed 'unnatural' amino acids) into proteins required *in vitro* methods such as solid phase peptide synthesis or the injection of chemically-acylated tRNAs into living cells.[67] These methods allowed for the development of novel proteins, but have various limitations. For instance, solid phase peptide synthesis can only yield proteins of limited size and the injection of chemically-misacylated tRNAs is limited by the stoichiometric amount of aminoacylated tRNA available as it cannot be reacylated *in vivo*.

Wang and Schultz reported in 2001 a genetically encoded method of unnatural amino acid (UAA) incorporation.[73] This method uses engineered suppressor tRNAs as the machinery to shuttle in unnatural amino acids during translation. Suppressor tRNAs incorporate natural amino acids in response to the stop codons UAA, UGA, UAG (ochre, opal and amber; respectively) instead of signaling translation termination. By evolving the suppressor tRNA and its corresponding aminoacyl-tRNA synthetase (aaRS) to selectively incorporate a UAA, new amino acids can be efficiently translated into proteins in response to stop codons. The amber stop codon, UAG, is the codon of choice for UAA incorporation as it is the least prevalent translation terminating codon in Nature. This prevents unwanted incorporation of UAAs at locations where translation should be terminated. An evolved tRNA/aminoacyl-tRNA synthetase pair from the archaea *Methanococuss jannaschii* was successfully used to incorporate *O*-methyltyrosine into *E. coli*. This amino acid and others have since been incorporated into eukaryotes.[74]

As a requirement for selective UAA incorporation, a tRNA/aaRS pair must be evolved to be orthogonal to all endogenous tRNA/aaRS pairs of the host cell (Figure 3.1). An orthogonal pair requires that the synthetase only charges its cognate tRNA with the unnatural amino acid of interest and the tRNA is not acylated by any other aaRS naturally occurring in the host organism.[75] The engineered synthetase will charge its tRNA with the unnatural amino acid for incorporation into the growing polypeptide chain by the ribosome, in response to an amber stop codon, UAG.

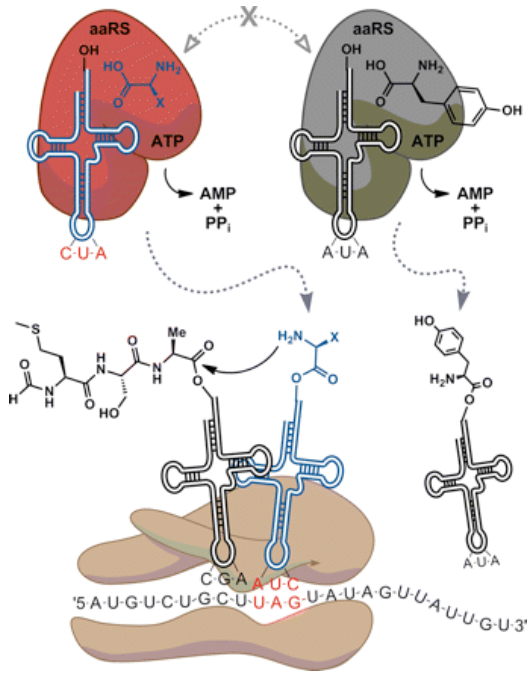


Figure 3.1. Amino-acylation of an orthogonal tRNA (blue) by an evolved cognate tRNA synthetase (red) for unnatural amino acid incorporation in response to an amber stop codon, UAG. X = unnatural amino acid side-chain. Adapted from *J. Biol. Chem.* **2010**, *285*, 11039.

To incorporate UAAs into proteins *in vivo*, a two-plasmid system is often used for prokaryotic systems.[76] One plasmid encodes the evolved synthetase and its corresponding suppressor tRNA. The second plasmid contains the gene of interest, usually under control of an inducible promoter, which has a codon mutated to the amber stop codon. Without amber suppression, the resulting protein will be truncated. To express full-length protein, the two plasmids must be co-transformed into cells and grown in media containing the unnatural amino acid.[76]

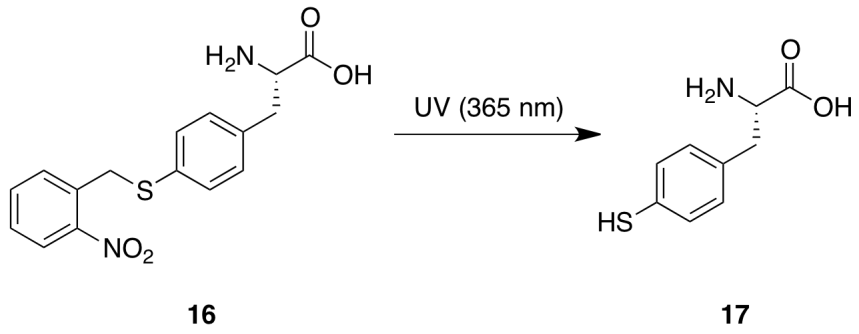
Previous UAAs that have been incorporated have introduced much new functionality to proteins including photocrosslinking moieties, fluorophores, bioconjugation handles and photocaged amino acids.[68] Various photocaged amino acids have been incorporated into *E. coli* and *S. cerevisiae.*[70,77] Using this technology, the Deiters lab has developed a variety of photocaged proteins that can be controlled with light.[51,78-80]

3.2 Incorporation of photocaged thio-tyrosine for novel bioconjugation reactions

The incorporation of the photocaged tyrosine into proteins has been used in our group as a means to control enzyme activity – the enzyme is rendered inactive when caged, but after decaging activity is restored.[51,78-80] Photocaging groups can also be used to function as a mask during the UAA incorporation process. Previously, this concept was used to incorporate fluorotyrosines into proteins.[81] Fluorotyrosine analogues could not be selectively incorporated into proteins, as their structure is not significantly different from natural tyrosine.[82] By photocaging various fluorotyrosines, an orthogonal tRNA/tRNA

synthetase pair could be engineered that would selectively incorporate the UAAs. After decaging, the fluorinated tyrosine residues were then revealed.

We hypothesize that by using a photocaging group as a mask; as well as a means for spatiotemporal control over protein function, we will be able to incorporate thio-tyrosine **17** into proteins. After irradiation of the photocaged thio-tyrosine **16**, a thio-tyrosine residue is generated (Scheme 3.1) Thio-tyrosine contains a thiophenol which is known to have an increased nucleophilicity and acidity compared to phenol.[83]

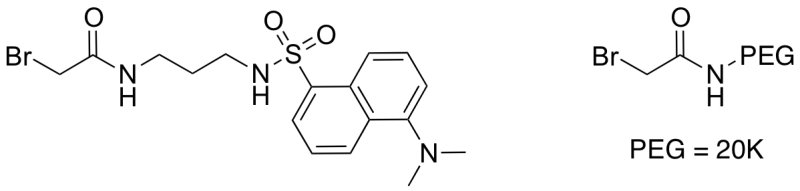


16 **17**

Scheme 3.1. Light activation of photocaged thio-tyrosine (PCSY) to yield thio-tyrosine **17**. The amino acid **16** was synthesized by Rajendra Uprety (Deiters Lab).

The introduction of the thio-phenol functionality will be applied to novel bioconjugation reactions for specific delivery of target molecules to cells, as well as a means to study protein-protein interactions through disulfide linkages. Reactions of a thio-phenol residue are also orthogonal to those of alkyl-thiol molecules (such as cysteine) and will allow for specific labeling of proteins. Two thio-phenol reactive compounds have been synthesized

in our lab by Rajendra Uprety: a fluorescent probe **18** for labeling and a PEG **19** for protein stabilization (Figure 3.2).



18 **19**

Figure 3.2. Thio-reactive linkers; fluorescent dansyl **18** and PEG **19**.

Thio-tyrosine was first incorporated into the small protein ubiquitin (8 kDa) using the pSUP-E10 plasmid. This plasmid, developed by the Schultz lab,[76] contains 6 copies of the tRNA gene and one copy of a mutated ONBY synthetase (evolved by the Cropp Lab at VCU).[81] To test protein expression, BL21(DE3)PLysS cells harboring pSUP-E10 and pETUbK48TAG were grown in LB media containing chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL) to log phase, then ONB-tyrosine (ONBY) or PCSY **16** were added to a 1 mM final concentration. Protein expression was induced with 1 mM IPTG and continued overnight. Cultures were harvested by centrifugation and the cell lysate was analyzed on a 16% SDS PAGE gel and visualized with coomassie blue (Figure 3.3).

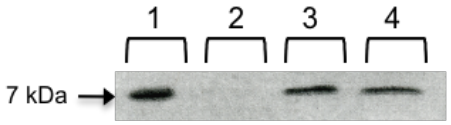


Figure 3.3. SDS-PAGE of ubiquitin expression. Lane 1: Molecular weight ladder. Lane 2: no amino acid added. Lane 3: 1 mM ONBY. Lane 4: 1 mM PCSY **16**.

Initial studies revealed that PCSY was incorporated with similar fidelity to ONBY. The E10 synthetase gene was then cloned into the more efficient pEVOL plasmid.[84] The pEVOL system differs from the original pSUP system in the number of synthetase genes and tRNA genes. pEVOL contains two copies of the synthetase gene; one inducible and one constitutively expressed, and only one copy of the tRNA gene. pEVOL-E10 was created by exchanging the two copies of the ONBY synthetase from pEVOL-ONBY for the E10 synthetase genes. Putative clones were sequenced to confirm both copies of the E10 gene were in proper alignment and orientation. A positive clone was then transformed into BL21G cells and made competent for further experiments.

We then attempted to incorporate **16** into myoglobin, another small (17 kDa) protein, with a TAG site engineered on the surface of the protein. A solvent accessible location for the thio-tyrosine residue is essential for bioconjugation reactions. pET24MyoD4TAG was transformed into BL21G cells harboring pEVOL-E10, and protein expression was conducted in LB media (pH = 7.4) containing chloramphenicol (34 μg/mL), kanamycin (50 μg/mL), and either ONBY or PCSY **16** (1 mM). Protein expression was induced at log phase ($OD_{600} \approx$ 0.6) with 0.02% arabinose and 0.1 mM IPTG, and continued overnight. Protein was then

purified on Ni-NTA resin, followed by analysis on a 12% SDS PAGE gel stained with coomassie blue (Figure 3.4).

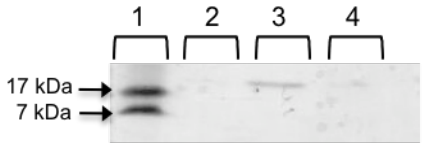


Figure 3.4. SDS-PAGE of myoglobin expression. Lane 1: Molecular weight ladder. Lane 2: no amino acid. Lane 3: 1 mM ONBY. Lane 4: 1 mM PCSY **16**.

Initial protein expression yielded very low amounts of myoglobin (<1 mg/L expression culture). Changes in plasmid systems (between pSUP-E10 and pEVOL-E10), expression times, temperature, IPTG concentration, amino acid concentration, pH, media and purification methods failed to give significant amounts of myoglobin.

We then used the pEVOL-E10 system to express T7 RNA polymerase (T7RNAP, 99 kDa). pBH161Y639TAG, a plasmid containing a 6X his-tagged T7RNAP mutated to contain an amber stop codon,[79] was transformed into BL21G cells containing pEVOL-E10. Cells harboring both plasmids were then used to inoculate 25 mL of LB media containing ampicillin (50 μg/mL) and chloramphenicol (34 μg/mL), and either ONBY or PCSY **16** (1 mM). Cells were grown to log phase (OD_{600} = 0.6), and protein expression was induced with 0.1 mM IPTG and 0.02% arabinose. Protein was expressed overnight, and cells were

harvested by centrifugation, lysed and protein was purified on a Ni-NTA resin. Samples were analyzed on a 12% SDS-PAGE gel stained with coomassie blue (Figure 3.5).

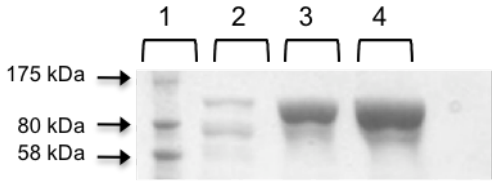


Figure 3.5. SDS-PAGE of pEVOL-E10 + pBH161Y639TAG T7RNAP protein expression. Lane 1: Molecular weight ladder, Lane 2: no amino acid, Lane 3: 1 mM ONBY, Lane 4: 1 mM PCSY **16**.

Though protein expression of photocaged T7RNAP yielded very good amounts of protein (5-6 mg/L expression culture), we did not pursue bioconjugation studies as the amber codon resides within the active site of the protein. The thio-tyrosine residue would not be solvent accessible after decaging and would likely not be able to bioconjugate. However, T7RNAP activity studies were conducted. Our photocaged T7RNAP was tested for its ability to synthesize RNA using radiolabeled ^{32}P-labeled ATP. After 20 minutes of UV irradiation, no T7RNAP activity was detected (Figure 3.6).

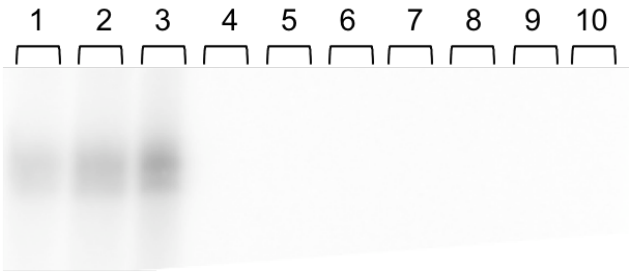


Figure 3.6. Radioactively labeled RNA synthesis by wild-type T7RNAP or photocaged T7RNAP. Lane 1-3: WT T7RNAP, Lane 4: photocaged T7RNAP 0 min UV, Lane 5: photocaged T7RNAP 1 min UV, Lane 6: photocaged T7RNAP 2 min UV, Lane 7: photocaged T7RNAP 5 min UV, Lane 8: photocaged T7RNAP 10 min UV, Lane 9: photocaged T7RNAP 15 min UV, Lane 10: photocaged T7RNAP 20 min UV.

We then returned to the protein ubiquitin for bioconjugation studies. The previous ubiquitin expression plasmid, pETUbK48TAG, did not contain a 6X his-tag for facile nickel affinity chromatography. A 6X his-tagged UbK48TAG was supplied to us from our collaborators in the Cropp lab (VCU). Unfortunately, protein expression levels of ubiquitin with PCSY were similar to that of myoglobin using either pSUP-E10 or pEVOL-E10 (Figure 3.7). Bioconjugation studies are being currently carried out with PCSY **16** in the Cropp lab.

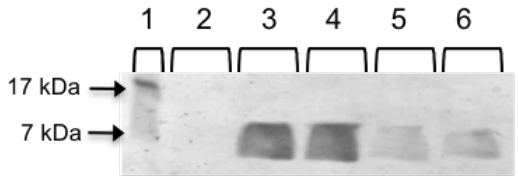


Figure 3.7. SDS-PAGE of pEVOL-E10 + pETUbK48TAG6XHis protein expression. Lane 1: Molecular weight ladder. Lane 2: no amino acid. Lane 3: 1 mM ONBY (0.1 mM IPTG induction). Lane 4: 1 mM ONBY (1 mM IPTG induction). Lane 5: 1 mM PCSY (0.1 mM IPTG induction). Lane 6: 1 mM PCSY (1 mM IPTG induction).

A likely cause of the continually low protein yields with PCSY is that the synthetase used did not undergo rounds of selection for PCSY, but is only a modified ONBY synthetase. The resulting E10 synthetase contains the following mutations compared to the ONBY RS; His70Met, Phe108Gly, Ile159Met.[81] Currently, PCSY is being incorporated with success into bacterial cells using a PylRS synthetase.

3.3 Expanding the eukaryotic genetic code

Interest has grown to expand the technology of unnatural amino acid incorporation to eukaryotes following the success in prokaryotes.[85-87] The *M. jannaschii* tyrosyl pair that is used in *E. coli* is unfortunately not orthogonal in *Saccharomyces cerevisiae* so a new system needed to be developed.[88] In 2003, Chin et al. reported a method to incorporate UAAs in *S. cerevisiae* using the orthogonal *E. coli* tyrosyl-tRNA synthetase (EcYRS)/$tRNA_{CUA}$ pair.[75] As much of the cellular machinery is conserved in eukaryotes, this system can also be used in mammalian cells.

The discovery of an orthogonal aaRS that is selective for a specific unnatural amino acid starts with a library of mutant *E. coli* tyrosyl or leucyl synthetases. The library is created by mutating up to 6 residues near the tyrosyl or leucyl synthetase active site to all possible natural amino acids. The resulting library is on the order of 10^8 clones. This library is subjected to multiple rounds of positive and negative selections to identify a new synthetase.

The system used to evolve an orthogonal pair from the library of mutant aaRS makes use of the Gal4 transcriptional activating protein. The Gal4 protein is comprised of two

domains, a DNA binding domain (BD) and a transcriptional activating domain (AD). For transcription to occur, the BD must bind DNA and be in close proximity to the AD. DNA binding alone does not initiate transcription, and likewise the AD domain alone cannot activate transcription. By placing an amber stop codon into the *GAL4* gene sequence this system can be used to screen for orthogonal amber suppressor tRNA/tRNA synthetase pairs. Expression of reporter genes under control of Gal4 responsive promoters can be correlated to full-length Gal4 expression and in turn, amber suppression (Figure 3.8).

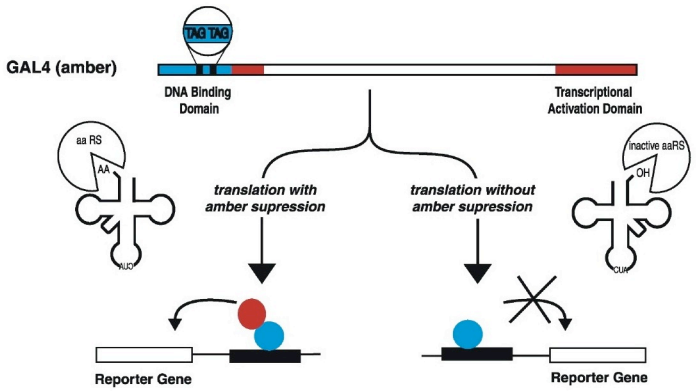


Figure 3.8. Schematic of amber suppression in the *GAL4* gene to activate transcription of reporter genes. Adapted from *Chem Biol,* **2003,** *10*, 511.

S. cerevisiae MaV203 cells containing pGADGAL4(2TAG) are used for selections.[75] This strain is auxotrophic in leucine and tryptophan biosynthesis for plasmids with selectable markers and contains deletions in *GAL4* and *GAL80* genes. MaV203 cells are

also auxotrophic in histidine and uracil biosynthesis and contain three reporter genes under control of Gal4 responsive promoters; *HIS3*, *URA3* and *lacZ*. *HIS3* and *URA3* are nutritional reporter genes whose products are involved in the biosynthesis of histidine and uracil; respectively. *LacZ*, as previously mentioned, expresses β-galactosidase and can be used in blue and white screening (Figure 3.9).

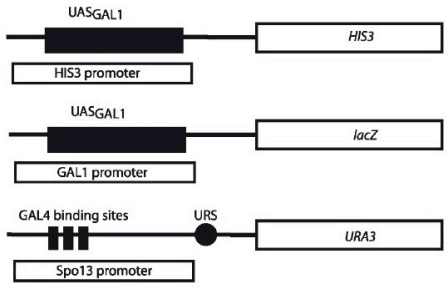


Figure 3.9. Reporter genes used in selections in *S. cerevisiae*. Gal4 recognizes all three of the promoters, but with varied affinity. Adapted from *Chem Biol,* **2003,** *10*, 511.

The pGADGAL4(2TAG) plasmid contains two amber stop codons within the *GAL4* sequence. Suppression of these stop codons will lead to expression of full length Gal4, and will activate transcription of *HIS3, URA3,* and *lacZ*. Expression of these nutritional reporters allows for growth on media lacking histidine or uracil. Blue and white screening can also be used on media containing X-Gal to detect expression of *lacZ*. Failure to suppress the amber codons will produce truncated Gal4, which will not activate transcription of the nutritional

reporter genes or *lacZ*. Cells will die on media lacking histidine or uracil when truncated Gal4 is expressed, or appear white on media containing X-Gal.

The selection process commences by transforming the tRNA synthetase library into MaV203 cells containing the pGADGAL4(2TAG) plasmid. These cells are then amplified and plated on positive selection media (–histidine or –uracil) in the presence of 1 mM UAA. Cells that grow on media lacking histidine are correctly suppressing amber stop codons with either the UAA or a natural amino acid. 3-aminotriazole (3-AT), a competitive inhibitor of *HIS3*, is also included in media lacking histidine. The addition of 3-AT increases the stringency of this selection in a dose dependent manner. Alternatively, cells can also be plated on media containing 1 mM UAA but lacking uracil. Growth on this media also indicates successful amber suppression with a UAA or a natural amino acid. The -uracil positive selection is considered more stringent than the –histidine media.

Clones that have survived the positive selection are collected, amplified and washed (to remove any media that may contain histidine or uracil). Cells are then plated onto media that *lacks* UAA, and contains 5-fluoroorotic acid (5-FOA). 5-FOA is converted to the toxic byproduct 5-fluorouracil by the enzyme encoded by *URA3*; orotidine-5-phosphate decarboxylase. Cells that survive this selection are *not* incorporating a natural amino acid into *GAL4*, which then does not activate transcription of *URA3*, and 5-FOA remains non-toxic. Clones that die in this selection are incorporating a natural amino acid into *GAL4*, and express *URA3*, which in turn converts 5-FOA to 5-FU and kills the cell.

The library undergoes three rounds of positive and negative selections before clones showing the correct phenotypes are isolated, sequenced, and then utilized for unnatural protein expression in *S. cerevisiae.*

3.3.1 Incorporation of photocaged histidine

We are using this selection strategy to evolve an orthogonal synthetase/tRNA pair that is selective for photocaged histidine. Though histidines are prevalent in many enzymes, few examples of modified or unnatural histidines are reported in literature.[89,90] Histidines have many biological functions and are found in the active site of many metal binding proteins such as hemoglobin, carbonic anhydrase and zinc-finger proteins.[91] Histidine binds metal ions such as Fe^{2+}, Co^{2+}, Cu^{2+}, Zn^{2+} and Ni^{2+} through its imidazole side chain.[92] The imidazole ring can also act as an acid or base at neutral pH to facilitate catalysis and is used as a purification tag for affinity chromatography. The incorporation of photocaged histidine will allow for spatial and temporal control over specific histidine binding events.

Photocaged histidines **20** and **21** were synthesized by Hrvoje Lusic (Figure 3.10). The two structurally similar analogues were synthesized to allow for flexibility in the incorporation process.

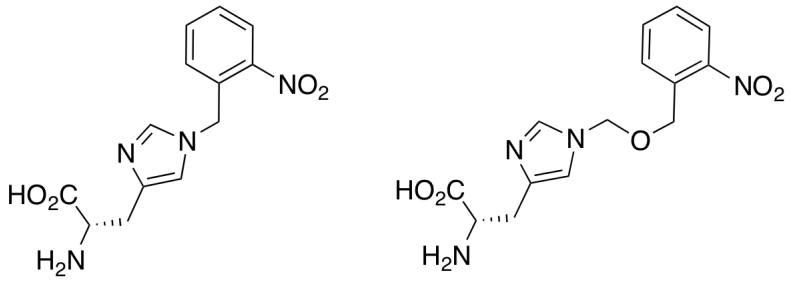


Figure 3.10. Photocaged histidines analogues **20** and **21** to be used for unnatural amino acid incorporation.

We are currently attempting to evolve an orthogonal tRNA/aaRS pairs for either **20** or **21** in both *E. coli* and *S. cerevisiae*. A full selection has been carried out once in prokaryotic cells and multiple times in the eukaryotic system but has consistently given inconclusive results. Currently, a control selection for the previously incorporated *O*-methyl tyrosine is being performed with the leucyl RS library and tyrosyl RS library in *S. cerevisiae*. Until the method can be proven to be reproducible for the positive control we cannot definitely say that an aaRS cannot be evolved for photocaged histidine. We have established our media to be correct by screening for expected phenotypes using the wild-type tyrosyl-synthetase (WT), an evolved *o*-methyl tyrosine tyrosyl-synthetase (OMeY YRS), an evolved *o*-methyl tyrosine leucyl-synthetase (OMeY LRS), and an inactive synthetase (A5). Also included was an inactive synthetase (C1), that was isolated from a previous *o*-methyl tyrosine selection. On positive selection media (Figure 3.11), OMeY YRS and OMeY LRS grow only in the presence of the unnatural amino acid. The WT synthetase grows on this media independent of UAA addition. The inactive A5 clone does not grow on any positive selection media, as it

does not incorporate any amino acid into Gal4. Our isolated C1 clone grows unselectively on all positive selection media. On negative selection media (Figure 3.11) OMeY YRS, OMeY LRS and A5 are the only clones that grow in the absence of amino acid. In the presence of amino acid on negative selection media, all clones surprisingly grow. We attribute this to a variation in the pH of this media, but as it is not a condition used in the selection process, this is not an immediate concern.

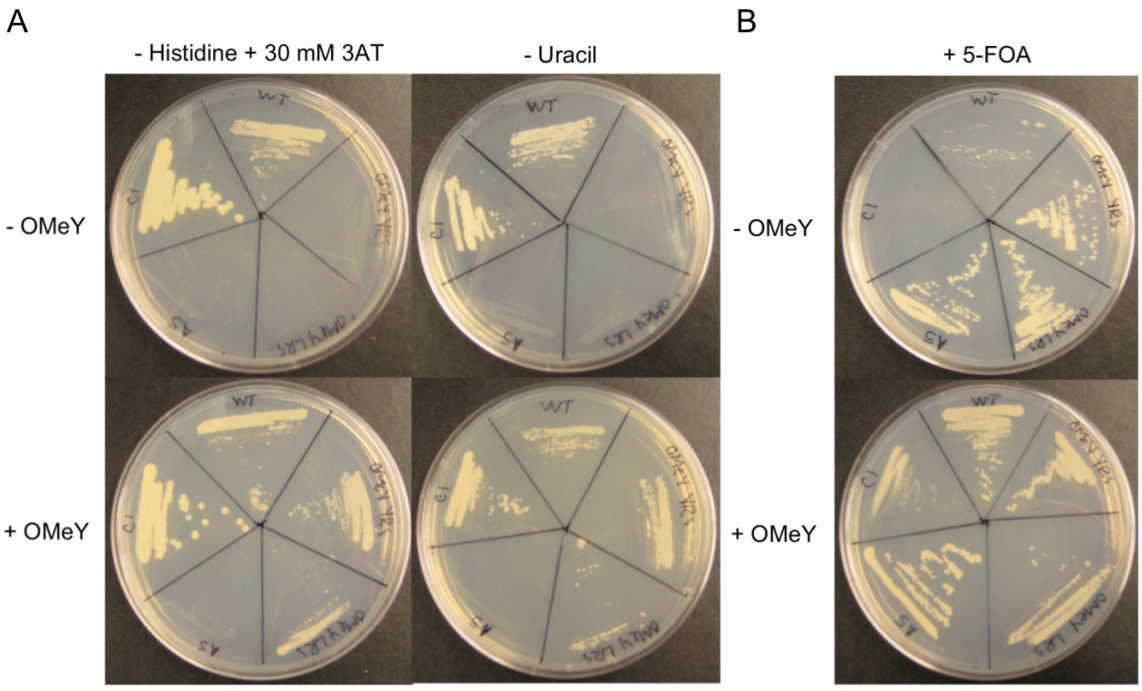


Figure 3.11. Phenotypes of known clones grown on selective media. (A) Various Synthetase clones grown on positive selection media in the absence (- OMeY) and in the presence (+ OMeY) amino acid. (B) Synthetase clones grown on negative selection media in the absence (- OMeY) and in the presence (+ OMeY) amino acid. Clones are plated clockwise from top: WT, OMeY YRS, OMeY LRS, A5, C1.

We then attempted to sequence the library to ensure it maintained full diversity. Initial attempts to isolate plasmid DNA from yeast were inefficient and the extracted DNA was low quality. After multiple attempts to sequencing the isolated DNA failed, we then began reconstruction of the original leucyl tRNA synthetase library. Homologous recombination has been used in yeast to construct libraries of $>10^7$ mutants through a simple cloning and transformation process.[93] Yeast will homologously recombine pieces of linear DNA into circular DNA when the linear portions contain 50 base pairs of overlap homology. We amplified the LRS gene in three portions to contain the mutations at positions Met40, Leu41, Tyr499 and Tyr527, generating a small library of ~1.6 x 10^5 clones. Each PCR product would contain 50 bp of overlap to with the next fragment or the pESC backbone. The mutations were introduced during PCR by using primers containing NNK codons. The fragments produced PCR products of approximately 200 bp, 1.1 kb and 1.3 kb in size, as expected (Figure 3.12).

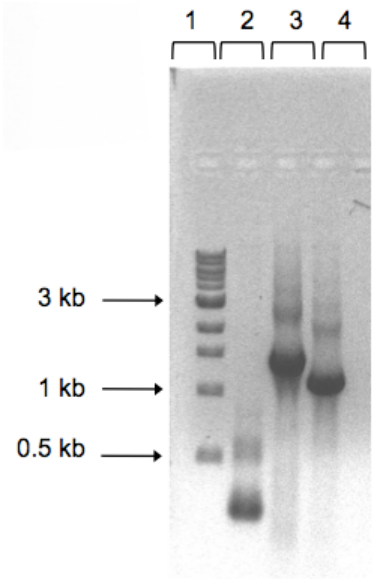


Figure 3.12. PCR products of LRS gene amplification. The gene was amplified in 3 fragments, a 200 bp, 1.3 kb and 1.1 kb portion. Each portion contained 50 bp of overlap to the portion of the gene flank it. As seen in this gel, most amplification was specific and very few bands were observed from non-specific binding.

After the linearized pESC backbone and PCR products were in hand, we then proceeded to transform the fragments into MaV203 cells following a protocol established by the Rao Lab (NC State University). The transformations did not yield high enough transformation efficiency for the library size. Currently the library is being reconstructed via our collaborators in the Cropp Lab (VCU).

3.4 Summary and Outlook

We have begun steps to incorporate novel photocaged unnatural amino acids into proteins in bacterial and yeast cells. The photocaged thio-tyrosine amino acid has been

successfully incorporated into multiple proteins (myoglobin, ubiquitin and T7RNAP), but with varying overall yields. Ideally photocaged thio-tyrosine will be efficiently incorporated into a protein at a location that is solvent accessible, such as on the surface of a protein. This amino acid will then be used for bioconjugation studies. Currently, photocaged thio-tyrosine is not efficiently incorporated into proteins with the E10 tyrosyl aaRS, but has been successfully incorporated into proteins with variants of the pyrrolsyl aaRS. Jihe Liu (Deiters Lab) is currently optimizing the expressions with photocaged thio-tyrosine before carrying out bioconjugation studies.

We have also begun the evolution of new amino-acyl tRNA synthetases in both bacterial and yeast cells. After addressing technical issues of the yeast selection such as media preparation, we have begun the construction of a new library of the leucyl tRNA synthetase. The library is currently being transformed into yeast cells, after which a new selection will begin with the photocaged histidine analogues. The incorporation of these two new classes of amino acids; photocaged thio-tyrosine and photocaged histidines, will allow for extensive manipulation of protein function *in vivo*.

3.5 Experimental Methods

Ubiquitin Expression with photocaged thio-tyrosine. Plasmids pSUPE10 and pETUbK48TAG were transformed into BL21(DE3)PLysS cells and grown overnight on media supplemented with chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL). A single

colony was then inoculated into 3 mL of LB media containing chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL) and grown overnight with shaking at 37 °C. In the morning, the culture was diluted 1:100 into 2XYT (10 mL) media containing chloramphenicol (34 μg/mL), ampicillin (50 μg/mL) and either ONBY or PCSY **16** (1 mM, pre-dissolved in hot 50% H_2O:DMSO) then grown to OD_{600} = 0.6. Protein expression was induced with 1 mM IPTG, and continued for 24 h at 37 °C. Cells were then harvested by centrifugation, resuspended in DI H_2O (1 mL) and a 10 μL aliquot was boiled with loading dye and run on a 16% SDS-PAGE gel. Protein bands were then visualized with Coomassie staining.

Cloning of pEVOLE10. pSUPE10 (1 μg) was digested with NdeI and PstI for 2 h at 37 °C, and the E10 aaRS insert was separated from the backbone on a 1% agarose gel (70V, 30 min). The E10 gene was excised from the gel and purified using a Qiagen gel purification kit. pEVOL-ONBY was digested with NdeI and PstI for 2 h at 37 °C, and the pEVOL backbone was separated from the ONBY aaRS insert on a 1% agarose gel (70V, 30 min). The pEVOL backbone was excised from the gel and purified using a Qiagen gel purification kit. The pEVOL backbone and the E10 insert were then ligated (1:6 backbone:insert) at 4 °C using T4 DNA ligase. The ligation was transformed into NovaBlue competent cells and plated on LB agar supplemented with chloramphenicol (34 μg/mL). The DNA from 3 colonies on the resulting ligation plate was isolated and digested with EcoR1 for 1.5 h at 37 °C, then analyzed on a 1% agarose gel (70V, 30 min). A positive clone showing two bands was used for further cloning steps.

The E10 aaRS gene was PCR amplified from pSUPE10 using Phusion DNA polymerase and primers that incorporated BglII and SalI restriction sites on the 5' and 3' ends, respectively. The PCR product was purified using a Qiagen nucleotide removal kit, and then digested with BglII and SalI for 3 h at 37 °C. The pEVOL backbone containing one copy of E10 was also digested with BglII and SalI for 3 h at 37 °C. and the backbone was separated from the ONBY aaRS insert on a 1% agarose gel (70V, 30 min). The pEVOL backbone was excised from the gel and purified using a Qiagen gel purification kit. . The backbone and the E10 insert were then ligated (1:6 backbone:insert) at 4 °C using T4 DNA ligase. The ligation was transformed into NovaBlue competent cells and plated on LB agar supplemented with chloramphenicol (34 μg/mL). Positive clones were confirmed by sequencing.

Myoglobin expression with photocaged thio-tyrosine. Plasmid pET24MyoD4TAG was transformed into BL21(DE3)PLysS cells harboring pEVOLE10 and grown overnight on media supplemented with chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL). A single colony was then inoculated into 3 mL of LB media containing chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL) and grown overnight with shaking at 37 °C. In the morning, the culture was diluted 1:100 into LB media (25 mL) media containing chloramphenicol (34 μg/mL), ampicillin (50 μg/mL) and either ONBY or PCSY **16** (1 mM, pre-dissolved in hot 50% H_2O:DMSO) then grown to OD_{600} = 0.6. Protein expression was induced with 0.1 mM IPTG and 0.02% arabinose, and continued for 24 h at 37 °C. Cells were then harvested by

centrifugation, resuspended in lysis buffer (300mM NaCl, 5mM NaH_2PO_4, 45mM Na_2HPO_4, 10mM Imidazole) containing lysozyme (4 μL, 1 mg/mL) and incubated on ice for 2 h. Cells were then lysed by sonication and the cleared lysate was applied to Ni-NTA resin (Qiagen) and incubated at 4 °C for 1 h. The resin was then washed with a low imidazole buffer (10-20 mM) and protein was eluted from the resin with a high imidazole buffer (250 mM). Purified protein was then run on a 12% SDS-PAGE gel and visualized with Coomassie stain. Protein concentration was determined by Bradford Assay (BioRad).

T7RNAP Expression with photocaged thio-tyrosine. Plasmid pBH161Y639TAG was transformed into BL21(DE3)PLysS cells harboring pEVOLE10 and grown overnight on media supplemented with chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL). A single colony was then inoculated into 3 mL of LB media containing chloramphenicol (34 μg/mL) and ampicillin (50 μg/mL) and grown overnight with shaking at 37 °C. In the morning, the culture was diluted 1:100 into LB media (25 mL) media containing chloramphenicol (34 μg/mL), ampicillin (50 μg/mL) and either ONBY or PCSY **16** (1 mM, pre-dissolved in hot 50% H_2O:DMSO) then grown to OD_{600} = 0.6. Protein expression was induced with 0.1 mM IPTG and 0.02% arabinose, and continued for 24 h at 37 °C. Cells were then harvested by centrifugation, resuspended in a lysis buffer containing lysozyme and incubated on ice for 2 h. Cells were then lysed by sonication and the cleared lysate was applied to Ni-NTA resin and incubated at 4 °C for 1 h. The resin was then washed with a low imidazole buffer (10-20 mM) and protein was eluted from the resin with a high imidazole buffer (250 mM). Purified

protein was then run on a 12% SDS-PAGE gel and visualized with Coomassie stain. Protein concentration was determined by Bradford Assay.

In vitro T7RNAP transcription assay. Purified T7RNAP (480 ng) was diluted in 100 μL of DPEC H_2O, and 6 μL aliquots were irradiated on a transilluminator in a PCR tube for 1 min, 2 min, 5 min, 10 min, 15 min an 20 min at 365 nm. After decaging, a master mix (4 μL) comprised of the oligonucleotide template (25 μM), RNase inhibitor (5 U), 5X T7RNAP buffer, NTP mix (12.5 mM each nucleotide) and α-^{32}P ATP was added to each 6 μL aliquot of decaged T7RNAP. The transcription reaction was performed at 37 °C for 2 h and then the product was separated on a 10% urea Tris-boric acid-DTA polyacrylamide gel (150V, 45 min). The gel was then exposed to a phosphor-imaging screen overnight after which the screen was imaged on a Typhoon phosphorimager (Typhoon FLA 7000, GE) and quantified.

Evolution of an orthogonal synthetase in E. coli. Plasmids pBK-JYRS, pREP/YC-J17 and pLWJ17B3 (available from P.G. Schultz) are used in the selection process. Electrocompetent cells containing either pREP/YC-J17 (the positive selection plasmid) or pLWJ17B3 (the negative selection plasmid) were prepared according to standard protocol [94] and had a transformation efficiency of ~1.0 x 10^8 CFU/μg DNA. The library plasmid (pBK-JYRS, 800 ng) was transformed into ten 50 μL aliquots of electrocompetent cells containing pREP/YC-J17, yielding ~10^9 independent transformants; 10-fold the size of the library. Transformants were recovered in SOC media for 1 h and then resuspended in 1 L LB media supplemented

with kanamycin (50 μg/mL) and tetracycline (5 μg/mL) and grown to OD_{600} = 1.0. 500 μL of culture was then plated onto 10 positive selection plates (GMML media, 5 μg/mL tetracycline, 50 μg/mL kanamycin, 34 μg/mL chloramphenicol and 1 mM UAA). Plates were incubated for 48 h at 37 °C. Surviving clones were scraped from the plate using 10 mL LB media and their plasmid DNA isolated using a Qiagen miniprep kit (84 ng/μL, 50 μL). The pBK-JYRS plasmid (~3 kb) was separated from the positive selection plasmid (~10 kb) by agarose gel electrophoresis, excised from the gel and purified using a Qiagen gel extraction kit (700 ng). The isolated clones were then transformed into electrocompetent cells containing pLWJ17B3, recovered for 1 h in SOC and then 500 μL was plated onto negative selection plates (LB agar, 50 μg/mL ampicillin, 50 μg/mL kanamycin and 0.2% L-arabinose) and incubated at 37 °C for 16 h. Surviving clones were scraped from the plate using 10 mL LB media and their plasmid DNA isolated using a Qiagen miniprep kit. The pBK-JYRS plasmid (~3 kb) was separated from the negative selection plasmid (~6 kb) by agarose gel electrophoresis, excised from the gel and purified using a Qiagen gel extraction kit. The library was subjected to 2 further rounds of selection. Individual colonies were picked from the last positive selection into a 96-deep well block and grown in 1 mL LB media overnight at 37 °C. The cultures were then plated (1 μL, using a Boekel replicator) onto selection plates (GMML media, 5 μg/mL tetracycline, 50 μg/mL, kanamycin, chloramphenicol ranging from 0-60 μg/mL, and +/- 1 mM UAA). Clones that grew on plates containing UAA and that did not grow on plates lacking UAA were sequenced to identify the active site mutations.

Evolution of an orthogonal synthetase in S. cerevisiae. One vial of MaV203 cells containing pGAD/Gal4(2TAG) and the pESC-EcYRS library was thawed and inoculated into SD-leu-trp media (100 mL, pH = 5.8) containing UAA (1 mM; a 100 mM stock solution of UAA was prepared in 0.1 M NaOH, and the pH was readjusted to 5.8 after addition of the amino acid). The culture was grown with shaking at 30 °C for 6 h, pelleted (1000g, 10 min) and washed with 0.9% sterile NaCl. The cell pellet was resuspended in 10 mL of 0.9% NaCl, and 500 μL was plated onto 4 positive selection plates (2 SD-leu-trp-his + 30 mM 3-aminotriazole + 1 mM UAA, and 2 SD-leu-trp-ura + 1 mM UAA). The plates were incubated at 30 °C for 48 h, then the surviving clones were collected with 10 mL of 0.9% NaCl, pooled and used to inoculate 500 mL of SD-leu-trp to an OD_{600} = 0.1. The culture was incubated overnight with shaking at 30 °C. The cells were pelleted by centrifugation (5000g, 5 min) and washed with 0.9% sterile NaCl. The cell pellet was resuspended in 10 mL of 0.9% NaCl and 500 μL was plated onto 4 negative selection plates (SD-leu-trp + 0.1% 5-fluoroorotic acid, 5-FOA was previously dissolved in 500 μL of DMSO to ensure solubility). The plates were incubated at 30 °C for 48 h, and surviving clones were then passed through two more rounds of positive and negative selection. Individual colonies were picked from the last positive selection into a 96-deep well block and grown in 1 mL SD-leu-trp overnight at 30 °C. The cultures were then plated (1 μL, using a Boekel replicator) onto selection plates (SD-leu-trp +/- 1 mM UAA, SD-leu-trp-his + 30 mM 3-AT +/- 1 mM UAA and SD-leu-trp-ura +/- 1 mM UAA) and incubated for 24 h at 30 °C. Clones that grew on selection plates

containing UAA and that did not grow on plates lacking UAA, but grew equally on SD-leu-trp plates, were sequenced to identify the active site mutations.

Transformation of plasmid DNA into pGAD/Gal42TAG maV203 yeast cells. A single colony containing pGAD/GAL42TAG (a gift from the Cropp Lab) was inoculated into SD-leu (1 mL) and vortexed to fully disperse the cells. This culture was then diluted to 50 mL in SD-leu and grown overnight at 30 °C. In the morning, the culture was diluted into SD-leu (300 mL) and grown with shaking at 30 °C until the OD_{600} = 0.5 (approximately 4 h). The culture was pelleted by centrifugation (5,000 x g for 5 minutes at room temperature). The supernatant was removed and the cells were washed with sterile DI H_2O, pelleted again and the supernatant was again removed. The yeast cells were resuspended in 1.5 mL of 1XTE/LiAc (1 mL of 10X TE, 1 mL of 1M LiAc, 8 mL H_2O) and placed on ice. In a prechilled eppendorf tube, DNA to be transformed (100 ng) was mixed with 5 μL carrier DNA (boiled and cooled salmon sperm DNA, 10 mg/mL). Cells (50 μL) were added to the chilled tube containing DNA and mixed by inversion. PEG/LiAc (500 μL, 1 mL of 10X TE, 1 mL of 1M LiAc, 8 mL of 50% w/v PEG) was added to the transformation mixture, and then mixed by inversion. The transformation was then incubated at 30 °C for 30 minutes with shaking. DMSO (20 μL) was added to the transformation and mixed by inversion, do NOT vortex at this point. The transformation was then heat shocked for 15 minutes at 42 °C and then pelleted at max speed for 15 seconds. The supernatant was removed and the cells were resuspended in sterile NaCl (1 mL, 0.9% w/v), plated onto SD-leu-trp plates in 1:10 and

1:100 dilutions, and incubated at 30 °C for 3 days. A control transformation is always performed using only carrier DNA. This procedure was adapted from a Clontech yeast transformation manual.

Assay for *E. coli* RS synthetase phenotypes in yeast. Cultures (3 mL, SD-leu-trp) of yeast cells containing the synthetase plasmid and pGAD/Gal42TAG were grown overnight at 30 °C with shaking. In the morning, each clone was streaked onto one-sixth of a selection plate (SD-leu-trp-his + 30 mM 3-AT +/- 1 mM UAA, SD-leu-trp-ura +/- 1 mM UAA, SD-leu-trp + 5-fluoroorotic acid +/- 1 mM UAA) and incubated at 30 °C for 2 days. Wild-type synthetase clones should grow on all positive selection plates regardless of UAA addition and should not grow on negative selection plates. The A5 mutant (an inactive synthetase that contains 5 alanines in the active site) should only grow on SD-leu-trp + 5-FOA plates, regardless of UAA. Evolved clones for UAA should only grow on positive selection plates in the presence of UAA, and on negative selection plates in the absence of UAA.

PCR amplification of EcLRS for homologous recombination. The PCR mix containing template DNA (pESC-EcLRS-OMeY, 100 ng), 1 μM forward primer, 1 μM reverse primer, 0.2 mM each dNTP, 1X Taq buffer and Taq (5 U, 1 μL) in a 50 μL reaction was reacted in a thermocycler with the following reaction conditions: 94 °C – 1min initial denaturing, 94 °C – 1min denaturing, 50 °C - 1 min annealing, 72 °C - 2 min extension, 72 °C – 10 min final

extension, with 30 cycles. PCR products were then purified with a Qiagen nucleotide removal kit and their size confirmed on a 1 % agarose gel.

CHAPTER 4 – SYNTHESIS OF MIRNA-21 INHIBITORS AS POTENTIAL THERAPEUTICS

4.1 Introduction to miRNAs and the miRNA pathway

MicroRNAs (miRNAs) are endogenous non-coding RNA molecules ranging from 21-23 nucleotides in length. miRNAs play an important role in eukaryotic gene regulation by binding to the 3′ untranslated region (UTR) of mRNA and either inhibit translation or cause degradation of the mRNA.[95] The misregulation of miRNAs has been linked to multiple forms of cancer, cardiovascular disease and immune disorders.[96] Small molecule inhibitors of miRNAs will serve as important tools for elucidating the biological roles of miRNAs and can lead to the discovery of potential therapeutic compounds.

MicroRNAs are transcribed in the nucleus by RNA Polymerase II to primary-miRNAs (pri-miRNAs), and are further processed by the enzyme Drosha to smaller stem-loop structures called precursor-miRNAs (pre-miRNAs). Pre-miRNAs are then exported to the cytoplasm by Exportin 5 and cleaved by Dicer to double stranded miRNAs. This double stranded duplex is unwound and one single-stranded mature miRNA is loaded into the RNA induced silencing complex (RISC) to target mRNA (Figure 4.1).[97] miRNAs do not require perfect complementarity to a target mRNAs which allows them to have multiple targets within cells.[95]

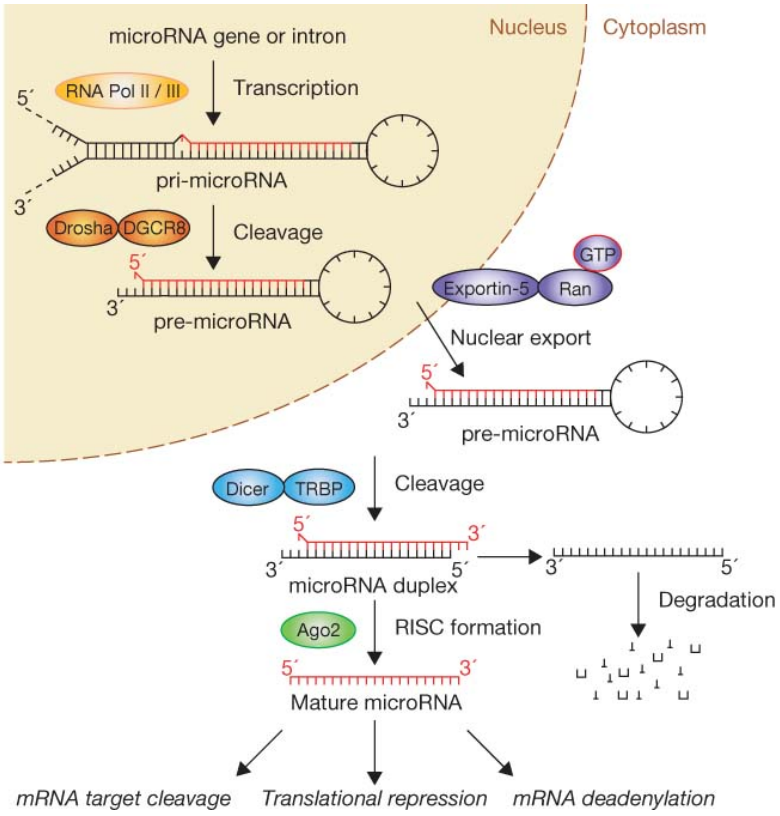


Figure 4.1. The miRNA processing pathway. Adapted from *Nature Cell. Biol.* **2009**, *11*, 228.

Cells have developed various gene regulation mechanisms to control the expression of miRNAs at different stages in the miRNA pathway. For example, mutations of the miRNA gene itself can regulate expression at the pre-transcriptional level, while transcription factors have been identified that control miRNA during transcription. Post-transcriptional regulation of miRNA expression through various protein interactions has also been reported.[97]

We are interested in the discovery of small molecule modifiers of miRNA expression and their application to explore miRNA biogenesis and to develop therapeutics for various diseases.

4.2 Introduction to small molecule inhibitors of miRNA-21

MicroRNA-21 (miR-21) is a miRNA that has been identified to be upregulated in pancreatic, breast, lung, and prostate cancers as well as forms of leukemia and glioblastomas.[98] Based on its prevalence in various forms of disease, there has been much interest in methods to control and inhibit expression of miR-21. Knockdown of miR-21 has been achieved with RNA antagomirs, synthetic complementary RNA molecules, in breast cancer tumors cells.[99] Though effective, antagomirs must be transfected into mammalian cells, are large in comparison to small molecules, and are more costly to produce.

Dr. Douglas Young (Deiters Lab), in collaboration with Dr. Qihong Huang of the Wistar Institute, developed the first assay to screen for small molecule modifiers of miR-21 function.[100] The assay is based on a lentiviral vector that contains the miR-21 target sequence downstream of a luciferase reporter gene. In the presence of functional miR-21, the miRNA binds its target sequence and inhibits translation of luciferase. Conversely, in the presence of a small molecule inhibitor of miR-21, the miRNA is unable to bind its target sequence and luciferase is expressed (Figure 4.2).

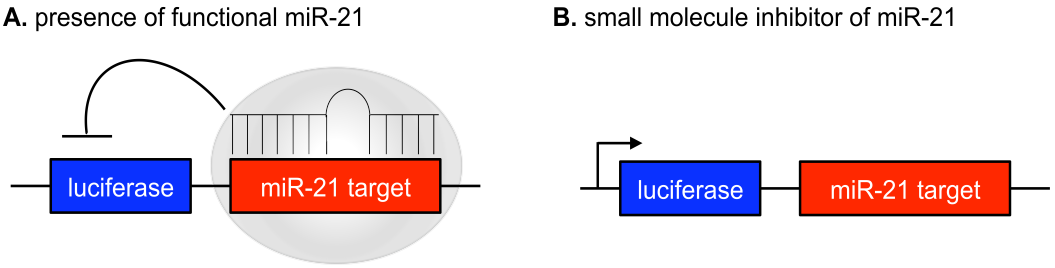
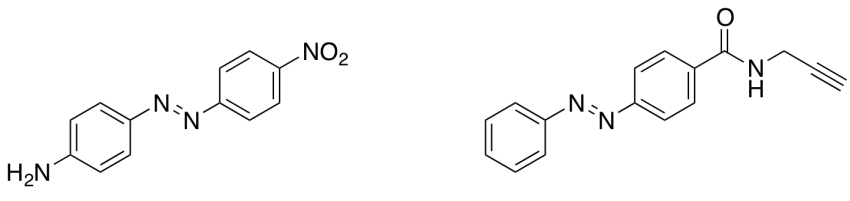


Figure 4.2. Schematic of assay used to screen for small molecule inhibitors of miR-21. **A.** In the presence of functional miR-21, the microRNA binds to its target sequence downstream of luciferase and decreases luminescence signal. **B.** In the presence of a small molecule inhibitor of miR-21, the microRNA is unable to bind the target sequence and an increase in luminescence signal is detected.[100] Adapted from *J. Am. Chem. Soc.* **2010**, *132*, 7976.

The initial screening of over 1000 small molecules yielded the diazobenzene compound **22** as a hit compound. When HeLa cells, a cervical cancer cell line that is known to have high levels of miR-21, were treated with **22** (10 μM) a 2.5-fold increase in luciferase expression over DMSO-treated cells was detected. A structure-activity relationship (SAR) study was then performed to further improve the efficacy of **22** The SAR yielded compound **23**, which increased luciferase expression 5-fold (Figure 4.3).[100]

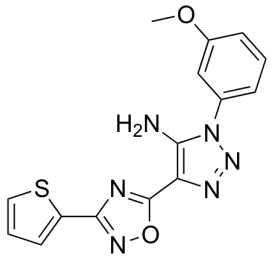


22 **23**

Figure 4.3. microRNA inhibitors of miR-21 **22** and **23**, discovered from a small molecule screen and subsequent SAR study.

4.3 Synthesis of new oxadiazole inhibitors of miRNA-21

We have recently explored other small molecule inhibitors of miR-21. A screen of 333,521 compounds was performed by the NIH Chemical Genomic Center using the luciferase assay for miR-21. From this initial screen, the oxadiazole containing **24** was identified as a hit compound (Figure 4.4).



24

Figure 4.4. Initial hit miR-21 inhibitor from the NIH library screen.

The dose response curve of oxadiazole **24** showed increased miR-21 inhibitory activity when compared to the original inhibitor **23** (Figure 31).

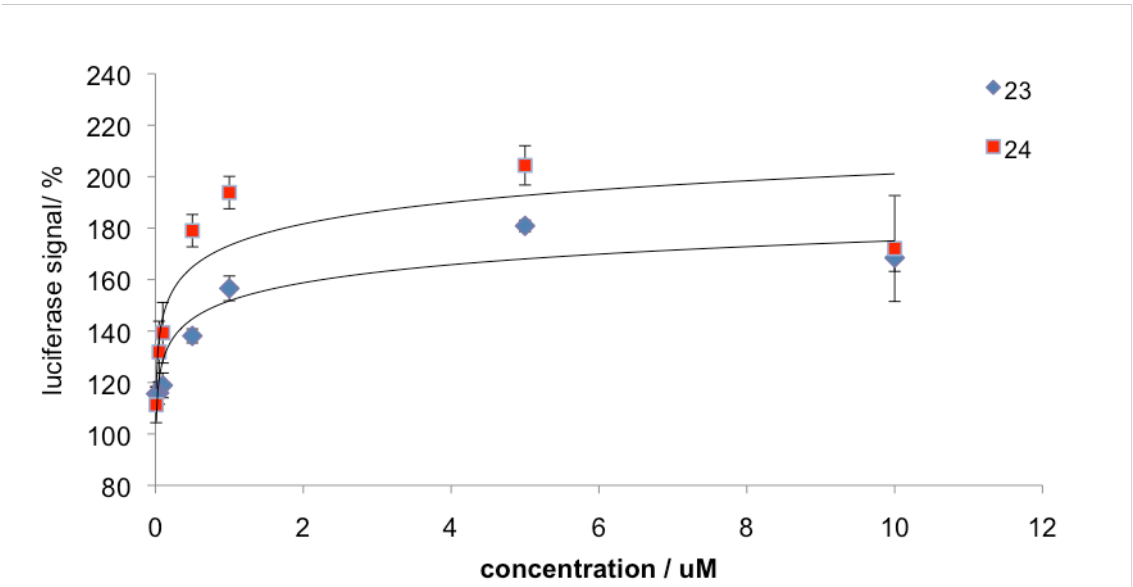
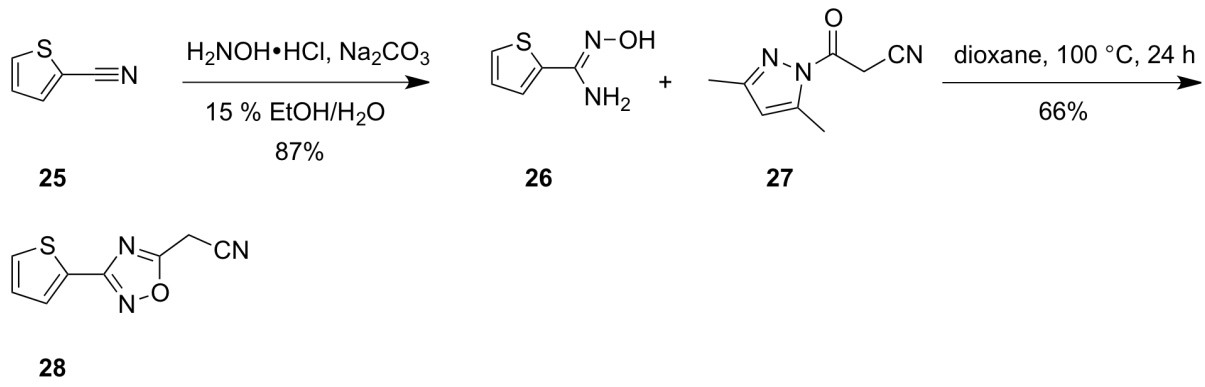


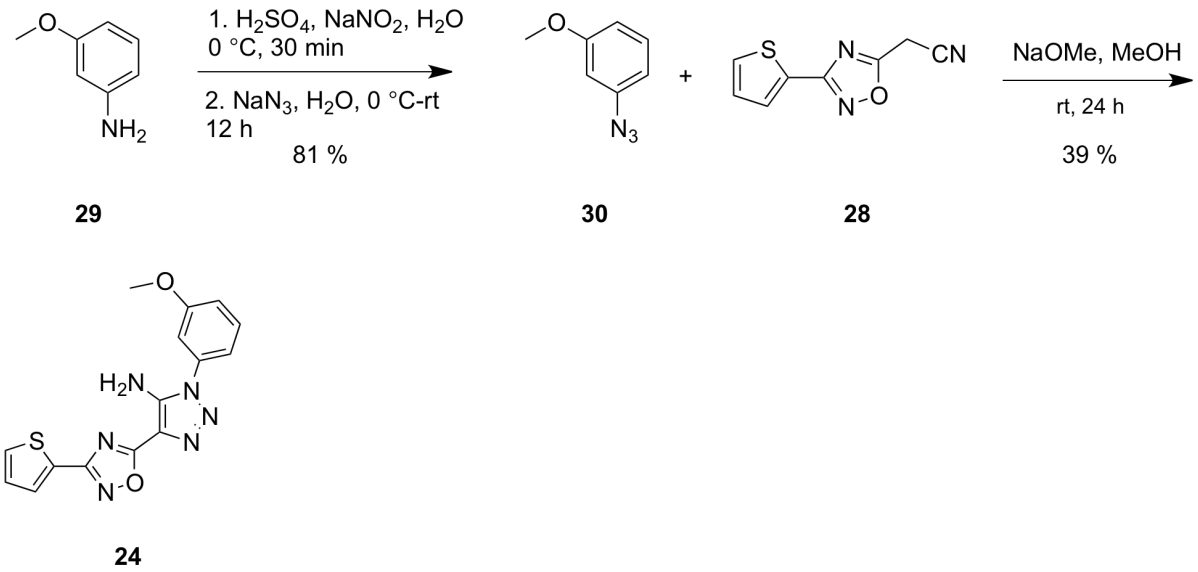
Figure 4.5. Dose-response curves for the original miR-21 inhibitor **23** and the new oxadiazole **24** miR-21 inhibitor**.**

We then pursued the synthesis of the oxadiazole **24** and various analogues to determine the structural requirements of a miR-21 inhibitor. The initial oxadiazole **24** was synthesized by reacting the commercially available 2-thiophenecarbonitrile (**25**) with hydroxylamine hydrochloride and sodium carbonate in 15% ethanol/water at room temperature to produce 2-thiopheneamidoxime (**26**) in 87% yield. Further reaction of the oxime **26** with commercially available 1-cyanoacetyl-3,5-dimethylpyrazole (**27**) in refluxing dioxane produced the oxadiazole **28** in 66% yield (Scheme 4.1).[101]



Scheme 4.1. Synthesis of the oxadiazole precursor **28**.

In order to produce the triazole ring core structure, we first converted the corresponding aniline, *m*-anisidine (**29**), to the diazonium salt by reaction with sulfuric acid and sodium nitrite in water. After 30 minutes, the diazonium salt was converted to the aryl azide **30** by the addition of sodium azide in water in 81% yield. The oxadiazole **28** was then reacted with the aryl azide **30** through a 1,3-dipolar cycloaddition in the presence of sodium methylate in methanol. The final product **24** was isolated in 39% yield after being precipitated from dichloromethane with water, filtered and washed with hexanes (Scheme 4.2).



Scheme 4.2. Synthesis of the miR-21 inhibitor **24**.

Oxadiazole **24** was then tested for its activity against miR-21. The original miR-21 inhibitor **23** was also tested to compare the activities of the two inhibitors. HeLa cells stably transfected with the miR21-Luc construct or the linker-Luc control construct (that contains no miRNA target sequence) were treated with compounds at 10 μM concentration, in triplicate, incubated 48 hours and then luminescence was quantified on a plate reader using a Bright Glo (Promega) assay (Figure 4.6).

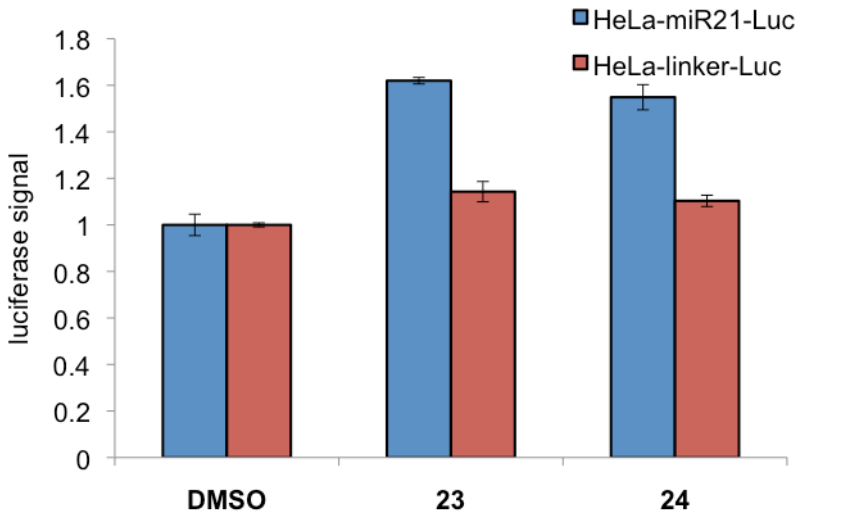
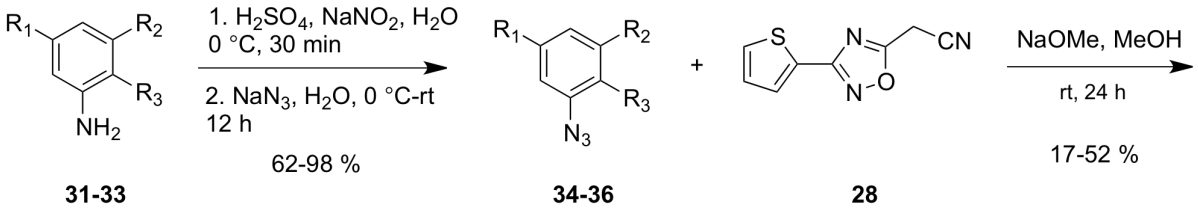
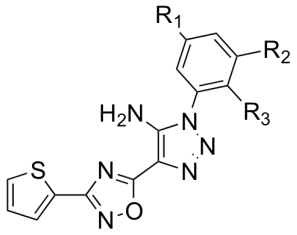


Figure 4.6. Bright Glo assay of miR-21 inhibitors **23** and **24**. At 10 μM concentration, the inhibitors induced a ~1.5-fold increase in luciferase signal compared to a DMSO control. Errors bars indicate standard deviation. This assay was conducted by Colleen Connelly.

The oxadiazole miR-21 inhibitor showed comparable activity to the original inhibitor **23**. Compounds **23** and **24** also do not cause any increase in luciferase signal for the HeLa-linker control. This indicates that our inhibitors are specific to miR-21 and do not target the luciferase enzyme. In an effort to further increase the activity of **24**, we conducted a preliminary SAR study. The first analogues of the original oxadiazole inhibitor contained various substitutions on the benzene ring. The corresponding aryl azides **34-36** were synthesized by the previously described electrophilic aromatic substitution procedure from the anilines **31-33** in 62-98% yield. The aryl azides were then subsequently reacted with the oxadiazole **26** in the presence of sodium methylate in methanol to generate the final analogues **37-39** in varying yield (Scheme 4.3).



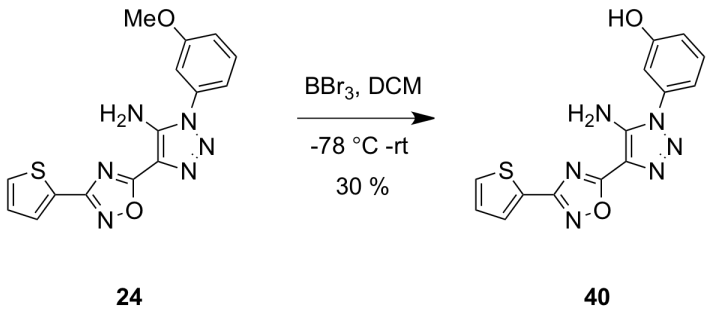


37-39

analogue	R1	R2	R3
31, 33, 37	H	Me	Me
32, 35, 38	OMe	OMe	H
33, 36, 39	H	H	H

Scheme 4.3. Synthesis of oxadiazole analogues **35-37**.

The phenol analogue **40** of the original inhibitor was synthesized through demethylation of **24** with boron tribromide in dichloromethane in 30% yield (Scheme 4.4).



24 **40**

Scheme 4.4. Synthesis of phenol oxadiazole **40**.

These analogues were tested for their activity and specificity against miR-21 in a Bright Glo Luciferase assay. Cells were treated with 10 μM concentration of compound, incubated 48 hours, and their luminescence was quantified on a plate reader (Figure 4.7). The results from this assay show that our modifications to the benzene ring lead to a reduction in miR-21 inhibitory activity and our original oxadiazole **24** remained the best inhibitor.

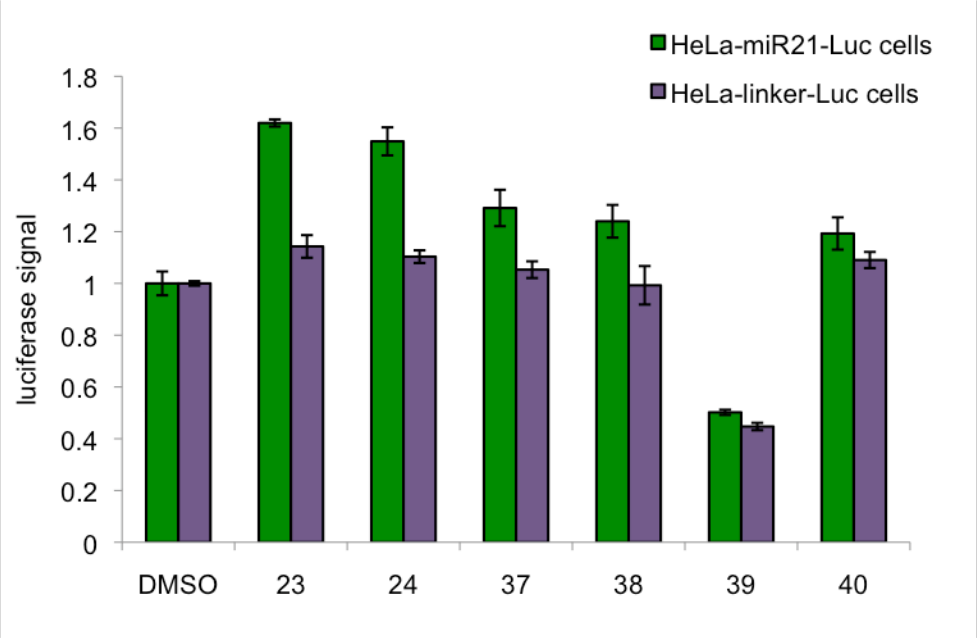
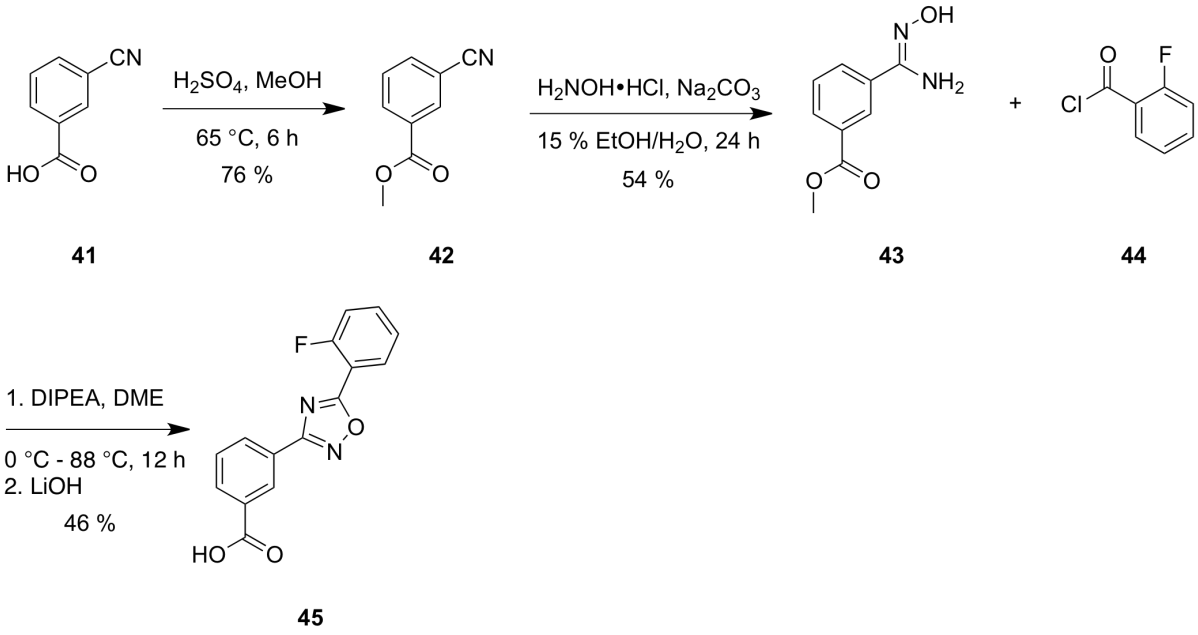


Figure 4.7. Bright Glo assay of miR-21 oxadiazole inhibitor analogues. Errors bars indicate standard deviation. This assay was conducted by Colleen Connelly.

A recent review article concerning firefly luciferase (FLuc) revealed that certain chemotypes of small molecules are apparent inhibitors of the enzyme and can lead to false positives in compounds screening.[102] Of these chemotypes, some oxadiazole containing compounds have been identified as inhibitors of FLuc. To ensure that our oxadiazole containing miR-21 inhibitors were specific for the miRNA and not for FLuc, we synthesized a known FLuc inhibitor, and compared its inhibitory FLuc activity to that of our known miR-21 inhibitors **23** and **24**.

To synthesize this FLuc inhibitor, 3-cyanobenzoic acid **41** was converted to methyl-3-cyanobenzoate **42** by reaction with sulfuric acid in methanol in 76% yield. The resulting methyl-3-cyanobenzoate was reacted with hydroxylamine hydrochloride and sodium

carbonate in 15% ethanol in water to generate the amidoxime **43** in 54% yield. Amidoxime **43** was further reacted with 2-fluorobenzoyl chloride (**44**) in refluxing dimethoxyethane in the presence of Hünig's base to produce the FLuc inhibitor **45** in 46% yield (Scheme 4.5).



Scheme 4.5. Synthesis of the known FLuc inhibitor **45**.

This FLuc inhibitor **45** as well as miR-21 inhibitors **23** and **24** were tested in an enzyme-based assay by Colleen Connelly for their ability to inhibit luciferase (Figure 4.8). The assay showed a dose-dependent decrease in luminescence signal when FLuc was treated with the known inhibitor **45**, but no significant decrease in luminescence signal for either miR-21 inhibitor, showing **23** and **24** are not Fluc inhibitors, but are specific to miR-21.

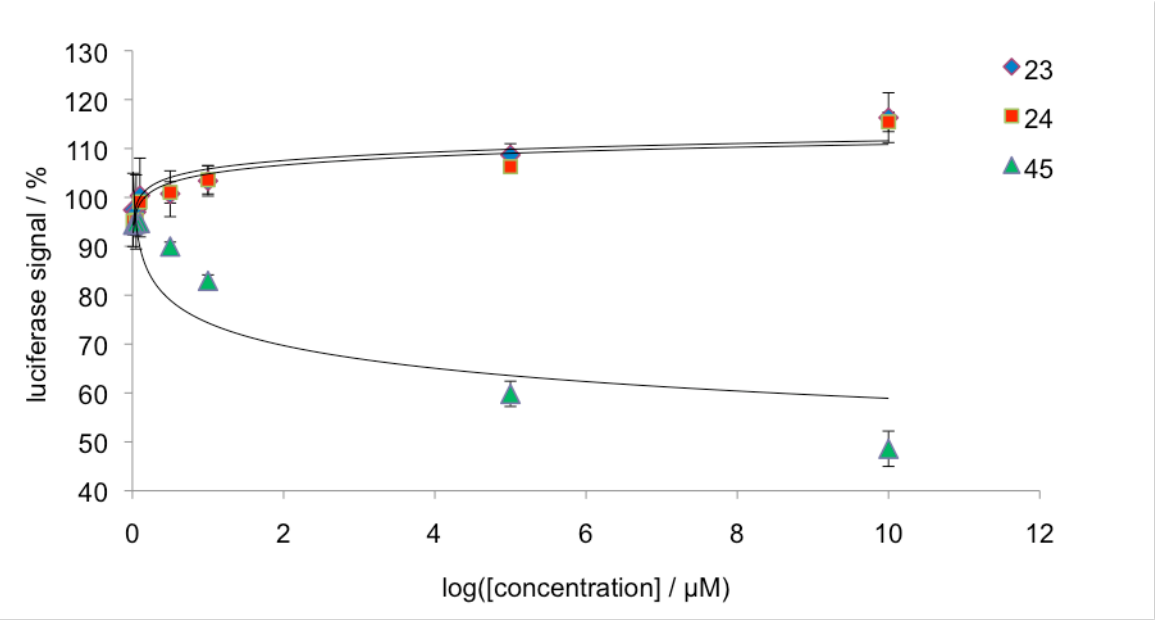


Figure 4.8. *In vitro* FLuc assay for **23** and **24**, compared to the known Fluc inhibitor **45**.

The inhibitor **45** was then tested for its ability to inhibit FLuc *in vivo* using the same reporter construct for previous miR-21 inhibitors (Figure 4.9). The results are consistent with observations in literature.[102]

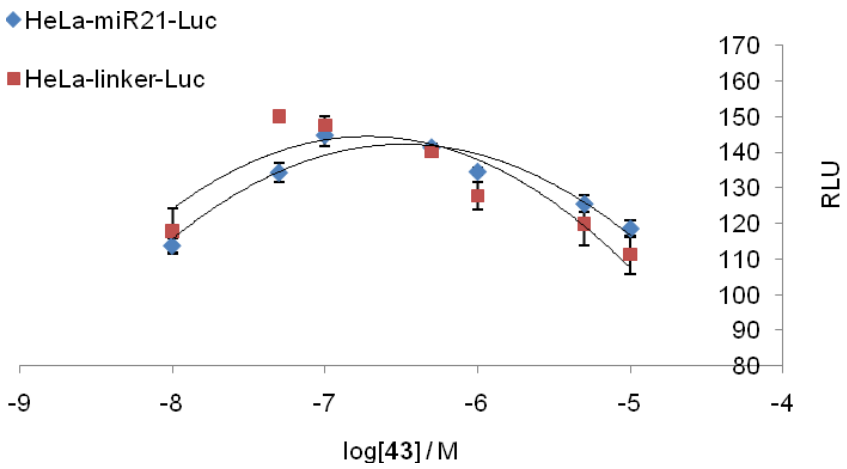


Figure 4.9. *In vivo* FLuc inhibition of known inhibitor **45**.

Inhibition of miR-21 expression by **24** was validated through quantitative RT-PCR to measure cellular levels of miR-21 using primers specific to the mature miRNA. Cells treated with **22** and **24** showed decreased levels of mature miR-21 compared to the DMSO control (Figure 4.10), suggesting the inhibitors regulate the expression of miR-21

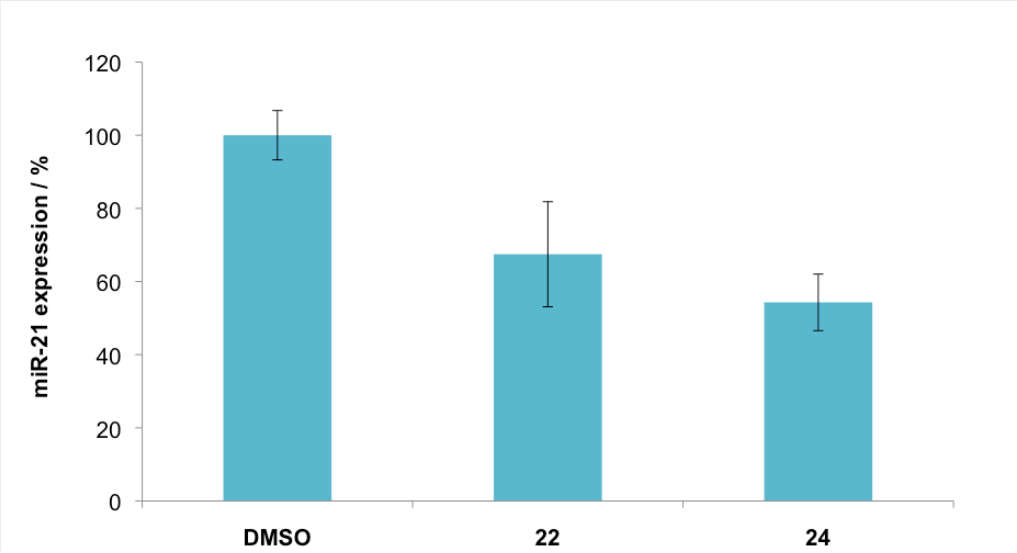


Figure 4.10. Levels of mature miR-21 in cells treated with **22** and **24** compared to a DMSO control. This assay was performed by Colleen Connelly.

Currently, further SAR studies are being carried out on the oxadiazole miR-21 inhibitor **24** to improve its efficacy against microRNA-21.

4.4 Summary and Outlook

We have successfully synthesized and assayed various analogues of the oxadiazole miR-21 inhibitor **24**. Initial SAR studies showed that the original oxadiazole **24** remained the most effective inhibitor of miR-21 activity in mammalian cells. Additionally, we have compared the activity of compound **24** to an inhibitor of luciferase activity to ensure that our oxadiazole analogues are acting upon the miRNA pathway and not on the reporter gene.

Currently, further SAR studies and RT-PCR experiments are being carried out by Colleen Connelly to further elucidate the mechanism by which **24** inhibits miR-21. Further

analogues of **24** have also been synthesized by Méryl Thomas for SAR studies. After the most potent miR-21 inhibitor is found, a pull down assay will be developed to identify the target of the inhibitor, which will reveal the mode of action of miRNA inhibition.

4.5 Experimental Methods

Synthesis of 2-thiopheneamidoxime (26). Na_2CO_3 (4.5 g, 43 mmol) and H_2N-OH•HCl (4.52 g, 64.5 mmol) were added to a solution of 2-thiophenecarbonitrile **25** (2 mL, 21.50 mmol) in 15% EtOH in H_2O (5 mL). The reaction mixture was stirred at r.t. for 24 h, then extracted with EtOAc (3 × 15 mL), and the combined organic layers were washed with saturated aqueous $NaHCO_3$ (15 mL), brine (15 mL) and dried over Na_2SO_4. The solvent was removed by rotary evaporation to yield the product as a white crystalline solid (2.67 g, 87%). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 4.92 (br, 2H), 7.11 (d, J = 30.3 Hz, 1H), 7.26-7.42 (m, 2H), 8.21 (br, 1H). LRMS: *m/z* calcd for ($C_5H_6N_2OS$) $[M+H]^+$: 143.02; found: 143.03.

Synthesis of oxadiazole (28). 1-Cyanoacetyl-3,5-dimethylpyrazole **27** (495 mg, 3.0 mmol) was added to a solution of amidooxime **26** (512 mg, 3.61 mmol) in dioxane (3 mL) and the reaction mixture was heated under reflux for 16 h. The reaction mixture was allowed to cool to room temperature and concentrated *in vacuo*. The crude intermediate was resuspended in DCM (10 mL), and the organic layer was washed with saturated aqueous $NaHCO_3$ (10 mL), brine (10 mL). The solvent was concentrated by rotary evaporation and the residue was

purified by silica gel chromatography, eluting with hexanes/EtOAc (2:1) to give the product **28** as a light orange solid (376 mg, 66% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 7.18 (dd, J^1 = 5.1 Hz, J^1 = 1.2 Hz, 1H), 7.55 (d, J = 6.0 Hz, 1H), 7.83 (d, J = 3.6 Hz, 1H).

General synthesis of aryl azides. 3-Methoxyaniline **29** (300 mg, 2.75 mmol) was dissolved in H_2O (2 mL) and the solution was cooled to 0 °C. H_2SO_4 (750 μL) was added to the reaction mixture dropwise at 0 °C. After 5 min, a solution of $NaNO_2$ (246 mg, 3.57 mmol) in H_2O (500 μL) was added to the reaction mixture dropwise at 0 °C. The reaction mixture was stirred for 30 min at 0 °C, after which a solution of NaN_3 (286 mg, 4.4 mmol) in H_2O (500 μL) was added to the mixture. The reaction was stirred for an additional 3 h, allowing the reaction to warm to r.t. The reaction mixture was poured into saturated aqueous $NaHCO_3$ (5 mL) and extracted with EtOAc (2 × 10 mL). The combined organic layers were washed with brine (10 mL), filtered and concentrated *in vacuo*. The crude product was then purified by silica gel chromatography, eluting with hexanes/EtOAc (4:1) to yield the azide **30** as an orange oil (252 mg, 81% yield).

(30). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 3.80 (s, 3H), 6.54 (s, 1H), 6.67 (dd, J^1 = 12.0 Hz, J^2 = 2.4 Hz, 2H), 7.25 (t, J = 7.35 Hz, 1H).

(34). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 2.13 (s, 3H), 2.23 (s, 3H), 6.96 (dd, J^1 = 14.4 Hz, J^2 = 11.1 Hz, 2H), 7.12 (t, J = 7.8 Hz, 1H).

(35). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 3.78 (s, 6H), 6.19 (s, 2H), 6.25 (s, 1H).

(36). The analytical data obtained matched reported literature (^{1}H NMR, $CDCl_3$).[103]

General synthesis of triazoles. NaOMe (126 mg, 2.34 mmol) in MeOH (1.3 mL) was added to a solution of oxadiazole **28** (117 mg, 0.79 mmol) in MeOH (2 mL). The reaction mixture was stirred for 15 min at r.t. Aryl azide **30** in MeOH (200 μL) was added and the reaction mixture was stirred 36 h at r.t. The reaction mixture was acidified with aqueous HCl (2 mL, 1 M), poured into brine (15 mL), extracted with DCM (3 × 5 mL), dried over anhydrous Na_2SO_4 and concentrated by rotary evaporation. The solution was then pipetted slowly into boiling hexanes (50 mL) and the product precipitated upon cooling. The solvent was decanted and **24** was isolated as a green solid (103 mg, 39% yield).

(24). ^{1}H NMR (300 MHz, d-6 DMSO): δ = 3.83 (s, 3H), 6.92 (br, 2H), 7.17-7.19 (m, 3H), 7.27 (dd, J^1 = 5.1 Hz, J^1 = 1.5 Hz, 1H), 7.53 (t, J = 8.1 Hz, 1 H), 7.88 (d, J = 5.1 Hz, 1H), 8.02 (d, J = 1.2 Hz, 1H). LRMS *m/z* calcd for ($C_{15}H_{12}N_6O_2S$) $[M+Na]^+$: 363.06; found: 363.10

(37). ^{1}H NMR (300 MHz, d-6 DMSO): δ = 1.92 (s, 3H), 2.33 (s, 3H), 6.72 (br, 2H), 7.22-7.41 (m, 3H), 7.42 (d, J = 7.8 1H), 7.86 (d, J = 5.1 Hz, 1H), 7.99 (d, J = 3.6 Hz, 1H). LRMS *m/z* calcd for ($C_{16}H_{14}N_6OS$) $[M+H]^+$: 339.09; found: 339.10

(38). ^{1}H NMR (300 MHz, d-6 DMSO): δ = 3.82 (s, 6H), 6.70 (t, *J* = 2.1 Hz, 1H), 6.80 (s, 2H), 6.93 (br, 2H), 7.28 (dd, J^1 = 5.1 Hz, J^2 = 1.2 Hz, 1H), 7.89 (d, *J* = 3.2 Hz, 1H), 8.02 (d, *J* = 2.6 Hz, 1H). LRMS *m/z* calcd for ($C_{16}H_{14}N_6O_3S$) $[M+Na]^+$: 393.07; found: 393.10.

(39). ^{1}H NMR (300 MHz, d-6 DMSO): δ = 6.92 (br, 2H), 7.29 (t, *J* = 3.6 Hz, 1H), 7.61-7.66 (m, 5H), 7.89 (d, *J* = 5.1 Hz, 1H), 8.02 (d, *J* = 2.3 Hz, 1H). LRMS *m/z* calcd for ($C_{14}H_{10}N_6OS$) $[M+Na]^+$: 333.05; found: 333.10.

Synthesis of oxadiazole (40). LG_157 **24** (20 mg, 0.059 mmol) was dissolved in DCM (600 μL) and cooled to -78 °C. BBr_3 (0.1 mmol, 1 M in DCM) was added to the solution dropwise at -78 °C. The reaction mixture was stirred for 36 h, and then diluted in NaOH (5 mL). The reaction mixture was then neutralized with aqueous HCl and extracted with Et_2O (3 × 3 mL). The combined organic layers were then dried over anhydrous Na_2SO_4 and solvent was removed *in vacuo* to give **40** as a light yellow solid (5.8 mg, 30% yield). ^{1}H NMR (300 MHz, d-6 DMSO): δ = 6.95 (br, 2H), 6.97-7.04 (m, 3H), 7.27 (dd, J^1 = 4.8 Hz, J^1 = 1.2 Hz, 1H), 7.42 (t, *J* = 8.1 Hz, 1H), 7.88 (d, *J* = 6.3 Hz, 1H), 8.01 (d, *J* = 3.6 Hz, 1H). LRMS *m/z* calcd for ($C_{14}H_{10}N_6O_2S$) $[M+Na]^+$: 349.05; found: 349.06

Synthesis of 5-cyanomethybenzoate (42). H_2SO_4 (275 μL, 5.16 mmol) was added to 5-cyanobenzoic acid **41** (500 mg, 3.40 mmol) in MeOH (5 mL) and the reaction mixture was heated under reflux for 6 h. The solvent was removed *in vacuo* and the crude reaction

mixture was resuspended in EtOAc (5 mL). The suspension was washed with H_2O (10 mL), and aqueous NaOH (10 mL, 1 M), and extracted with EtOAc (3 × 5 mL). The combined organic layers were washed with brine (15 mL), dried over anhydrous Na_2SO_4 and the solvent removed by rotary evaporation to yield **42** as a light green solid (415 mg, 76% yield). The analytical data obtained matched reported literature (^{1}H NMR, $CDCl_3$). [104]

Synthesis of amidoxime (43). Nitrile **42** (200 mg, 1.24 mmol), Na_2CO_3 (263 mg, 2.50 mmol) and H_2N-OH•HCl (260 mg, 3.71 mmol) were dissolved in 15% EtOH/H_2O (3.5 mL) and stirred at r.t. for 24 h. The reaction mixture was then extracted with EtOAc (3 × 15 mL), and the combined organic layers were washed with saturated aqueous $NaHCO_3$ (15 mL), brine (15 mL) and dried over anhydrous Na_2SO_4. The solvent was removed by rotary evaporation to yield the product as a light green oil (132 mg, 54% yield). The analytical data obtained matched reported literature (^{1}H NMR, $CDCl_3$).[104]

Synthesis of FLuc inhibitor (45). DIPEA was added to a solution of amidoxime **43** (75 mg, 0.387 mmol) in DME (1 mL) at 0 °C. 2-fluorobenzoylchloride **44** (46 μL, 0.387 mmol) was added to the reaction mixture. The reaction was stirred for 4 h and then heated under reflux for 12 h. The reaction mixture was concentrated *in vacuo* and then resuspended in THF/H_2O (1 mL, 1:1). LiOH (1 mL, 2M) was added to the solution and the reaction mixture was stirred at r.t. for 24 h. The reaction was quenched with aqueous HCl (5 mL, 1M) and the precipitate

was filtered out to give the product **45** as a light green solid (51 mg, 46% yield). The analytical data obtained matched reported literature (^{1}H NMR, $CDCl_3$).[104]

CHAPTER 5 – CHEMICAL BIOLOGY APPLICATIONS OF THE HALOTAG PROTEIN

5.1 Introduction to the HaloTag protein

The HaloTag protein (Promega) is a mutant haloalkane dehalogenase that forms a covalent bond with chloroalkanes. This protein, developed in 2005, was derived from the bacterial *Rhodococcus erythropolis* dehalogenase (DhaA).[105] DhaA removes halogens from aliphatic hydrocarbons through a catalytic aspartate residue that nucleophilic attacks the halocarbon in an S_N2 reaction.[106] A nearby histidine residue facilitates the hydrolysis of the aspartic ester intermediate to generate an alcohol and regenerate the aspartate residue for further dehalogenation (Figure 5.1 A). Mutating this histidine residue to a phenylalanine made the ester bond formation irreversible and created a new protein that can be covalently labeled with chloroalkane molecules.[105]

HaloTag expression plasmids and a line of HaloTag ligands that allow for multi-wavelength labeling, as well as affinity tags for purification, are available from Promega (Figure 5.1 B).

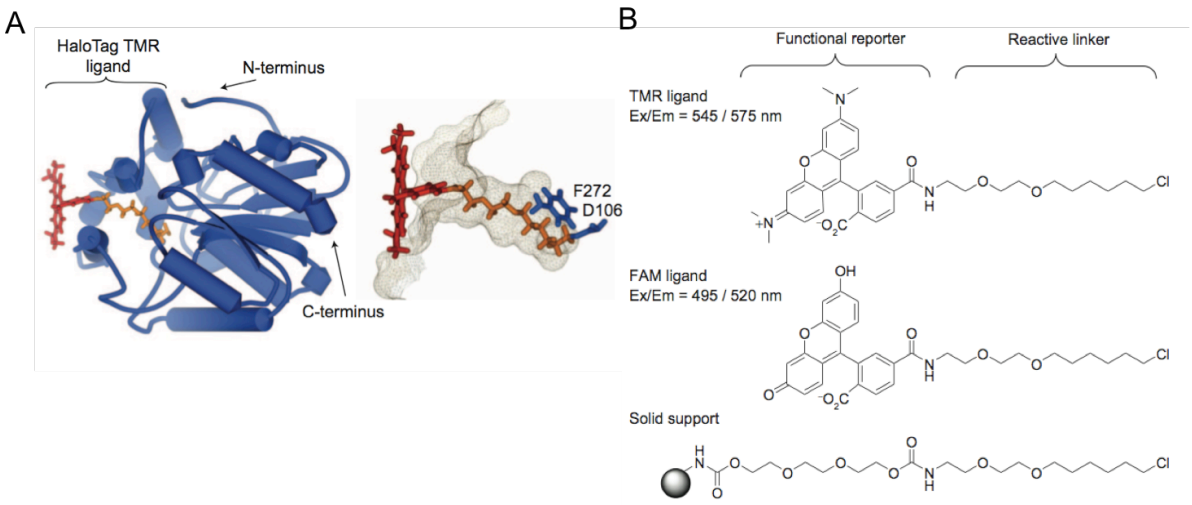


Figure 5.1. (A) The HaloTag protein forms a covalent bond with HaloTag ligands through active site residue D106. (B) Commercially available HaloTag ligands for fluorescent imaging and immobilization to a solid phase. Adapted from *ACS Chem. Biol.* **2008**, *3*, 373.

Since its introduction in 2005, the HaloTag protein has been utilized for various applications. Harwood, et al. used the HaloTag protein to aid in the photochemical control of Rho GTPase activation.[107] The GTPase Cdc42 was initially photocaged at a cysteine residue within the GTP binding site through alkylation with a traditional ONB group. The caging group alone was unable to effectively block GTP from binding with Cdc42 and it was discovered a much bulkier ligand was required to completely block this interaction. An alternative ONB group **46** was then synthesized to contain the reactive HaloTag linker. After Cdc42 was alkylated with the new photocaging group **46**, the exposed HaloTag linker was reacted with the HaloTag protein (Figure 5.2). This new method of sterically blocking the GTP binding site successfully allowed for the photochemical control over GTPase activation.

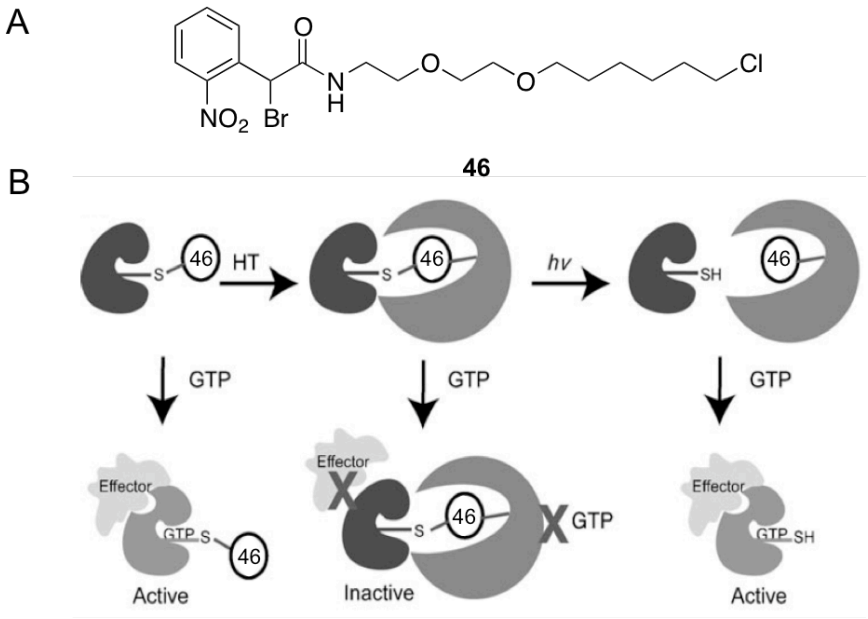


Figure 5.2. (A) ONB-HaloTag linker **46**. (B) Photochemical activation of Cdc42 activity. Reaction with **46** alone fails to block GTP binding. After the HaloTag protein is recruited, GTP is unable to bind and Cdc42 remains inactive. Activity is restored after UV irradiation removes **46**. Adapted from *ChemBioChem*. **2009**, *10*, 2855.

In 2011, Ohana, et al. used the HaloTag protein to isolate high-levels of recombinant proteins expressed from mammalian cells.[108] Various human kinases were expressed in high-yield from HEK293 cells as fusions to the HaloTag protein. After expression, the fusion proteins were purified from the cell lysate using the HaloLink resin and then the kinases were cleaved from the resin by TEV proteolytic cleavage. When compared to other methods of protein purification (FLAG, 3xFLAG, and His_6Tag), the HaloTag provided the highest levels of pure protein (Figure 5.3).

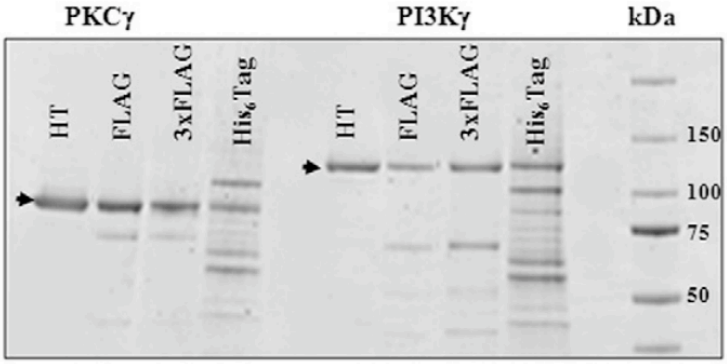
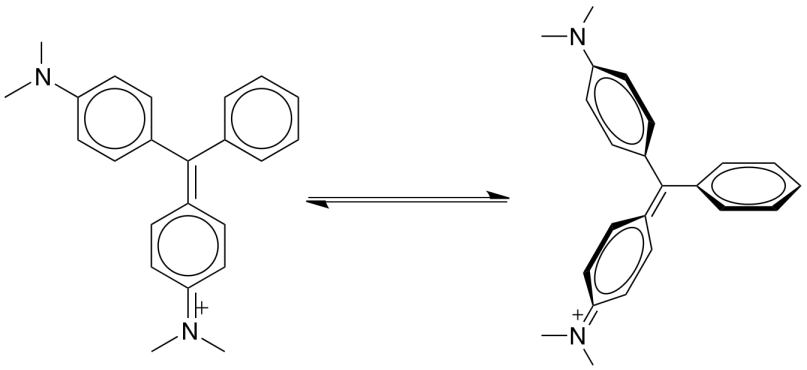


Figure 5.3. Purification of the human kinases PKCγ and PI3Kγ by affinity tagging methods. The arrow represents the expected molecular weight of the kinase after proteolytic cleavage. Adapted from *Protein Express Purif*. **2011**, *76*, 154.

5.2 Development of a Malachite-Green:HaloTag viscosity sensor

In an effort to further the application of the HaloTag protein as a chemical biology tool, we hypothesized that the synthesis of a malachite green HaloTag ligand can be used as a protein-labeled viscosity sensor. Malachite green (**47**) is a triphenylmethane dye, whose phenyl rings twist away from one another in low-viscosity environments to relieve steric strain. Triphenylmethane dyes are poor fluorophores as they often can be relaxed through vibrational modes.[109] Malachite green has a low quantum yield of 7.9×10^{-5} in solution.[110] When malachite green is in a high-viscosity environment, or bound in such a way that the phenyl rings must lay planar, the malachite green can act as a fluorophore (Scheme 5.1). This has been demonstrated by using triphenylmethane dyes and their corresponding RNA aptamers.[111]



Scheme 5.1. Malachite green (**47**) can fluctuate between a planar state when in a high viscosity environment and a propeller shape when in a low viscosity environment.

To express the 34 kDa HaloTag protein, pET21-HaloTag2 (Miller Lab, UMass Medical School) was transformed into BL-21G cells and a resulting colony was used to inoculate 25 mL of LB containing kanamycin (50 μg/mL). The culture was grown to OD_{600} = 0.6 and protein expression was induced with 0.1 mM IPTG. Protein was expressed at 37 °C for 6 hours, cells were harvested by centrifugation, lysed and the supernatant was purified on Ni-NTA agarose resin. The purity was checked by SDS-PAGE analysis (Figure 5.4) and the protein was then dialyzed into phosphate buffered saline, and stored at 4 °C.

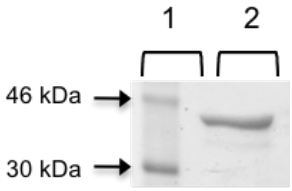
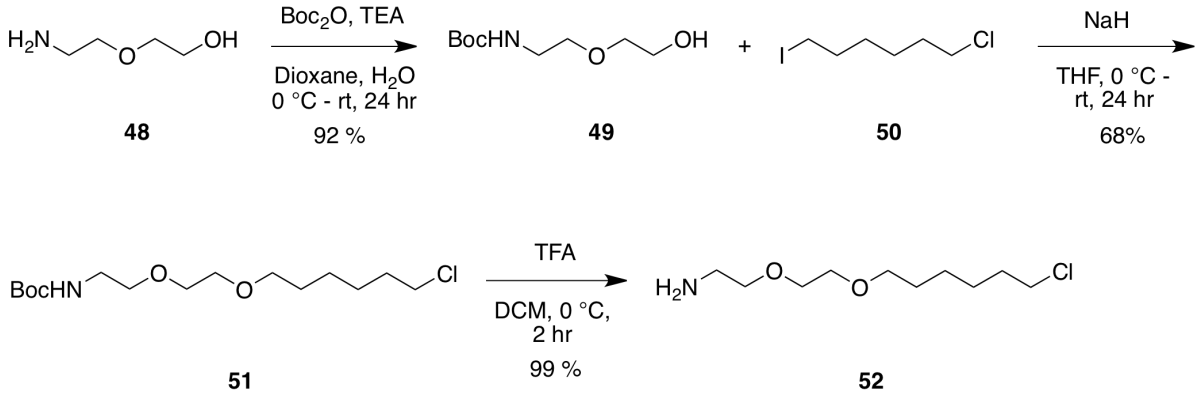


Figure 5.4. SDS-PAGE of purified HaloTag protein. Lane 1: Molecular weight ladder. Lane 2: HaloTag protein.

In order to synthesize a malachite green HaloTag linker, 2-(2-aminoethoxy)ethanol (**48**) was reacted with di-*tert*-butyl dicarbonate (Boc_2O) in the presence of triethylamine in a 1:1 dioxane/water mixture to yield the boc-protected aminoalcohol **49** in 92% yield. Further reaction of **49** with 1-chloro-6-iodohexane (**50**) in the presence of sodium hydride in THF gave the boc-protected HaloTag linker **51** in 68%. The boc-protected linker was deprotected by reaction with trifluoroacetic acid in dichloromethane to give the HaloTag linker **52** in 99% yield (Scheme 5.2).



Scheme 5.2. Synthesis of HaloTag linker **52**.

We then synthesized a fluorescent linker to test labeling of the HaloTag protein. The amine linker **52** was reacted with dansyl chloride (**53**), in the presence of triethylamine in dichloromethane to yield the dansyl HaloTag linker **54** in 97% yield (Scheme 5.3).

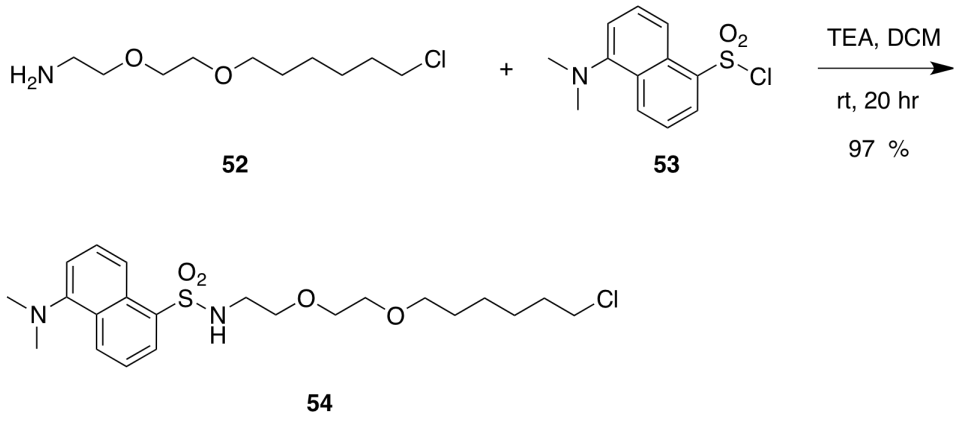
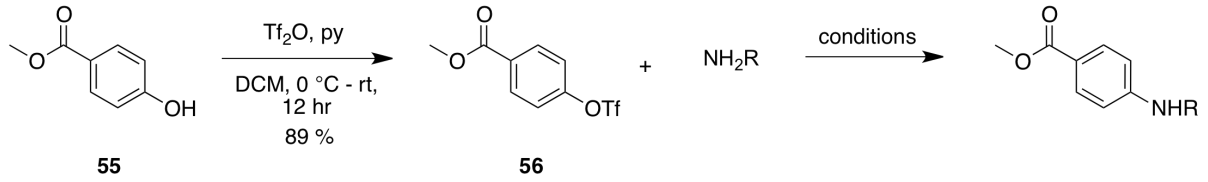


Figure 5.3. Synthesis of fluorescent HaloTag linker **54**.

Our initial route to synthesize the malachite green HaloTag linker was to couple the amine linker to a hydroxy benzoate, then reduce the benzoate with two equivalents of a phenyl Grignard reagent to fabricate the triphenylmethane moiety. To begin, we attempted to couple the amine linker **52** to a phenyl ring through a palladium catalyzed cross coupling reaction.[112] In order to synthesize the triflate **56**, methyl-4-hydroxybenzoate (**55**) was reacted with trifluoromethanesulfonic anhydride (Tf_2O) in the presence of pyridine in dichloromethane in 89% yield. The triflate **56** was then reacted with butylamine as well as

the aminoalcohol **48** under various conditions to optimize the carbon-nitrogen bond formation reaction. However, no conditions yielded the desired product (Scheme 5.4).



55 **56**

R	conditions	results
C_4H_9	$Pd(dba)_2$, DPPF, NaOtBu, 100 °C, 5 hr	45%
C_4H_9	$Pd(dba)_2$, BINAP, NaOtBu, 100 °C, 5 hr	no reaction
C_4H_9	$Pd(dba)_2$, DPPF, NaOtBu, 300 W, 30 min	16%
$C_4H_9O_2$	$Pd(dba)_2$, DPPF, NaOtBu, 100 °C, 5 hr	no reaction
$C_4H_9O_2$	$Pd(dba)_2$, DPPF, NaOtBu, 300 W, 30 m	no reaction

Scheme 5.4. Buchwald-Hartwig reaction to form the carbon-nitrogen bond.

We chose to switch to an amide linkage between the triphenylmethane dye and the HaloTag linker after we failed to label the HaloTag protein with our dansyl HaloTag linker **54**, but were able to label the protein with the TMR HaloTag linker **57** from Promega (Figure 5.5). No fluorescence was detected when the gel was visualized at 365 nm.

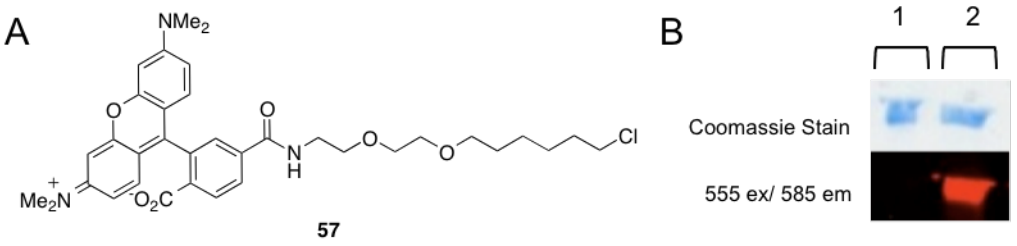
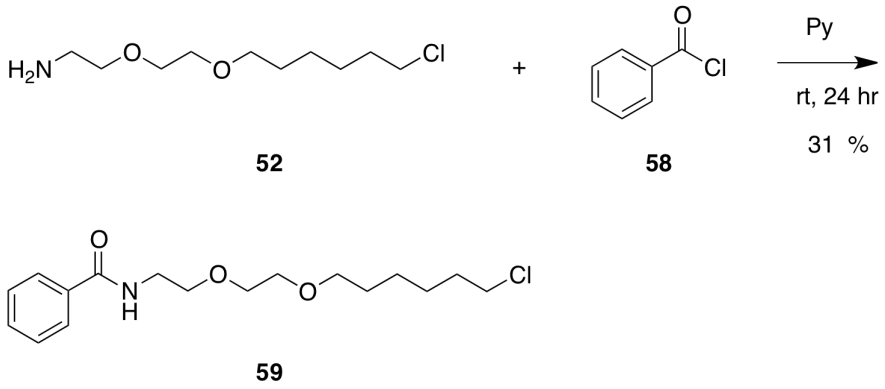


Figure 5.5. Labeling of HaloTag protein with fluorescent ligands; Lane 1: 1 hour labeling of HaloTag with dansyl-HaloTag linker **54**, Lane 2: 1 hour labeling of HaloTag with TMR linker **57**.

To further test the necessity of an amide linkage for labeling the HaloTag protein, we synthesized the simple benzamide linker **59** in one step by reacting the HaloTag linker **52** with benzoyl chloride (Scheme 5.5).



Scheme 5.5. Synthesis of the bezamide HaloTag linker **59**.

The benzamide HaloTag linker **59** was then tested for its ability to label the HaloTag protein in a competitive binding assay with the TMR HaloTag ligand. Purified HaloTag

protein that had been dialyzed into PBS was reacted with **59** and **57** in the following reaction conditions; (Figure 5.6 Lane 2) reagents were reacted with the protein at the same time, 10-fold excess of **59**, (Figure 5.6, Lane 3) **59** was reacted with protein for 1 hour prior to addition of **57**, 10-fold excess of **59**, (Figure 5.6, Lane 4) reagents were reacted with the protein at the same time, same concentration of each reagent, (Figure 5.6, Lane 5) **59** was reacted with protein for 1 hour prior to addition of **57**, same concentration of each reagent, (Figure 5.6, Lane 6) the protein was reacted with **57** only. After the addition of the final reagent, the reactions were carried out at room temperature for 1 hour, boiled and run on a 12 % SDS-PAGE gel. TMR fluorescence was detected at 555 nm and then the gel was stained with Coomassie blue.

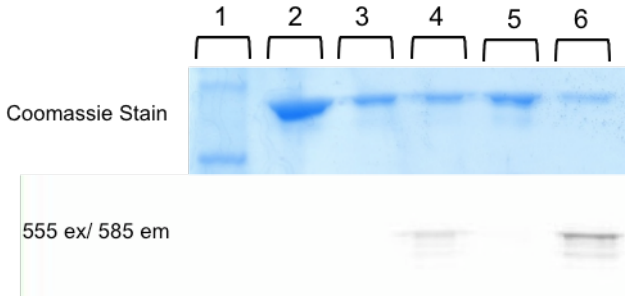
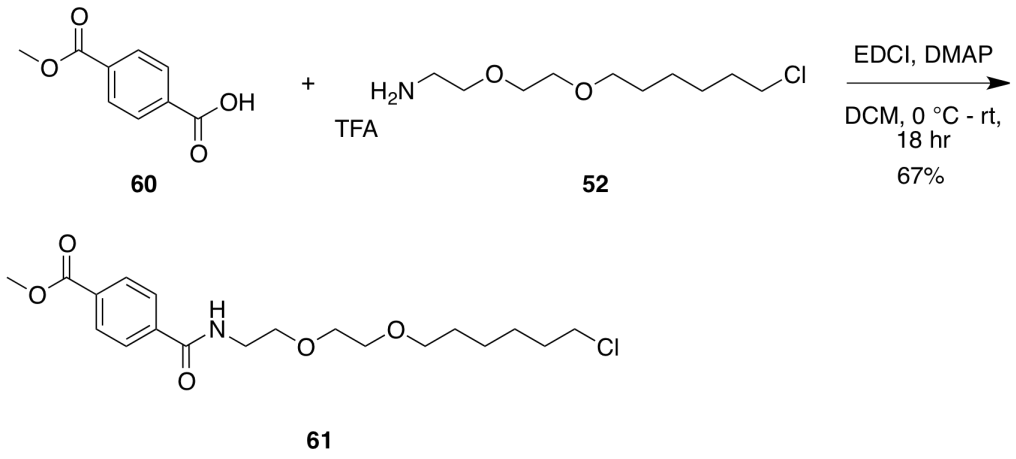


Figure 5.6. Competitive binding assay for **59**. Lane 1: molecular weight marker, Lane 2: simultaneous reaction of reagents with 10-fold excess **59**, Lane 3: sequential reaction of reagents with 10-fold excess **59**, Lane 4: simultaneous reaction of reagents at equimolar concentrations, Lane 5: sequential reaction of reagents at equimolar concentrations, Lane 6: reaction of HaloTag with **57** only.

As shown in the gel in lanes 3 and 5, **59** successfully blocked labeling of the HaloTag protein with the TMR ligand when the protein was reacted with **59** first. Additionally lane 2

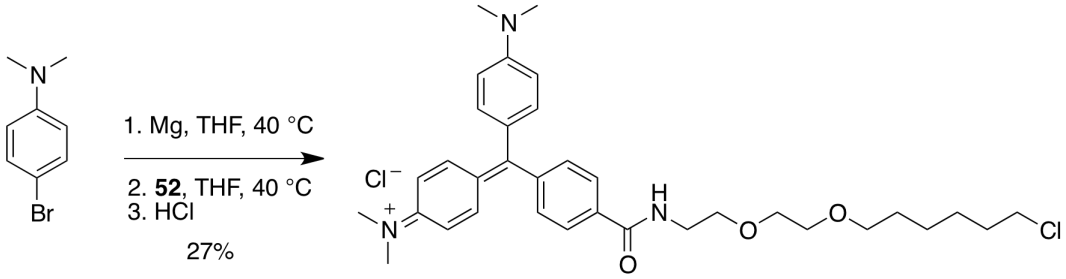
shows that a large excess of **59** will also result in no apparent TMR labeling when the two reagents are in competition with each other.

The amide linkage was then introduced into the malachite green HaloTag linker by coupling of **52** with the commercially available 4-(methoxycarbonyl)-benzoic acid (**60**) via 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide to produce the amide **61** in 67% yield (Scheme 5.6).



Scheme 5.6. Carbodiimide coupling of HaloTag linker to benzoic acid.

We were then able to synthesize the final dye by reaction of **61** with two equivalents of the Grignard reagent **62** followed by acidic work up in 27% yield (Scheme 5.7).



62 **63**

Scheme 5.7. Grignard reaction to produce the triphenylmethane dye **63**.

The blue-green dye exhibited an absorbance maximum at 605 nm (Figure 5.7) and is highly soluble in water and methanol.

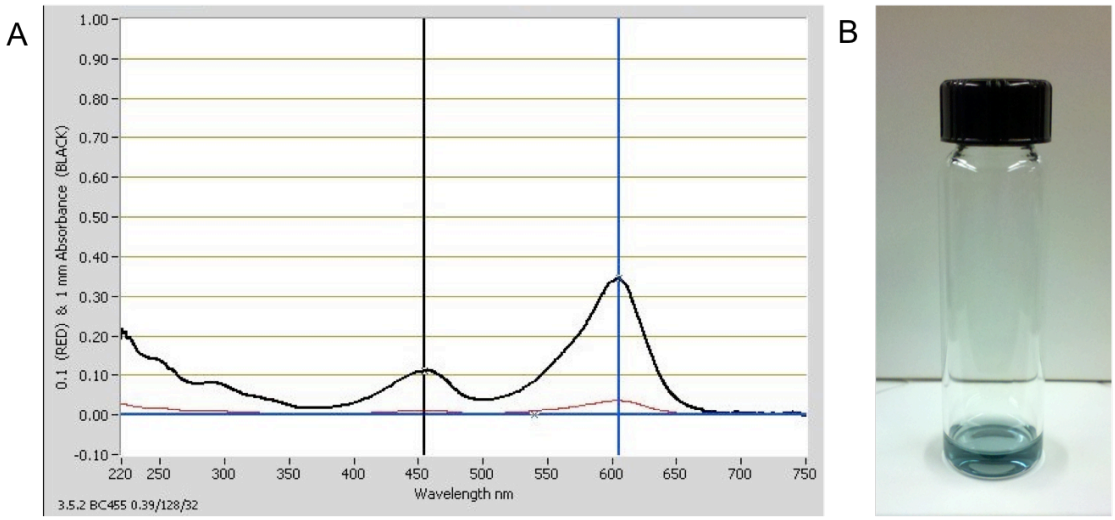


Figure 5.7. (A) UV-visible absorption spectrum for **63**, λ_{max} = 605 nm. (B) **63** in solution.

Malachite Green is commonly used by the fishing industry as an antifungal agent to quell fungal growth in fish hatcheries,[113] but it has also shown to be hazardous to rats, mice[114] and zebrafish.[115] Crystal Violet, a close structural analogue of Malachite Green has long been known to harbor antibacterial activity.[116] We tested the toxicity of **63** in *E. coli* at varying concentrations by monitoring the change in optical density over time. Due to the fact that **63** has strong absorbance near 600 nm, we subtracted the measured OD of cultures that were *not* inoculated with cells, but contained the same concentration of **63** and graphed the change in optical density over a 24 hour period (Figure 5.8).

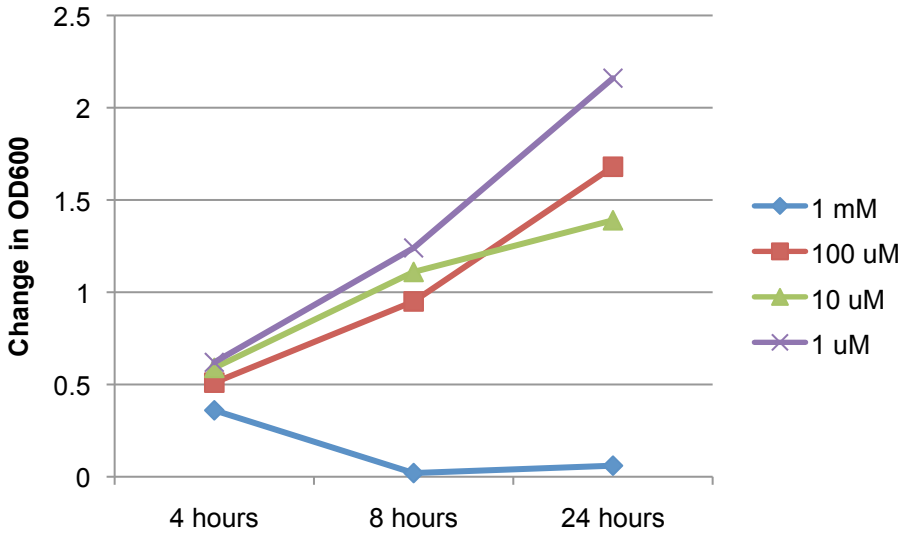


Figure 5.8. Toxicity study of **63** on *E. coli* at varying concentrations.

We observed that concentrations of **63** were not toxic in the micromolar range; the range at which we plan to use **63** for *in vivo* assays. We are now currently in the process of developing *in vivo* assays to utilize **63** as a viscosity sensor in *E. coli*.

5.3 Control of protein degradation with hydrophobic HaloTag ligands

The HaloTag protein has further been applied to controlling cellular processes in eukaryotes. In 2011, the Crews lab used the HaloTag protein to selectively control protein degradation in mammalian cells with a small molecule.[117] Fusion proteins of the HaloTag and a reporter gene (EGFP or luciferase) were selectively degraded after reaction with the hydrophobic HaloTag ligand **64** (Figure 5.9). The researchers hypothesize that the hydrophobic tag on the fusion protein would be perceived as protein unfolding by the cell, and cause the fusion to be further targeted for degradation.

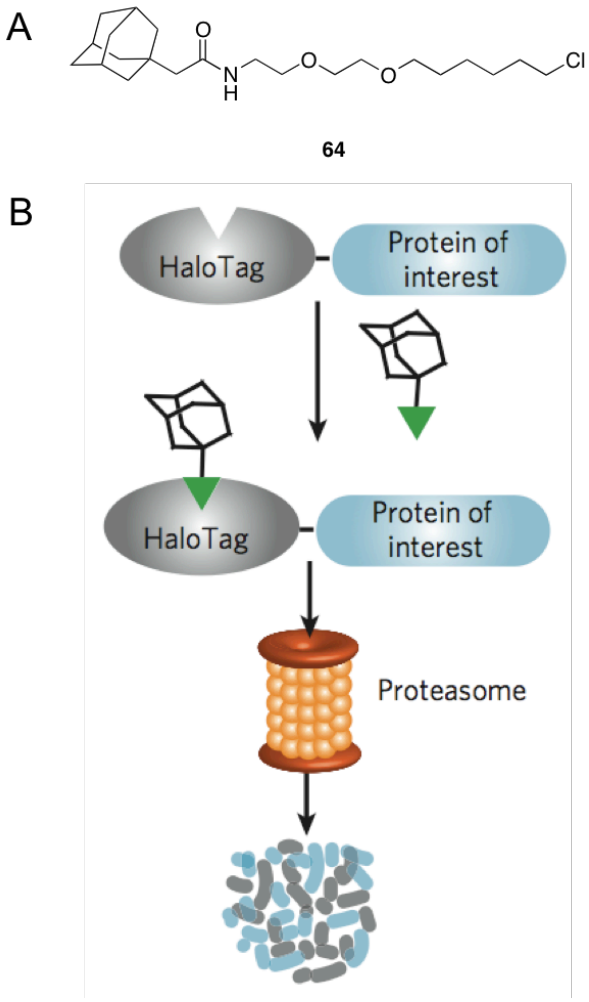
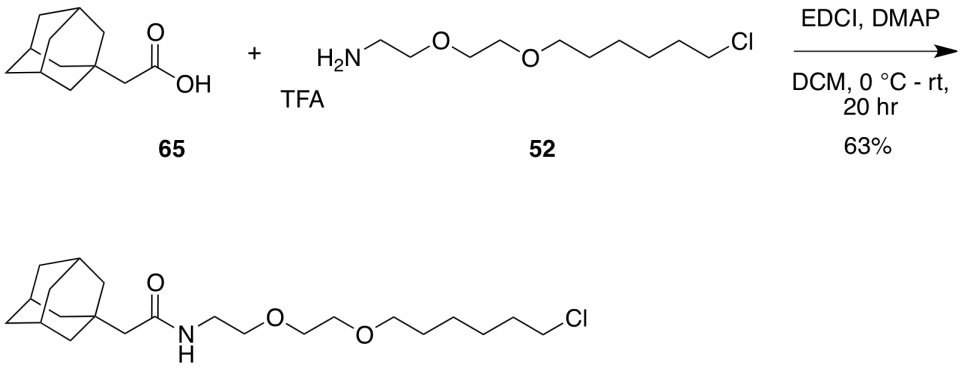


Figure 5.9. (A) Hydrophobic HaloTag ligand **64**. (B) General procedure for controlling protein degradation with a small molecule. A fusion of HaloTag:POI is reacted with **64** and then the fusion protein is targeted for degradation by the proteosome. Adapted from *Nat. Chem. Biol.* **2011**, *7*, 538.

Though this method offers an excellent way to achieve temporal control over protein degradation, it does not provide any spatial control. We aim to achieve spatiotemporal control over protein degradation in mammalian cells through the synthesis of a photocaged analogue of **64**. We first synthesized the ligand reported by the Crews lab to reconfirm their

results. The hydrophobic HaloTag ligand **64** was synthesized in 63% yield by carbodiimide coupling of the carboxylic acid **65** and the reactive HaloTag linker **52** (Scheme 5.8).



Scheme 5.8. Synthesis of the hydrophobic HaloTag ligand **64**.

We also constructed a HaloTag:EGFP fusion construct to test the ligands. *HaloTag* was PCR amplified from pFN18A (Promega) to contain EcoRI and BamHI restriction sites. The PCR product was ligated into pEGFP-N1 that had been digested with EcoRI and BamHI, and the circularized plasmid was transformed into NovaBlue *E. coli* cells. Positive clones were confirmed by sequencing. The plasmid pEGFP:HT was transfected into human embryonic kidney (HEK293T) cells, treated with **64** at 1 μM concentration, and imaged 24 hours post-transfection (Figure 5.10). DsRed was used as a transfection control.

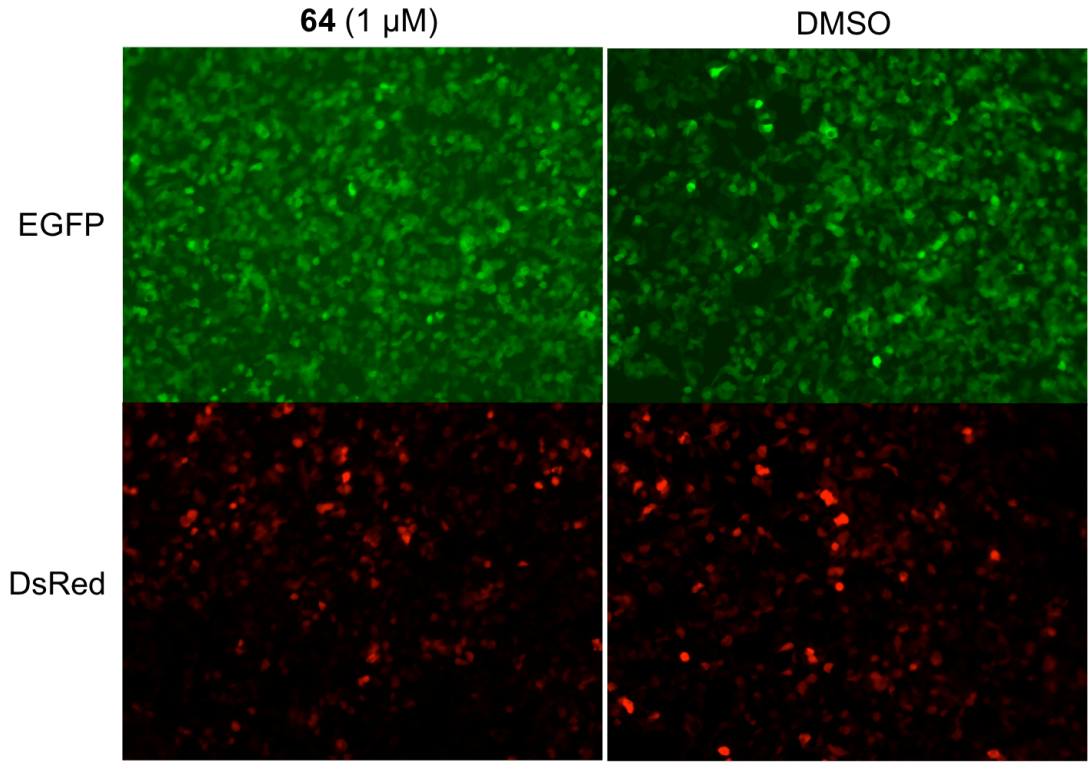
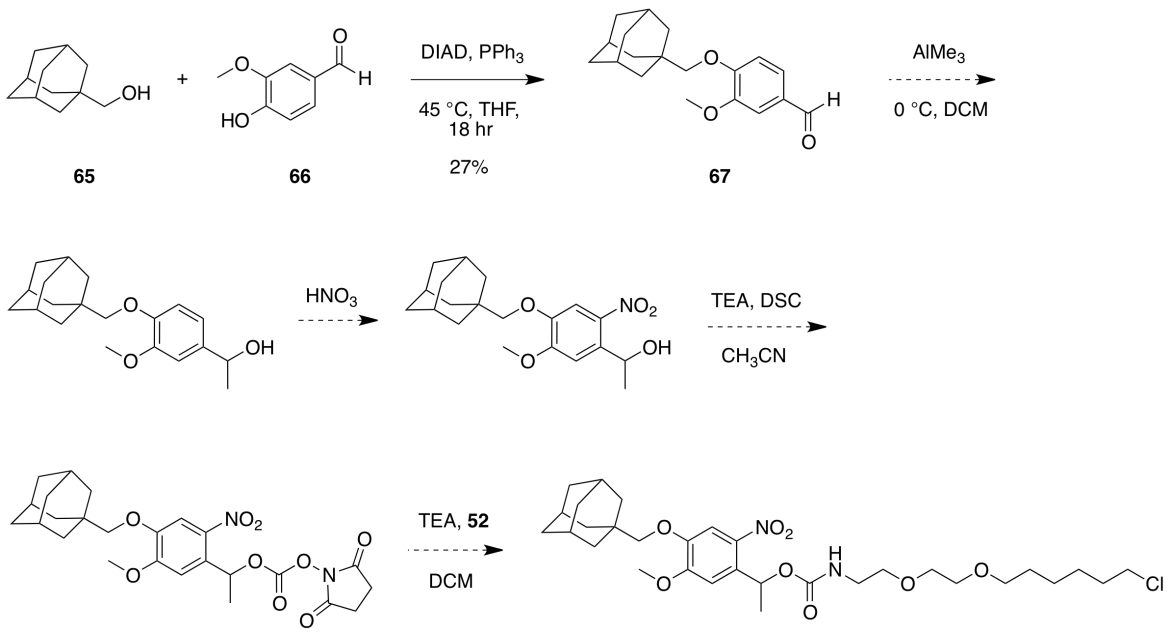


Figure 5.10. Fluorescent images of HEK293T cells treated with **64** (1μM) or DMSO. This experiment was performed by Kalyn Brown.

From the micrographs we did not detect any degradation of EGFP in HEK293T cells when treated with the hydrophobic HaloTag ligand **64**. After the sequence of the plasmid and the structure of **64** were reconfirmed, we discovered the variant of HaloTag in the pFN18A vector is HaloTag7, an improved form of the HaloTag2 protein used in the original Crews research, that has multiple mutations to increase the stability of the protein.[118] The two enzymes differ by 22 mutations through out the protein sequence. This marked change between HaloTag2 and HaloTag7 stability to hydrophobic tags was also very recently

confirmed by the Crews lab.[119] We are currently in the process of cloning an EGFP:HT fusion with the HaloTag2 variant of the protein.

The synthesis of a photocaged hydrophobic HaloTag linker is currently underway. Our synthetic design places the photocaging group between the hydrophobic adamantly group and the HaloTag linker so after decaging, the free amine will remain. We have begun the synthesis by reacting the commercially available adamantly methanol (**65**) with vanillin (**66**) using Mitsunobu reaction conditions to generate **67** in 27% yield. From there we plan to alkylate the aldehyde **67** with trimethylaluminum, perform nitration with concentrated nitric acid, followed by activation of the resulting alcohol with disuccinimidyl carbonate. The final step will be reaction of the HaloTag linker **52** with the activated ester to produce the final carbamate (Scheme 5.9).



Scheme 5.9. Proposed route to synthesize the photocaged hydrophobic HaloTag linker.

Additionally, we propose the synthesis of various other HaloTag linkers to further elucidate the structural requirements of the linker. We could successfully label the HaloTag protein with the amide **59**, but not with the sulfonamide **54**; which leads us to believe this linkage is involved in some interaction with the protein. We plan to further explore the structural requirements by synthesizing analogues that contain carbamate, carbonate, and urea functional groups (Figure 5.11).

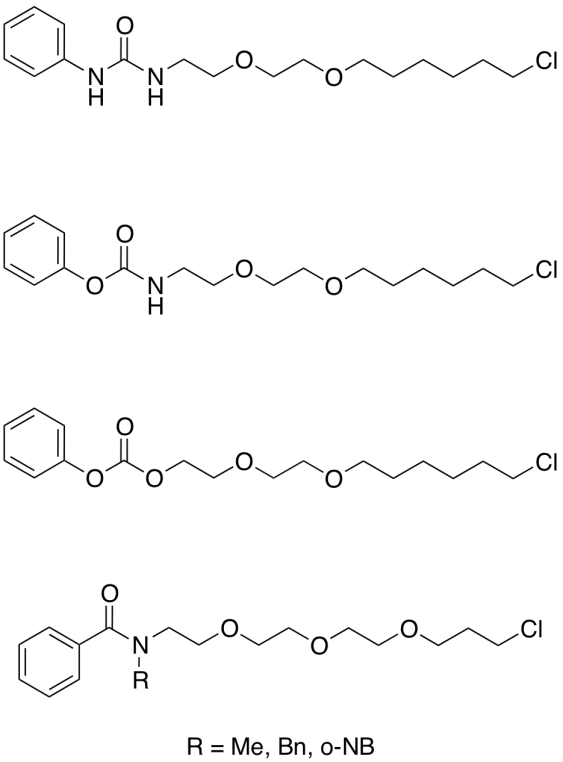


Figure 5.11. Proposed HaloTag ligands to study the necessity of an amide linkage.

We also plan to synthesize a substituted amide to see how the N-H bond is related to HaloTag ligand binding. Substitutions will range from groups of small steric demand (-CH_3) to larger groups (benzyl) as well as the ONB group for photochemical control over –NH bonding.

5.4 Summary and Outlook

The HaloTag protein has been used for a diverse set of applications. We have successfully synthesized a Malachite-Green HaloTag ligand that will be used as a 'turn-on' fluorescence sensor of viscosity *in vivo*. The Malachite-Green HaloTag ligand has been found to be non-toxic at concentrations that would be used in cellular assays. The next step of this project would be to design a fluorescence assay to measure differences in viscosity of various environments, and then use this assay for *in vivo* testing.

Additionally, we have synthesized a fluorescent HaloTag ligand, **54**, that reacted poorly with the enzyme. Through various competitive binding assays we were able to explore the required structural features of a HaloTag ligand, namely that an amide linkage is necessary for efficient HaloTag labeling. We plan to synthesize other HaloTag ligands to further probe the structures accommodated by the protein, including carbamate, carbonate, urea and ONB functional groups.

We have also begun the synthesis of a photocaged hydrophobic HaloTag ligand for photochemical control over protein degradation in mammalian cells. We have successfully constructed a HaloTag7:EGFP reporter system for use in mammalian cells, and are currently in the process of cloning a HaloTag2:EGFP reporter system. We then plan to utilize a photocaged hydrophobic HaloTag ligand to achieve spatial and temporal control over the regulated process of protein degradation.

5.5 Experimental Methods

Synthesis of *tert*-butyl (2-(2-hydroxyethoxy)ethyl)carbamate (49). Triethylamine (2.78 mL, 20 mmol) was added to a solution of 2-(2-aminoethoxy)ethanol **48** (1 mL, 10 mmol) in dioxane/H_2O (2 mL, 1:1) at 0 °C and stirred. After 15 min, Boc_2O (2.6 g, 12 mmol) was added and the reaction mixture was stirred at rt. After 24 h, the reaction mixture was diluted in saturated aqueous $NaHCO_3$ (10 mL) and extracted with EtOAC (2 × 15 mL). The combined organic layers were washed with brine (10 mL), dried over anhydrous Na_2SO_4 and solvent was removed by rotary evaporation to yield **49** as a clear oil (1.9 g, 92% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[120]

Synthesis of *tert*-butyl (2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)carbamate (51). NaH (60% in mineral oil, 60 mg, 1.47 mmol) was added to a solution of **49** (200 mg, 0.98 mmol) in anhydrous THF (2 mL) at 0 °C. The reaction mixture was stirred for 1 h at 0 °C, after which 1-chloro-6-iodohexane **50** was added (446.5 μL, 2.94 mmol). The reaction was allowed to warm to room temperature and stirred overnight. After 24 h, the reaction was quenched with saturated aqueous NH_4Cl (5 mL), extracted with EtOAc (3 × 10 mL) and the combined organic layers were washed with brine (15 mL). The crude product was dried over anhydrous Na_2SO_4, concentrated *in vacuo* and purified via silica gel chromatography, eluting with hexanes/EtOAc (2:1) to yield **51** as a yellow oil (215 mg, 68% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[120]

Synthesis of the 2-(2-((6-chlorohexyl)oxy)ethoxy)ethanamine (52). Trifluoroacetic acid (500 μL) was added dropwise to a solution of **51** (212 mg, 0.66 mmol) in DCM (2 mL) at 0 °C. After 2 h, the reaction was quenched with ice cold H_2O (2 mL) and the aqueous layer was decanted. The organic layer was dried with anhydrous Na_2SO_4 and filtered through a cotton plug to yield **52** as a light yellow oil (225 mg, 99% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[120]

Synthesis of *N*-(2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)-5-(dimethylamino)naphthalene-1-sulfonamide (54). Triethylamine (61 μL, 0.44 mmol) was added to a solution of **52** (30 mg, 0.089 mmol) in DCM (500 μL) at 0 °C. Dansyl chloride **53** (39 mg, 0.147 mmol) was added and the reaction mixture was stirred for 24 h, warming slowly to room temperature. The reaction mixture was then diluted in DCM (5 mL), washed with brine (5 mL), dried over anhydrous Na_2SO_4, concentrated *in vacuo* and purified by silica gel chromatography eluting with hexanes/EtOAc (5:3) to give **54** as a light green-yellow fluorescent oil (38 mg, 97% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 1.39-1.45 (m, 4H), 1.59-1.63 (m, 2H), 1.74-1.79 (m, 2H), 2.89 (s, 6H), 3.11 (q, J = 4.8 Hz, 2H), 3.37-3.34 (m, 8H), 3.52 (t, J = 6.6 Hz, 2H), 7.18 (d, J = 3.9 Hz, 1H), 7.51-7.56 (m, 2H), 8.27 (dd, J^1 = 18.3, J^2 = 9.6, 2H), 8.58 (d, J = 3.9 Hz, 1H).

Synthesis of methyl 4-(((trifluoromethyl)sulfonyl)oxy)benzoate (56). Triflluoromethanesulfonic anhydride (333 μL, 1.97 mmol) was added at 0 °C to a solution of methyl-4-hydroxybenzoate **55** (250 mg, 1.65 mmol) in pyridine (500 μL, 6.2 mmol) in DCM (2 mL). The reaction mixture was stirred for 12 h, warming slowly to room temperature. The reaction mixture was quenched with ice water, acidified with aqueous HCl (1M, 5 mL), and extracted with EtOAc (2 × 10 mL). The combined organic layers were washed with brine, dried over anhydrous Na_2SO_4, and the solvent was removed *in vacuo* to yield the product **56** as an orange oil (420 mg, 89% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ =3.95 (s, 3H), 7.36 (d, J = 9.0 Hz, 2H), 8.15 (d, J = 9.0 Hz, 2H).

HaloTag protein expression in *E. coli* cells. Plasmid pET21-HT2 was transformed into BL21G cells and grown overnight on media supplemented with kanamycin (50 μg/mL). A single colony was then inoculated into 3 mL of LB media containing kanamycin (50 μg/mL) and grown overnight at 37 °C with shaking. In the morning, the culture was diluted 1:100 into LB media (25 mL) containing kanamycin (50 μg/mL) and grown to OD_{600} = 0.6. Protein expression was induced with 0.1 mM IPTG, and continued for 6 h at 37 °C. Cells were harvested by centrifugation, resuspended in a lysis butter (300 mM NaCl, 5 mM HaH_2PO_4, 45 mM Na_2HPO_4, 10 mM imidazole) containing lysozyme (4 μL, 1 mg/mL) and incubated on ice for 2 h. Cells were then lysed by sonication and the cleared lysate was applied to Ni-NTA resin (Qiagen) and incubated at 4 °C for 1 h. The resin was then washed with a low imidazole butter (10-20 mM) and protein was eluted from the resin with a high imidazole

butter (250 mM). Purified protein was then run on a 12% SDS-PAGE gel and visualized with Coomassie stain. Protein concentration was determined by Bradford assay (BioRad).

HaloTag labeling assay. Purified HaloTag protein (2.5 μL, 1 μg) that had been dialyzed into PBS buffer was reacted with the **54** (2 μL, 50 μM or 5μM) and **57** (2 μL, 50 μM) in 10 μL reaction mixtures and allowed to proceed for 1 h at room temperature. Reaction mixtures were then boiled with a loading dye and run on a 12% SDS-PAGE gel. The gel was visualized for TMR fluorescence (Typhoon FLA 7000, GE; filter 535 nm/[580]) and then stained with Coomassie blue.

Synthesis of *N*-(2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)benzamide (59). HaloTag linker **52** (60 mg, 0.16 mmol) was dissolved in anhydrous pyridine (500 μL) and cooled to 0 °C. Benzoyl chloride (50 μL, 0.43 mmol) was added to the solution and the reaction mixture was stirred for 24 h. The reaction mixture was quenched with aqueous HCl (1 M, 2 mL) and extracted with EtOAc (2 × 5 mL). The combined organic layers were washed with brine, dried over anhydrous Na_2SO_4 and concentrated. The crude product was purified via silica gel chromatography, eluting with hexanes/EtOAc (2:1) to give **59** as a yellow solid (19.1 mg, 31% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 1.23-1.49 (m, 6H), 1.71-1.76 (m, 2H), 3.32 (t, J = 6.6 Hz, 2H), 3.42 (dd, J^1 = 5.1 Hz, J^2 = 1.8 Hz, 2H), 3.50 (t, J = 6.6 Hz, 2H), 3.59 (dd, J^1 = 6.6 Hz, J^2 = 2.1 Hz, 2H), 3.89 (t, J = 5.4 Hz, 2H), 4.25 (t, J = 5.7 Hz, 2H), 7.13-7.18 (m,

2H), 7.22-7.27 (m, 1H), 7.43-7.47 (m, 2H). ^{13}C NMR (400 MHz, d-6 DMSO): δ = 25.6, 26.8, 29.7, 32.7, 46.0, 47.2, 68.2, 69.3, 70.1, 70.4, 70.9, 129.1, 129.9, 132.3, 133.5, 136.7, 174.3.

HaloTag competitive binding assay. Purified HaloTag protein (2.5 μL, 1 μg) that had been dialyzed into PBS buffer was reacted with the **59** (2 μL, 50 μM or 5μM) and **57** (2 μL, 50 μM) in 10 μL reaction mixtures. In reactions where reagents were added at the same time, the reaction was allowed to proceed for 1 h at room temperature, and then stored at 4 °C. In the reaction where **59** was added first, the initial reaction was carried out for 1 h at room temperature, then **57** was added and the reaction was carried out for 1 h at room temperature. Reaction mixtures were then boiled with a loading dye and run on a 12% SDS-PAGE gel. The gel was visualized for TMR fluorescence (Typhoon FLA 7000, GE; filter 535 nm/[580]) and stained with Coomassie blue.

Synthesis of methyl 4-((2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)carbamoyl)benzoate (61). 4-(methoxycarbonyl)benzoic acid **60** (119 mg, 0.66 mmol), **52** (213 mg, 0.66 mmol) and catalytic DMAP were dissolved in DCM (2 mL) and cooled to 0 °C. EDCI (189 mg, 0.99 mmol) was added to the solution and the reaction mixture was stirred for 20 h, warming to room temperature. The reaction was diluted in H_2O (3 mL), extracted with EtOAc (2 × 10 mL) and the combined organic layers were washed with brine. The crude product was dried over anhydrous Na_2SO_4, concentrated and purified by silica gel chromatography, eluting with MeOH/DCM (1:19) giving **61** as an off white solide (170 mg, 67% yield). ^{1}H NMR (300

MHz, $CDCl_3$): δ = 1.32-1.41 (m, 4H), 1.54-1.57 (m, 2H), 1.71-1.75 (m, 2H), 3.45 (t, 5.1 Hz, 4H), 3.50 (t, 5.1 Hz, 4H), 3.59 (dd, J^1 = 4.2 Hz, J^2 = 2.4 Hz, 2H), 3.67 (dd, J^1 = 8.4 Hz, J^2 = 1.8 Hz, 2H), 3.94 (s, 3H), 7.84 (d, J = 6.0 Hz, 2H), 8.09 (d, J = 6.6 Hz, 2H). LRMS *m/z* calcd for ($C_{19}H_{28}ClNO_5$) $[M+H]^+$: 386.17, found: 386.07.

Synthesis of Malachite Green:HaloTag linker (63). 4-Bromo-N,N-dimethylaniline (555 mg, 2.77 mmol) and magnesium metal (74 mg, 3 mmol) were dissolved in anhydrous THF (2 mL), heated to 40 °C, and stirred for 1 h. Amide **61** (100 mg, 0.25 mmol) was dissolved in THF (750 μL) and added to the solution of Grignard reagent at room temperature. The reaction mixture was stirred overnight, after which it appeared deep green in color, and then cooled to 0 °C. The dye was generated by acidic work up with aqueous HCl (1M, 200 μL). The reaction mixture was diluted in H_2O (5 mL) and then extracted with DCM (2 × 5 mL). The organic layer was dried over anhydrous Na_2SO_4, concentrated and purified via silica gel chromatography, eluting with MeOH/DCM (1:49) producing the final product **63** as a blue-green solid (41 mg, 27% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 1.32-1.41 (m, 4H), 1.54-1.57 (m, 2H), 1.71-1.75 (m, 2H), 2.93 (s, 12H), 3.44-3.50 (m, 6H), 3.57-3.59 (m, 2H), 3.62-3.65 (m, 4H), 6.64 (d, J = 8.4 Hz, 4H), 7.082 (d, J = 8.4 Hz, 4H), 7.40 (d, J = 8.1 Hz, 2H), 7.69 (d, J = 8.4 Hz, 2H). ^{13}C NMR (400 MHz, $CDCl_3$) δ = 25.1, 26.7, 30.0, 33.5, 39.9, 41.0, 44.7, 70.1, 70.2, 72.1, 112.1, 114.8, 121.8, 127.5, 128.2, 129.5, 133.9, 135.7, 141.0, 149.9, 152.0, 167.9. LRMS *m/z* calcd for ($C_{34}H_{45}ClN_3O_3^+$) [M+H]+: 379.31; found: 378.25.

HaloTag-Malachite Green Toxicity Assay. A saturated culture of *E. coli* (10 μL, DH5α, NEB) was diluted into 1 mL of LB media containing **63** (1 mM, 100 μM, 10 μM and 1 μM concentrations). Control solutions of LB containing **63** (1 mM, 100 μM, 10 μM and 1 μM concentrations) lacking *E. coli* were also prepared. Samples were incubated in a 96-deep well block at 37 °C for 24 hours. The OD_{600} was recorded after 4 h, 8 h and 24 h for all samples, and the value of the control solutions was subtracted from the samples inoculated with cells and plotted against time.

Synthesis of hydrophobic HaloTag linker (64). 1-adamantylacetic acid **65** (74 mg, 0.38 mmol), **52** (121 mg, 0.38 mmol) and DMAP (92 mg, 0.76 mmol) were dissolved in DCM (1.5 mL) and cooled to 0 °C. EDCI (108 mg, 0.57 mmol) was added to the solution, and the reaction mixture was stirred for 20 h, warming to room temperature. The reaction mixture was diluted in H_2O, extracted with EtOAc (2 × 10 mL), and the combined organic layers were washed with aqueous HCl (1M, 5 mL) and brine (5 mL). The crude product was dried over anhydrous Na_2SO_4, concentrated *in vacuo* and purified by silica gel chromatography, eluting with hexanes/EtOAc (2:1) to give **64** as a light yellow oil (95 mg, 63% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[117]

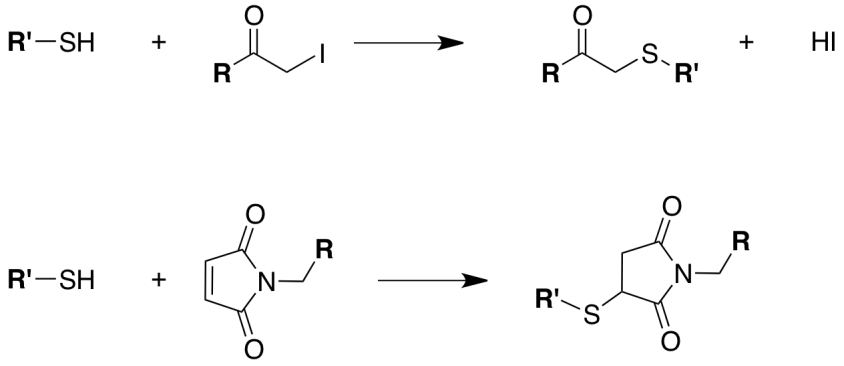
Synthesis of 4-(-adamantan-1-ylmethoxy)-3-methoxybenzaldehyde (67). Adamantyl methanol **65** (1 g, 6.01 mmol), vanillin **66** (1.47 g, 7.23 mmol), DIAD (2.1 mL, 10.82 mmol) and triphenylphosphine (2.8 g, 10.82 mmol) were dissolved in THF (15 mL) and heated to 45

°C with stirring for 18 h. The reaction mixture was cooled to room temperature, the solvent was evaporated, and the crude mixture was resuspended in DCM (15 mL), washed with H_2O (10 mL) and brine (10 mL). The organic layer was dried over anhydrous Na_2SO_4, concentrated and purified by silica gel chromatography, eluting with hexanes/EtOAc (4:1) to give **67** as a light yellow solid (567 mg, 27% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 1.72 (d, *J* = 13.8 Hz, 12H), 2.03-2.05 (m, 3H), 3.62 (s, 2H), 3.91 (s, 3H), 6.95 (d, *J* = 7.8 Hz, 1H), 7.26-7.44 (m, 2H), 9.83 (s, 1H).

CHAPTER 6 – SYNTHESIS OF A PHOTOCAGED CYSTEINE CROSSLINKER

6.1 Introduction to Cysteine Bioconjugation

Cysteine residues within proteins play an important role in protein folding, catalysis and enzyme chemistry. The nucleophilic amino acid has been used extensively in protein labeling technology, FRET, as well as reactions with metal colloids.[121] Many of the modifications of cysteine exploit the nucleophilic nature of the amino acid. Cysteine readily reacts with haloacetyl functional groups in substitution reactions and with maleimides through addition reactions (Scheme 6.1).



Scheme 6.1. Examples of cysteine biochemistry; reaction with haloacetyl groups (top) and maleimide derivatives (bottom) to form thio-ether bonds.

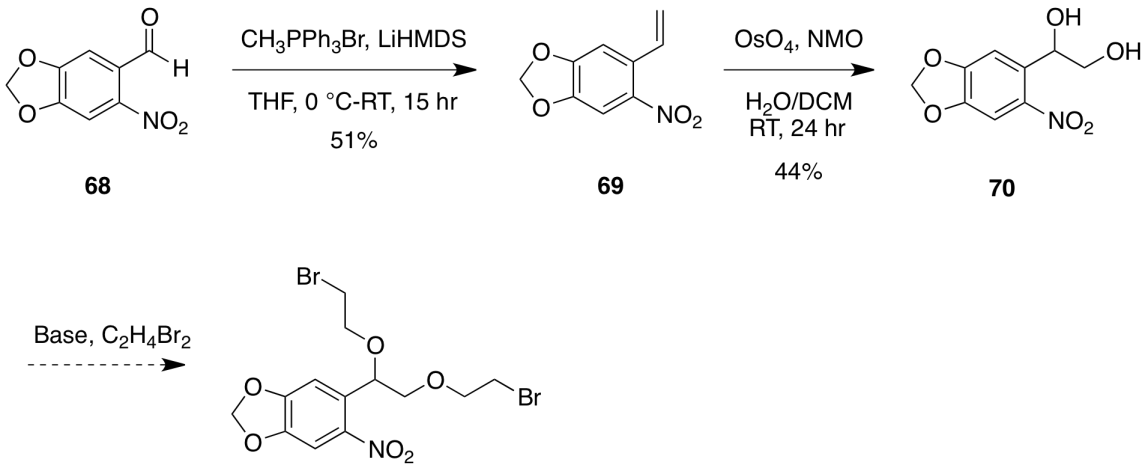
Thiol-maleimide chemistry is of particular use in biological systems as the reaction is specific to thiol groups at pH's between 6.5 and 7.5.[121] The Ha group (UCIC) uses thio-maleimide chemistry to label proteins for single molecule FRET studies.[122] In particular,

single molecule FRET was used to study conformational changes of the *E. coli* helicase enzyme Rep during ATP hydrolysis. Rep exists as a monomer made up of 4 subdomains that together bind and unwind dsDNA. The function of the subdomain 2B however remains unclear. To further elucidate the function of subdomain 2B and the dsDNA substrate, Rep was mutated to change all cysteine residues to aliphatic amino acids, and then single cysteine mutations were introduced. The single-cysteine Rep mutants were labeled with a Cy3 maleimide fluorophore and reacted with a Cy5 labeled nucleotide for FRET studies. The FRET data suggested that there are at least two confirmations of Rep and partial-duplex DNA, and that it is more likely Rep protein itself is in a closed confirmation than an open confirmation.

The development of a photocaged cysteine crosslinker would further the study of enzyme confirmations, such as Rep helicase. A photolabile crosslinker would allow for dynamic studies of confirmation changes and enzyme activity. Prior to UV irradiation the enzyme would be locked into a more rigid confirmation and be either activated or deactivated. After UV irradiation, the enzyme can achieve another confirmation and become inactivated or activated. The Ha group has successfully used a dimaleimide linker to control RepA activity (unpublished results). We propose the synthesis of a photocaged dimaleimide linker to achieve spatiotemporal control over helicase activity. The enzyme will be active in the closed, linked form. After irradiation the linker will be cleaved, allowing the enzyme to fold to an open inactive conformation.

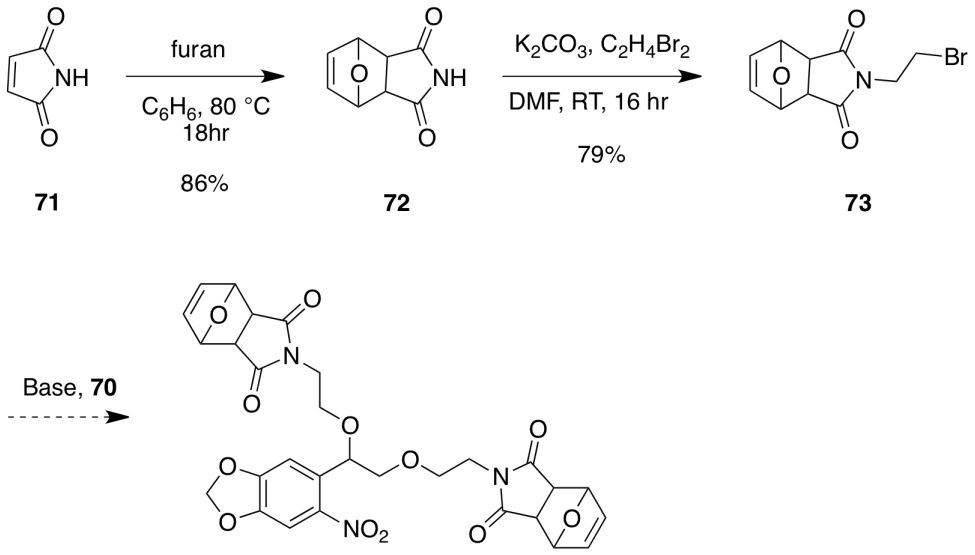
6.2 Synthesis of a photocaged cysteine crosslinker

To synthesize the photocaged cysteine crosslinker, 6-nitropiperonal **68** was reacted with methyl triphenylphosphonium bromide under Wittig conditions to generate the styrene **69** in 51% yield. The styrene **69** was oxidized to the diol **70** by osmium tetroxide. We then attempted to alkylate **70** with 1,2-dibromoethane, but were unsuccessful (Scheme 6.2).



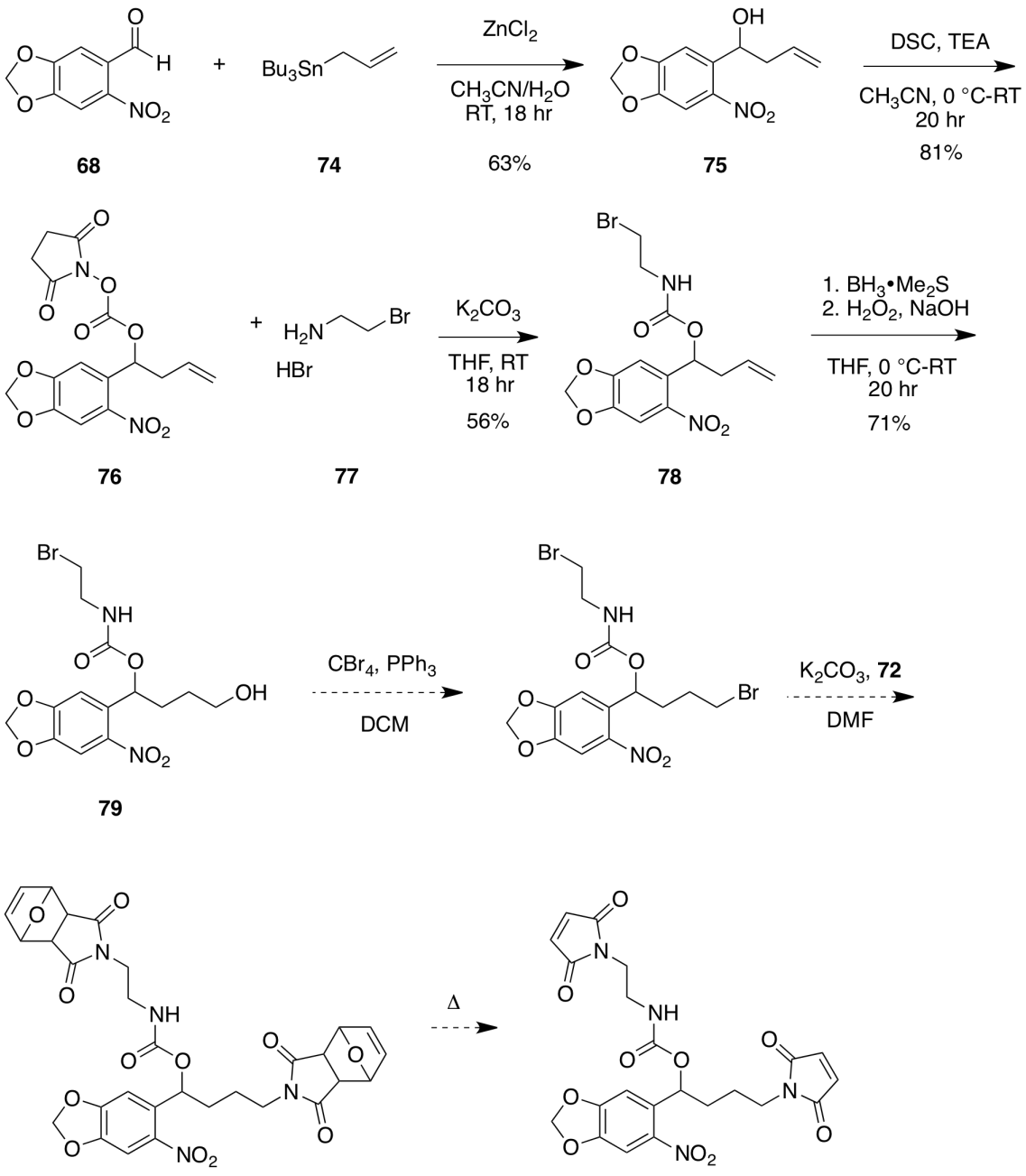
Scheme 6.2. Initial synthetic route to synthesize a photocaged cysteine crosslinker

Variations in base and solvent did not lead to the desired product. We then attempted to alkylate **70** with alkyl bromide **73**. First, maleimide **71** was protected with furan through a Diels-Alder reaction to give **72** in 86% yield. The protected maleimide was then alkylated with 1,2-dibromoethane to give **73**. We again attempted to react the diol **70** with the alkyl bromide **73** under various reaction conditions but were unable to obtain the desired product (Scheme 6.3).



Scheme 6.3. Second attempted synthetic route to synthesize a dimaleimide cysteine crosslinker.

Currently we are in the process of synthesizing the crosslinker through another synthetic route (Scheme 6.4). We began with the allylation of **68** with tributyl stannane **74** to generate the alcohol **75**. The alcohol **75** was activated with disuccinimidyl carbonate to provide the activated ester **76** in 81% yield. We then reacted **76** with 2-bromoethylamine hydrobromide **77** in the presence of potassium carbonate in THF to give **78**. The alkene then underwent a hydroboration-oxidation reaction with borane-methyl sulfide in anti-Markovnikov conditions to give the alcohol **79** in 71% yield. Our next steps are to convert the alcohol to the bromide through an Appel reaction, followed by Sn_2 reaction of the bromides with the protected maleimide **72**. The final step of the synthesis will be a reverse Diels-Alder reaction to remove the furan groups on the maleimides.



68 **74** **75** **76** **77** **78** **79**

Scheme 6.4. Current synthetic route to synthesize the photocaged cysteine crosslinker.

After synthesis of the photocaged cysteine crosslinker is completed, we will use the compound in *E. coli* expressing a RepA cysteine mutant to test its ability to photoregulate DNA helicase activity.

6.3 Summary and Outlook

We have begun the synthesis of a photocaged cysteine crosslinker for the photochemical control over the DNA helicase RepA. We plan to continue the synthesis outlined in Scheme 6.4, after which the crosslinker will be utilized *in vivo* to control RepA helicase activity. We will design an assay that correlates helicase activity to reporter gene function and use this assay to exhibit spatial and temporal control over RepA with the photocaged cysteine crosslinker.

6.4 Experimental Methods

Synthesis of 5-nitro-6-vinylbenzo[*d*][1,3]dioxole (69). Methyltriphenylphosphonium bromide (475 mg, 1.33 mmol) was dissolved in THF (2 mL) and cooled to 0 °C. LiHMDS (1.33 mL, 1M in THF) was added to the solution and the reaction mixture was stirred at 0 °C to generate the ylide. Nitropipperonal **68** (200 mg, 1.03 mmol) was added and the reaction mixture was stirred for 15 h, allowed to warm to room temperature. The reaction mixture was diluted into H_2O (10 mL), extracted with EtOAc (3 × 10 mL), washed with brine, dried over anhydrous Na_2SO_4, concentrated and purified by silica gel chromatography; eluting with

hexanes/EtOAc (2:1) to give **69** as a yellow solid (100 mg, 51% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[123]

Synthesis of 1-(6-nitrobenzo[*d*][1,3]dioxol-5-yl)ethane-1,2-diol (70). Alkene **69** (167 mg, 0.87 mmol) and NMO (131 mg, 1.12 mmol) were dissolved in H_2O/DCM (1:3) and stirred vigorously. One crystal of OsO_4 was added and the reaction mixture was stirred vigorously for 24 h at room temperature. The reaction mixture was then diluted in DCM (10 mL), washed with saturated $NaHCO_3$ (10 mL), saturated $NaHSO_3$ (2 × 10 mL). The combined aqueous layers were extracted with DCM (2 × 10 mL) and the combined organic layer was dried over anhydrous Na_2SO_4. The solvent was then removed *in vacuo* to produce the diol **70** as a dark solid (85 mg, 44% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[123]

Synthesis of protected maleimide (72). Maleimide **71** (500 mg, 5.15 mmol) and furan (560 μL, 7.7 mmol) were dissolved in anhydrous benzene (3 mL) and heated under reflux overnight. The resulting solid was filtered from the reaction mixture, washed with Et_2O and dried *in vacuo* to give **72** as a white solid (736 mg, 86%). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$).[124]

Synthesis of alkylbromide (73). Protected maleimide **72** (200 mg, 1.21 mmol) and K_2CO_3 (501 mg, 3.63 mmol) were dissolved in DMF (2 mL) and stirred. After 10 min, 1,2-

dibromoethane (209 μL, 2.42 mmol) was added to the reaction mixture, which was then stirred for 16 hr. The solvent was removed by rotary evaporation, and then the reaction mixture was dissolved in H_2O, extracted with EtOAc (2 × 10 mL), the combined organic layers were washed with brine (10 mL), and dried over anhydrous Na_2SO_4 and purified by silica gel chromatography, eluting with hexanes/EtOAc (2:1) to yield **73** as a light red solid (260 mg, 79% yield). The analytical data obtained matched reported literature (^{1}H NMR, ^{13}C NMR, $CDCl_3$). [125]

Synthesis of 1-(6-nitrobenzo[*d*][1,3]dioxol-5-yl)but-3-en-1-ol (75). Aldehyde **68** (300 mg, 1.54 mmol) was dissolved in CH_3CN/H_2O (5:1). Allyltributyl stannane **74** (714 μL, 2.31 mmol) and $ZnCl_2$ (314 mg, 2.31 mmol) were added to the solution and the reaction mixture was stirred overnight. After 18 h, the solvent was removed *in vacuo* and the reaction mixture was diluted in H_2O (5 mL), extracted with DCM (3 × 10 mL), and the combined organic layers were washed with saturated $NaHCO_3$, brine and dried over anhydrous Na_2SO_4. The product was purified by silica gel chromatography, eluting with hexanes/EtOAc (4:1) to give **75** as a light yellow solid (230 mg, 65% yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ = 2.52 (dd, J^1 = 119.2 Hz, J^2 = 8.8 Hz, 2H), 5.219 (dd, J^1 = 16.0 Hz, J^2 = 4.0 Hz, 2H), 5.38 (d, J = 8.4 Hz, 1H), 5.931 (m, 1H), 6.13 (s, 2H), 7.27 (s, 1H), 7.50 (s, 1H).

Synthesis of DSC-activated 1-(6-nitrobenzo[*d*][1,3]dioxol-5-yl)but-3-en-1-ol (76). Alcohol **75** (62 mg, 0.26 mmol) and TEA (100 μL, 0.78 mmol) were dissolved in CH_3CN (1

mL) at 0 °C. DSC (200 mg, 0.78 mmol) was added to the solution and the reaction mixture was stirred overnight. After 20 h, the solvent was removed by rotary evaporation and purified by silica gel chromatography, eluting with hexanes/EtOAc (1:1). The product **76** was isolated at a light yellow oil (80 mg, 81% yield). ^{1}H NMR (300 MHz, $CDCl_3$): δ =2.60 (d, J = 8.8 Hz, 2H), 2.79 (s, 2H), 5.22 (dd, J^1 = 16.8 Hz, J^2 = 3.3 Hz, 2H), 5.93 (m, 1H), 6.15 (s, 2H), 6.41 (m, 1H), 7.06 (s, 1H), 7.53 (s, 1H).

Synthesis of carbamate (78). DCS-activated alcohol **76** (80 mg, 0.21 mmol), K_2CO_3 (58 mg, 0.42 mmol) and bromoethylamine hydrobromide (87 mg, 0.42 mmol) were dissolved in THF (1.5 mL) and stirred overnight at room temperature. After 18 h, the reaction was acidified with aqueous HCl (1M, 3 mL), extracted with EtOAc (2 × 10 mL), the combined organic layers washed with brine (10 mL), dried over anhydrous Na_2SO_4, and dried *in vacuo* to give the product as a light yellow oil (48 mg, 56%). ^{1}H NMR (300 MHz, $CDCl_3$): δ = 2.54 (m, 1H), 2.65 (d, J = 6.9 Hz, 2H), 3.42 (t, J = 5.7 Hz, 2H), 3.52 (t, J = 5.7 Hz, 2H), 5.11 (m, 2H), 5.24 (br, NH), 5.84 (m, 1H), 6.10 (s, 2H), 6.28 (m, 1H), 6.96 (s, 1H), 7.49 (s, 1H).

Synthesis of alcohol (79). Alkene **78** (100 mg, 0.26 mmol) was dissolved in THF (2 mL) and cooled to 0 °C. Borane-methyl sulfide (100 μL, 2M in THF) was added to the solution and the reaction mixture was stirred overnight, allowing to warm to room temperature. After 20 h, 30% H_2O_2 (500 μL) and NaOH (500 μL, 3M) was added to the reaction mixture, and then stirred for an additional 24 h. The reaction mixture was then diluted in H_2O, extracted with

DCM (2 × 10 mL), the combined organic layers washed with brine (10 mL), dried over anhydrous Na_2SO_4 and dried *in vacuo* to yield **79** as a yellow oil (74 mg, 71% yield).

REFERENCES

1. Agapakis CM, Silver PA: **Synthetic biology: exploring and exploiting genetic modularity through the design of novel biological networks.** *Mol Biosyst* 2009, **5**:704-713.

2. Tanouchi Y, Pai A, You L: **Decoding biological principles using gene circuits.** *Mol Biosyst* 2009, **5**:695-703.

3. Weber W, Fussenegger M: **Molecular diversity--the toolbox for synthetic gene switches and networks.** *Curr Opin Chem Biol* 2011, **15**:414-420.

4. Andrianantoandro E, Basu S, Karig DK, Weiss R: **Synthetic biology: new engineering rules for an emerging discipline.** *Mol Syst Biol* 2006, **2**:2006.0028.

5. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M: **A tunable synthetic mammalian oscillator.** *Nature* 2009, **457**:309-312.

6. Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in Escherichia coli.** *Nature* 2000, **403**:339-342.

7. Tan C, Marguet P, You L: **Emergent bistability by a growth-modulating positive feedback circuit.** *Nat Chem Biol* 2009, **5**:842-848.

8. Clackson T: **Controlling mammalian gene expression with small molecules.** *Curr Opin Chem Biol* 1997, **1**:210-218.

9. Buskirk AR, Liu DR: **Creating small-molecule-dependent switches to modulate biological functions.** *Chem Biol* 2005, **12**:151-161.

10. Athavankar S, Peterson BR: **Control of gene expression with small molecules: biotin-mediated acylation of targeted lysine residues in recombinant yeast.** *Chem Biol* 2003, **10**:1245-1253.

11. Mayer G, Heckel A: **Biologically active molecules with a "light switch".** *Angew Chem Int Ed Engl* 2006, **45**:4900-4921.

12. Kaplan JH, Forbush B, Hoffman JF: **Rapid photolytic release of adenosine 5'-triphosphate from a protected analogue: utilization by the Na:K pump of human red blood cell ghosts.** *Biochemistry* 1978, **17**:1929-1935.

13. Cruz F, Koh J, Link K: **Light-activated gene expression**. *Journal of the American Chemical Society* 2000, **122**:8777-8778.

14. Adams SR, Tsien RY: **Controlling cell chemistry with caged compounds.** *Annu Rev Physiol* 1993, **55**:755-784.

15. Lin W, Albanese C, Pestell RG, Lawrence DS: **Spatially discrete, light-driven protein expression.** *Chem Biol* 2002, **9**:1347-1353.

16. Young DD, Deiters A: **Photochemical activation of protein expression in bacterial cells.** *Angew Chem Int Ed Engl* 2007, **46**:4290-4292.

17. Young DD, Garner RA, Yoder JA, Deiters A: **Light-activation of gene function in mammalian cells via ribozymes.** *Chem Commun (Camb)* 2009:568-570.

18. Cambridge SB, Geissler D, Keller S, Cürten B: **A caged doxycycline analogue for photoactivated gene expression.** *Angew Chem Int Ed Engl* 2006, **45**:2229-2231.

19. Umeda N, Ueno T, Pohlmeyer C, Nagano T, Inoue T: **A photocleavable rapamycin conjugate for spatiotemporal control of small GTPase activity.** *J Am Chem Soc* 2011, **133**:12-14.

20. Karginov A, Zou Y, Shirvanyants D, Kota P, Dokholyan N, Young D, Hahn K, Deiters A: **Light Regulation of Protein Dimerization and Kinase Activity in Living Cells Using Photocaged Rapamycin and Engineered FKBP**. *Journal of the American Chemical Society* 2011, **133**:420-423.

21. Sadovski O, Jaikaran AS, Samanta S, Fabian MR, Dowling RJ, Sonenberg N, Woolley GA: **A collection of caged compounds for probing roles of local translation in neurobiology.** *Bioorg Med Chem* 2010, **18**:7746-7752.

22. Cambridge SB, Geissler D, Calegari F, Anastassiadis K, Hasan MT, Stewart AF, Huttner WB, Hagen V, Bonhoeffer T: **Doxycycline-dependent photoactivated gene expression in eukaryotic systems.** *Nat Methods* 2009, **6**:527-531.

23. Sauers DJ, Temburni MK, Biggins JB, Ceo LM, Galileo DS, Koh JT: **Light-activated gene expression directs segregation of co-cultured cells in vitro.** *ACS Chem Biol* 2010, **5**:313-320.

24. Newell CA, Gray JC: **Binding of lac repressor-GFP fusion protein to lac operator sites inserted in the tobacco chloroplast genome examined by chromatin immunoprecipitation.** *Nucleic Acids Res* 2010, **38**:e145.

25. Cronin CA, Gluba W, Scrable H: **The lac operator-repressor system is functional in the mouse.** *Genes Dev* 2001, **15**:1506-1517.

26. Miller MB, Bassler BL: **Quorum sensing in bacteria.** *Annu Rev Microbiol* 2001, **55**:165-199.

27. Leclercq R, Courvalin P: **Intrinsic and unusual resistance to macrolide, lincosamide, and streptogramin antibiotics in bacteria.** *Antimicrob Agents Chemother* 1991, **35**:1273-1276.

28. Weisblum B: **Erythromycin resistance by ribosome modification.** *Antimicrob Agents Chemother* 1995, **39**:577-585.

29. Almofti YA, Dai M, Sun Y, Haihong H, Yuan Z: **Impact of erythromycin resistance on the Virulence properties and fitness of Campylobacter jejuni.** *Microb Pathog* 2011.

30. Noguchi N, Emura A, Matsuyama H, O'Hara K, Sasatsu M, Kono M: **Nucleotide sequence and characterization of erythromycin resistance determinant that encodes macrolide 2'-phosphotransferase I in Escherichia coli.** *Antimicrob Agents Chemother* 1995, **39**:2359-2363.

31. Zheng J, Sagar V, Smolinsky A, Bourke C, LaRonde-LeBlanc N, Cropp TA: **Structure and function of the macrolide biosensor protein, MphR(A), with and without erythromycin.** *J Mol Biol* 2009, **387**:1250-1260.

32. Noguchi N, Takada K, Katayama J, Emura A, Sasatsu M: **Regulation of transcription of the mph(A) gene for macrolide 2'-phosphotransferase I in Escherichia coli: characterization of the regulatory gene mphR(A).** *J Bacteriol* 2000, **182**:5052-5058.

33. Nam G, Kang TW, Shin JH, Choi KI: **Design, synthesis, and anti-Helicobacter pylori activity of erythromycin A (E)-9-oxime ether derivatives.** *Bioorg Med Chem Lett* 2006, **16**:569-572.

34. Pandey D, Katti SB, Haq W, Tripathi CK: **Synthesis and antimicrobial activity of erythromycin-A oxime analogs.** *Bioorg Med Chem* 2004, **12**:3807-3813.

35. Miller JH: **Experiments in Molecular Genetics**. Edited by. Cold Spring Harbor: Cold Spring Harbor Laboratory; 1972.

36. Tsien RY: **The green fluorescent protein.** *Annu Rev Biochem* 1998, **67**:509-544.

37. Muramatsu S, Kinbara K, Taguchi H, Ishii N, Aida T: **Semibiological molecular machine with an implemented "AND" logic gate for regulation of protein folding.** *J Am Chem Soc* 2006, **128**:3764-3769.

38. Pu F, Liu Z, Yang X, Ren J, Qu X: **DNA-based logic gates operating as a biomolecular security device.** *Chem Commun (Camb)* 2011, **47**:6024-6026.

39. Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, Voigt CA, Ellington AD: **A synthetic genetic edge detection program.** *Cell* 2009, **137**:1272-1281.

40. Lou C, Liu X, Ni M, Huang Y, Huang Q, Huang L, Jiang L, Lu D, Wang M, Liu C, et al.: **Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch.** *Mol Syst Biol* 2010, **6**:350.

41. de Silva AP, Uchiyama S: **Molecular logic and computing.** *Nat Nanotechnol* 2007, **2**:399-410.

42. Anderson JC, Voigt CA, Arkin AP: **Environmental signal integration by a modular AND gate.** *Mol Syst Biol* 2007, **3**:133.

43. Shenoi BA: ***Introduction to Digital Signal Processing and Filter Design***: Wiley, John & Sons, Incorporated; 2005.

44. Weber W, Stelling J, Rimann M, Keller B, Daoud-El Baba M, Weber CC, Aubel D, Fussenegger M: **A synthetic time-delay circuit in mammalian cells and mice.** *Proc Natl Acad Sci U S A* 2007, **104**:2643-2648.

45. Weber W, Fux C, Daoud-el Baba M, Keller B, Weber CC, Kramer BP, Heinzen C, Aubel D, Bailey JE, Fussenegger M: **Macrolide-based transgene control in mammalian cells and mice.** *Nat Biotechnol* 2002, **20**:901-907.

46. Sauer B, Henderson N: **Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1.** *Proc Natl Acad Sci U S A* 1988, **85**:5166-5170.

47. Sherratt DJ, Wigley DB: **Conserved themes but novel activities in recombinases and topoisomerases.** *Cell* 1998, **93**:149-152.

48. Grindley ND, Whiteson KL, Rice PA: **Mechanisms of site-specific recombination.** *Annu Rev Biochem* 2006, **75**:567-605.

49. Guo F, Gopaul DN, Van Duyne GD: **Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse.** *Proc Natl Acad Sci U S A* 1999, **96**:7143-7148.

50. Link KH, Shi Y, Koh JT: **Light activated recombination.** *J Am Chem Soc* 2005, **127**:13088-13089.

51. Edwards WF, Young DD, Deiters A: **Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells.** *ACS Chem Biol* 2009, **4**:441-445.

52. Kennedy MJ, Hughes RM, Peteya LA, Schwartz JW, Ehlers MD, Tucker CL: **Rapid blue-light-mediated induction of protein interactions in living cells.** *Nat Methods* 2010, **7**:973-975.

53. Voziyanov Y, Stewart AF, Jayaram M: **A dual reporter screening system identifies the amino acid at position 82 in Flp site-specific recombinase as a determinant for target specificity.** *Nucleic Acids Res* 2002, **30**:1656-1663.

54. Buchholz F, Angrand PO, Stewart AF: **Improved properties of FLP recombinase evolved by cycling mutagenesis.** *Nat Biotechnol* 1998, **16**:657-662.

55. Schleif R: **Regulation of the L-arabinose operon of Escherichia coli.** *Trends Genet* 2000, **16**:559-565.

56. Hasan A, Stengele K, Giegrich H, Cornwell P, Isham K, Sachleben R, Pfleiderer W, Foote R: **Photolabile protecting groups for nucleosides: Synthesis and photodeprotection rates**. *Tetrahedron* 1997, **53**:4247-4264.

57. Bhushan K, DeLisi C, Laursen R: **Synthesis of photolabile 2-(2-nitrophenyl)propyloxycarbonyl protected amino acids**. *Tetrahedron Letters* 2003, **44**:8585-8588.

58. Khlebnikov A, Skaug T, Keasling JD: **Modulation of gene expression from the arabinose-inducible araBAD promoter.** *J Ind Microbiol Biotechnol* 2002, **29**:34-37.

59. Khlebnikov A, Risa O, Skaug T, Carrier TA, Keasling JD: **Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture.** *J Bacteriol* 2000, **182**:7029-7034.

60. Rautio J, Kumpulainen H, Heimbach T, Oliyai R, Oh D, Järvinen T, Savolainen J: **Prodrugs: design and clinical applications.** *Nat Rev Drug Discov* 2008, **7**:255-270.

61. Lavis LD: **Ester bonds in prodrugs.** *ACS Chem Biol* 2008, **3**:203-206.

62. Aich U, Campbell CT, Elmouelhi N, Weier CA, Sampathkumar SG, Choi SS, Yarema KJ: **Regioisomeric SCFA attachment to hexosamines separates metabolic flux from cytotoxicity and MUC1 suppression.** *ACS Chem Biol* 2008, **3**:230-240.

63. Sampathkumar SG, Jones MB, Meledeo MA, Campbell CT, Choi SS, Hida K, Gomutputra P, Sheh A, Gilmartin T, Head SR, et al.: **Targeting glycosylation pathways and the cell cycle: sugar-dependent activity of butyrate-carbohydrate cancer prodrugs.** *Chem Biol* 2006, **13**:1265-1275.

64. Gauthier C, Legault J, Lavoie S, Rondeau S, Tremblay S, Pichette A: **Synthesis of two natural betulinic acid saponins containing alpha-L-rhamnopyranosyl-(1 -> 2)-alpha-L-arabinopyranose and their analogues**. *Tetrahedron* 2008, **64**:7386-7399.

65. Rodriguez-Perez T, Lavandera I, Fernandez S, Sanghvi Y, Ferrero M, Gotor V: **Novel and efficient chemoenzymatic synthesis of D-glucose 6-phosphate and molecular modeling studies on the selective biocatalysis**. *European Journal of Organic Chemistry* 2007:2769-2778.

66. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: *Molecular Biology of The Cell*. Edited by Anderson M, Granum S. New York, NY: Garland Science; 2008.

67. Liu CC, Schultz PG: **Adding new chemistries to the genetic code.** *Annu Rev Biochem* 2010, **79**:413-444.

68. Young TS, Schultz PG: **Beyond the canonical 20 amino acids: expanding the genetic lexicon.** *J Biol Chem* 2010, **285**:11039-11044.

69. Chin JW: **Reprogramming the genetic code.** *EMBO J* 2011, **30**:2312-2324.

70. Wu N, Deiters A, Cropp TA, King D, Schultz PG: **A genetically encoded photocaged amino acid.** *J Am Chem Soc* 2004, **126**:14306-14307.

71. Lemke EA, Summerer D, Geierstanger BH, Brittain SM, Schultz PG: **Control of protein phosphorylation with a genetically encoded photocaged amino acid.** *Nat Chem Biol* 2007, **3**:769-772.

72. Gautier A, Deiters A, Chin JW: **Light-activated kinases enable temporal dissection of signaling networks in living cells.** *J Am Chem Soc* 2011, **133**:2124-2127.

73. Wang L, Brock A, Herberich B, Schultz PG: **Expanding the genetic code of Escherichia coli.** *Science* 2001, **292**:498-500.

74. Chin JW, Cropp TA, Anderson JC, Mukherji M, Zhang Z, Schultz PG: **An expanded eukaryotic genetic code.** *Science* 2003, **301**:964-967.

75. Chin JW, Cropp TA, Chu S, Meggers E, Schultz PG: **Progress toward an expanded eukaryotic genetic code.** *Chem Biol* 2003, **10**:511-519.

76. Ryu Y, Schultz PG: **Efficient incorporation of unnatural amino acids into proteins in Escherichia coli.** *Nat Methods* 2006, **3**:263-265.

77. Deiters A, Groff D, Ryu Y, Xie J, Schultz PG: **A genetically encoded photocaged tyrosine.** *Angew Chem Int Ed Engl* 2006, **45**:2728-2731.

78. Chou C, Young DD, Deiters A: **A light-activated DNA polymerase.** *Angew Chem Int Ed Engl* 2009, **48**:5950-5953.

79. Chou C, Young DD, Deiters A: **Photocaged t7 RNA polymerase for the light activation of transcription and gene function in pro- and eukaryotic cells.** *Chembiochem* 2010, **11**:972-977.

80. Chou C, Deiters A: **Light-activated gene editing with a photocaged zinc-finger nuclease.** *Angew Chem Int Ed Engl* 2011, **50**:6839-6842.

81. Wilkins BJ, Marionni S, Young DD, Liu J, Wang Y, Di Salvo ML, Deiters A, Cropp TA: **Site-specific incorporation of fluorotyrosines into proteins in Escherichia coli by photochemical disguise.** *Biochemistry* 2010, **49**:1557-1559.

82. Pal PP, Bae JH, Azim MK, Hess P, Friedrich R, Huber R, Moroder L, Budisa N: **Structural and spectral response of Aequorea victoria green fluorescent proteins to chromophore fluorination.** *Biochemistry* 2005, **44**:3663-3672.

83. Tandon V, Maurya H: **Water-promoted unprecedented chemoselective nucleophilic substitution reactions of 1,4-quinones with oxygen nucleophiles in aqueous micelles**. *Tetrahedron Letters* 2010, **51**:3843-3847.

84. Young TS, Ahmad I, Yin JA, Schultz PG: **An enhanced system for unnatural amino acid mutagenesis in E. coli.** *J Mol Biol* 2010, **395**:361-374.

85. Cropp TA, Schultz PG: **An expanding genetic code.** *Trends Genet* 2004, **20**:625-630.

86. Wang Q, Parrish AR, Wang L: **Expanding the genetic code for biological studies.** *Chem Biol* 2009, **16**:323-336.

87. Hancock SM, Uprety R, Deiters A, Chin JW: **Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair.** *J Am Chem Soc* 2010, **132**:14819-14824.

88. Fechter P, Rudinger-Thirion J, Tukalo M, Giegé R: **Major tyrosine identity determinants in Methanococcus jannaschii and Saccharomyces cerevisiae tRNA(Tyr) are conserved but expressed differently.** *Eur J Biochem* 2001, **268**:761-767.

89. Wimalasena DS, Cramer JC, Janowiak BE, Juris SJ, Melnyk RA, Anderson DE, Kirk KL, Collier RJ, Bann JG: **Effect of 2-fluorohistidine labeling of the anthrax protective antigen on stability, pore formation, and translocation.** *Biochemistry* 2007, **46**:14928-14936.

90. Ma Z, Cowart DM, Ward BP, Arnold RJ, DiMarchi RD, Zhang L, George GN, Scott RA, Giedroc DP: **Unnatural amino acid substitution as a probe of the allosteric coupling pathway in a mycobacterial Cu(I) sensor.** *J Am Chem Soc* 2009, **131**:18044-18045.

91. Lu Y, Yeung N, Sieracki N, Marshall NM: **Design of functional metalloproteins.** *Nature* 2009, **460**:855-862.

92. Sundberg R, Martin R: **Interactions of histidine and other imidazole derivatives with transition-metal ions in chemical and biological-systems**. *Chemical Reviews* 1974, **74**:471-517.

93. Swers JS, Kellogg BA, Wittrup KD: **Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display.** *Nucleic Acids Res* 2004, **32**:e36.

94. **High Efficiency Transformation by Electroporation Short Protocols in Molecular Biology**. edn Second Edition. Edited by. New York: Green Publishing Association and John Wiley & Sons.

95. Carthew RW: **Gene regulation by microRNAs.** *Curr Opin Genet Dev* 2006, **16**:203-208.

96. Sevignani C, Calin GA, Siracusa LD, Croce CM: **Mammalian microRNAs: a small world for fine-tuning gene expression.** *Mamm Genome* 2006, **17**:189-202.

97. Deiters A: **Small molecule modifiers of the microRNA and RNA interference pathway.** *AAPS J* 2010, **12**:51-60.

98. Garzon R, Marcucci G, Croce CM: **Targeting microRNAs in cancer: rationale, strategies and challenges.** *Nat Rev Drug Discov* 2010, **9**:775-789.

99. Si ML, Zhu S, Wu H, Lu Z, Wu F, Mo YY: **miR-21-mediated tumor growth.** *Oncogene* 2007, **26**:2799-2803.

100. Gumireddy K, Young DD, Xiong X, Hogenesch JB, Huang Q, Deiters A: **Small-molecule inhibitors of microrna miR-21 function.** *Angew Chem Int Ed Engl* 2008, **47**:7482-7484.

101. Destistov OS, Zhuravel IO, Orlov VD: **Cylic systems containing 1,2,4,-oxadiazole ring. 1. Synthesis of 5-(1H-1,2,3-triazol-4-yl)-1,2,4-oxadizoles - potential bioactive compound**. *Visnik Kharkivs'kogo Natsional'nogo Universitetu im. V. N. Karazina* 2007, **770**:232-238.

102. Thorne N, Inglese J, Auld DS: **Illuminating insights into firefly luciferase and other bioluminescent reporters used in chemical biology.** *Chem Biol* 2010, **17**:646-657.

103. Diez-Gonzalez S, Correa A, Cavallo L, Nolan S: **(NHC)copper(I)-catalyzed [3+2] cycloaddition of azides and mono- or disubstituted alkynes**. *Chemistry-a European Journal* 2006, **12**:7558-7564.

104. Auld DS, Thorne N, Maguire WF, Inglese J: **Mechanism of PTC124 activity in cell-based luciferase assays of nonsense codon suppression.** *Proc Natl Acad Sci U S A* 2009, **106**:3585-3590.

105. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, et al.: **HaloTag: a novel protein labeling technology for cell imaging and protein analysis.** *ACS Chem Biol* 2008, **3**:373-382.

106. Janssen DB: **Evolving haloalkane dehalogenases.** *Curr Opin Chem Biol* 2004, **8**:150-159.

107. Harwood KR, Miller SC: **Leveraging a small-molecule modification to enable the photoactivation of rho GTPases.** *Chembiochem* 2009, **10**:2855-2857.

108. Ohana RF, Hurst R, Vidugiriene J, Slater MR, Wood KV, Urh M: **HaloTag-based purification of functional human kinases from mammalian cells.** *Protein Expr Purif* 2011, **76**:154-164.

109. Oster G, Nishijima Y: **Fluorescence and internal rotation - their dependence on viscosity of the medium**. *Journal of the American Chemical Society* 1956, **78**:1581-1584.

110. Xu W, Lu Y: **Label-Free Fluorescent Aptamer Sensor Based on Regulation of Malachite Green Fluorescence**. *Analytical Chemistry* 2010, **82**:574-578.

111. Babendure JR, Adams SR, Tsien RY: **Aptamers switch on fluorescence of triphenylmethane dyes.** *J Am Chem Soc* 2003, **125**:14716-14717.

112. Louie J, Driver M, Hamann B, Hartwig J: **Palladium-catalyzed amination of aryl triflates and importance of triflate addition rate**. *Journal of Organic Chemistry* 1997, **62**:1268-1273.

113. Culp SJ, Beland FA: **Malachite Green: A Toxicological Review**. *Journal of the American College of Toxicology* 1996, **15**:219-238.

114. Culp SJ, Mellick PW, Trotter RW, Greenlees KJ, Kodell RL, Beland FA: **Carcinogenicity of malachite green chloride and leucomalachite green in B6C3F1 mice and F344 rats.** *Food Chem Toxicol* 2006, **44**:1204-1212.

115. Jang GH, Park IS, Lee SH, Huh TL, Lee YM: **Malachite green induces cardiovascular defects in developing zebrafish (Danio rerio) embryos by blocking VEGFR-2 signaling.** *Biochem Biophys Res Commun* 2009, **382**:486-491.

116. Adams E: **The antibacterial action of crystal violet.** *J Pharm Pharmacol* 1967, **19**:821-826.

117. Neklesa TK, Tae HS, Schneekloth AR, Stulberg MJ, Corson TW, Sundberg TB, Raina K, Holley SA, Crews CM: **Small-molecule hydrophobic tagging-induced degradation of HaloTag fusion proteins.** *Nat Chem Biol* 2011, **7**:538-543.

118. Ohana RF, Encell LP, Zhao K, Simpson D, Slater MR, Urh M, Wood KV: **HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification.** *Protein Expr Purif* 2009, **68**:110-120.

119. Tae HS, Sundberg TB, Neklesa TK, Noblin DJ, Gustafson JL, Roth AG, Raina K, Crews CM: **Identification of hydrophobic tags for the degradation of stabilized proteins.** *Chembiochem* 2012, **13**:538-541.

120. Zhang Y, So MK, Loening AM, Yao H, Gambhir SS, Rao J: **HaloTag protein-mediated site-specific conjugation of bioluminescent proteins to quantum dots.** *Angew Chem Int Ed Engl* 2006, **45**:4936-4940.

121. Hermanson GT: *Bioconjugation Techniques* edn 2nd: Elsevier; 2008.

122. Balci H, Arslan S, Myong S, Lohman TM, Ha T: **Single-molecule nanopositioning: structural transitions of a helicase-DNA complex during ATP hydrolysis.** *Biophys J* 2011, **101**:976-984.

123. Ellisdavies G, Kaplan J: **A new class of photolabile chelators for the rapid release of divalent-cations - generation of caged-Ca and caged-Mg**. *Journal of Organic Chemistry* 1988, **53**:1966-1969.

124. Struga M, Miroslaw B, Pakosinska-Parys M, Drzewiecka A, Borowski P, Lossakowski J, Koziol A: **Synthesis, characterization and supramolecular synthons in crystals of new derivatives of 10-oxa-4-azatricyclo[5.2.1.0(2,6)]dec-8-ene-3,5-dione**. *Journal of Molecular Structure* 2010, **965**:23-30.

125. Reetz MT, Rentzsch M, Pletsch A, Taglieber A, Hollmann F, Mondière RJ, Dickmann N, Höcker B, Cerrone S, Haeger MC, et al.: **A robust protein host for anchoring chelating ligands and organocatalysts.** *Chembiochem* 2008, **9**:552-564.