

Expanding the Scope of Terminal Alkynes in Chemical Biology

Johnathan Charles Maza

Purcellville, Virginia

B.S. in Chemistry, The College of William and Mary, 2015

A Thesis presented to the Graduate Faculty  
of the College of William and Mary in Candidacy for the Degree of  
Master of Science

Chemistry

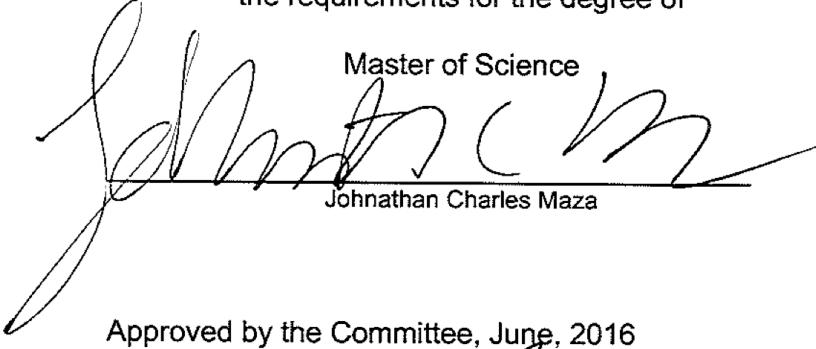
The College of William and Mary  
August, 2016

© Copyright by Johnathan C. Maza 2016

## APPROVAL PAGE

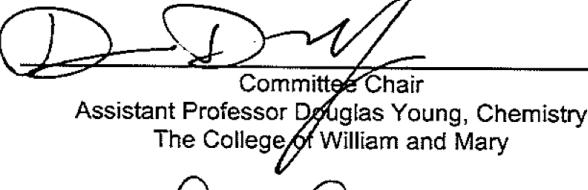
This Thesis is submitted in partial fulfillment of  
the requirements for the degree of

Master of Science



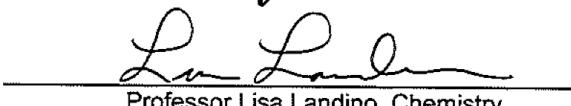
Johnathan Charles Maza

Approved by the Committee, June, 2016



Committee Chair

Assistant Professor Douglas Young, Chemistry  
The College of William and Mary



Professor Lisa Landino, Chemistry

The College of William and Mary



Assistant Professor William McNamara, Chemistry

The College of William and Mary

## COMPLIANCE PAGE

Research approved by

Institutional Biosafety Committee

Protocol number(s): IBC-2015-11-23-10690-dyoung01

Date(s) of approval: 2015-11-23

## ABSTRACT

The terminal alkyne is one of the most widely used chemical moieties in chemical biology. Thanks to the relative absence of this functional group in biology, it has become a widespread functional handle for a plethora of biorthogonal chemistries. Our group has extended the scope of this functionality by developing a biological variant of the Glaser-Hay coupling, which brings together two terminal alkynes to form a diyne linkage. However, our initial findings revealed that the chemistry is plagued by protein degradation due to a deleterious copper(II) hydroxyl intermediate. Herein we extend the scope of the terminal alkyne by developing a biological variant of the Cadiot-Chodkiewicz coupling, which brings together a bromoalkyne and a terminal alkyne to form a diyne linkage. Unlike the Glaser-Hay, the Cadiot-Chodkiewicz is thought to be net redox neutral. We found that this chemistry is biologically compatible and does indeed reduce copper(II)-mediated cytotoxicity while increasing rates of diyne formation. Finally, we extend our findings on diyne-forming chemistries by applying this conjugation to a variety of projects, ranging from undergraduate teaching labs, bioconjugate preparation, and protein function, thereby expanding the scope of this conjugation strategy.

## TABLE OF CONTENTS

Acknowledgements	ii
List of Tables	iii
List of Figures	iv
Chapter 1. Introduction to Unnatural Amino Acids and Bioconjugations	1
Chapter 2. Synthesis and Incorporation of Unnatural Amino Acids to Probe and Optimize Protein Bioconjugations	25
Chapter 3. Expanding the Scope of Alkyne-Mediated Bioconjugations through a Biological Cadiot-Chodkiewicz Reaction	60
Chapter 4. Using UAAs to Illustrate and Alter Protein Structure-Function Relationships	82
Chapter 5. Immobilization of the RNA-Guided Engineered Nuclease Cas9	125
Chapter 6. Towards the Development of Glaser-Hay “Turn-On” Probes	143

## ACKNOWLEDGEMENTS

I would like to thank Dr. McNamara and Dr. Landino for once again serving on my committee.

I would also like to thank my lab members, who have put up with my moodiness through this process. I particularly enjoyed getting to mentor those of you who shadowed under me, and know that you all will continue to do great work in lab and science!

Finally, I would like to thank Doug who has been amazing mentor for five years. Thank you so much for believing in me and helping me realize that science is what I wanted to do with my life. I'll always remember the lab fondly, and look forward to seeing what you all continue to produce!

## LIST OF TABLES

1.	Overall Yields for Tethered UAA Syntheses	31
2.	Protein Expression Yields for the Tethered UAAs	32
3.	Comparison of Standard GFP Laboratories Versus the UAA Laboratory	93

## LIST OF FIGURES

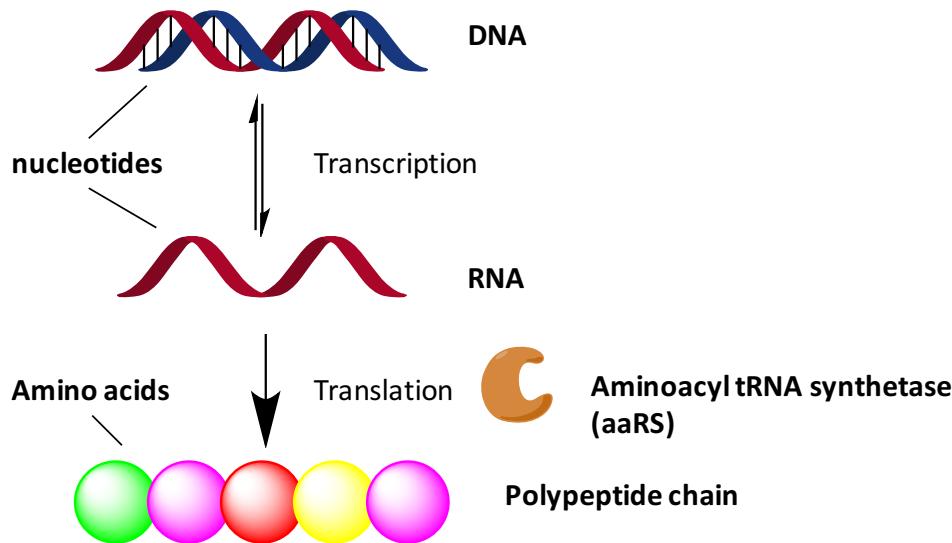
4.	The Central Dogma	2
5.	Protein immobilization on a solid support	4
6.	Twenty canonical amino acids	6
7.	Unnatural Amino Acids	7
8.	Whole cell replacement of a canonical amino acid with a synthetically similar analogue (UAA)	10
9.	Mechanism for genetic incorporation of unnatural amino acids	11
10.	General scheme for the selection of an aminoacyl-tRNA synthetase that recognizes and incorporates a specific unnatural amino acid	13
11.	Labeling of pIF-containing OmpC using a fluorescen boronic acid via a biological Suzuki coupling	18
12.	Surface labeling of E. coli using a an alkyne containing UAA in the presence of an iodo containing fluorogenic dye via the biological Sonogashira coupling	19
13.	Expression and conjugation of GFP-151 mutants with an AlexaFluor488 dye	34
14.	Verification of alkyne functiaonlized resin	35
15.	Click immobilization of GFP-151 mutants on a sepharose resin with either an azide tethered tyrosine UAAs or a p-azidophenylalanine	37
16.	Glaser-Hay reactions of alkyne-containing GFP mutants	38
17.	Immobilization of bromide-containing UAAs on a derivatized sepharose resin	40
18.	Validation that the sepharose resin was derivatized with an amine functional group using the FITC dye	53
19.	The Cadiot-Chodkiewicz reaction	61

20.	Proposed reaction intermediates in the Glaser-Hay coupling reaction	63
21.	Alkynyl and Bromo-alkynyl bearing UAAs	64
22.	Expression of UAA-containing GFP151 using the promiscuous pCNF aaRS	66
23.	Bioorthogonal Cadiot-Chodkiewicz reaction	67
24.	Copper(I) salt optimization for the Cadiot-Chodkiewicz reaction	68
25.	Bioorthogonal Cadiot-Chodkiewicz reaction Reaction profile at 4 °C over a 24 hr time period	70
26.	SDS-PAGE of Cadiot-Chodkiewicz time course with pBrPrF containing GFP and at 4°C	70
27.	Bioorthogonal Cadiot-Chodkiewicz reaction reaction profile at 37 °C over a 24 hr time period	70
28.	Bioorthogonal Cadiot-Chodkiewicz reaction using the aromatic pBrEtF UAA contaiing GFP	71
29.	Coupling profile of the pPrF Glaser-Hay at 4°C	72
30.	Comparison of alkyne reactions	73
31.	Structure of GFP fluorophore	84
32.	Coomassie stained SDS-PAGE of the UAA-containing GFP <sub>66</sub> produced by the biochem students	89
33.	Fluorescent imaging of SDS-PAGE of the UAA-containing GFP <sub>66</sub> produced by the biochem students	90
34.	Fluorescence results for the expression of GFP mutants	92
35.	Modification of GFP function via UAA mutagenesis	117
36.	Structures of the fluorophores generated when performing a Glaser-Hay reaction with the two UAAs	119
37.	Fluorescent profile obtained after reacting various alkyne containing partners with pPrf or pEtF-containing GFP <sub>TAG66</sub>	121

38.	Overview of Cas9 guide RNA structure	128
39.	Positional effect of mismatches between the guide RNA and the DNA template as detected via a functional Ub-GFP assay	130
40.	FRET experiments involve attaching two fluorophores to a protein	131
41.	Cas9 with HNH (magenta) and RuVCIII (cyan) highlighted	134
42.	Results from sequencing and alignment for our PCR mutagenesis	135
43.	Protein expression experiment for pET-Cas9Y1265X and pEVOL-WT	136
44.	Expression of a pAzF-containing Cas9Y1265X	137
45.	Alkyne-containing profluorophores to be synthesized to probe the application of the Glaser-Hay towards turn-on probes	146

## **CHAPTER 1: INTRODUCTION TO UNNATURAL AMINO ACIDS AND BIOCONJUGATIONS**

In the sequential arrangement of amino acids that form proteins, nature has achieved elegant macromolecules with properties chemists have yet to fully replicate (see Figure 1).<sup>1</sup> Based on their biological functions, proteins can both bind small molecules with incredibly high affinities and catalyze reactions with ease that often require harsh conditions *in vitro*. Moreover, proteins have both synthetic and structural roles within the cell, making them essential to life. Based on these characteristics, harnessing the power of proteins would allow scientists to develop novel diagnostic and catalytic tools that would advance science. Additionally, these unique properties of proteins make them responsible for the myriad of cellular functions that enable living systems to exist. As such, molecular biology has been revolutionized by the ability to study and observe proteins *in vivo*. Thus, the study of proteins and their development towards novel technologies is a crucial area of science that constantly requires additional innovations.



*Figure 1. The Central Dogma. DNA is comprised a triple nucleotide code, that is transcribed into RNA. This triplet nucleotide code in the RNA is then translated into proteins via tRNAs that recognize the triplet code in the RNA. These tRNAs are charged with the appropriate amino acid by aminoacyl tRNA synthetases.*

Protein bioconjugates, in which a protein is conjugated to another molecule (i.e. biomolecule, surface, fluorophore, therapeutic agent, etc), represent a broad area of research with huge potential to impact basic science and medicine. For example, a serious issue facing researchers wishing to fight cancer is discovering new ways to accurately detect the presence of cancer within patients. A report in *Nature* detailed the use of fluorescent probes as a means to live image cancer cells during surgery based on the presence and detection of certain proteins.<sup>2</sup> The results of such technology will enhance removal of malignant tissue while decreasing relapse rates and damage to patient's healthy tissue. Accordingly, fluorescent molecular imaging is proving to be a key player in the war against cancer, and while it has come a long way, there are limitations that must be overcome before such specialized applications become widespread.

Another emerging field involving proteins is the development of protein biochips. These biochips contain dense spots, or microarrays of thousands of diverse or

homogenous proteins immobilized on a solid support.<sup>3,4</sup> They offer the potential to harness the unique power of protein chemistry. For example, such technology can allow for the rapid and high-throughput screening of thousands of biologically interesting molecules, allowing for the rapid discovery of new drug candidates. Similar to the *Nature* paper previously discussed, such protein-based technology can be directly applied to diagnostics, such as detecting cancer markers in blood.<sup>5,6</sup> Due to their high affinity for their substrate, protein chips would allow for the rapid and sensitive detection of biomolecules and toxins.

Despite their promise, protein-based chips and protein conjugates are still a nascent technology, being highly limited by current immobilization and conjugation techniques. Many of these involve simple adsorption.<sup>4</sup> In this technique, proteins are adsorbed onto a surface via ionic or hydrophobic interactions. When performed on a surface, the potential for multiple interacting sites often results in a heterogeneous layer with a random protein orientation that may also be easily reversed. This is disadvantageous, as the protein may not be correctly oriented to perform the desired function. Covalent bonding to form protein conjugates is often the most ideal mechanism to ensure an irreversible attachment. By reacting a functionalized partner with key amino acid residues on the protein, such as lysine or cysteine residues that respectively contain amine and thiol functionalities, more robust attachments can be made. However, these covalent linkages on amino acids can occur at multiple sites, if more than one reactive amino acid is present, resulting in proteins that may either be conformationally inactive or also inappropriately oriented to bind molecules of interest.<sup>4</sup> Bioaffinity based attachment, in which key protein structures or properties (i.e. an antibody) are exploited

to bind a high affinity ligand, is an option that can allow for the binding of proteins in specific orientations. However, this technique relies on non-covalent binding, and is thus less robust. Furthermore, it limits the number of proteins that can be bound to the surface, since proteins must have particular structures or properties in order to interact with the ligand.

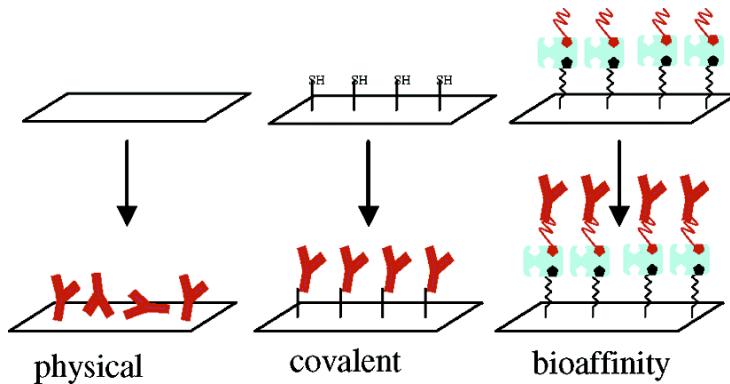


Figure 2. There are numerous methods for immobilizing proteins on a solid support. Covalent attachment of the protein represents the most robust technique, although it is often complicated by the presence of multiple reaction partners on a single peptide. Adapted Rusmini et al.<sup>4</sup>

Before this biochip technology can become widespread, scientists must solve these immobilization problems. Additionally, the need to find new ways to probe protein structure and function also creates the need to develop new ways to manipulate protein structure and function. A potential alternative to both of these areas involves the application of unnatural amino acid technology.

Unnatural amino acids (UAAs) allow scientists to expand the rather limited biochemical functionality found in the 20 canonical amino acids and introduce novel reactivity not typically found in proteins.<sup>7,8</sup> Incorporated via suppression of a TAG stop codon in translation, which can be engineered into the DNA at a desired residue position of the target protein, “chemical handles” can be introduced on the surface of the protein. These functional handles then serve as sites of chemoselective reactivity. This strategy

avoids the issue of multiple reaction partners on a single protein (which is typical when trying to conjugate on a cysteine/lysine residue), and allows the chemist to access new chemistries typically not available from the canonical 20 amino acids. Furthermore, compared to the immobilization strategies discussed above, this technique opens up the potential to exploit any protein that the chemist desires in a robust way.

### **Unnatural Amino Acids**

Given the large number of functions proteins serve within an organism, as well as the myriad of ways in which they could be combined, it is reasonable to assume that evolution would have dictated the need for amino acids with diverse functional groups. However, a quick glance at the natural 20 amino acids reveals only four main chemical functionalities: nonpolar, polar, acidic, and basic.<sup>8,9</sup>

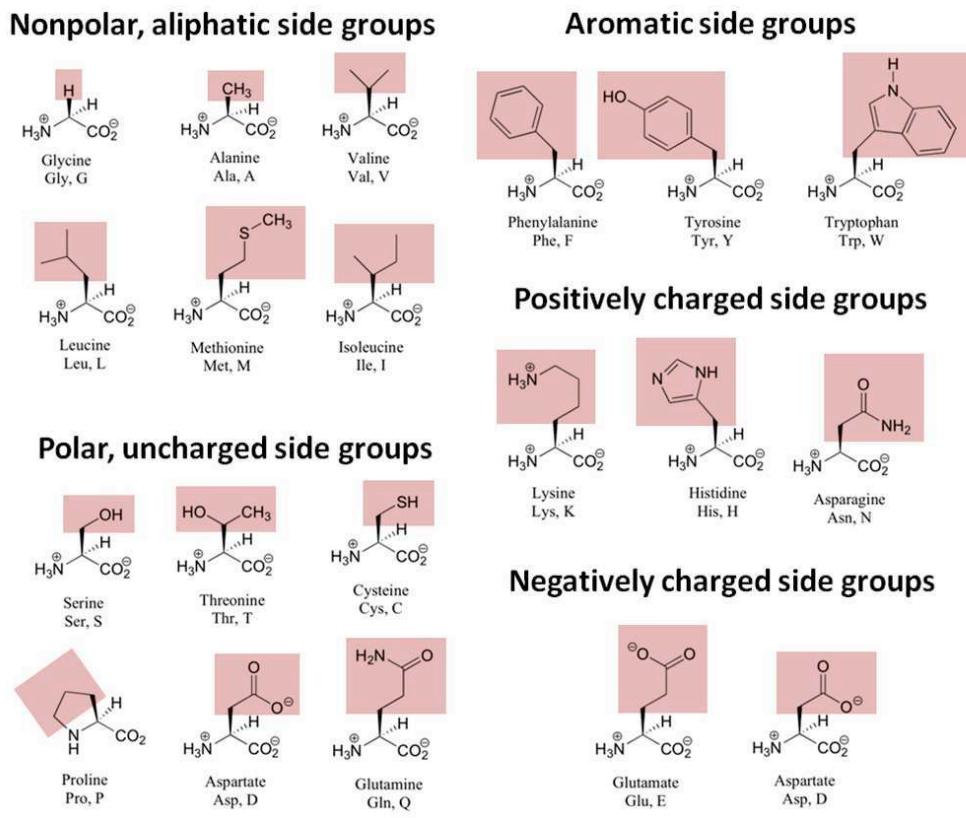
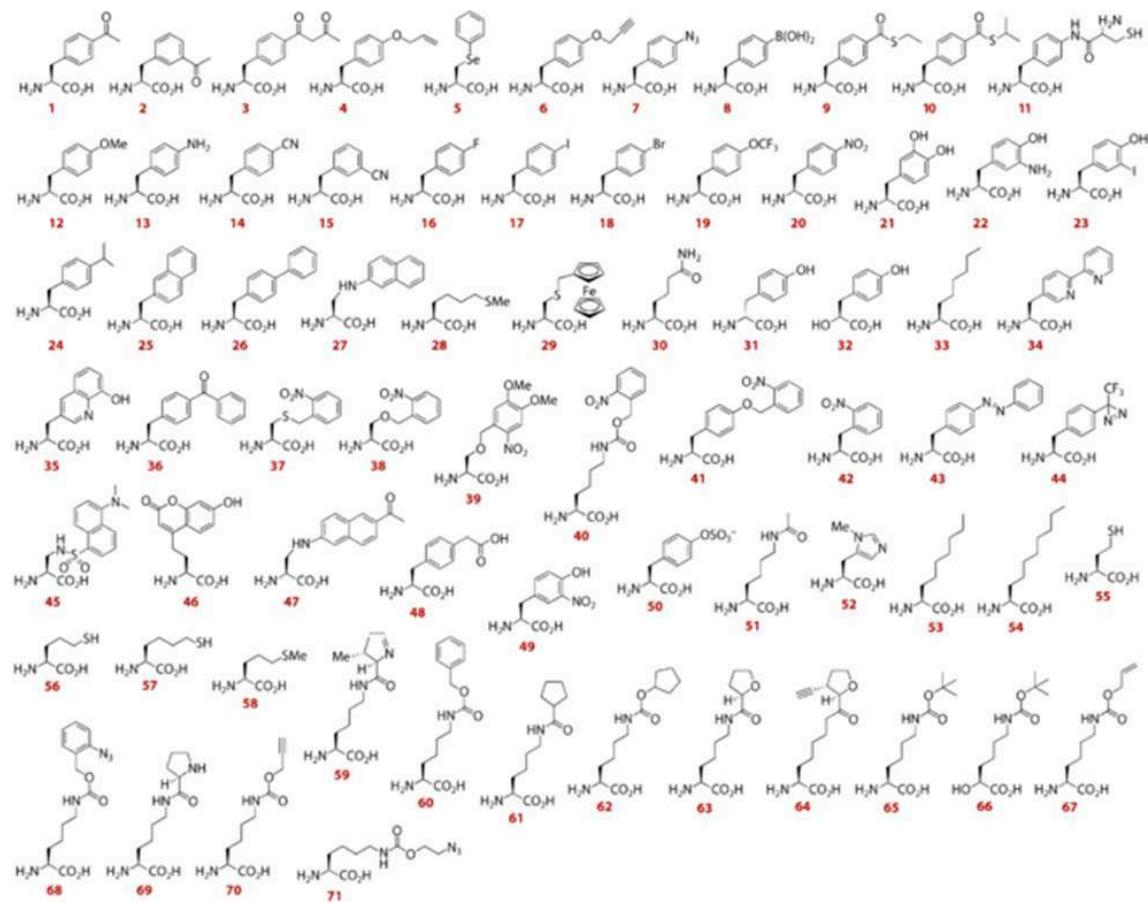


Figure 3. There are 20 naturally occurring amino acids. Only five elements are represented in the canonical amino acids, as well as a very limited set of chemical functionality. Adapted from chemwiki.ucdavis.edu

Advances in chemical biology have led to the development of UAAs that can be synthesized with novel biochemical properties not available in the 20 canonical amino acids. Such syntheses are relatively simple, as amino acids can be purchased with their carboxyl and amino ends masked with protecting groups. The result is a selectively reactive R-group that can be chemically modified. In particular, my research has focused on derivatizing the amino acid tyrosine (Tyr) that has a phenol functionality that can be activated by the addition of base. The deprotonated hydroxyl group can then undergo nucleophilic substitution reactions with molecules that possess good leaving groups. The result is a UAA that can serve as a starting point to introduce a myriad of novel biochemical functionality into proteins.



A Liu CC, Schultz PG. 2010.  
B Annu. Rev. Biochem. 79:413–44

*Figure 4. Unnatural amino acids allow chemists to access novel functionality and elements not present in the 20 canonical amino acids.*

### Generation of an Unnatural Amino Acid-Containing Protein

While the synthesis of UAAs is feasible, in order to capitalize on this technology it is necessary to be able to introduce these UAAs into proteins of interest. Methods have been developed to site-specifically incorporate these UAAs into proteins via insertion in response to a TAG stop codon in the genome, giving researchers very precise control over the placement and number of these UAAs in a protein's amino acid sequence.<sup>8</sup>

Early methodology relied on incorporating UAAs *in vitro*. By utilizing cell-free protein assembly machinery, researchers were able to incorporate unnaturals into specific sites of the protein via a nonsense codon. The associated tRNA for this nonsense codon was then chemically aminoacylated with the UAA of interest. The result is a system that ensured the incorporation of the UAA into a single, predefined residue of the polypeptide. In particular, the amber suppressor codon (TAG) was typically employed, as its frequency in the *Escherichia coli* genome, the organism of choice for many protein expressions, is relatively low. This ensures the incorporation of a single UAA into a single site.<sup>10,11</sup> While this approach allowed for the site-specific incorporation of a UAA into a protein's primary sequence, the use of *in vitro* technology has serious limitations. Many proteins require the aid of additional proteins *in vivo* before they can acquire the correct conformation. As such, while this technique is ideal for the preparation of small peptides, polypeptides may not be able to be generated in a robust manner. This severely limits the biological applicability of UAAs. Furthermore, the synthesis of mutated proteins *in vitro* limits the use of this technology beyond biochemical studies. For example, novel methods of protein imaging *in vivo* could not be developed since this technique required cell-free machinery.

An alternative strategy for UAA incorporation that allowed for the *in vivo* incorporation of UAAs involved the complete replacement of a particular amino acid in the organism's proteome with a UAA (see Figure 5). In this non-site specific UAA incorporation technique, the canonical amino acids methionine, leucine, isoleucine, phenylalanine, tryptophan, or proline are replaced by an amino acid that resembles these amino acids in structure.<sup>12</sup> Typically, cells transformed with a protein of interest are

grown in media containing all 20 amino acids, pelleted, to remove the media prior to protein expression. Upon the induction of protein expression, the cells are resuspended in a new media that lacks one of these canonical amino acids. In place of the amino acid, an unnatural amino acid is included in the broth. By synthesizing amino acids that resemble the removed amino acid, scientists can trick the cell into incorporating their UAA in place of the canonical amino acid in the protein of interest. Tirrell and Finn demonstrated the utility of this approach by synthesizing methionine analogs. In place of the thioether functionality, the group included azide and alkyne functionalities in their UAAs. As a result, they were able to demonstrate that in place of methionine, *E. coli* expressing a bacteriophage protein capsid could generate capsids labeled with either the azide or alkyne functionality in the residues that were originally methionine. While this approach also allows researchers to incorporate UAAs in a semi-site specific fashion, the frequency of UAA placement in a protein can be highly variable, due to the redundancy inherent in the 20 amino acid lexicon. The result can be complete dysfunction of the protein if the presence of multiple UAAs alters the protein's structure too much. Additionally, this technique impacts the organism's proteome, introducing the UAA in place of the canonical amino acid in every protein expressed. The result could be highly deleterious for the organism, which is not beneficial for *in vivo* applications.<sup>12</sup>

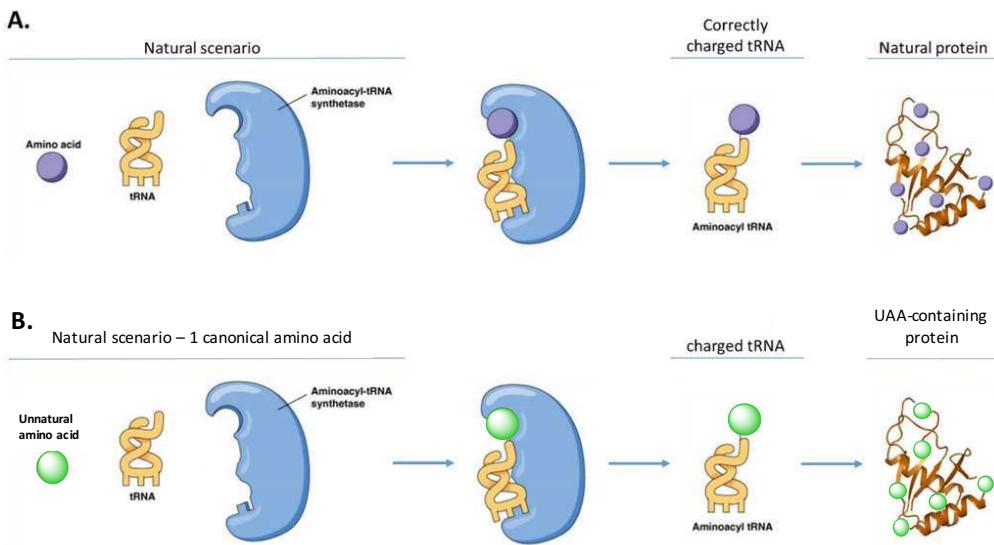
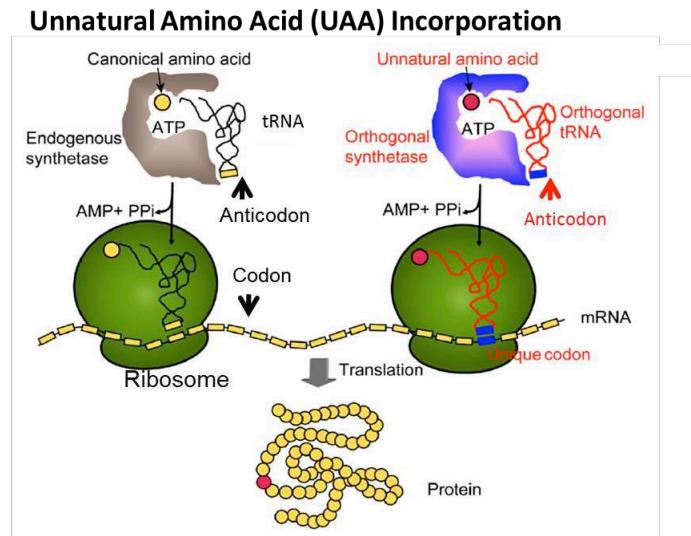


Figure 5. Whole cell replacement of a canonical amino acid with a synthetically similar analogue (UAA). The resulting cell incorporates the UAA in place of the canonical amino acid throughout its entire proteome, resulting in UAA incorporation in a non-specific manner. Adapted from <sup>13</sup>

Since these early *in vitro* studies, more recent work has focused on generating technology that would allow for the *in vivo* incorporation of UAAs. Additional challenges come into play when considering such techniques. First, as mentioned above, the incorporation of the UAA must be site specific, and it must occur only in the residue sites that have been previously determined by the researcher. Should the UAA be incorporated into additional non-specified sites, the result could be inactivation of the protein of interest, or subsequent death of the organism. Additionally, the tRNA machinery employed must act solely on the UAA of interest. By translating unnatural amino acid technology towards *in vivo* incorporation, the added problem of having to deal with three elements arises: the UAA, the tRNA, and the aminoacyl-tRNA synthetase (aaRS) that is responsible for charging the tRNA with the UAA. In order for successful incorporation to occur, the aminoacyl-tRNA synthetase should only pair with the UAA and with the corresponding suppressor tRNA. Should this orthogonality to the

endogenous translation machinery not be maintained, endogenous amino acids could be incorporated into proteins in place of the UAA and vice versa. Finally, the UAA must be cell permeable and non-cytotoxic in order to be incorporated.<sup>8,14</sup>



*Figure 6. Mechanism for genetic incorporation of unnatural amino acids. An orthogonal synthetase must be introduced that does not cross-react with any natural tRNA or amino acid. This is coupled with a new tRNA that harbors an anticodon for the TAG codon. When the ribosome encounters the TAG codon, a new amino acid is incorporated, rather than the discontinuation of translation. Adapted from Chem. Biol. 2009 16, 323-336.*

Early studies with *E. coli* models attempted to evolve an endogenous tRNA synthetase/tRNA pair that was orthogonal to a UAA of interest utilizing *E. coli*'s translational machinery. However, such tactics were wrought with misaminoacetylation, and subsequently terminated.<sup>15</sup> Instead, researchers focused on examining the protein translation machinery of other organisms, which would not interact with the endogenous aaRS/tRNA pairs in *E. coli*. *In vitro* studies found that a tyrosine aaRS/tRNA pair from the archaea *M. jannaschii* was the first of such pairs used to successfully incorporate a Tyr in response to a TAG stop codon in a manner completely orthogonal to the host *E. coli*'s translational machinery.<sup>16</sup>

With acceptable aaRS/tRNA machinery from *M. jannaschii* in hand, researchers next focused on a means to alter the specificity of the pair to recognize different UAAs. This was done through a double-sieve selection method. First, a library of tRNA synthetases was constructed based off the Tyr-aaRS of *M. jannaschii*. Within this library, five to six residues within the binding site of the aaRS were genetically mutated to be one of the 20 amino acids, generating a library size on the order of  $10^6$  to  $10^7$  mutant aaRS. This library size is still viable for conventional double transformation procedures. Next, the library underwent a positive selection. In this step, *E. coli* were co-transformed with the library of mutated synthetases and an enzyme encoding for a mutated-chloramphenicol degrading enzyme that contained an early TAG residue. Following double-transformation, the transformed cells were grown on a plate in the presence of a novel UAA and chloramphenicol. Survival implied the organism was able to use the foreign aaRS machinery to charge the suppressor tRNA and thus produce a functional chloramphenicol degrading enzyme.<sup>17</sup> However, suppression does not guarantee that the aaRS charged the suppressor tRNA using the UAA, as canonical amino acids could also have been recognized by the aaRS and charged to the tRNA. As such, those cells that grew were then isolated, their plasmids purified, and the remaining plasmid library was then transformed for a negative selection. In the negative selection the toxic barnase gene was employed. Barnase is a bacterial ribonuclease, and will degrade cellular RNA, preventing protein synthesis and leading to cell death. The gene was mutated to harbor three stop codons, and grown in the absence of unnatural amino acid. If the organism can use the exogenous aaRS/tRNA pair to charge endogenous amino acids and suppress the three TAG stop codons, it will produce the barnase protein and die.<sup>17</sup> Thus, the end of

this two round protocol selected for organisms that possessed an aaRS that could selectively incorporate the UAA to some degree. By iterating this process for multiple rounds with the UAA of interest, it was possible to evolve an aaRS/tRNA pair that works selectively for the UAA of interest. This is the evolution procedure used today for *in vivo* incorporation of UAA. These site-specifically incorporated UAAs can now serve as functional “handles” for the development of new protein conjugates.

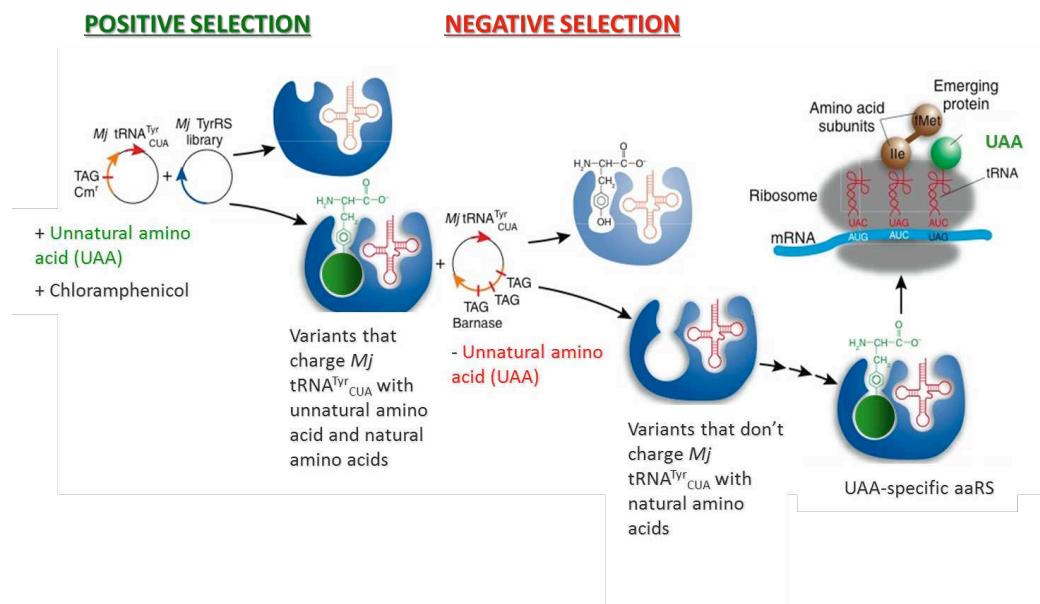


Figure 7. General scheme for the selection of an aminoacyl-tRNA synthetase that recognizes and incorporates a specific unnatural amino acid.

## Biorhogonal Chemistry

A growing area of interest in chemical biology is the ability to perform chemistry, particularly conjugation reactions, within living systems. In order for such chemistry to be feasible though, certain conditions must be met by the chemical reaction employed. Most importantly, the chemistry must be biologically chemoselective, and not interfere with any endogenous biological processes. Furthermore, the reaction must proceed at

relatively rapid rates under physiological conditions (typically at 37°C and ~pH 7) as well as being inert to the myriad of functionalities encountered in biological systems.<sup>18</sup>

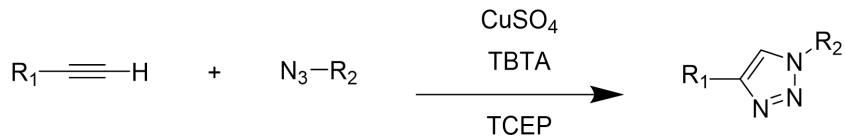
Should the above conditions be met, the chemistry is deemed bioorthogonal. However, when considering the numerous functionalities found within living systems, as well as the limitations of concentrations and temperatures, the task of identifying bioorthogonal chemistries seems daunting.

### **Cu(I)-1,3-dipolar Cycloaddition or Cu(I) “Click” Chemistry**

Click chemistry, a subset of reactions that can proceed to completion in relatively mild conditions to produce a single product with minimal to no side reactions, has emerged as a means for performing bioorthogonal chemistry. In particular, the [3+2] cycloaddition of azides and alkynes, termed a Huisgen-1,3-dipolar cycloaddition, has flourished as a means to tackle the complexity of bioorthogonality.<sup>19</sup> Not only do these reactions produce a stable, covalent triazole moiety, thus ensuring efficient and stable conjugation between the two functionalities *in vivo* where bonds may readily decompose, but the azide and alkyne functionalities are also absent from biological systems. Furthermore, the two molecules react selectively with each other, thus side-reactions with other biological molecules is not a concern. Thus, the 1,3-dipolar cycloaddition using azides and alkynes has the potential to be an optimal chemistry for the development of bioconjugation reactions. However, the initial reaction reported in the early 20<sup>th</sup> century required temperatures upwards of 98°C in order to achieve useful reaction rates, well above the temperature range of physiological viability.<sup>19</sup>

The viability of this chemistry was restored when the Sharpless and Meldal groups concurrently, yet separately, developed a copper(I) catalyzed variant of the cycloaddition.<sup>20,21</sup> Reduction of copper(II) to copper(I) in the presence of a terminal alkyne allows for the formation of an activated copper acetylide intermediate, which has an increased reactivity with azides to form the traizole ring product. This copper-catalyzed azide-alkyne 1,3-dipolar cycloaddition (CuAAC) exhibited an increased reaction rate of seven orders of magnitude, as well as being capable of reacting at room temperature and even 4°C.<sup>22</sup> The fact that the reaction can be made even more favorable by the addition of tetradentate ligands, such as tris[(1-benzyl-*I*H-1,2,3-trizol-4-yl)methyl]amine (TBTA), that help stabilize the copper(I) catalyst resulted in this chemistry rising to the forefront of bioconjugation reactions.<sup>23</sup>

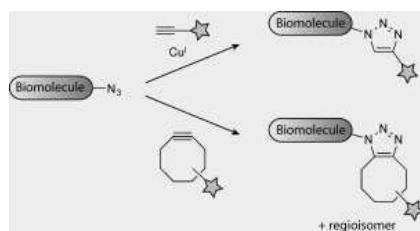
*Scheme 1*



While the CuAAC reaction afforded desirable temperature ranges and reaction rates, it possessed a single hindrance to its use in biological systems. The reaction's dependence on copper was ultimately its greatest flaw, as copper is cytotoxic to most organisms. Bacterial cells have been shown to halt cell division following exposure to 100 μM CuBr, and mammalian cells can only endure very low concentrations of copper (below 500 μM) for 1 hour.<sup>18</sup> As such, the Bertozzi group sought to develop a copper free variant that may proceed with similar favorability by capitalizing on strained ring systems. In particular, cyclooctyne, the smallest ring that can contain a triple bond, was used to activate the alkyne towards the cycloaddition via strain. The result was a

cycloaddition, however, the reaction proceeded at rates considerably slower than CuAAC.<sup>24</sup> By addition of electron withdrawing fluorines, the Bertozzi group was able to generate a difluorinated cyclooctyne (termed DIFO) capable of undergoing copper free click reactions at rates comparable to CuAAC. Thus, the cyclooctyne reagent is nontoxic and can be derivitized to include additional functionalities, such as conjugated fluorescent moieties.<sup>25</sup>

*Scheme 2.*<sup>18</sup>



The CuAAC and the strain promoted azido-alkyne cycloaddition (SPAAC) both serve as useful reactions for the bioconjugation of proteins bearing UAAs with other molecules. Because the placement of UAAs within a protein can be site-specifically controlled, UAAs can be inserted into proteins in unique residues that minimize changes to the protein's folding while allowing for maximal exposure of the UAA. These surface exposed residue sites allow for UAAs to serve as unique biochemical handles for the targeted promotion of these bioconjugation reactions. UAAs synthesized with azide or alkyne functionality can be correspondingly clicked to another molecule bearing the appropriate cognate handle. Such strategies allow for the directed control of bioorthogonal chemistry towards proteins of interest.

Indeed, both CuAAC and SPAAC have already begun to receive attention as novel methodologies to probe biological systems. Research on intracellular pathogens is often complicated due to the difficulty of probing the pathogen in their endosomal

environment.<sup>26,27</sup> Methods such as antibody labeling are complicated by the need to pass through multiple membrane structures. As such, Siegrist *et al.* demonstrated that UAA analogs of D-alanine, which harbor an alkyne functional group, can be used to probe peptidoglycan dynamics of intracellular bacterial pathogens *in vivo*, a technique that was not previously accessible using more traditional antibody labeling approaches.<sup>26</sup> In particular, the authors demonstrated that the incorporation of the alkyne or azide derivitized D-alanine UAAs allowed the researchers to perform CuAAC reactions using the orthogonal probe on a fluorescent dye. Because the D-alanine analogs are used exclusively by bacteria in the synthesis of their peptidoglycan layer, the researchers were able to label intracellular pathogens, such as *Listeria monocytogenes*, in an *in vivo* infection model and then observe peptidoglycan dynamics in real time.

### Biorthogonal Transition Metal Chemistry

Many transition metal cross coupling reactions are biorthogonal, as the reactive groups and metals involved are not found in biological systems. While limitations include the toxicity and oxidative damage induced by the use of the transition metal, as well as the prohibitive costs, overall these technologies are proving to be viable bioorthogonal reactions. Through the use of 3-mercaptopropionic acid (3-MprAc) as a small and effective palladium scavenger, Spicer and Davis were able to utilize *p*-iodophenylalanine (*p*IF) to allow for introduction of an aryl halide for Suzuki-Miyaura cross coupling.<sup>28–32</sup> This chemistry was then employed for the installation of hydrophobic groups on the surface of a protein without requiring organic solvents or denaturing conditions. Full conversion of the protein substrate was achieved and is notable when compared to the 2%

and 25% conversion achieved via the Heck reaction and Sonogashira coupling respectively.<sup>32</sup> This technology was also used to tag porin channels via OmpC protein monomers on the outside of cell surfaces. Using a fluorescent boronic acid as a label and by measuring fluorescent intensity, a dramatic increase in labeling via Suzuki-Miyaura coupling was achieved once a threshold concentration of 300  $\mu\text{M}$  of Pd was utilized.<sup>31</sup> A corresponding Suzuki coupling has been performed with the boronate-containing UAA; however, this reaction required both a high temperature (70 °C) and a high pH (8.5), thus requiring further optimization.<sup>33</sup>

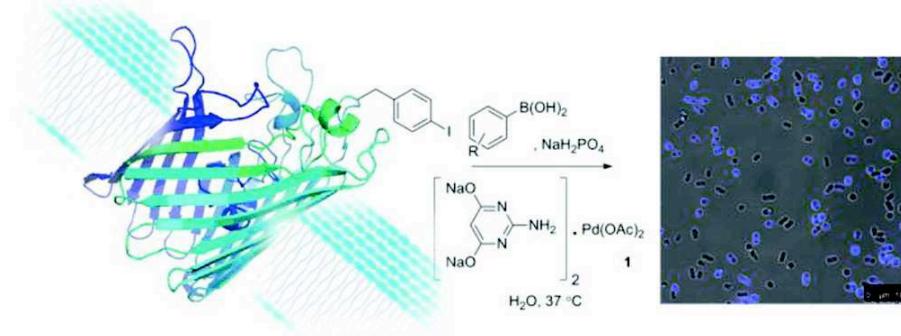
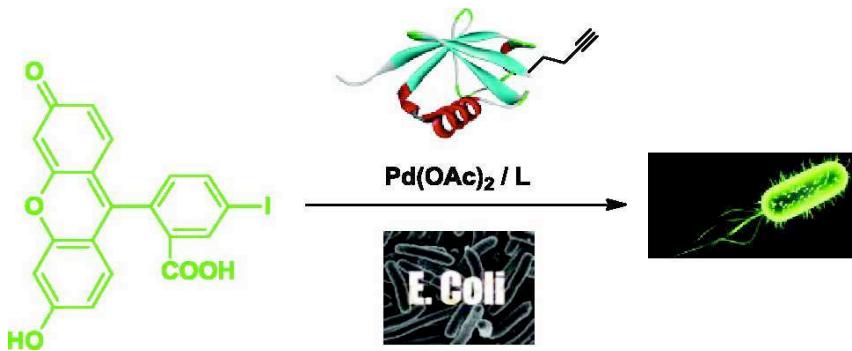


Figure 8. Labeling of pIF-containing OmpC using a fluororescent boronic acid via a biological Suzuki coupling.

Another palladium-based bioorthogonal chemistry is the Sonogashira cross coupling reaction. It was recently reported that an alkyne encoded ubiquitin was employed to facilitate the attachment of aryl or vinyl containing groups.<sup>34</sup> The reaction reached completion in 30 minutes and did not occur in the absence of the palladium catalyst, demonstrating the efficiency and Pd-mediated regulation of the reaction. This coupling has been accomplished via the incorporation of a variety of UAAs including pIF and lysine derivatized with an alkyne and iodo handle.<sup>34,35</sup>



*Figure 9.* Surface labeling of *E. coli* using a an alkyne containing UAA in the presence of an iodo containing fluorogenic dye via the biological Sonogashira coupling.

Though palladium mediated reactions that alter cell-surface functionality are biologically viable at low levels of palladium, intracellular use of this reaction is limited by the differing effects of palladium within the cell. Therefore, optimizing such reactions for intracellular purposes can broaden the scope of such bioorthogonal chemistries.

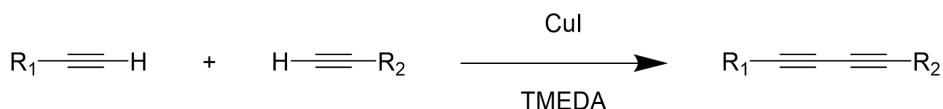
As such, a recent report developed a Pd-mediated intracellular bioconjugation employing lysine-derivatized UAAs containing either an alkynyl or aryliodo functionality. These UAAs were readily incorporated into intracellular gram negative bacterial proteins *in vivo* using an evolved pyrolysine tRNA synthetase/tRNA system. In the presence of a simple  $\text{Pd}(\text{NO}_3)_2$  catalyst, the authors were able to show efficient intracellular conjugation of a dye at room temperature ( $25^\circ\text{C}$ ) within 40 mins. Furthermore, the use of this Pd-catalyst was biocompatible, with neglible cytotoxicity at concentrations below  $200\mu\text{M}$ .<sup>35</sup>

### Development of a Biological Glaser-Hay Coupling

As demonstrated by the CuAAC, SPAAC, and the transition metal chemistries above, it is apparent that reactions involving terminal alkynes are advantageous for use in biological systems due to the relative absence of triple bonds within these systems. The

Glaser-Hay coupling is a reaction that involves conjugating two terminal alkynes via a copper catalyst to afford a conjoined diyne product<sup>36,37</sup> Salts of copper(I) are typically used as the catalyst, and it has been proposed that the copper(I) complexes with the terminal alkyne, activating the functionality towards the formation of a carbon-carbon bond with another terminal alkyne.<sup>38,39</sup>

*Scheme 3.*



The coupling of terminal alkynes was first described by Carl Glaser in 1869.<sup>36</sup> However, the reaction was not suited for biological settings. Glaser's initial reaction, however, involved an explosive copper(I)-phenylacetylide catalyst that required tedious isolation following the reaction.<sup>36</sup> Then, in 1962, Allan Hay published results using the bidentate ligand N,N,N',N'-tetramethylethylenediamine (TMEDA) in conjunction with the copper catalyst. This methodology improved reaction times, reduced temperature requirements, and reduced the amount of copper necessary to allow the reaction to proceed.<sup>37</sup> It is proposed that the bidentate ligand TMEDA stabilizes the Cu(I) species, opening up the complex to favor addition of the activated alkyne species to one end of the copper species.<sup>38,39</sup> The Glaser-Hay reaction today, in which a Cu(I) salt in the presence of TMEDA in an organic solvent is added to two terminal alkynes, allows for the rapid and facile synthesis of conjugated diynes. Furthermore, recent work has demonstrated the reaction can proceed in good yield under biological temperatures (4°C to 37°C), and has enhanced reaction rates.<sup>40</sup> While little work has been done to take advantage of these unique features of this chemistry, these characteristics make the Glaser-Hay coupling

ideal for the utilization of the bioorthogonal alkyne moiety for bioconjugation applications.

## Conclusion

UAA-mutagenesis technology allows the researcher to create site-specific functional “handles” on proteins of interest. Herein we propose applying UAA-mutagenesis for the development of novel biorthogonal chemistries, biological probes, protein technologies, and immobilization techniques. In particular, employing the terminal alkyne moiety throughout our studies to further expand the scope of its role in chemical biology. By combining this technique with bioorthogonal chemistries, we hope to manipulate biological systems and proteins in new ways that can further advance basic science and medicine.

## References

1. Berg, J. M., Tymoczko, J. L., Gatto, G. J. & Stryer, L. *Biochemistry*. (MacMillan, 2015).
2. Nguyen, Q. T. & Tsien, R. Y. Fluorescence-guided surgery with live molecular navigation--a new cutting edge. *Nat. Rev. Cancer* **13**, 653–62 (2013).
3. Zhu, H. & Snyder, M. Protein chip technology. *Curr. Opin. Chem. Biol.* **7**, 55–63 (2003).
4. Rusmini, F., Zhong, Z. & Feijen, J. Protein immobilization strategies for protein biochips. *Biomacromolecules* **8**, 1775–1789 (2007).
5. Tan, W. *et al.* Optical protein sensor for detecting cancer markers in saliva. *Biosens. Bioelectron.* **24**, 266–271 (2008).
6. Jokerst, J. V. *et al.* Nano-bio-chips for high performance multiplexed protein detection: Determinations of cancer biomarkers in serum and saliva using quantum dot bioconjugate labels. *Biosens. Bioelectron.* **24**, 3622–3629 (2009).

7. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
8. Young, T. S. & Schultz, P. G. Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *J. Biol. Chem.* **285**, 11039–11044 (2010).
9. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249 (2006).
10. Cornish, V. W. & Schultz, P. G. Mutagenesis an. (1995).
11. Martin, a B. & Schultz, P. G. Opportunities at the interface of chemistry and biology. *Trends Cell Biol.* **9**, M24–M28 (1999).
12. Strable, E. *et al.* Unnatural amino acid incorporation into virus-like particles. *Bioconjugate ...* 866–875 (2008).
13. Bezerra, A. R., Guimarães, A. R. & Santos, M. A. S. Non-Standard Genetic Codes Define New Concepts for Protein Engineering. *Life (Basel, Switzerland)* **5**, 1610–1628 (2015).
14. Mehl, R. a. *et al.* Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.* **125**, 935–939 (2003).
15. Liu, D. R., Magliery, T. J., Pastrnak, M. & Schultz, P. G. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10092–10097 (1997).
16. Wang, L., Brock, A., Hererich, B. & Schultz, P. G. Expanding the genetic code of E. Coli. *Science* **292**, 498–500 (2001).
17. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8**, 883–890 (2001).
18. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chemie - Int. Ed.* **48**, 6974–6998 (2009).
19. Huisgen, R. Kinetics and Mechanism of 1,3-Dipolar Cycloadditions. *Angew. Chemie Int. Ed. English* **2**, 633–645 (1963).
20. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. Supporting Information for ‘A Stepwise Huisgen Cycloaddition Process Catalyzed by Copper (I): Regioselective Ligation of Azides and Terminal Alkynes’. *Angew. Chemie Int. Ed. English* **114**, 2708–2711 (2002).

21. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–3064 (2002).
22. Himo, F. *et al.* Copper(I)-catalyzed synthesis of azoles. DFT study predicts unprecedented reactivity and intermediates. *J. Am. Chem. Soc.* **127**, 210–216 (2005).
23. Chan, T. R., Hilgraf, R., Sharpless, K. B. & Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* **6**, 2853–2855 (2004).
24. Agard, N. J., Prescher, J. a & Bertozzi, C. R. A Strain-Promoted [ 3 + 2 ] Azide - Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. 15046–15047 (2004).
25. Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16793–16797 (2007).
26. Siegrist, M. S. *et al.* D-amino acid chemical reporters reveal peptidoglycan dynamics of an intracellular pathogen. *ACS Chem. Biol.* **8**, 500–505 (2013).
27. Swarts, B. M. *et al.* Probing the mycobacterial trehalome with bioorthogonal chemistry. *J. Am. Chem. Soc.* **134**, 16123–16126 (2012).
28. Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22**, 1297–301 (2004).
29. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun.* **47**, 1698–1700 (2011).
30. Dumas, A. *et al.* Self-ligated Suzuki-Miyaura coupling for site-selective protein PEGylation. *Angew. Chemie - Int. Ed.* **52**, 3916–3921 (2013).
31. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki-Miyaura cross-coupling. *Chem. Commun.* **49**, 2747–2749 (2013).
32. Spicer, C. D., Triemer, T. & Davis, B. G. Palladium-mediated cell-surface labeling. *J. Am. Chem. Soc.* **134**, 800–803 (2012).
33. Brustad, E. *et al.* A Genetically Encoded Boronate Amino Acid. ... (*International ed.* ... 8344–8347 (2008). doi:10.1002/ange.200803240
34. Li, N., Lim, R. K. V., Edwardraja, S. & Lin, Q. Copper-free sonogashira cross-coupling for functionalization of alkyne-encoded proteins in aqueous medium and in bacterial cells. *J. Am. Chem. Soc.* **133**, 15316–15319 (2011).

35. Li, J. *et al.* Ligand-free palladium-mediated site-specific protein labeling inside gram-negative bacterial pathogens. *J. Am. Chem. Soc.* **135**, 7330–7338 (2013).
36. Glaser, C. Glaser describes formation of diyne. *Ber. Dtsch. Chem. Ges.* **2**, 422–424 (1869).
37. Hay, A. Oxidative Coupling of Acetylenes. II1. *J. Org. Chem.* **3**–4 (1962). doi:10.1021/jo01056a511
38. Vilhelmsen, M. H., Jensen, J., Tortzen, C. G. & Nielsen, M. B. The Glaser-Hay reaction: Optimization and scope based on <sup>13</sup>C NMR kinetics experiments. *European J. Org. Chem.* **3**, 701–711 (2013).
39. Fomina, L., Vazquez, B., Tkatchouk, E. & Fomine, S. The Glaser reaction mechanism. A DFT study. *Tetrahedron* **58**, 6741–6747 (2002).
40. Lampkowski, J. S., Villa, J. K. & Young, D. D. Development and Optimization of Glaser-Hay Bioconjugations. *Angew. Chemie Int. Ed. English (pending Revis.* (2015).

## **CHAPTER 2: SYNTHESIS AND INCORPORATION OF UNNATURAL AMINO ACIDS TO PROBE AND OPTIMIZE PROTEIN BIOCONJUGATIONS**

Multiple techniques exist for the introduction of UAAs that range from completely synthetic to completely biological via exploitation of endogenous translational machinery.(20-22) The evolution of orthogonal aminoacyl-tRNA synthetase/tRNA pairs has found a wide degree of success both in bioconjugations and various other applications.(22) Specifically, several examples of the use of unnatural amino acids to generate well defined bioconjugates including the reaction of an azide containing calmodulin onto carbon nanotubes and nanoparticles via the Staudinger-Bertozzi ligation,(23, 24) the immobilization of other proteins to M-20 Dynabeads in a cell-free system with a click reaction,(25, 26) and the labeling of proteins with an agarose resin and azido UAA for secretome MS analysis; (however, in this case the UAA was not site-specifically incorporated into the pool of proteins).(27) More recently, we demonstrated the optimization of protein immobilization utilizing an incorporated azido UAA, conferring increased stability to organic solvents.(28) The combination of these studies demonstrates the utility of site-specific bioconjugations, and the stabilizing benefits of the solid support. Other studies have used similar techniques to label proteins with molecular probes or generate novel therapeutic bioconjugates, demonstrating the wide applicability of site-specific bioconjugations. (4, 29) Given the power of this technology, we aimed to expand the utility of UAA bioconjugations through the alteration of the reactive UAAs to optimize reaction parameters.

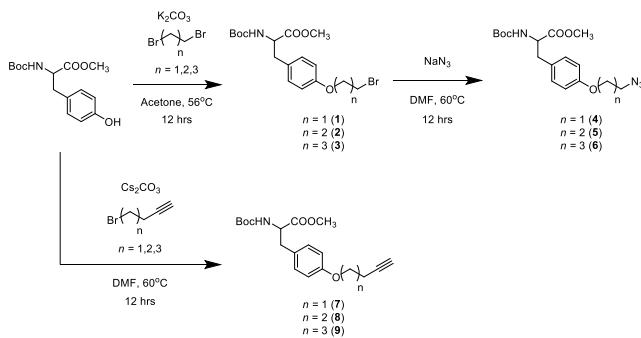
Initial investigations attempted to further utilize UAAs to optimize bioconjugations through several different bioorthogonal reactions. Our previous studies immobilizing GFP demonstrated that the UAA residue location within the protein dictated its immobilization efficiency.(28) Thus, the spatial configuration of the UAA handle is vital for its reactivity with its conjugation partner. While it is possible to alter where the UAA can be incorporated, this requires extensive genetic manipulation. An alternative approach is to simply vary the distance of the UAA reactive group from the protein's surface. By extending the distance from the protein, the handle is less hindered by the protein's steric bulk, which may increase its propensity to react with its orthogonal partner. Moreover, removing the reactive partner from conjugation with the aromatic ring of the phenylalanine moiety may alter its chemical reactivity. Thus, it becomes important to ascertain the ideal distance of the functional handle from the protein for efficient reactivity. In order to probe novel bioorthogonal reactions utilizing UAA technology, a variety of UAAs with new functional groups were synthesized and designed with variable tether lengths to separate the reactive species from the bulk of the protein. These groups included an azide, an alkyne, and a bromide respectively, all of which can be derived from a modified tyrosine (Scheme 1). These functionalities afford access to the Cu(I)-catalyzed 1,3-dipolar cycloaddition, the Glaser-Hay coupling, and a nucleophilic substitution respectively.

## Results and Discussion

To optimize the immobilization process, UAAs were synthesized with varied methylene units extending from a tyrosine precursor. Derivatives containing 2, 3, and 4 methylenes

of the azide, alkyne, and bromide functionalities previously mentioned were readily accessible via a S<sub>N</sub>2 reaction (Scheme 1).

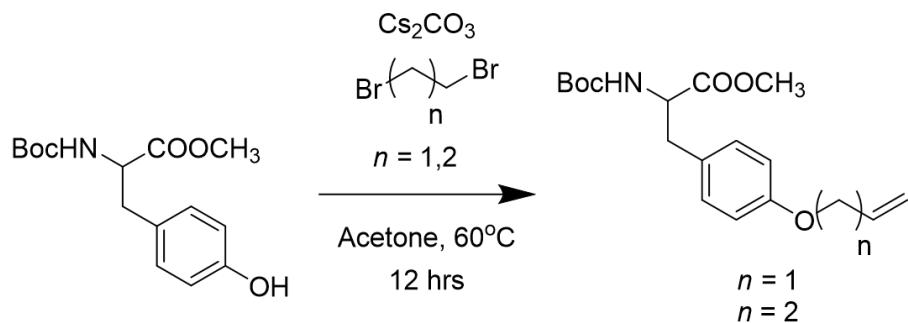
*Scheme 4.*



Synthesis of the nine UAAs began from a common N-Boc-OMe protected tyrosine precursor. The goal was to synthesize a bromotyrosine derivative via a substitution reaction with either dibromoethane, dibromopropane, or dibromobutane. These bromotyrosines (1-3) would then possess alkyl chains of varying lengths, and a good leaving group for future reactions. Then, a secondary S<sub>N</sub>2 reaction with either an azide (4-6) or alkynyl nucleophile would install the desired novel reactivity to the tyrosine derivatives. However, synthesis of the alkynyl-tyrosine series proved more difficult than initially anticipated, and the series was prepared directly via S<sub>N</sub>2 reactions commercially available with bromo-alkynes (see Experimental).

Initial reactions between the protected tyrosine and the dibromobutane in dimethylformamide (DMF) yielded the desired tyrosine derivative 3. However, reactions involving the dibromopropane and dibromoethane in DMF did not proceed as expected towards the desired tyrosine derivatives 1 and 2 respectively. Instead, starting material or the alkene tyrosine derivative, the product of an elimination reaction, was obtained for both reactions.

Scheme 5.



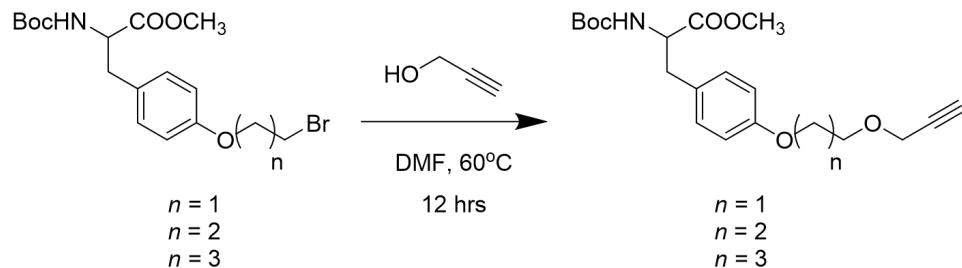
Initially, the concern was that the DMF was not dry, and the presence of water in the solvent may have led to increased elimination of the bromide. However, switching to dry solvent systems did not improve the reaction, nor did it reduce the amount of elimination product obtained. To optimize the reactions, we attempted adding a catalytic amount of potassium iodide to both the dibromopropane and dibromoethane reactions. The iodide can serve as a catalyst in this nucleophilic substitution, replacing the bromine with iodine, a better leaving group, thereby facilitating the S<sub>N</sub>2 addition by the phenolate of tyrosine. However, only starting material was recovered from these reactions, even when reaction times and temperatures were varied. Ultimately, a literature search revealed a similar compound had been successfully prepared using acetone as the solvent system.<sup>37,38</sup> Thus, all three reactions were attempted in acetone, with greater success. All three bromotyrosine derivatives were successfully synthesized using acetone as the solvent in the presence of base and potassium iodide as a catalyst. We hypothesize that the additional polar amine functionality in the DMF allowed for increased solvent-mediated stability of the free bromine via dipole-dipole interactions following an elimination reaction. Thus, DMF aided in the formation of the eliminated product. Furthermore, electronics are clearly involved, as the shorter the alkyl chain became, the closer in proximity the large, electronegative bromine groups were, increasing the

propensity to eliminate. This was observed by a higher rate of elimination product using dibromoethane when compared to dibromopropane. The use of acetone, a less polar solvent, decreases the solvent-mediated dipole-dipole interactions with the free bromine, and thus disfavors the elimination reaction. This was observed as the bromotyrosine derivatives were able to be successfully synthesized under these modified conditions.

Once the bromotyrosine compounds 1 to 3 had been synthesized, they were then divided and reacted with either sodium azide or propargyl alcohol. Reaction with sodium azide in DMF resulted in a nucleophilic substitution reaction between the azide and the bromotyrosine, affording the azidotyrosine derivatives 4 through 6 following silica gel purification and NMR confirmation. The rest was reacted with propargyl alcohol in DMF in order to perform a nucleophilic substitution between the hydroxyl group and the bromotyrosine, resulting in the formation of the alkynyltyrosine products.

The alkyne-derivitized tyrosine series was not successfully synthesized under these conditions (see Scheme 3), and instead yielded either the previously described elimination product or the starting material.

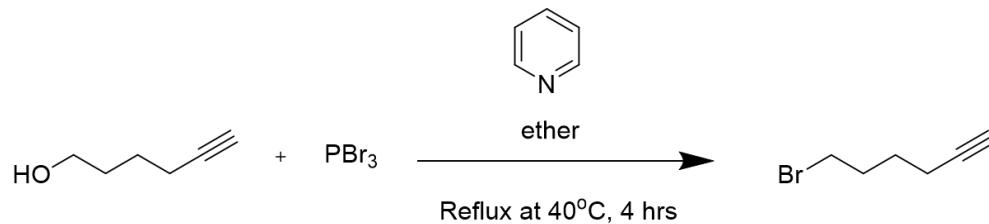
*Scheme 6.*



Initially, potassium tert-butoxide ( $\text{pK}_\text{a} \sim 17$ ) was used as the activating base in the reaction. We hypothesized that the propargyl alcohol ( $\text{pK}_\text{a} \sim 13.6$ ) may not be fully deprotonated under these conditions, perhaps reducing the overall yield of the reaction. As such, a small amount of sodium hydroxide was used as the activating base with little

success. Ultimately, due to the inability to obtain derivatized tyrosine product, an alternative synthesis strategy was adopted. Reaction of 4-bromo-1-butyne and 5-bromo-1-pentyne with protected tyrosine in acetone with a catalytic amount of potassium iodide was allowed to progress overnight (see scheme 1). Following column chromatography, the desired alkyne products 7 and 8 were obtained, with the proper length alkyl chains (either 3 or 2 methyl) for proper comparison to the azido- and bromo- derivatives of similar lengths. In addition, 6-bromo-1-hexyne could be readily synthesized by reacting 6-hydroxy-1-hexyne with phosphorus (III) bromide (Scheme 4), followed by immediate reaction with the protected tyrosine in acetone. Again, following column chromatography the desired compound 9 was isolated.

*Scheme 7.*



The syntheses described above afford the derivatized tyrosines still harboring protecting groups. To prevent reactivity of the amino group of the amino acid, an acid labile tert-butyloxy (Boc) group was employed. Additionally, to protect the carboxylic acid functional group of the amino acid backbone, a base labile methyl ester was used. Successive stirring in 1M lithium hydroxide to remove the methyl ester and 50% TFA in DCM to remove the Boc group, afforded the nine deprotected UAAs in moderate overall yields (see Table 1).

*Table 1. Overall Yields for Tethered UAA Syntheses*

<b>X</b>	<b>n</b>	<b>% Yield</b>
	1 (10)	5.3% (p2BrY)
	2 (11)	8.2% (p3BrY)
	3 (12)	10.8% (p4BrY)
	1 (13)	32.0% (p2AzY)
	2 (14)	58.0% (p3AzY)
	3 (15)	43.3% (p4AzY)
	1 (16)	2.6% (p2yneY)
	2 (17)	18.7% (p3yneY)
	3 (18)	3.1% (p4yneY)

Typically with the new synthesis of UAAs a unique aaRS must be evolved to recognize and incorporate the UAA; however, based on the previous demonstration of the promiscuity of the *p*CNF-aaRS in *E. coli*,(30, 31) it was our hope that the structural similarity of the nine prepared UAAs could facilitate incorporation without the need of a double sieve aaRS selection. With respect to the bromo-derivatives, a previous report used a similar approach to their incorporation in mammalian cells.(32, 33) Based on our earlier studies we employed green fluorescent protein (GFP) as a model protein for

incorporation due to its well-documented fluorescent properties facilitating rapid visualization and quantitation. Our previous findings indicated that residue 151 conferred the best conjugation efficiencies, and was thus selected for this study.(28) BL21(DE3) *E. coli* were co-transformed with pEVOL-*p*CNF-aaRS and pET-GFP-151-TAG plasmids to introduce both the GFP protein and the translational machinery necessary for TAG suppression with UAAs. Following protein expression (30 °C, 16 h) in the presence or absence of the UAA, cells were lysed and GFP was purified via Ni-NTA chromatography. Gratifyingly, the application of the polyspecific aaRS was successful, as the nine UAAs were able to be incorporated into GFP, albeit with differential expression levels (Table 2). The incorporation of the UAA was visualized both by fluorescence and SDS-PAGE analysis. As the *p*CNF synthetase was evolved to recognize *p*CNF, it was hypothesized that the *p*2yneY-, *p*3yneY-, and *p*4yneY-UAA series would be most easily incorporated due to the structural similarity between the cyano group and alkyne group in size and polarizability.

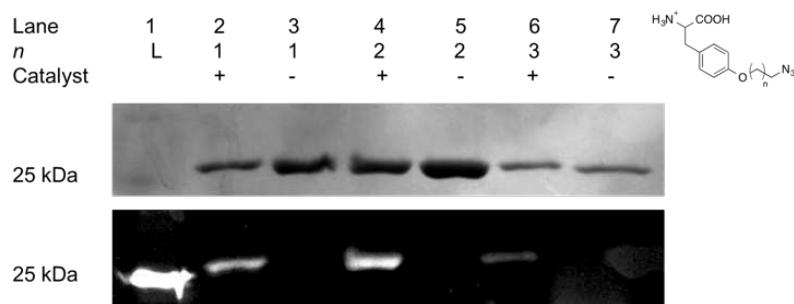
*Table 2. Protein Expression Yields for the Tethered UAAs*

<i>n</i>	<b>X</b>		
	Br	N <sub>3</sub>	C≡CH
1	0.704 mg/mL	0.347 mg/mL	0.702 mg/mL
2	0.291 mg/mL	0.685 mg/mL	0.767 mg/mL
3	0.277 mg/mL	0.246 mg/mL	0.447 mg/mL

This result was observed, with the *p*3yneY (*n* = 2) producing a larger quantity of mutant-GFP, and the *p*4yneY producing the lowest quantity of mutant-GFP within the alkyne series. These qualitative observations were in accordance with our predictions based on the evolution of the synthetase. We hypothesize that the increased tether distance may have more difficulty fitting into the binding pocket of the tRNA synthetase relative to smaller derivatives, preventing the UAA from being incorporated into the polypeptide chain. Thus, a tradeoff may exist between the advantages of distancing the functional group from the core of the protein and the reduced protein yields of these same compounds. This trend was observed for all of the UAAs, as the tether length increased, the expression yield decreased. However, all 9 UAAs were capable of being incorporated using a single aaRS, not only obviating the necessity of a tedious selection, but also increasing the versatility of this approach as only a single cotransformation is required to produce 9 different UAA-containing proteins.

Given the ability to incorporate 3 different functionalities with variable tether lengths into GFP, we next sought to employ these new UAAs in bioorthogonal conjugation reactions. Studies were initiated with the azido-GFP series, as the cycloaddition reaction is one of the most common and well-understood bioorthogonal reactionS. Moreover, the various tethers could be directly compared to the commonly utilized *p*-azidophenylalanine (*p*AzF) that has no methylene tether and contains the azido group in direct conjugation with the aromatic ring. We hypothesized that in addition to providing distance between the protein and the conjugation partner, the removal of the azido conjugation may also alter its reactivity. The various azido-containing GFP mutants were conjugated with a fluorescent Alexafluor 488 alkyne dye using a CuSO<sub>4</sub>/TBTA/TCEP catalyst system based

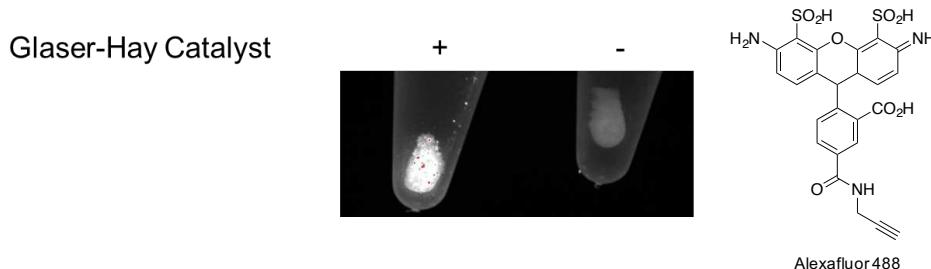
on previously reported conditions.<sup>40,41</sup> Gratifyingly, all tethered derivatives could be conjugated to the dye, as determined by both fluorimetry and SDS-PAGE analysis (Figure 1). Control reactions were also performed in the absence of catalyst system to ensure that protein labeling was not due to non-specific reactions or non-covalent association by the dye. These controls yielded no fluorescent product following protein denaturation, and thus indicated that any fluorescence is indeed a result of the coupling to form the triazole ring. Interestingly, the *p*3AzY handle appears to have the best conjugation efficiency, compared to the more sterically hindered *p*2AzY and the more flexible *p*4AzY handles (Figure 1).



*Figure 10. Expression and conjugation of GFP-151 mutants with an AlexaFluor488 dye. The Coomassie stained gel (top) indicates the successful incorporation of all three azides in the series as an appropriate band is observed for GFP at ~26 kDa. The fluorescence of the same gels (bottom) indicate each azido-GFP mutant also was able to undergo a successful bioconjugation with an AlexaFluor-488 alkyne. No coupling was observed in the absence of copper catalyst despite the inclusion of the fluorophore alkyne, as no fluorescence is detected in the absence of catalyst.*

With successful bioorthogonal reaction conditions in hand, we next sought to apply the chemistry for immobilization purposes. As mentioned previously, the most robust and ideal mechanisms for protein immobilization involve covalent linkages. However, this technique is often confounded by the potential for multiple reaction sites on a given protein, which could render proteins that are oriented incorrectly for their function. UAA mutagenesis allows for the development of a bioorthogonal handle in a predetermined

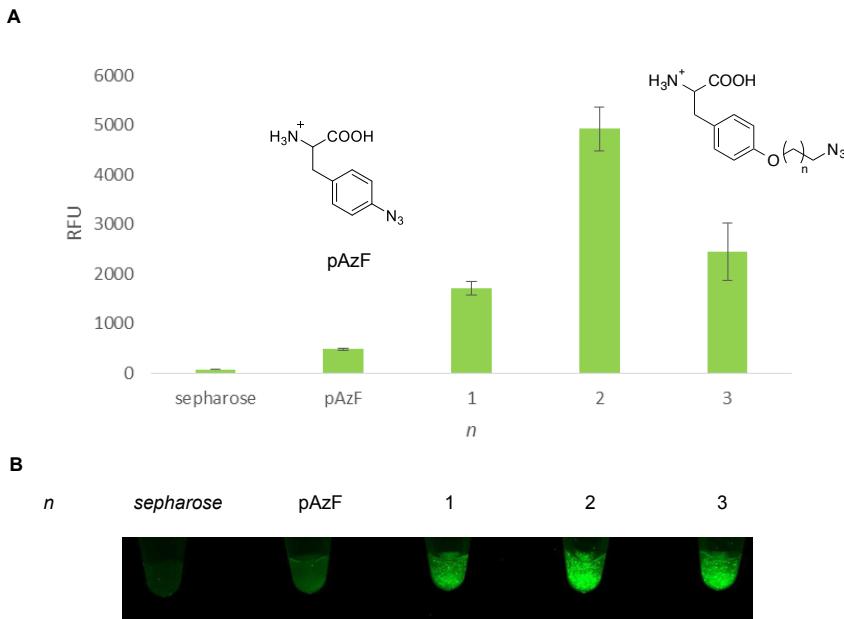
site.(34) Based on previous UAA immobilization results, an epoxy-activated sepharose resin was derivatized with propargyl alcohol in the presence of trimethylamine (TEA) by stirring overnight at room temperature.(28) This afforded a resin with an alkyne functionality that could be employed as a reaction partner with the azido-GFP series (see Figure 2).



*Figure 11. Verification of alkyne functiaonlized resin. Prepared alkyne sepharose resin was reacted with an alkyne containing fluorophore (Alexafluore 488) in the presence and absence of Glaser-Hay conjugating conditions (50 mM Cul and 50 mM TMEDA), which has been shown to conjugate two terminal alkynes. Following successive rinsing with PBS, fluorescence was only observed in the presence of catalyst, demonstrating that the fluorophore was bound to the resin via generation of a diyne linkage and not through non-specific interaction. This reaction verifies that the sepharose resin was indeed derivatized with a terminal alkyne handle for immobilization experiments.*

It was expected that by extending the tethering distance of the functional handle of the UAA, and thus the site of bioorthogonal reactivity, the steric hindrance due to the protein's bulk could be decreased and immobilization efficiency could be improved. As such, the *p*2AzY-, *p*3AzY -, and *p*4AzY -mutated GFPs were exposed to copper click conditions (50 mM CuSO<sub>4</sub>, 50 mM TCEP, and 5 mM TBTA) at fixed concentration of GFP (roughly 0.250 mg/mL). As a positive control, the *p*AzF-GFP, which has no tether between the aromatic ring and the azide functional group, was also subjected to the same copper click conditions. We have previously immobilized the *p*-azidophenylalanine mutated GFP on the alkyne-functionalized sepharose resin using similar conditions.(28)

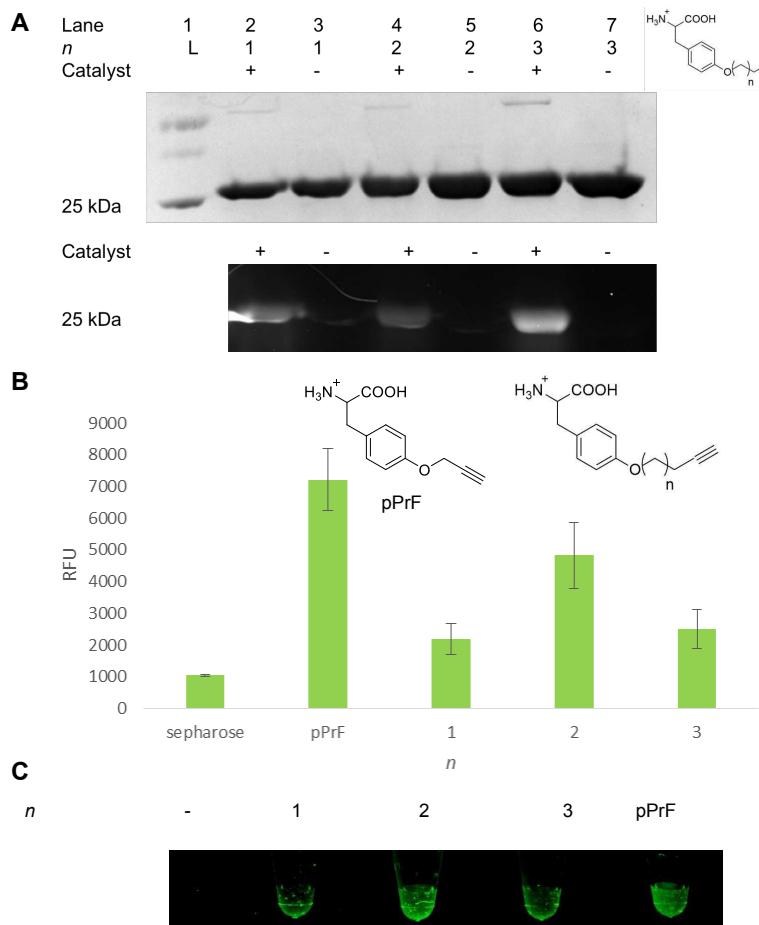
As a negative control, wild-type GFP, in which there is no azido-functional group present, was also subjected to the same copper click conditions. This control provided a means to rule out non-specific interactions as the reason for immobilization rather than the selective copper click reaction. Results indicated that the copper click-based immobilization strategy was effective for all of the mutants in the series (Figure 3). Interestingly, the tether distance did factor into the efficiency of the immobilization reaction but not as expected, as increased distance from the protein did not necessarily lead to improved immobilization as measured by relative fluorescence intensity on a microplate reader (Figure 3). However, confirming the results observed when coupling to the Alexafluor488 dye, the *p*3AzY (*n* = 2) derivative again had the highest coupling and immobilization efficiency, indicating that the interplay of structural rigidity and distance is important when optimizing the bioconjugation reaction. This 3-carbon tether seems to exploit the best properties of each factor, conferring a high coupling and immobilization efficiency. The *p*4AzY (*n* = 3) had decreased immobilization efficiency. We hypothesize that the increased flexibility afforded to this variant may decrease the propensity of its azido handle to orient in the proper geometry in space for reaction.



*Figure 12. Click immobilization of GFP-151 mutants on a sepharose resin with either an azide tethered tyrosine UAAs or a p-azidophenylalanine. A) The immobilization of GFP at a fixed concentration of 0.246mg/mL of each mutant protein. Fluorescence signal at 528nm on a plate reader indicates that the n=2 GFP-151 azide tethered variant performed best for immobilization (immobilizations conducted in triplicate, average standard deviation +/- 296). B) Image of the immobilized resins as irradiated at 365 nm on a transilluminator. Overall, the tethered variants appeared to have a greater immobilization efficiency than the traditional pAzF.*

Based on these results, we next investigated the role of the methylene tether on alternate reactions. Recently, we reported the first Glaser-Hay bioconjugation, linking two terminal alkyne functionalities to yield a conjugated diyne. (18) Consequently, this reaction was selected to explore the reactivity of the alkyne-GFP series. The alkyne-containing GFPs were first conjugated to the same Alexafluor 488 alkyne dye previously employed in the click reactions. All mutant proteins were subjected to conjugation conditions in the presence of CuI (50 mM) and TMEDA (50 mM) and a control in the presence of the alkyne fluorophore but in the absence of catalyst system. After 4 hours at 4 °C , the proteins were denatured at 98 °C for 15 mins and then analyzed via SDS-PAGE (Figure 4A). These couplings were all successful, as demonstrated by a fluorescent band at ~26

kDa; however, the different tethered derivatives reacted with different efficiencies. Furthermore, no coupling was observed in the absence of the Glaser-Hay catalyst system, even in the presence of the fluorophore, indicating that the bioorthogonal Glaser-Hay chemistry is responsible for the conjugation between the fluorophore and the dye.



**Figure 13.** Glaser-Hay reactions of alkyne-containing GFP mutants. A) Bioconjugation of mutant GFPs with AlexaFluor488 under Glaser-Hay conditions. The Coomassie stained gel (top) indicates the incorporation of all tethered mutants into GFP-151. Moreover, the successful Glaser-Hay coupling with the fluorophore was confirmed via fluorescent imaging of the same gel (bottom), as AlexaFluor fluorescence is only observed in reactions containing the fluorophore and the CuI/TMEDA system (immobilizations conducted in triplicate, average standard deviation  $\pm 774$ ). B) Immobilization of alkyne-GFP mutants on a sepharose resin under Glaser-Hay conditions. Measuring resin fluorescence on a plate-reader (Ex 395/Em 512), all GFPs were able to be immobilized. C) Immobilization was also confirmed by GFP fluorescent imaging on a transilluminator.

The alkyne series was also reacted with the alkyne-derivatized sepharose resin to assess coupling efficiencies. All of the alkynyl-UAA containing GFP mutants could be immobilized, as indicated by the presence of fluorescence on the resin after reaction with CuI/TMEDA (Figure 4B and C). However, it appears that there is no relative benefit to increasing the tether length, in contrast to the trends observed with the azide tethers. Interestingly, the *p*-propargylphenylalanine, which has an alkynyl moiety closest to the phenyl ring of phenylalanine, performed the best. Based on the proposed mechanism for the Glaser-Hay reaction, we hypothesize the enhanced rigidity of the *p*PrF helps properly orient the protein-resin interactions as the copper binds in an  $\eta^1$  fashion. Moreover, the enhanced flexibility afforded by the novel alkynyl-UAs may diminish their propensity to react, as the steric bulk of the protein may be more predisposed to interact with the resin. An odd/even effect is observed as the *p*3yneY-containing GFP performed the best amongst the UAA-mutated GFPs in the immobilization experiment (Figure 4C). This suggests that there may be some benefit, although slight, to increasing the distance of the protein from the site of reactivity. This increase drops once the alkyl chain gets too flexible, as evidenced by the drop in the *p*4yneY-containing GFP.

Interestingly, within the series, the conjugation partner also plays a role in efficiency as the Glaser-Hay reaction with the small fluorophore was most efficient with the longer tether (Figure 4A), whereas, reaction with the bulky resin required decreased flexibility.

Finally, conjugation with the bromo-GFP series was investigated utilizing a S<sub>N</sub>2 reaction. The incorporation of this series with a different synthetase, followed by reaction with thiols has previously been reported in a mammalian system;(32, 33) however, we were interested to see if these synthetic intermediates could also be incorporated with the

*p*CNF aaRS. Due to the presence of many nucleophilic functional groups within biological settings, this reaction is not truly bioorthogonal as cross-reactivity is likely. Nevertheless, we reasoned that successful nucleophilic substitution with an amine resin, immobilizing the GFP, was an ideal means to demonstrate the successful incorporation of the bromo-UAAs and the generation of the *p*2BrY-, *p*3BrY-, *p*4BrY-GFPs. An amino-functionalized sepharose resin was generated using ethylenediamine (see Experimental). This resin was then subjected to the bromo-GFP series in the presence of a carbonate buffer (pH = 9.5). As a control, wild-type GFP was subjected to the same conditions. Gratifyingly, all of the bromo-containing GFPs successfully immobilized, with *p*4BrY performing the best (Figure 5 and 6). Compared to the other functionalities and reactions, in which the three carbon alkyl chain variant of the UAA consistently performed the best, this was the only case where we found a four-carbon alkyl chain to increase the propensity to immobilize. This is to be expected, as this bromo-tethered variant is the least sterically hindered by the protein's bulk and will therefore perform the best in a nucleophilic substitution reaction. Furthermore, the wild-type GFP was not immobilized, as indicated by minimal fluorescence on a microplate reader, verifying that the presence of the bromine-functionality was responsible for the immobilization, and thereby demonstrating that we had successfully synthesized and incorporated bromo-tyrosine UAAs.

*Figure 14. Immobilization of bromide-containing UAAs on a derivatized sepharose resin. A) Fluorescence was determined on a platereader (Ex 395/Em 512), with the increased tether increasing immobilization efficiency (immobilizations conducted in triplicate, average standard deviation +/- 1401). B) Image of bromo-UAAs immobilized on the resin.*

## **Conclusion**

In conclusion, nine, novel UAAs with various functional handles and alkyl chains were synthesized. Furthermore, all UAAs were successfully incorporated into GFP using a single promiscuous aaRS. The novel functional handles present in the UAAs allowed access to a range of different bioorthogonal chemistries. Additionally, the presence of alkyl chains of different lengths seemed to successfully distance the site of reaction from the protein surface, having differential effects on the outcome of the different bioorthogonal reactions. In particular, we have elaborated upon a previously reported immobilization technique using an azido-containing GFP. We improved the immobilization efficiency by increasing the distance between the protein and the site of the CuAAC reaction. In addition, we have also demonstrated the utility of the Glaser-Hay reaction for protein immobilization.

Based on these results, we have illustrated the potential of UAA-mutagenesis to generate unique, site-specific functional handles in proteins for the preparation of various bioconjugates. This technology has downstream molecular imaging applications for the study of protein dynamics *in vivo*, as well as applications involving immobilized enzymes and the synthesis of well-defined therapeutic agents.

## **Experimental**

General Solvents and reagents, including the AlexaFluor 488 Alkyne, were obtained from Sigma Aldrich or Acros Organics and used without further purification. The Boc-OMe-Tyr was purchased from Matrix Scientific and the 5-Bromo-1-pentyne was

procured from Enamine. Epoxy-activated Sepharose 6B was obtained from GE Healthcare. Plasmids were provided by the laboratory of Dr. Peter Schultz at The Scripps Research Institute. Reactions were conducted under ambient atmosphere with solvents directly from the manufacturer without further purification. All proteins were purified according to manufacturer's protocols using a Qiagen Ni-NTA Quik Spin Kit. Compound purities were assessed by NMR and found to be 90% or greater for all compounds. Fluorescence measurements were acquired on a Synergy HT microplate reader, with a 485 nm excitation filter and a 528 nm emission filter. All NMRs were acquired on an Agilent Technologies 400 MHz NMR. Compound purities were assessed by NMR and found to be 90% or greater for all compounds. MS analysis was conducted on a Thermo Finnigan LCQ Deca Quadropole Ion Trap via direct injection of samples at 100  $\mu$ M in a 1:1 H<sub>2</sub>O/MeOH solution.

Synthesis of p-Bromo-ethyl-Boc-OMe-Tyrosine (1) A solution of Boc-Tyrosine-OMe (0.500 g, 1 eq, 1.693 mmol) in acetone (5 mL) was prepared in a flame-dried vial. To this solution, potassium carbonate (0.70 g, 3 eq, 5.08 mmol) was added, followed by the addition of dibromoethane (292  $\mu$ L, 2 eq, 3.38 mmol). The reaction stirred overnight at 55 °C, then cooled to room temperature, and filtered into a round-bottom flask. The solvent was removed from the filtrate by rotatory evaporation. The solid was then purified by silica gel chromatography (gradient of 9:1 hexanes/ethyl acetate to 7:1 hexanes/ethyl acetate to 5:1 hexanes/ethyl acetate) and afforded the desired product as a clear oil (36 mg, 0.089 mmol, 5.3% yield). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  7.02 (d, J = 8 Hz, 2H), 6.82 (d, J = 8 Hz, 2H), 4.96 (d, J = 8 Hz, 1H), 4.53 (d, J = 8 Hz, 1H), 4.01 (t, J = 4 Hz, 2 H), 3.70 (s, 3 H), 3.51 (t, J = 8 Hz, 2 H), 2.94 (m, 2 H), 1.41 (s, 9 H)

Synthesis of p-Bromo-propyl-Boc-OMe-Tyrosine (2) A solution of Boc-Tyrosine-OMe

(0.50 g, 1 eq, 1.69 mmol) in acetone (5 mL) was prepared in a flame-dried vial. To this solution, potassium carbonate (0.70 g, 3 eq, 5.08 mmol) was added, followed by the addition of 1,3- dibromopropane (345  $\mu$ L, 2 eq, 3.38 mmol). The reaction stirred overnight at 55 °C, then cooled to room temperature, and filtered into a round-bottom flask. The solvent was removed from the filtrate by rotatory evaporation. The solid was then purified by silica gel chromatography (gradient of 9:1 hexanes/ethyl acetate to 7:1 hexanes/ethyl acetate to 5:1 hexanes/ethyl acetate) and afforded the desired product as a clear oil (18 mg, 0.283 mmol, 33.5% yield).  $^1$ H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  6.99 (d, J = 8 Hz, 2 H), 6.78 (d, J = 8 Hz, 2 H), 5.07 (d, J = 4 Hz, 1 H), 4.59 (d, J = 4 Hz, 1 H), 4.01 (t, J = 4 Hz, 2 H), 3.65 (s, 3 H), 3.53 (t, J = 4 Hz, 2 H), 2.97 (m, 2 H), 2.24 (quintet, J = 4 Hz, 2 H), 1.37 (s, 9 H)

Synthesis of p-Bromo-butyl-Boc-OMe-Tyrosine (3) A solution of Boc-Tyrosine-OMe

(0.50 g, 1 eq, 1.69 mmol) in acetone (5 mL) was prepared in a flame-dried vial. To this solution, potassium carbonate (0.70 g, 3 eq, 5.08 mmol) was added, followed by the addition of 1,4- dibromobutane (369  $\mu$ L, 2 eq, 3.38 mmol). The reaction stirred overnight at 55 °C, then cooled to room temperature, and filtered into a round-bottom flask. The solvent was removed from the filtrate by rotatory evaporation. The solid was then purified by silica gel chromatography (gradient of 9:1 hexanes/ethyl acetate to 7:1 hexanes/ethyl acetate to 5:1 hexanes/ethyl acetate) and afforded the desired product as a white crystal (342mg, 0.795 mmol, 46.9% yield).  $^1$ H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  7.01 (d, J = 8 Hz, 2 H), 6.79 (d, J = 8 Hz, 2 H), 4.97 (d, J = 8 Hz, 1 H), 4.52 (d, J = 8 Hz, 1 H),

3.95 (t,  $J = 4$  Hz, 2 H), 3.69 (s, 3 H), 3.47 (t,  $J = 4$  Hz, 2 H), 3.01 (m, 2 H), 2.04 (m, 2 H), 1.92 (m, 2 H), 1.40 (s, 9 H)

Synthesis of p-Azido-propyl-Boc-OMe-Tyrosine (5) Dry dimethylformamide (DMF, 5 mL) was added to a vial containing 2 (60 mg, 1 eq, 0.137 mmol). This solution was then transferred to a flame dried vial, and sodium azide (11 mg, 1.2 eq, 165 mmol) was added. The reaction stirred overnight at 65°C. The organic layer was quenched with brine and then extracted with methylene chloride (DCM, 10 mL x 3). The organic layer was back extracted with brine (10 mL). The organic layer was then dried with magnesium sulfate, filtered, and rotovapped. Column chromatography (silica gel, 1:1 hexanes ethyl acetate) afforded the desired product as a crystal (32 mg, 0.085 mmol, 61.7% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.03 (d,  $J = 12$  Hz, 2 H), 6.82 (d,  $J = 12$  Hz, 2 H), 4.96 (d,  $J = 8$  Hz, 1 H), 4.53 (d,  $J = 8$  Hz, 1 H), 4.01 (t,  $J = 4$  Hz, 2 H), 3.71 (s, 3 H), 3.46 (t,  $J = 8$  Hz, 2 H), 3.02 (m, 2H), 2.03 (quintet,  $J = 4$  Hz, 2 H), 1.41 (s, 9 H).  $^{13}\text{C}$  (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  172.4, 157.7, 155.1, 130.3, 128.2, 114.5, 79.9, 64.4, 54.5, 52.2, 48.2, 37.4, 28.8, 28.3.

Synthesis of p-Azido-butyl-Boc-OMe-Tyrosine (6) Tyrosine derivative 3 (252 mg, 1 eq, 0.59 mmol) was dissolved in a solution of DMF (12 mL) in a flame-dried vial. Sodium azide (76 mg, 2 eq, 1.17 mmol) was then added to the solution. The reaction was stirred overnight at 65 °C, then quenched brine and extracted with methylene chloride (DCM, 10 mL x 3). The organic layer was then back extracted with brine (10 mL). The organic layer was dried with magnesium sulfate, filtered, and rotovapped. Column chromatography (silica gel, 3:1 hexanes ethyl acetate) afforded the desired product (132

mg, 0.349 mmol, 57% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.03 (d,  $J = 8$  Hz, 2 H), 6.84 (d,  $J = 8$  Hz, 2 H), 4.97 (d,  $J = 8$  Hz, 1 H), 4.54 (d,  $J = 8$  Hz, 1 H), 3.95 (t,  $J = 8$  Hz, 2 H), 3.72 (s, 3 H), 3.36 (t,  $J = 8$  Hz, 2 H), 3.05 (m, 2 H), 1.80 (m, 4 H), 1.43 (s, 9 H).  $^{13}\text{C}$  (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  172.4, 157.9, 130.3, 114.5, 67.1, 54.5, 52.2, 51.1, 37.4, 28.3, 26.5, 25.7.

Synthesis of p-Alkynyl-ethyl-Boc-OMe-Tyrosine (7) To a flame dried vial, DMF (5 mL) and Boc-Tyrosine-OMe (295 mg, 1 eq, 1.00 mmol) was added without stirring followed by the slow addition of potassium tert-butoxide (337 mg, 3 eq, 3.00 mmol). The reaction immediately turned dark green. This mixture was allowed to stir at 100 °C for 10 mins, then 4-bromo-1-butyne was added (94  $\mu\text{L}$ , 133 mg, 1 eq, 1.00 mmol). The mixture immediately turned a lighter, more brownish green. The mixture was then capped and allowed to stir at 80 °C overnight. Following reaction, brine (10 mL) was added followed by the addition of ether (10 mL). The aqueous layer was then drained and the ether layer was washed with brine (10 mL x 3). The ether layer was collected and then the aqueous layer was back extracted with additional ether. The combined ether layers were then dried and rotovapped. Column chromatography (silica gel, 5:1 hexanes/ethyl acetate) yielded the desired product as an oil (23 mg, 0.066 mmol, 3.9% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  6.99 (d,  $J = 8$  Hz, 2 H), 6.75 (m, 2 H), 4.47 (m, 1 H), 4.45 (m, 1 H), 4.07 (t,  $J = 4$  Hz, 2 H), 3.70 (s, 3 H), 3.54 (m, 2 H), 3.31 (m, 2 H), 2.66 (td,  $J = 8$  Hz,  $J = 4$  Hz, 2 H), 2.05 (s, 1 H), 1.43 (s, 9 H).  $^{13}\text{C}$  (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  171.2, 155.0, 131.3, 130.2, 115.5, 115.3, 81.7, 58.84, 28.5, 27.7, 26.1, 25.5, 19.5.

Synthesis of p-Alkynyl-propyl-Boc-OMe-Tyrosine (8) Boc-Tyrosine-OMe (114 mg, 2 eq, 0.385 mmol) was added to a flame-dried vial. Cesium carbonate (254 mg, 3 eq, 0.578 mmol) was then added, followed by dry DMF (3 mL). This mixture was stirred at 100 °C for 20 mins. 5- Bromo-1-pentyne (20 µL, 1 eq, 0.193 mmol) was then added to the mixture, as well as a catalytic potassium iodide. The reaction was stirred overnight at 100 °C, then cooled to room temperature and washed with brine (10 mL) and ether (10 mL). The ether layer was then washed with brine (10 mL x 3). The brine layer was then back-extracted with ether (10 mL). The organic layers were combined, dried with magnesium sulfate, filtered, and rotovapped. Column chromatography (silica gel, 5:1 hexanes/ethyl acetate) to yield the desired product as a white crystal (22 mg, 0.061 mmol, 31.6% yield).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 7.02 (d, J = 12 Hz, 2 H), 6.82 (d, J = 12 Hz, 2 H), 4.95 (d, J = 8 Hz, 1 H), 4.53 (d, J = 8 Hz, 1 H), 4.03 (t, J = 4 Hz, 2 H), 3.71 (s, 3 H), 3.02 (m, J = 8 Hz, 1 H), 2.39 (t, J = 4 Hz, 2 H), 1.97 (m, J = 8 Hz, 2 H), 1.55 (s, 1 H), 1.41 (s, 9 H).

<sup>13</sup>C (400 MHz, CDCl<sub>3</sub>): δ 172.4, 157.9, 130.3, 127.9, 114.5, 83.5, 79.9, 68.8, 66.0, 54.5, 52.2, 37.4, 28.3, 28.2, 21.1, 15.1.

Synthesis of p-Alkynyl-butyl-Boc-OMe-Tyrosine (9) Ether (10 mL) was added to a flame- dried round bottom flask followed by pyridine (18 µL, 16.2 mg, 0.21 mmol) and 6-hexyne-1-ol (342 µL, 300 mg, 3 eq, 3.06 mmol) then heated to 40 °C. Phosphorous (III) bromide (97 µL, 276 mg, 1 eq, 1.02 mmol) was then slowly added through the reflux condenser, producing a white gas and precipitate in the process. The reaction was then refluxed for 2 hours, cooled in an ice bath and diluted with ether (25 mL). The organic mixture was then washed twice with a solution of saturated sodium bicarbonate (20 mL x

2) and a single wash with brine (20 mL). The organic layer was then dried and rotovapped, to yield a clear, pungent oil that was the 6-Bromo-hexyne product (295 mg).

A solution of Boc-Tyrosine-OMe (1.089g, 2 eq, 3.69 mmol) in acetone (1 mL) was then prepared in a flame-dried vial. The 6-Bromohexyne product (295 mg, 1 eq, 1.84 mmol) was then dissolved in 4 mL of acetone, and transferred to the vial containing the tyrosine solution. To this mixture, cesium carbonate was added (1.802g, 3 eq, 5.53 mmol). The reaction was stirred overnight at 60 °C then cooled to room temperature and filtered. The precipitate was washed with ethyl acetate and the collected filtrate was concentrated by rotatory evaporation. Column chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate to 5:1 hexanes/ethyl acetate to 1:1 hexanes/ethyl acetate) afforded the desired product as a white crystal (62 mg, 0.165 mmol, 9.0% yield).  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  7.01 (d,  $J = 8$  Hz, 2 H), 6.81 (d,  $J = 8$  Hz, 2 H), 4.95 (d,  $J = 8$  Hz, 1 H), 4.53 (d,  $J = 8$  Hz, 1 H), 3.95 (t,  $J = 8$  Hz, 2 H), 3.70 (s, 3 H), 3.02 (m, 2 H), 2.27 (td,  $J = 8$  Hz, 4 Hz, 2 H), 1.96 (t,  $J = 4$  Hz, 1 H), 1.90 (m, 2 H), 1.71 (m, 2 H), 1.41 (s, 9 H).  $^{13}\text{C}$  (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  172.4, 158.0, 130.7, 127.8, 114.5, 84.1, 68.6, 67.2, 52.2, 37.4, 28.3, 25.0, 18.1.

General Protocol for Removal of Methyl Protecting Group To deprotect the C-terminus of the unnatural amino acids, 500  $\mu\text{L}$  of dioxane and 500  $\mu\text{L}$  of 1M lithium hydroxide were added to a vial containing the unnatural amino acid on ice. The vial was then stirred at room temperature for 2 hours. The dioxane was removed by rotatory evaporation and 6M hydrochloric acid was added to dropwise until a pH of 4 was obtained. The product was then extracted with ethyl acetate into a vial, and the organic layer was dried with

magnesium sulfate and filtered and rotovapped.

General Protocol for Removal of tert-Butyloxy Protecting Group To the methyl-free unnatural amino acid, 1 mL of 2% trifluoroacetic acid (200  $\mu$ L trifluoroacetic acid into 10 mL of DCM) was added on ice to deprotect the N-terminus of the unnatural amino acid. The reaction was stirred at room temperature for 1 hour, and the DCM was then removed by rotatory evaporation, affording the final deprotected unnatural amino acid, typically in as a white crystalline solid.

Synthesis of p-Bromoethyltyrosine (p2BrY, 10) According to the general protocol outlined above, 1 was deprotected to yield a white crystalline powder (26 mg, 0.090 mmol, ~100% yield).  $^1$ H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.11 (d, J = 8 Hz, 2H), 6.82 (d, J = 8 Hz, 2H), 4.23 (t, J = 4 Hz, 2 H)

Synthesis of p-Bromopropyltyrosine (p3BrY, 11)<sup>1</sup> According to the general protocol outlined above, 2 was deprotected to yield a white crystalline powder (21 mg, 0.069 mmol, 24.5% yield).  $^1$ H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.10 (d, J = 8 Hz, 2 H), 6.81 (d, J = 8 Hz, 2 H), 3.57 (t, J = 4 Hz, 2 H), 2.23 (quintet, J = 4 Hz, 2 H). Values corroborated by literature (1) 1 Xiang, Z.; Lacey, V.; Ren, H.; Xu, J.; Burban, D.; Jennings, P.; Wang, L., Proximity-Enabled Protein Crosslinking through Genetically Encoding Haloalkane Unnatural Amino Acids. Angew Chem Int Ed 2014, 53, 2190-2193.

Synthesis of p-Bromobutyltyrosine (p4BrY, 12)<sup>1</sup> According to the general protocol outlined above, 3 was deprotected to yield a white crystalline powder (58 mg, 0.183

mmol, 23.0% yield).  $^1\text{H}$  NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.09 (d, J = 8 Hz, 2 H), 6.79 (d, J = 8 Hz, 2 H), 3.94 (t, J = 4 Hz, 2 H), 2.80 (t, J = 8 Hz, 1 H), 1.98 (m, 2 H), 1.86 (m, 2 H).

Values corroborated by (1)

Synthesis of p-Azidopropyltyrosine (p3AzY, 14)<sup>1</sup> According to the general protocol outlined above, 5 was deprotected to yield a white crystalline powder (21 mg, 0.08 mmol, 94.1% yield).  $^1\text{H}$  NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.10 (d, J = 8 Hz, 2 H), 6.81 (d, J = 8 Hz, 2 H), 4.90 (s, 1 H), 4.16 (t, J = 4 Hz, 2 H), 1.97 (m, J = 4 Hz, 2 H).  $^{13}\text{C}$  (400 MHz, CD<sub>3</sub>OD):  $\delta$  172.5, 156.1, 128.37, 127.7, 112.5, 110.0, 62.8, 26.9, 25.7; MS: calcd for C<sub>12</sub>H<sub>17</sub>N<sub>4</sub>O<sub>3</sub>: 265.293, found 265.290. Values corroborated by (1).

Synthesis of p-Azidobutyltyrosine (p4AzY, 15) According to the general protocol outlined above, 6 was deprotected to yield a white crystalline powder (28 mg, 0.100 mmol, 30.0% yield).  $^1\text{H}$  NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.09 (d, J = 12 Hz, 2 H), 6.79 (d, J = 12 Hz, 2 H), 3.93 (t, J = 4 Hz, 2 H), 1.99 (m, 2 H), 1.87 (m, 2 H).  $^{13}\text{C}$  (400 MHz, CD<sub>3</sub>OD):  $\delta$  172.4, 156.3, 154.8, 128.3, 127.5, 112.4, 77.5, 65.0, 31.1, 27.8, 26.1, 25.7; MS: calcd for C<sub>13</sub>H<sub>19</sub>N<sub>4</sub>O<sub>3</sub>: 279.145, found 279.142

Synthesis of p-Alkynylethyltyrosine (p2yneY, 16) According to the general protocol outlined above, 7 was deprotected to yield a white crystalline powder (15 mg, 0.064 mmol, 97% yield).  $^1\text{H}$  NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.00 (d, J = 4 Hz, 2 H), 6.67 (d, J = 4 Hz, 2 H), 4.25 (m, 2 H), 2.59 (td, J = 4 Hz, J = 8 Hz, 2 H), 2.27 (t, J = 4 Hz, 2 H).  $^{13}\text{C}$

(400 MHz, CD<sub>3</sub>OD): δ 172.5, 170.0, 128.2, 126.2, 113.1, 77.5, 58.5, 52.0, 49.9, 43.7, 41.7, 25.7, 17.8, 11.4; MS: calcd for C<sub>13</sub>H<sub>16</sub>NO<sub>3</sub>: 234.113, found 234.113

Synthesis of p-Alkynylpropyltyrosine (p3yneY, 17) According to the general protocol outlined above, 8 was deprotected to yield a white crystalline powder (13 mg, 0.053 mmol, 86.2% yield). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ 7.53 (d, J = 12 Hz, 2 H), 7.24 (d, J = 12 Hz, 2 H), 4.44 (t, J = 4 Hz, 2 H), 2.76 (td, J = 8 Hz, 4 Hz, 2 H), 2.63 (t, J = 4 Hz, 1 H), 2.34 (m, 2 H). <sup>13</sup>C (400 MHz, CD<sub>3</sub>OD): δ 172.4, 128.3, 112.4, 66.9, 64.5, 26.5, 25.7, 12.7; MS: calcd for C<sub>14</sub>H<sub>18</sub>NO<sub>3</sub>: 248.128, found 248.129

Synthesis of p-Alkynylbutyltyrosine (p4yneY, 18) According to the general protocol outlined above, 9 was deprotected to yield a white crystalline powder (15 mg, 0.057 mmol, 34.5% yield). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ 7.09 (d, J = 8 Hz, 2 H), 6.79 (d, J = 8 Hz, 2 H), 3.93 (t, J = 4 Hz, 2 H), 2.21 (td, J = 8 Hz, 4 Hz, 2 H), 2.18 (t, J = 4 Hz, 2 H), 1.83 (m, 2 H), 1.64 (m, 2 H). <sup>13</sup>C (400 MHz, CD<sub>3</sub>OD): δ 172.4, 156.3, 128.2, 127.4, 112.4, 81.7, 77.5, 66.7, 65.4, 26.3, 25.7, 23.3, 15.8; MS: calcd for C<sub>15</sub>H<sub>20</sub>NO<sub>3</sub>: 262.144, found 262.142

General GFP Expression A pET-GFP-TAG plasmid (0.5 μL) was co-transformed with a pEVOL-pCNF plasmid (0.5 μL) into Escherichia coli BL21(DE3) cells using an Eppendorf eporator. The cells were then plated and grown on LB agar in the presence of chloramphenicol (34 mg/mL) and ampicillin (50 mg/mL) at 37°C overnight. One colony was then used to inoculate LB media (4 mL) containing both ampicillin and

chloramphenicol. The culture was incubated at 37 °C overnight and used to inoculate an expression culture (10 mL LB media, 50 mg/mL Amp, 34 mg/mL Chlr) at an OD600 0.1. The cultures were incubated at 37 °C to an OD600 between 0.6 and 0.8, and protein expression was induced by addition of UAA (100 µL, 100 mM) and 20 % arabinose (10 µL) and 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; 10 µL). The cultures were incubated at 30 °C for 16-20 h then centrifuged at 5,000 rpm for 10 minutes and stored at -80 °C for 3 hours. The cell pellet was resuspended using 500 µL of Bugbuster (Novagen) containing lysozyme, and incubated at 37 °C for 20 minutes. The solution was transferred to an Eppendorf tube and centrifuged at 15,000 rpm for 10 minutes, then the supernatant was added to an equilibrated His- pur Ni-NTA spin (Qiagen) column and GFP was purified according to manufacturer's protocol. Purified GFP was analyzed by SDS-PAGE (BioRad 10% precast gels, 150V, 1.5h), and employed without further purification. Protein concentrations were determined both by a BCA assay and fluorescence measurements.

Biological Cu(I)-1,3-dipolar Cycloaddition Reaction Clicks on azido-tyrosine GFP-mutants were performed in 1.5mL Eppendorf tubes. To the experimental tube, in order, the following were added: 2 µL of 50mM CuSO<sub>4</sub>, 20 µL of 5mM Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA), 5 µL of azido-tyrosine mutated GFP (with either an ethyl, propyl, or butyl tether) in PBS, 10 µL of 488-Alexafluorophore-alkyne, 50 mM of tris(2- carboxyethyl)phosphine (TCEP), and 11 µL of phosphate buffer solution (PBS; 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>, 2.7 mM KCl, 137 mM NaCl, pH = 7.4) to bring the reaction to a 50 µL volume. Control reactions were performed with 20 µL azido-tyrosine mutated GFP samples (either ethyl, propyl, or butyl) and 10 µL 1 mM

488-Alexafluorophore in the absence of the Cu- 1,3-dipolar cycloaddition catalysts (CuSO<sub>4</sub>, TBTA, and TCEP) to ensure that successful reactions were not the result of non-specific interactions between the protein and fluorophore. The control reaction was brought to 50 µL volume using PBS. All mixtures were reacted for 10 hours at 4°C. The reactions were then transferred to concentrator column (10k MWCO, Corning Spin-X) pre-wet with PBS and spun at maximum speed in a tabletop centrifuge for 2 minutes. They were then washed with 100 µL portions of PBS and spun for another 2 minutes. This process was repeated for a total of seven washes, and the final spin time was adjusted to bring the samples to a concentration of ~25 µL.

Biological Glaser-Hay Coupling Glaser-Hay couplings on alkyne-tyrosine mutated GFP were performed in 1.5mL Eppendorf tubes. To the experimental tube, the following was added in order: 3 µL of 50mM CuI, 3 µL 50 mM trimethylethylenediamine (TMEDA), 10 µL of 1 mM 488-Alexafluorophore alkyne, 20 µL of alkyne-tyrosine mutated GFP (with either ethyl, propyl, or butyl tether) in PBS for a total reaction volume of 36 µL. Control reactions were performed with 6 µL PBS, 10 µL 488-Alexafluorophore alkyne, and 20 µL of alkyne-tyrosine mutated GFP (with either ethyl, propyl, or butyl tether) to ensure that successful reactions were not the result of non-specific interactions between the protein and fluorophore. All mixtures were reacted for 4 hours at 4°C. The reactions were then transferred to concentrator column (10k MWCO, Corning Spin-X) pre wet with PBS and spun at maximum speed in a tabletop centrifuge for 2 minutes. They were then washed with 100 µL portions of PBS and spun for another 2 minutes. This process was repeated for a total of seven washes, and the final spin time was adjusted to bring the samples to a concentration of ~25 µL (as indicated by the marking on the column).

### Immobilization of Alkynyl- or Amino- Functionality onto Epoxy Sepharose 6B Resin

Epoxy-activated 6B Sepharose (GE Healthcare, 200 mg) was added to a filter syringe and washed with distilled water (5 washes, 3 mL). Alkyn-ol (700  $\mu$ mol) or ethylenediamine (700  $\mu$ mol) and coupling buffer (3.5 mL, 1x PBS brought to pH 13.0 w/ NaOH) was added to a vial followed by the washed resin. The mixture was shaken at room temperature for 16 h. The resin was transferred to a filter syringe and washed with coupling buffer (4 mL). The sepharose was transferred to a vial and capped with ethanolamine (3.5 mL). The resin was incubated at 30 °C for four hours then washed in a filter syringe with 0.1 M acetate buffer (10 mM) and 0.1 M Tris- HCl buffer (pH 4 and pH 8 respectively, 3 alternating washes with 3 mL each).

To verify that we had successfully immobilized ethelyenediamine on the sepharose resin, amino-, alkynyl-, and unreacted sepharose resin were placed in carbonate buffer (0.2 M  $\text{Na}_2\text{CO}_3$ , 0.2M  $\text{NaHCO}_3$ , pH = 9.) in the presence of 5  $\mu$ L of FITC. After washing, imaging with a transilluminator demonstrated the successful derivitization of the sepharose resin with ethylenediamine as little fluorescence is observed on the alkynyl resin and practically no fluorescence was observed on the unreacted sepharose (see Figure 7).

*Figure 15. Validation that the sepharose resin was derivatized with an amine functional group using the FITC dye, which contains an isothiocyanate group susceptible to nucleophilic attack by an amine. Because the coupling buffer is generated using a hydroxide ion, we wanted to verify the propensity of reacted coupling buffer on the resin to react with the FITC dye (alkynyl-). As can be seen, the amino-resin is much more fluorescent than the alkynyl resin, due to the presence of the nucleophilic amine group. This verifies that we did indeed generate an amine resin for our immobilization studies.*

Protocol for Immobilization of Azido-Tyrosine Mutated GFP Series All azido-tyrosine

GFP mutants were brought to a final concentration of 0.250 mg/mL using PBS. To a 1.5 mL Eppendorf tube, 15  $\mu$ L of each azido-tyrosine GFP mutant was added (with either an ethyl, propyl, or butyl tether). To this, the Cu(I)-1,3-dipolar cycloaddition reaction conditions were then added. In the following order, 5  $\mu$ L of 4 mM TCEP, 6  $\mu$ L 1 mM CuSO<sub>4</sub>, 15  $\mu$ L PBS, and 20  $\mu$ L 5 mM TBTA were then added to the mix. This liquid mixture was then mixed and transferred to a 1.5 mL Eppendorf that contained 15 mg pre-weighed alkyne-derivatized sepharose resin. A positive control consisting of the conditions above using p-azidophenylalanine-GFP, which was previously demonstrated to immobilize on the alkyne-derivatized sepharose resin, was also prepared. All samples were allowed to react overnight at 4 °C, then transferred to spin columns by adding PBS (200  $\mu$ L x 2) to the reaction in order to aid the transfer. The reaction was then purified at 1.1 rcf on a table-top microcentrifuge for 1 minute, and washed two times with 100  $\mu$ L portions at 1.1 rcf for 1 minute each. The protein-immobilized sepharose resin was then transferred to a clear Eppendorf tube for visualization and fluorescence quantification on a microplate reader.

Protocol for Immobilization of Alkynyl-Tyrosine Mutated GFP Series All alkynyl-tyrosine GFP mutants were brought to a final concentration of 0.400 mg/mL using PBS. To a 1.5 mL Eppendorf tube, 5  $\mu$ L of 50 mM CuI, 5  $\mu$ L of 50 mM TMEDA and 10  $\mu$ L of PBS were added. Followed by 20  $\mu$ L of the alkynyl-GFP. This liquid mixture was then mixed and transferred to a 1.5 mL Eppendorf tube that contained 20 mg pre-weighed alkyne-derivatized sepharose resin. A positive control consisting of the conditions above using p-propargylphenylalanine-GFP, which was previously demonstrated to immobilize

on the alkyne-derivatized sepharose resin, was prepared. All samples were allowed to react at 4 °C for 4 hrs, then transferred to spin columns by adding PBS (200 µL x 2) to the reaction in order to aid the transfer. The reaction was then purified at 1.1 rcf on a table-top microcentrifuge for 1 minute, and washed two times with 100 µL portions at 1.1 rcf for 1 minute each. The protein-immobilized sepharose resin was then transferred to a clear Eppendorf tube for visualization and fluorescence quantification on a microplate reader.

Protocol for Immobilization of the Bromo-Tyrosine Mutated GFP Series To a 1.5 ml  
Eppendorf tube, 15 mg of amino-functionalized sepharose resin was added. To the sepharose resin, 10 µL of a carbonate buffer solution (pH = 9.5) and 40 µL of bromo-containing GFP in an imidazole elution buffer were added. A negative control consisting of the conditions above using wild type GFP was also prepared. The reactions were incubated overnight at 4 °C, then transferred to spin columns by adding PBS (200 µL x 2) to the reaction in order to aid the transfer. The reaction was then purified at 1.1 rcf on a table-top microcentrifuge for 1 minute, and washed two times with 100 µL portions at 1.1 rcf for 1 minute each. The protein- immobilized sepharose resin was then transferred to a clear Eppendorf tube for visualization and fluorescence quantification on a microplate reader.

## References

- (1) Schmid, A., Dordick, J., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature* 409, 258-268.

- (2) Comfort, D., Chhabra, S., Conners, S., Chou, C., Epting, K., Johnson, M., Jones, K., Sehgal, A., and Kelly, R. (2004) Strategic biocatalysis with hyperthermophilic enzymes. *Green Chem* 6, 459-465.
- (3) Montiel, C., Quintero, R., and Aburto, J. (2009) Petroleum biotechnology: Technology trends for the future. *AfJ Biotechnol* 8, 7228-7240.
- (4) Hutchins, B., Kazane, S., Staflin, K., Forsyth, J., Felding-Habermann, B., Smider, V., and Schultz, P. (2011) Selective Formation of Covalent Protein Heterodimers with an Unnatural Amino Acid. *Chem Biol* 18, 299-303.
- (5) Vo-Dinh, T., and Cullum, B. (2000) Biosensors and biochips: advances in biological and medical diagnostics. *J Anal Chem* 366, 540-51.
- (6) Steen Redeker, E., Ta, D. T., Cortens, D., Billen, B., Guedens, W., and Adriaensens, P. (2013) Protein engineering for directed immobilization. *Bioconjug Chem* 24, 1761-77.
- (7) Mateo, C., Palomo, J., Fernandez-Lorente, G., Guisan, J., and Fernandez-Lafuente, R. (2007) Improvement of enzyme activity, stability and selectivity via immobilization techniques. *Enz Microb Technol* 40, 1451-1463.
- (8) Elgren, T., Zadvornyy, O., Brecht, E., Douglas, T., Zorin, N., Maroney, M., and Peters, J. (2005) Immobilization of active hydrogenases by encapsulation in polymeric porous gels. *Nano Letters* 5, 2085-2087.
- (9) Sletten, E., and Bertozzi, C. (2009) Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew Chem Int Ed* 48, 6974-6998.
- (10) Hermanson, G. T. Bioconjugate techniques. 3rd edition. ed.; Academic Press: London, UK, 1996..

- (11) Dirksen, A., and Dawson, P. (2008) Rapid Oxime and Hydrazone Ligations with Aromatic Aldehydes for Biomolecular Labeling. *Bioconjug Chem* 19, 2543-2548.
- (12) Kolb, H., and Sharpless, K. (2003) The growing impact of click chemistry on drug discovery. *Drug Disc Today* 8, 1128-1137.
- (13) Baskin, J., and Bertozzi, C. (2007) Bioorthogonal click chemistry: Covalent labeling in living systems. *Combi Sci* 26, 1211-1219.
- (14) Chou, C., Uprety, R., Davis, L., Chin, J., and Deiters, A. (2011) Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem Sci* 2, 480-483.
- (15) Dibowski, H., and Schmidtchen, F. (1998) Bioconjugation of peptides by palladium-catalyzed C-C cross-coupling in water. *Angew Chem Int Ed* 37, 476-478.
- (16) Chalker, J., Wood, C., and Davis, B. (2009) A Convenient Catalyst for Aqueous and Protein Suzuki-Miyaura Cross-Coupling. *J Am Chem Soc* 131, 16346.
- (17) Li, N., Lim, R., Edwardraja, S., and Lin, Q. (2011) Copper-Free Sonogashira Cross-Coupling for Functionalization of Alkyne-Encoded Proteins in Aqueous Medium and in Bacterial Cells. *J Am Chem Soc* 133, 15316-15319.
- (18) Lampkowski, J. S., Villa, J. K., Young, T. S., and Young, D. D. (2015) Development and Optimization of Glaser-Hay Bioconjugations. *Angew Chem Int Ed* 54, 9343-9346.
- (19) Young, T. S., and Schultz, P. G. (2010) Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *J Biol Chem* 285, 11039-11044.
- (20) Mendel, D., Cornish, V., and Schultz, P. (1995) Site Directed Mutagenesis with an Expanded Genetic Code. *Ann Rev Biophys Biomol Struct* 24, 435-462.

- (21) Wang, L., and Schultz, P. G. (2004) Expanding the genetic code. *Angew Chem Int Ed* 44, 34-66.
- (22) Liu, C., Schultz, P., Kornberg, R., Raetz, C., Rothman, J., and Thorner, J. (2010) Adding New Chemistries to the Genetic Code. *Ann Rev Biochem*, Vol 79 79, 413-444.
- (23) Yoshimura, S. H., Khan, S., Ohno, S., Yokogawa, T., Nishikawa, K., Hosoya, T., Maruyama, H., Nakayama, Y., and Takeyasu, K. (2012) Site-specific attachment of a protein to a carbon nanotube end without loss of protein function. *Bioconjug Chem* 23, 1488-93.
- (24) Ikeda-Boku, A., Kondo, K., Ohno, S., Yoshida, E., Yokogawa, T., Hayashi, N., and Nishikawa, K. (2013) Protein fishing using magnetic nanobeads containing calmodulin site-specifically immobilized via an azido group. *J Biochem* 154, 159-65.
- (25) Smith, M. T., Wu, J. C., Varner, C. T., and Bundy, B. C. (2013) Enhanced protein stability through minimally invasive, direct, covalent, and site-specific immobilization. *Biotechnol Prog* 29, 247-54.
- (26) Seo, M. H., Han, J., Jin, Z., Lee, D. W., Park, H. S., and Kim, H. S. (2011) Controlled and oriented immobilization of protein by site-specific incorporation of unnatural amino acid. *Anal Chem* 83, 2841-5.
- (27) Eichelbaum, K., Winter, M., Diaz, M., Herzig, S., and Krijgsveld, J. (2012) Selective enrichment of newly synthesized proteins for quantitative secretome analysis. *Nat Biotechnol* 30, 984.
- (28) Raliski, B. K., Howard, C. A., and Young, D. D. (2014) Site-specific protein immobilization using unnatural amino acids. *Bioconjug Chem* 25, 1916-20.

- (29) Hutchins, B., Kazane, S., Staflin, K., Forsyth, J., Felding-Habermann, B., Schultz, P., and Smider, V. (2011) Site-Specific Coupling and Sterically Controlled Formation of Multimeric Antibody Fab Fragments with Unnatural Amino Acids. *J Mol Biol* 406, 595-603.
- (30) Young, D., Jockush, S., Turro, N., and Schultz, P. (2011) Synthetase polyspecificity as a tool to modulate protein function. *Bioorg Med Chem Lett* 21, 7502-7504.
- (31) Young, D., Young, T., Jahnz, M., Ahmad, I., Spraggon, G., and Schultz, P. (2011) An Evolved Aminoacyl-tRNA Synthetase with Atypical Polysubstrate Specificity. *Biochem* 50, 1894-1900.
- (32) Chen, X., Xiang, Z., Hu, Y., Lacey, V., Cang, H., and Wang, L. (2014) Genetically Encoding an Electrophilic Amino Acid for Protein Stapling and Covalent Binding to Native Receptors. *ACS Chem Biol* 9, 1956-1961.
- (33) Xiang, Z., Lacey, V., Ren, H., Xu, J., Burban, D., Jennings, P., and Wang, L. (2014) Proximity-Enabled Protein Crosslinking through Genetically Encoding Haloalkane Unnatural Amino Acids. *Angew Chem Int Ed* 53, 2190-2193.
- (34) Brustad, E., Lemke, E., Schultz, P., and Deniz, A. (2008) A General and Efficient Method for the Site-Specific Dual-Labeling of Proteins for Single Molecule Fluorescence Resonance Energy Transfer. *J Am Chem Soc* 130, 17664.

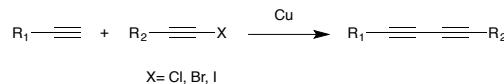
**CHAPTER 3: EXPANDING THE SCOPE OF ALKYNE-MEDIATED  
BIOCONJUGATIONS THROUGH A BIOLOGICAL CADIOT-CHODKIEWICZ  
REACTION**

The Glaser-Hay coupling of two terminal alkynes in the presence of CuI and TMEDA affords a diyne.<sup>29, 30</sup> The reaction brings together two terminal alkynes to form a diyne linkage; however, is not redox neutral and has chemoselectivity issues. These issues arise as there is not a means to differentiate the two terminal alkynes, resulting in 3 possible products. Recently, we demonstrated a bioorthogonal variant of the Glaser-Hay reaction that can be conducted in an aqueous setting under physiological conditions.<sup>22</sup> Using a terminal alkyne-containing UAA, we were able to demonstrate that the reaction proceeds to completion within approximately 6 hrs at 4 °C. As previously mentioned, the Glaser-Hay coupling of terminal alkynes has a chemoselectivity issue when the terminal alkynes differ, resulting in the formation of unwanted homodimers. However, due to the steric bulk of the protein, we found that the homodimerization of the reaction was mostly inhibited, leading to primarily the desired protein heterodimer product. As such we were able to demonstrate that the Glaser-Hay reaction could be employed as a novel bioorthogonal chemistry, yielding stable conjugates with well-defined geometries. However, the reaction was limited by the oxidative damage of the protein, due to the mechanistic cycling of the copper

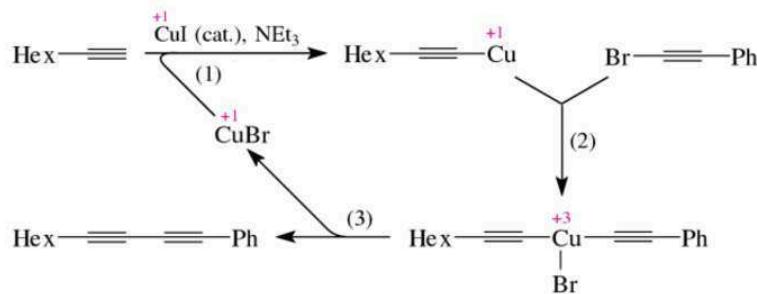
through three different redox states. As a result, reactions proceeding for longer than 6 hrs resulted in oxidative damage and protein degradation.

An alternative to the Glaser-Hay coupling is the Cadiot-Chodkiewicz reaction, which affords a conjugated diyne via the reaction of a terminal alkyne and a haloalkyne (Figure 1). In the presence of a copper(I) salt and a monodentate nitrogenous ligand, usually triethylamine (TEA), the reaction proceeds to form a covalent linkage in the form of a conjugated diyne in a relatively chemoselective fashion.<sup>23, 24</sup> Furthermore, the overall reaction is net redox neutral, as a single copper(I) catalyst goes through a series of oxidative additions and reductive eliminations with the bromoalkyne and terminal alkyne reactant to yield the conjugated diyne.<sup>25</sup>

### A



### B



*Figure 16. The Cadiot-Chodkiewicz reaction. A) Overall reaction scheme. B) Proposed reaction mechanism for the reaction. Adapted from Organic Mechanisms: Reactions, Stereochemistry, and Synthesis by Reinhard Bruckner.*

The Cadiot-Chodkiewicz reaction has diverse applications to many areas of chemistry. It can be utilized in polymerization reactions, such as in the formation

of the backbone of a solid-state polymer crystal, or in the fabrication of polymerized monolayer assemblies.<sup>26, 27</sup> Additionally, several acetylenic natural products exhibiting valuable biological properties can be obtained via a Cadiot-Chodkiewicz reaction.<sup>28</sup>

A key component of the Glaser-Hay mechanism is the formation of a copper(II)-hydroxyl intermediate, which has been shown to produce hydroxyl radicals that are deleterious to living systems (Figure 2).<sup>33</sup> Because the copper (I) of the Cadiot-Chodkiewicz reaction is not thought to utilize copper(II) intermediates, we reasoned that the chemistry could be employed in a biological context to minimize previously observed oxidative damage. Furthermore, the reaction is more chemoselective, as the use of a halo-alkyne minimizes the formation of homodimer side products by differentiating the two alkynes. Ultimately, this has the potential to increase the yield of the conjugated protein product. Based on these facts and the limitations of the bioorthogonal Glaser-Hay, we sought to develop a bioorthogonal variant of the Cadiot-Chodkiewicz reaction that could be conducted in an aqueous setting and under physiological conditions.

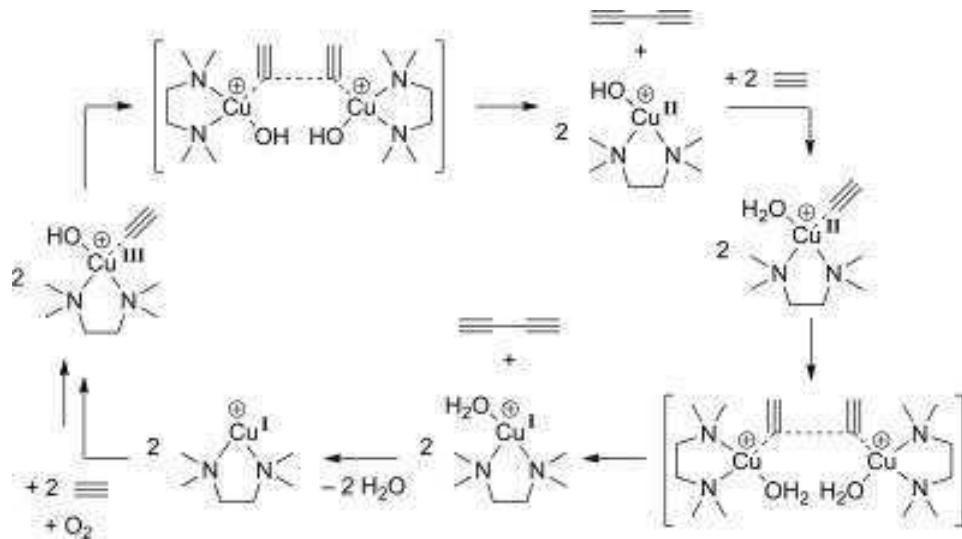
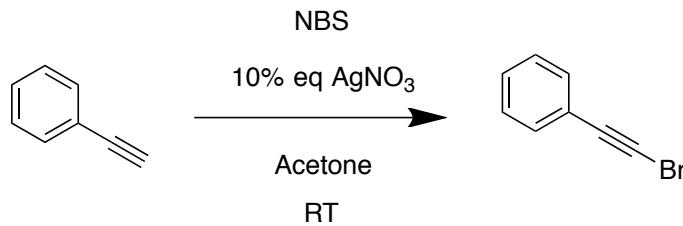


Figure 17. Proposed reaction intermediates in the Glaser-Hay coupling reaction. Adapted from (31).

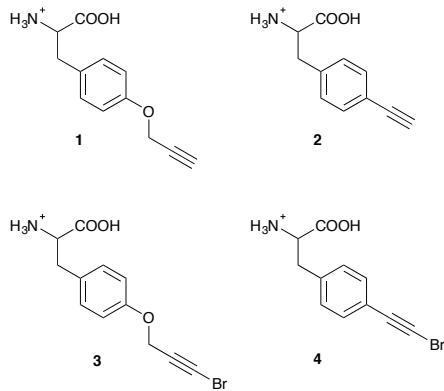
## Results and Discussion

In order to conduct and optimize a Cadiot-Chodkiewicz bioconjugation, new UAAs harboring a terminal haloalkyne needed to be synthesized and incorporated into a protein. In order to probe UAA-dependent effects on the reaction, aliphatic and aromatic brominated alkynyl UAAs were prepared from the previously reported protected *p*-propargyloxyphenylalanine (*p*PrF, **1**) and the *p*-ethynylphenylalanine (*p*EtF, **2**) respectively.<sup>34, 35</sup> Gratifyingly, the well-established bromination of phenylacetylene using N-bromosuccinamide (NBS) and silver nitrate worked well for the synthesis of both UAAs in moderate yields.<sup>36, 37</sup>

Scheme 8.



Following deprotection, the final UAAs, *p*-bromo-propargyloxyphenylalanine (*p*BrPrF, **3**) and *p*-bromo-ethynylphenylalanine (*p*BrEtF, **4**), were recovered in overall good yields (67% and 34% respectively) (Figure 1).

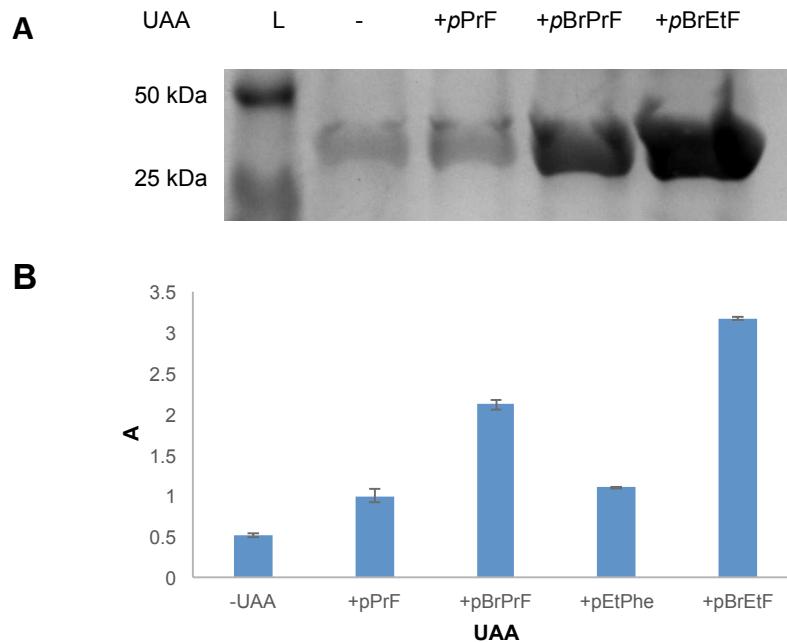


*Figure 18. p*-Propargyloxyphenylalanine (1, *p*PrF), *p*-ethynylphenylalanine (2, *p*EtF), *p*-bromopropargyloxyphenylalanine (3, *p*BrPrF), *p*-bromoethynylphenylalanine (4, *p*BrEtF) incorporated into GFP-151 for Glaser-Hay and Cadiot-Chodkiewicz reactions.

With the *p*BrPrF and *p*BrEtF in hand, it was imperative to incorporate these UAAs into a model protein. Due to both its fluorescent properties and well-documented prior use in UAA development technologies, green fluorescent protein (GFP) was selected as a model system. Specifically, attempts were made to incorporate newly synthesized UAAs at residue 151 by suppressing the amber stop codon. Furthermore, our previous work immobilizing GFP revealed that this surface exposed site is ideal for UAA placement, as the rigidity of the residue helps orient the bioorthogonal functional handle.<sup>38</sup> In lieu of undergoing a tedious aaRS selection process, we hoped to incorporate the brominated UAAs using the previously described promiscuous *p*CNF aaRS.<sup>39, 40</sup> The *p*CNF aaRS was investigated first to incorporate *p*BrPrF and *p*BrEtF due to their structural similarity to other UAAs that the *p*CNF aaRS incorporates.

BL21(DE3) *E. coli* were co-transformed with a pEVOL-*p*CNF plasmid and a pET-GFP-TAG151 plasmid and used to initiate an expression culture at OD<sub>600</sub> =0.1 which was grown to an OD<sub>600</sub> =0.7. The culture was subsequently centrifuged and the cell pellet was resuspended in 4 mL of LB broth supplemented with antibiotics, IPTG, arabinose, and the presence or absence of a UAA.<sup>34</sup> This previously reported expression protocol allowed for the minimization of the amount of UAA employed, and was found to be very effective. After 18-20 h at 30 °C, cells were pelleted and the expressed GFP was purified.

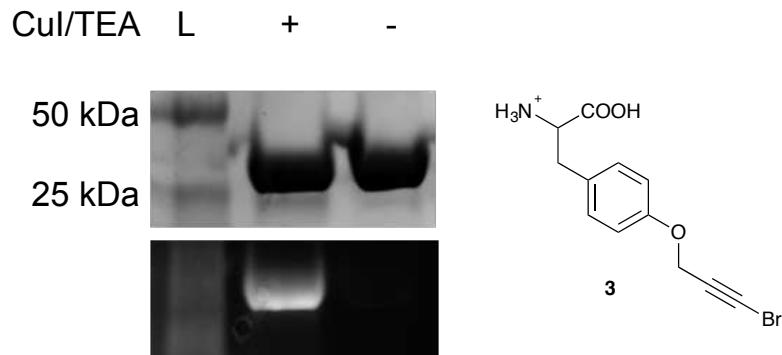
Gratifyingly, the promiscuous *p*CNF aaRS incorporated both brominated-UAA variants with a higher fidelity than the simple terminal alkyne analogs (Figure 2). As is to be expected, the smaller *p*BrEtF UAA, had a higher incorporation than the *p*BrPrF. We hypothesize that the 1-bromo-alkyne moiety provides a degree of hydrophobic character to the UAA, making the interaction between the amino acid and the hydrophobic binding pocket of the aaRS more favourable.



*Figure 19. Expression of UAA-containing GFP151 using the promiscuous pCNF aaRS. A) A stained SDS-PAGE gel indicated successful incorporation of **3** and **4** over background (-). A positive control, **1**, was also utilized in an expression. B) Data for overall incorporation of UAAs via the pCNF aaRS as measured via absorbance at 280 nm on a nanodrop spectrophotometer ( $\epsilon/1000 = 20$ , MW = 26.80 kDa). Data obtained were normalized to the pPrF UAA (1) to demonstrate the difference in incorporation efficiency over the previously reported UAA.*

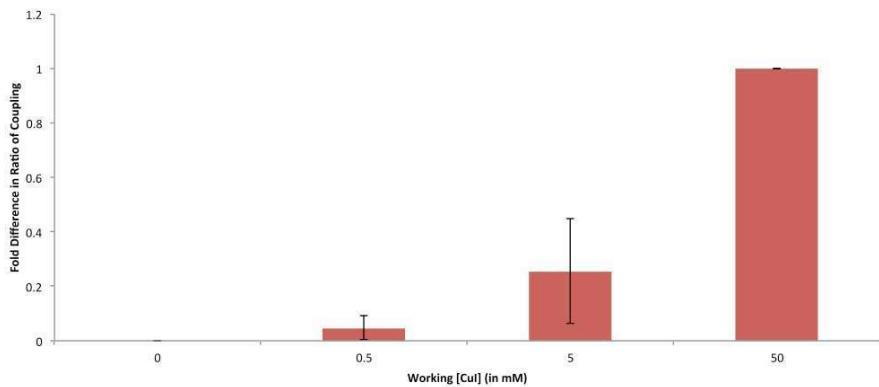
With both brominated alkyne UAAs in hand, it was feasible to develop a bioorthogonal Cadiot-Chodkiewicz reaction. Initial studies employed the *pBrPrF*-containing GFP variant, to mimic previous Glaser-Hay *pPrF*-GFP studies and provide an effective comparison. The Cadiot-Chodkiewicz reaction was carried out in PBS (pH = 7) using copper iodide and triethylamine (TEA) (both at a final concentration of 5 mM) in the presence of a terminal alkyne-containing fluorophore (AlexaFluor 488 alkyne) at 4 °C for 6 hrs. The reaction was successful, as fluorescence could be detected on a denatured SDS-PAGE gel only when protein and fluorophore were exposed to the CuI/TEA system (Figure 5). Even more exciting was the minimal protein degradation relative to the previously

reported Glaser-Hay reaction. Also, due to the chemoselective nature of the reaction no protein dimerization was detected, and fluorophore dimerization was minimal and easily removed.



*Figure 20.. Bioorthogonal Cadiot-Chodkiewicz reaction. The reaction performed on GFP (0.2 mg/mL) and AlexaFluor 488 (0.2 mM) in the presence of 5 mM of Cul and 5 mM TEA at 4 °C,. Fluorescence is only observed in the presence of the Cul/TEA system.*

In an attempt to further optimize the reaction, both copper concentrations and temperatures were varied. A 5 mM working Cu(I) concentration was found to be ideal, which represents a marked improvement over the ~50 mM concentrations required for the Glaser-Hay reaction (see Figure 6).



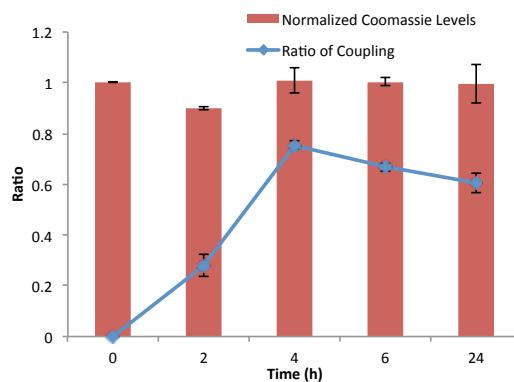
*Figure 21. Copper(I) salt optimization for the Cadiot-Chodkiewicz reaction. Data are normalized to the best performing working concentration (50 mM).*

Previous reports have indicated that *in vivo* use of copper-mediated bioorthogonal chemistries required working concentrations of near 0.1 mM of copper(I) salt to minimize cytotoxicity.<sup>10, 41</sup> Thus, the minimized copper concentrations help bring - bioorthogonal conjugated diyne chemistry into the range of *in vivo* use. These copper concentrations also had no impact on GFP fluorescence as determined by control reactions.

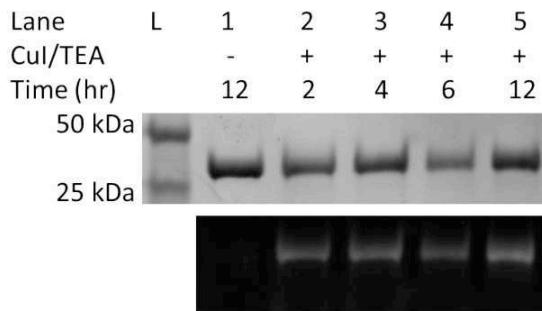
Additionally, the optimal temperature profile for the reaction was also investigated. To do so, reactions were performed over a time course at both 37 °C and 4 °C by starting with the 24 hr reaction and ending with the 2 hour reaction before running a denatured SDS-PAGE. We found that it was not necessary to perform any washing steps, as the free fluorophore was run beyond the length of the gel. This allowed us to visualize protein bound to fluorophore, and upon staining with Coomassie blue, to image protein degradation relative to a zero time point control.

To analyze our data, we had to develop a meaningful interpretation that could illustrate the interplay between conjugation efficiency and protein degradation. As protein degradation increased, our Coomassie signal decreased, artificially inflating the amount of conjugation observed (based on fluorescent signal alone). To overcome this, fluorescent signals were first corrected to the zero time point, in which free fluorophore was eliminated from the gel. This allowed us to correct for higher fluorescent signals based on the non-covalent dye interactions to the gel and not the protein. To correct for protein degradation, we normalized our Coomassie signals to the zero time point in which no degradation could have occurred. These corrected fluorescent and Coomassie signals were able to be used to gauge relative conjugation efficiency, taking into account increased protein degradation.

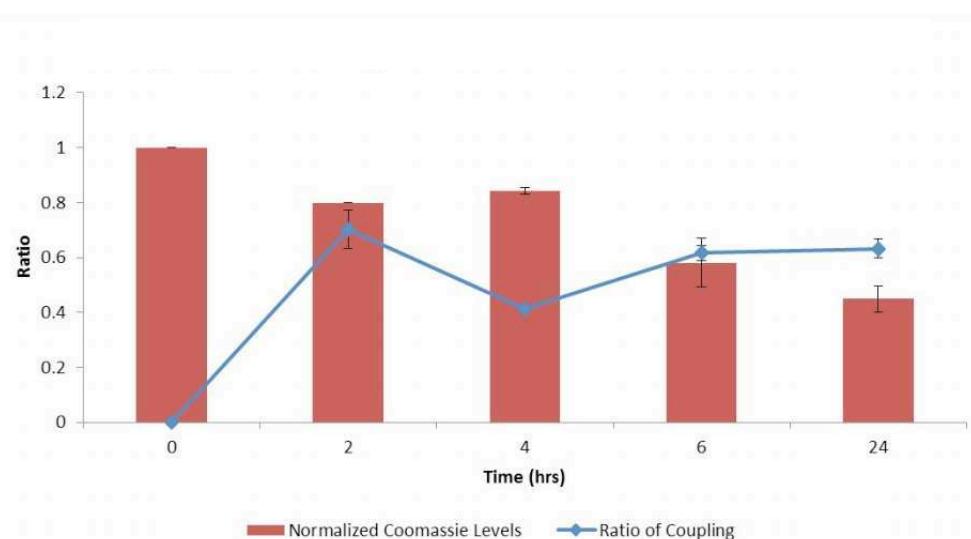
After performing a time course of the reaction at both 37 °C and 4 °C, we were able to determine very little difference in degradation between either temperature at early time-points. However, as the reaction was extended to 24 h, greater protein degradation at 37 °C occurred, most likely due to an increase in the rate of disproportionation of the Cu(I) catalyst at this temperature, producing a reactive copper(II) species (Figures 7-9).



*Figure 22. Bioorthogonal Cadiot-Chodkiewicz reaction Reaction profile at 4 °C over a 24 hr time period. Following analysis via SDS-PAGE, fluorescent imaging, and staining with Coomassie blue, protein levels were normalized to the 0 time point control. The protein levels indicate minimal to no protein degradation occurs, even at longer time points. Analysis of coupling efficiency was determined by calculating the ratio of fluorescence to Coomassie staining for each time point. All reactions were performed in triplicate.*



*Figure 23. SDS-PAGE of Cadiot-Chodkiewicz time course with pBrPrF containing GFP and at 4 °C. As can be seen, conjugation efficiency increases over time.*



*Figure 24. . Bioorthogonal Cadiot-Chodkiewicz reaction Reaction profile at 37 °C over a 24 hr time period. Following analysis via SDS-PAGE, fluorescent imaging, and staining with Coomassie blue, protein levels were normalized to the 0 time point control. The protein levels indicate minimal to no protein degradation occurs, even at longer time points. Analysis of coupling efficiency was determined by calculating the ratio of fluorescence to Coomassie staining for each time point. All reactions were performed in triplicate.*

However, for the 4 °C temperature profile the reaction reached approximately 86% completion in 4h with minimal protein degradation, indicating that the bioorthogonal Cadiot-Chodkiewicz reaction can be performed quickly and in a relatively mild conditions. Extended times and temperatures resulted in higher yields, however were accompanied by protein degradation.

We next sought to explore the effects of an aromatic variant of the pBrPrF. As such, *p*BrEtF-GFP151 was expressed, and subjected to identical coupling conditions at 4 °C in the presence of an alkyne fluorophore. Once again a successful conjugation was observed as determined by SDS-PAGE (Figure 10).

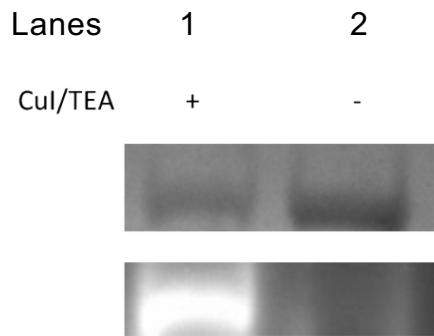
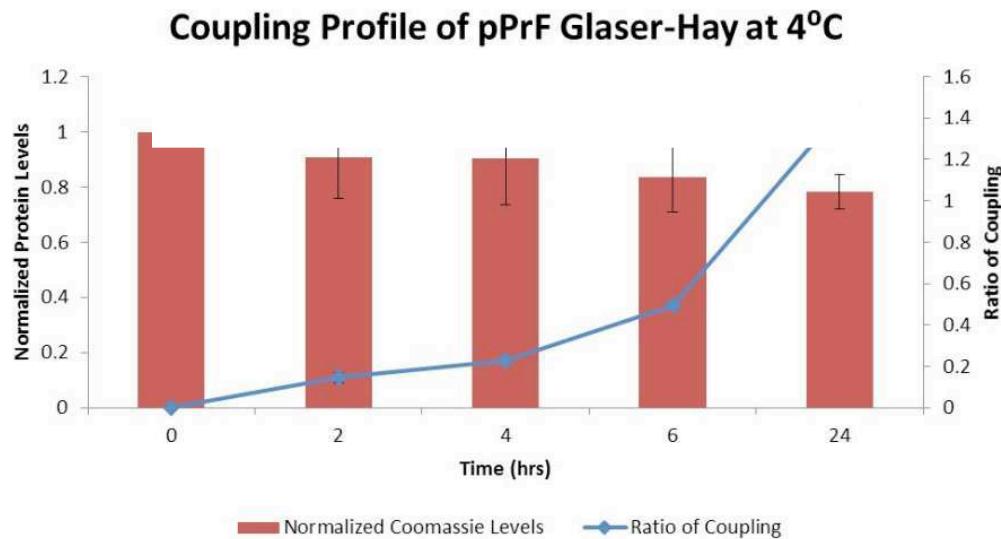


Figure 25. Bioorthogonal Cadiot-Chodkiewicz reaction using the aromatic *p*BrEtF UAA containing GFP. The reaction performed on GFP (0.2 mg/mL) and AlexaFluor 488 (0.2 mM) in the presence of 5 mM of CuI and 5 mM TEA at 4 °C overnight,. Fluorescence is only observed in the presence of the CuI/TEA system.

Only samples exposed to the CuI/TEA system exhibited fluorescence while other controls did not, indicating the fluorescence was not due to non-specific interactions. Interestingly, the use of an aromatic containing bromoalkyne appears to be less effective in the Cadiot-Chodkiewicz reaction than its aliphatic analog.

Next, we became interested in exploring how the novel biological Cadiot-Chodkiewicz conjugation compared to our previously described Glaser-Hay reactions. In direct comparison, the Cadiot-Chodkiewicz exhibited far less protein

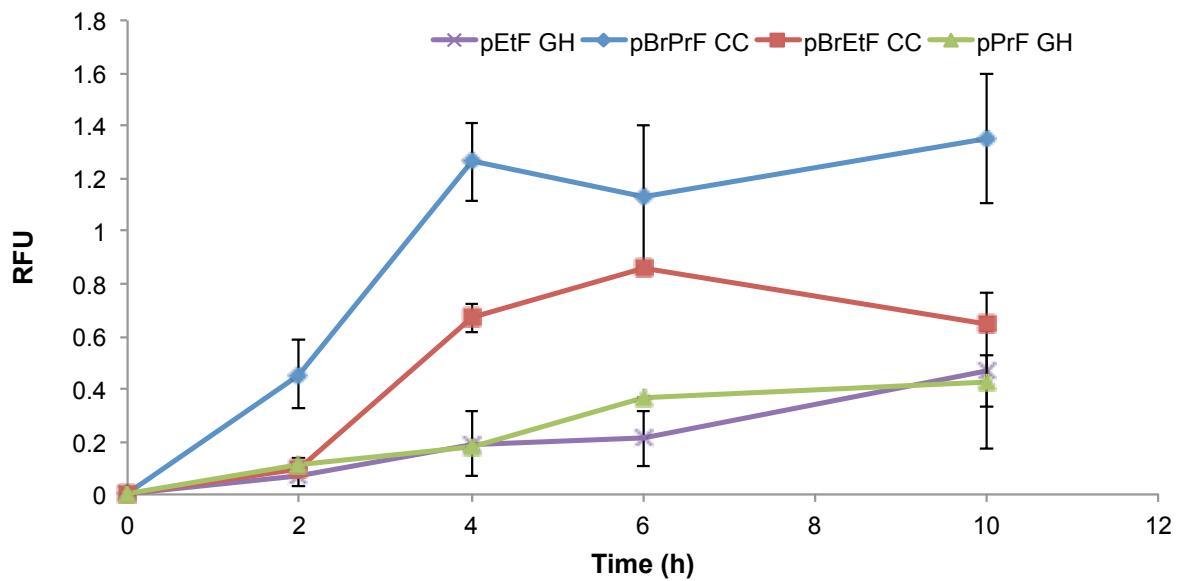
degradation as compared to the Glaser-Hay versions of either an aliphatic (*p*PrF) or aromatic (*p*EtF) terminal alkyne containing UAAs.



*Figure 26. Coupling profile of the pPrF Glaser-Hay at 4°C. Conjugation yields drastically increase after 6 hr, but are accompanied by a significant degree of protein degradation.*

Furthermore, the data indicates that the biological Cadiot-Chodkiewicz reaction proceeds at a faster rate than the Glaser-Hay, with the aliphatic version (*p*BrPrF) of the coupling reaching completion the fastest in 4 hrs (Figure 12). Gratifyingly, these results correlate well with the mechanistic understanding of both reactions, as the Cadiot-Chodkiewicz requires a single copper atom, while the Glaser-Hay necessitates two copper-alkyne conjugates to form the diyne product. Moreover, the Cu(I)/(III) redox couple of the Cadiot-Chodkiewicz reaction most likely aids in the minimized protein oxidation relative to the Glaser-Hay coupling that involves a reactive Cu(II) intermediate. However, it is important to note that the Cadiot-Chodkiewicz requires brominated UAAs, requiring additional synthetic preparation. In comparison to other bioconjugation techniques, the Cadiot-

Chodkiewicz reaction may be slower; however, it employs more synthetically accessible UAAs and results in a well-defined linear geometry primed for further reactions. Thus, the selection of the Cadiot-Chodkiewicz reaction may be dependent on the downstream application and available resources.



*Figure 27. Comparison of alkyne reactions. Reactions were conducted either under the described Cadiot-Chodkiewicz conditions or the Glaser-Hay conditions at 4 °C depending on incorporated UAA. Due to differences in protein levels resulting from oxidative damage, the ratio of fluorophore coupling for each data set was normalized to the 24 hour time point in order to compare coupling trends between reactions.*

## Conclusion

Overall, we have demonstrated the successful application of the Cadiot-Chodkiewicz reaction to a biological context. Furthermore, we have demonstrated that the reaction can be performed with minimal protein oxidation. Finally, we have demonstrated that the Cadiot-Chodkiewicz variant requires less harsh copper(I) concentrations, bringing the reaction near the range for *in vivo* use.

Future work will involve optimization of conditions to increase the compatibility of the reaction with biological systems, and extension of the reaction towards *in vivo* applications.

## Experimental

General All solvents and reagents, including the AlexaFluor 488 Alkyne, were obtained from Sigma Aldrich or Acros Organics and used without further purification. Plasmids were provided by the laboratory of Dr. Peter Schultz at The Scripps Research Institute. Reactions were conducted under ambient atmosphere with solvents directly from the manufacturer without further purification. All proteins were purified according to manufacturer's protocols using a Qiagen NiNTA Quik Spin Kit. Compound purities were assessed by NMR and found to be 90% or greater for all compounds. All NMRs were acquired on an Agilent Technologies 400 MHz NMR. MS analysis was conducted on a Thermo Finnigan LCQ Deca Quadropole Ion Trap via direct injection of samples at 100  $\mu\text{M}$  in a 1:1 H<sub>2</sub>O/MeOH solution. Unless indicated otherwise, all solutions were prepared in deionized water (pH ~ 7).

Synthesis of p-bromopropargyloxyphenylalanine (pBrPrF) p-Propargyloxyphenylalanine (pPrF, 0.050g, 0.150 mmol) was synthesized according to the literature<sup>1</sup>, and dissolved in acetone (5 mL). This solution was then transferred to a vial containing AgNO<sub>3</sub> (0.03g, 0.015 mmol) and N-bromosuccinamide (NBS, 0.030g, 0.165 mmol). The mixture was allowed to stir for 3 hours at room temperature, after which it was diluted with hexanes (10 mL) and the resulting crystals were filtered off. The filtrate was then evaporated under reduced pressure and purified via flash chromatography (5:1 hexanes/ethyl acetate)

yielding the desired product as a white crystal (0.042 g, 0.102 mmol, 68% yield). The product was then subjected to 1M LiOH in dioxanes (500  $\mu$ L each) on ice, and then stirred for two hours at room temperature. The dioxanes were subsequently removed by rotatory evaporation. Additional deionized water (1 mL) was then added to the product, and this solution was then brought to a pH = 4 with 6 N HCl. The solution was extracted in ethyl acetate, dried with MgSO<sub>4</sub>, filtered, and evaporated under pressure to dryness. A 1 mL solution 2% trifluoroacetic acid (TFA) in dimethylene chloride (DCM) was then added to the vial, and the mixture was allowed to stir at room temperature for 1 hour, prior to evaporation under reduced pressure to yield the final product as a yellowish crystal (0.030g, 0.101 mmol, 99% yield, 67% overall yield). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  7.19 (d, J = 8 Hz, 2H), 6.93 (d, J = 8 Hz, 2H), 4.91 (s, 1 H), 4.14 (m 1 H), 3.06 (m, 2 H). <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD):  $\delta$  168.4, 155.7, 128.6, 125.4, 113.5, 73.4, 54.4, 52.3, 45.4, 25.4; MS: calcd for C<sub>12</sub>H<sub>13</sub>BrNO<sub>3</sub>: 299.14, found 298.1 and 300.1.

|

Synthesis of p-bromoethynylphenylalanine (pBrEtF) p-ethynylphenylalanine (pEtF, 0.052g, 0.171 mmol) was synthesized according to the literature protocol<sup>2</sup>, and dissolved in acetone (5 mL). This solution was then transferred to a vial containing AgNO<sub>3</sub> (~0.03g, 0.0171 mmol) and N-bromosuccinamide (NBS, 0.034g, 0.189 mmol). The mixture was allowed to stir for 3 hours at room temperature, after which it was diluted with hexanes (10 mL) and the resulting crystals were filtered off. The filtrate was then evaporated under reduced pressure and purified via flash chromatography (3:1 hexanes/ethyl acetate) yielding the desired product as a yellow crystal (0.024 g, 0.063 mmol, 38% yield). The product was then subjected to 1M LiOH in dioxanes (500  $\mu$ L

each) on ice, and then stirred for two hours at room temperature. The dioxanes were subsequently removed by rotatory evaporation. Additional deionized water (1 mL) was then added to the product, and this solution was then brought to a pH = 4 with 6 N HCl. The solution was extracted in ethyl acetate, dried with MgSO<sub>4</sub>, filtered, and evaporated under pressure to dryness. A 1 mL solution 2% trifluoroacetic acid (TFA) in dimethylene chloride (DCM) was then added to the vial, and the mixture was allowed to stir at room temperature for 1 hour, prior to evaporation under reduced pressure to yield the final product as a yellowish crystal (0.015g, 0.056 mmol, 89% yield, 34% overall yield). <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ 7.35 (d, J = 8 Hz, 2H), 7.21 (d, J = 8 Hz, 2H), 4.34 (m, 1H), 3.16 (m, 2H), 2.90 (m, 2H). <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD): δ 173.63, 138.42, 132.16, 131.4, 129.2, 79.2, 54.6, 37.2, 27.22; MS: calcd for C<sub>11</sub>H<sub>11</sub>BrNO<sub>2</sub>: 269.12, found 268.1 and 270.1.

Expression of UAA-containing GFP-151 Escherichia coli BL21(DE3) cells were co-transformed with a pET-GFP-TAG-151 plasmid (0.5 μL) and a pEVOL-pCNF plasmid (0.5 μL) using an Eppendorf electroporator. Cells were then plated on LB-agar plates supplemented with ampicillin (50 mg/mL) and chloramphenicol (34 mg/mL) and grown at 37°C. After 16 h, a single colony was selected and used to inoculate LB media (4 mL) supplemented with ampicillin and chloramphenicol. The culture was grown to confluence at 37°C over 16 h. This culture was then used to initiate an expression culture in LB media (20 mL) at OD<sub>600</sub> = 0.1, then incubated at 37°C until an OD<sub>600</sub> ~0.7 was obtained. At this point, the cells were then pelleted at 4°C at 5000rpm for 10 mins. The supernatant was removed and the cells were gently resuspended in 4 mL of LB media.

Ampicillin and chloramphenicol (4  $\mu$ L each), 1M IPTG (4  $\mu$ L), 20% arabinose (4  $\mu$ L), and 100mM of the UAA (40  $\mu$ L) were then added. Induced cells were grown for an additional 16 h at 37°C, then harvested via centrifugation (10 min at 5000 rpm). The media was decanted and the cell pellet was placed in the -80°C freezer for 20 min. Purification of the UAA-containing GFP was then performed using commercially available Ni-NTA spin columns according to the manufacturer's protocol. Protein yield and purity was then assessed by SDS-PAGE and spectrophotometrically by Nanodrop spectrophotometry. Protein was then transferred into phosphate-buffered saline solution (PBS, pH = 7) using 10k MWCO spin columns prior to use in the biological Cadiot-Chodkiewicz coupling.

General Protocol for the Biological Cadiot-Chodkiewicz Coupling To a sterile 1.5 mL Eppendorf tube, the following was added in order: 10  $\mu$ L of bromo-alkyne UAA-containing GFP151 (~0.5 mg/mL), 5  $\mu$ L of a vigorously shaken CuI solution (50 mM), 2  $\mu$ L of trimethylamine (TEA, 50 mM), 5  $\mu$ L of 488-Alexafluor alkyne (1 mM in DMSO), and 3  $\mu$ L of PBS (pH = 7). Control reactions were set up in the absence of CuI and TEA and with the addition of a total of 10  $\mu$ L of PBS. Reactions were performed at various temperatures and times as indicated in the main text, with the 4°C for 6 hr variant being the ideal reaction conditions. Denatured SDS-PAGE were run following reaction, without cleaning or condensing the reaction, as it was found that the fluorophore would run off the gell and stil allow for adequate imaging.

General Protocol for the Biological Glaser-Hay Coupling To a sterile 1.5 mL Eppendorf tube, the following was added in order: 10 µL of terminal alkyne UAA-containing GFP151 (~0.5 mg/mL), 5 µL of a vigorously shaken CuI solution (500 mM), 2 µL of trimethylethylenediamine (TMEDA, 500 mM), 5 µL of 488-Alexafluor alkyne (1 mM in DMSO), and 3 µL of PBS (pH = 7). Control reactions were set up in the absence of CuI and TMEDA and with the addition of a total of 10 µL of PBS. Reactions were performed at 4°C for various times as indicated in the main manuscript. Denatured SDS-PAGE were run following reaction, without cleaning or condensing the reaction, as it was found that the fluorophore would run off the gel and still allow for adequate imaging.

## References

- 1 G. T. Hermanson, *Bioconjugate Techniques*, 3<sup>rd</sup> Ed., Academic Press: London, 1996.
- 2 K. Lang and J. Chin, *Chem Rev*, 2014, **114**, 4764-4806.
- 3 H. Zhu and M. Snyder, *Curr Opin Chem Biol*, 2003, **7**, 55-63.
- 4 W. Tan, L. Sabet, Y. Li, T. Yu, P. R. Klokkevold, D. T. Wong and C. M. Ho, *Biosens Bioelectron*, 2008, **24**, 266-271.
- 5 D. Brady and J. Jordaan, *Biotechnol Lett*, 2009, **31**, 1639-1650.
- 6 F. Rusmini, Z. Zhong and J. Feijen, *Biomacromolecules*, 2007, **8**, 1775-1789.
- 7 C. Liu, P. Schultz, R. Kornberg, C. Raetz, J. Rothman and J. Thorner, *Ann Rev Biochem*, Vol 79, 2010, **79**, 413-444.
- 8 T. S. Young and P. G. Schultz, *J Biol Chem*, 2010, **285**, 11039-11044.
- 9 L. Wang and P. G. Schultz, *Angew Chem Int Ed Engl*, 2004, **44**, 34-66.

- 10 E. Sletten and C. Bertozzi, *Angew Chem Int Ed Engl*, 2009, **48**, 6974-6998.
- 11 N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, *ACS Chem Biol*, 2006, **1**, 644-648.
- 12 Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless and M. G. Finn, *J Am Chem Soc*, 2003, **125**, 3192-3193.
- 13 V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew Chem Int Ed Engl*, 2002, **41**, 2596-2599.
- 14 H. Kolb and K. Sharpless, *Drug Discov Today*, 2003, **8**, 1128-1137.
- 15 N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J Am Chem Soc*, 2004, **126**, 15046-15047.
- 16 J. C. Jewett and C. R. Bertozzi, *Chem Soc Rev*, 2010, **39**, 1272-1279.
- 17 J. Baskin and C. Bertozzi, *Qsar Combi Sci*, 2007, **26**, 1211-1219.
- 18 J. Kalia and R. Raines, *Curr Org Chem*, 2010, **14**, 138-147.
- 19 J. Kalia and R. T. Raines, *Angew Chem Int Ed Engl*, 2008, **47**, 7523-7526.
- 20 B. Hutchins, S. Kazane, K. Staflin, J. Forsyth, B. Felding-Habermann, V. Smider and P. Schultz, *Chem Biol*, 2011, **18**, 299-303.
- 21 S. A. Kazane, J. Y. Axup, C. H. Kim, M. Ciobanu, E. D. Wold, S. Barluenga, B. A. Hutchins, P. G. Schultz, N. Winssinger and V. V. Smider, *J Am Chem Soc*, 2013, **135**, 340-346.
- 22 J. S. Lampkowski, J. K. Villa, T. S. Young and D. D. Young, *Angew Chem Int Ed Engl*, 2015.
- 23 W. Chodkiewicz, P. Cadot and A. WillemartI, *Comp Ren Hebdo Sean*, 1957, **245**, 2061-2062.

- 24 J. Montierth, D. DeMario, M. Kurth and N. Schore, *Tetrahedron*, 1998, **54**, 11741-11748.
- 25 J. Berná, S. M. Goldup, A. L. Lee, D. A. Leigh, M. D. Symes, G. Teobaldi and F. Zerbetto, *Angew Chem Int Ed Engl*, 2008, **47**, 4392-4396.
- 26 M. Mowery and C. Evans, *Tet Lett*, 1997, **38**, 11-14.
- 27 H. Matsuda, H. Nakanisi, T. Hosomi and M. Kato, *Macromol*, 1988, **21**, 1238-1240.
- 28 A. Shun and R. Tykwiński, *Angew Chem Int Ed Engl*, 2006, **45**, 1034-1057.
- 29 C. Glaser, *Ber. Dtsch. Chem. Ges.*, 1896, **2**, 422-424.
- 30 A. Hay, *J Org Chem*, 1962, **27**, 3320.
- 31 M. Vilhelmsen, J. Jensen, C. Tortzen and M. Nielsen, *European J Org Chem*, 2013, 701-711.
- 32 A. Hay, *J Polym Sci A*, 1962, **58**, 581-&.
- 33 J. Ueda, Y. Shimazu and T. Ozawa, *Free Radic Biol Med*, 1995, **18**, 929-933.
- 34 S. I. Lim, Y. Mizuta, A. Takasu, Y. H. Kim and I. Kwon, *PLoS One*, 2014, **9**, e98403.
- 35 A. Deiters and P. G. Schultz, *Bioorg Med Chem Lett*, 2005, **15**, 1521-1524.
- 36 X. Nie and G. Wang, *J Org Chem*, 2006, **71**, 4734-4741.
- 37 J. P. Marino and H. N. Nguyen, *J Org Chem*, 2002, **67**, 6841-6844.
- 38 B. K. Raliski, C. A. Howard and D. D. Young, *Bioconjug Chem*, 2014, **25**, 1916-1920.
- 39 D. Young, T. Young, M. Jahnz, I. Ahmad, G. Spraggon and P. Schultz, *Biochemistry*, 2011, **50**, 1894-1900.

40 D. Young, S. Jockush, N. Turro and P. Schultz, *Bioorg Med Chem Lett*, 2011, **21**, 7502-7504.

41 A. J. Link, M. K. Vink and D. A. Tirrell, *J Am Chem Soc*, 2004, **126**, 10598-10602.

## **CHAPTER 4: USING UAAS TO ILLUSTRATE AND ALTER PROTEIN STRUCTURE-FUNCTION RELATIONSHIPS**

*Part I: Utilizing Unnatural Amino Acids to Illustrate Protein Structure-Function Relationships: An Experiment Designed for an Undergraduate Biochemistry Laboratory*

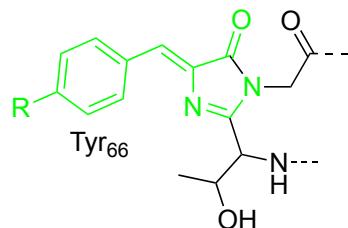
A learning outcome in most biochemistry courses is instilling an understanding of protein structure-function relationships. In order to master this concept, students must be able to recognize how changing even a single amino acid in the protein's primary sequence directly impacts the structure of the protein. Moreover, linked to this concept is the central dogma, a core tenant of any molecular biology or biochemical program. To understand how alterations in a protein's primary sequence arise, students must recognize that the genetic material coding each amino acid of a protein is stored in the DNA. This DNA then gets transcribed into RNA, which is ultimately translated into the amino acid sequence. Thus, an understanding of protein structure and function is intrinsically tied to an understanding of the central dogma. That being said, it is often difficult to illustrate this connection in a purely classroom setting. Consequently, we have developed a laboratory experiment to highlight these connections simultaneously.

Unnatural amino acids (UAAs) represent a unique means to convey all of these concepts.<sup>1,2</sup> In particular, the ability to introduce novel chemical functionality unavailable in the 20 canonical amino acids allows students to observe how altering amino acids can affect the microenvironment around an amino acid residue, drastically affecting how the

protein folds and behaves. This represents an advantage over simply changing one naturally-occurring amino acid for another, as these changes tend to be more nuanced, and thus make it harder for students to fully comprehend the chemical alteration they have performed. Furthermore, utilization of the Schultz methodology of UAA incorporation, whereby successful incorporation of a UAA hinges on the suppression of a TAG point mutation, encourages students to engage with the transcription and translational processes (Figure 1).<sup>1-3</sup> In doing so, they can better visualize how mutations at the DNA level are carried through into a protein's final structure and function. Moreover, due to the necessity to introduce orthogonal translational machinery (a tRNA and aminoacyl tRNA synthetase pair capable of recognizing and charging the UAA to the tRNA), the laboratory also provides a mechanism to introduce multiple molecular cloning techniques to the students.<sup>3</sup> Moreover, the use of UAAs in biological chemistry is becoming increasingly prevalent in multiple applications (therapeutics, bioconjugates, protein probes, etc), and thus exposing students to this valuable technique has pedagogical advantages beyond the scope of the laboratory.

Herein we report a laboratory that uses the Schultz methodology to incorporate commercially available UAAs into Tyr66 of green fluorescent protein (GFP; Figure 1). This tyrosine, along with Ser67 and Gly68, is essential in fluorophore formation, and thus is directly linked to protein function.<sup>4,5</sup> Consequently, students can directly visualize the alteration in UAA structure as a change in the GFP fluorescence spectrum, illustrating how protein function is altered by relatively minor changes to the amino acid sequence. Furthermore, compared to other methods to illustrate structure-function relationship using GFP, such as PCR mutagenesis, the Schultz methodology employing UAAs is more

direct and requires less genetic manipulation on the part of the student given the time constraints of a laboratory setting.<sup>5,6</sup> As a result, it is more likely to produce a functional mutated-protein product, and therefore concretely illustrate for the students the effects of altering even a single amino acid on protein function.



*Figure 28. Structure of GFP fluorophore. Tyrosine-66 (green) will be replaced with different unnatural amino acids to alter spectral properties.*

## Learning Objectives

The utilization of UAAs is a cutting edge technology, and is an excellent mechanism to foster discussion involving protein structure-function relationships. To the best of our knowledge, there has been no reported undergraduate teaching lab incorporating UAA technology.<sup>7,8</sup> As a result, we sought to generate a simple and inexpensive laboratory exercise that would introduce students to this rapidly expanding technology while simultaneously reinforcing the principles of protein structure and function and the central dogma. Moreover, this laboratory confers practical experience in bacterial transformation, protein expression, protein purification, and fluorescence spectrometry. The laboratory also seeks to solidify key concepts from the classroom, specifically involving translational processes.

## Overview

In this experiment the expression and purification of a UAA-containing GFP as well as a wild type GFP will be performed. While typically a unique aminoacyl-tRNA synthetase (aaRS) must be evolved for a specific UAA, recent reports have demonstrated that the *pCNF* aaRS is capable of incorporating up to 20 different UAAs.<sup>9</sup> This is optimal for this experiment to minimize the storage, purification, and organization of multiple aaRS plasmids. The other requirement for UAA incorporation is a gene containing a TAG codon at the residue desired for UAA incorporation, allowing for the suppression of the stop codon by UAA integration if the system is properly functioning. Using plasmids harboring the previously reported promiscuous aaRS, *pCNF* aaRS, and GFP<sub>TAG66</sub>, students will perform a co-transform of competent *E. coli*. This can be accomplished by either electroporation or chemical transformation with competent cells, facilitating discussions on mechanisms of bacterial transformation, transformation efficiency, and the requisite antibiotics to facilitate plasmid uptake and maintenance. The transformed cells will then be used to induce GFP expression in the presence of an assigned UAA. Depending on laboratory size, each pair of students may be assigned a different commercially available UAA. Additionally, students will transform and express wild-type GFP for comparison of the fluorescence emission profile of their UAA-containing GFP versus the wild-type. All plasmids required for the experiment are available from Addgene (Plasmid #48215), or can be obtained from the authors, and the UAAs are all commercially available.

Protein purification may be carried out by numerous methods. Hydrophobic interaction columns, which are commercially available and simple for the undergraduate

student, will be used to partially purify both UAA-containing GFP and wild-type GFP. Our results indicate that this procedure affords GFP in good yield with only minor contaminants, and with approximately no effect on the fluorescence profile of the protein.<sup>9</sup>

This experiment will incorporate several key biochemistry laboratory skills for the undergraduate researcher including electroporation, protein expression and purification, fluorimetry, and UAA-mutagenesis.

### **Laboratory Set-Up and Procedure**

A wild-type GFP transformation and expression protocol was adapted from the commercially available pGLO plasmid kit (Bio-rad; Cat # 166-0003EDU).

UAA-mutagenesis using the Schultz methodology requires the preparation of plasmid stocks harboring the pEVOL-*p*CNF-aaRS plasmid with chloramphenicol-resistance, the pET-GFP<sub>WT</sub> plasmid with ampicillin resistance, and the pET-GFP<sub>TAG66</sub> plasmid with ampicillin resistance. Additionally, 100mM stocks of *p*-azidophenylalanine (*p*AzF), *p*-biphenylphenylalanine (*p*BiPhF), *p*-acetylphenylalanine (*p*AcF), *p*-iodophenylalanine (*p*IF) in deionized water should be prepared prior to the start of the laboratory period (see Experimental).

Students work in pairs over the course of three weeks, and require typically two-hours for each lab meeting. Week 1 of the laboratory focuses on bacterial transformation, and mutant GFP expression. Week 2 involves both wild-type and mutant GFP purification and fluorescence. Week 3 entails a “GFP symposium” facilitating discussion about the techniques utilized in lab and student presentations.

### Week 1

On the first week, students perform the transformations and induce expression of GFP mutants harboring an unnatural amino acid. The chemical transformation of pET-GFP<sub>WT</sub> are performed in direct accordance with the Bio-Rad pGLO kit (substituting 1 mM IPTG for arabinose in the agar plates). This plasmid encodes the wild-type GFP and does not require the aaRS machinery. The plasmid can also be transformed by electroporation to provide a direct comparison to the chemical transformation. This portion of the lab emphasizes the transformation of plasmid DNA into *E. coli* for heterologous expression. If available, students may also perform an optional double transformation using 0.5 µL of pEVOL-pCNF plasmid and 0.5 µL of pET-GFP<sub>TAG66</sub> plasmid, both at approximately 100 ng/µL with an electroporator at 1800V. These transformations should be recovered at 37 °C for 1 hr, then plated on plates containing ampicillin and chloramphenicol.

While transformations are being performed, students also induce the expression of a mutant GFP with their specified UAA using a second bacterial culture that was previously prepared. Two days prior to the laboratory, the instructor must co-transform the pEVOL-pCNF plasmid and pET-GFP<sub>TAG66</sub> as described above. The next day, a colony should be selected and grown in 10 mL of LB Media containing ampicillin and chloramphenicol at 37 °C overnight. The day of the lab, approximately 1-2 hours before the scheduled time, a 10 mL expression culture of the cells should be inoculated for each group at OD<sub>600</sub> 0.1. During the actual laboratory students will begin expressing their UAA-containing GFP using an expression culture that has grown to log phase (OD<sub>600</sub> =

0.6 – 0.8). Expression of the GFP<sub>TAG66</sub> is induced with the addition of 10 µL each of 1000x stocks of IPTG and 20% arabinose. Also, at this time students add 100 µL of their assigned 100 mM UAA stock. This laboratory period facilitates a discussion of typical transformation procedures (students are exposed to both heat shock and electroporation in the exercise) as well as a discussion of molecular genetics for gene expression. Finally, the use of the Schultz methodology for incorporation of the UAA into GFP<sub>TAG66</sub> facilitates a discussion of gene transcription and protein translation.

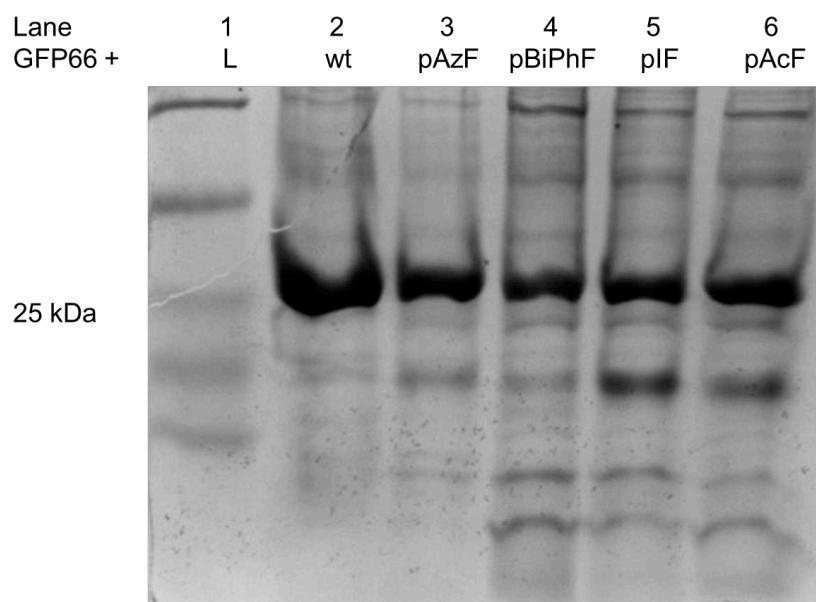
The next day the laboratory supervisor centrifuges the expression cultures of the UAA-containing GFP (5000 g for 10 min). The cell pellets are then stored at -80 °C until the next laboratory session. Additionally, at some point in-between lab periods students observe the fluorescence of the *E. coli* cell pellet, and count the colonies on their transformation plates. This allows them to identify the differences between electroporation (UAA plasmids, and possibly the wild-type plasmid) and heat-shock (wild-type plasmid), as well as single and double transformations. These colony counts are used to calculate the transformation efficiency for the single transformations versus the double transformations.

### Week 2

In the second week students use hydrophobic interaction columns (HIC) to purify both their wild-type and UAA-containing GFP. This week follows the protocol described in the Green Fluorescent Protein Chromatography Kit (Bio-Rad; Cat # 166-0005EDU). Using their partially purified protein at a concentration of 0.5 to 1 mg/mL, students prepare a sample for fluorescence detection by diluting 10 µL of their protein in 3 mL of 1 x PBS. Students then observe the fluorescence spectra of their wild-type and UAA-

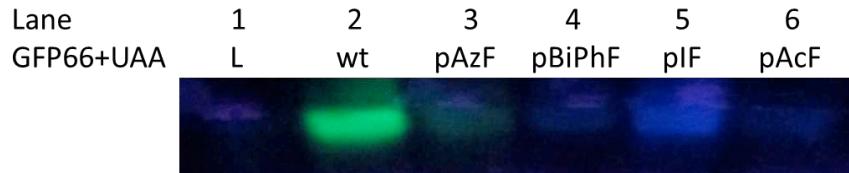
containing GFPs using a fluorimeter set to excite at 395 nm and scan between 410 nm to 600 nm.

This week also has the potential to facilitate a discussion on protein purification techniques, as the HIC process pulls down many contaminating proteins along with the UAA-mutated GFP. As such, column fractions can be analyzed via SDS-PAGE and stained with Coomassie blue to highlight the limitations of this purification technique (see Figure 2).



*Figure 29. Coomassie stained SDS-PAGE of the UAA-containing GFP<sub>66</sub> produced by the biochem students. The gel highlights that the hydrophobic interaction column is an adequate means to purify the GFP and facilitates a conversation about what the properties of the other contaminants. Our fluorescence data indicates that the contaminants did not affect the spectra.*

Furthermore, if time is not available to stain and destain with Coomassie blue, the GFP fluorescence can be directly observed through the SDS-PAGE gel using a UV transluminator or lamp (see Figure 3).



*Figure 30. Fluorescent imaging of SDS-PAGE of the UAA-containing GFP<sub>66</sub> produced by the biochem students. Gel was excited at 302 nm for optimal fluorescent detection of all UAA-variants. Samples were treated with a 2x SDS-dye mix prior to loading, which retained protein structure allowing for the detection of the GFP via fluorescence, and thus obviating the need for Coomassie staining.*

### Week 3

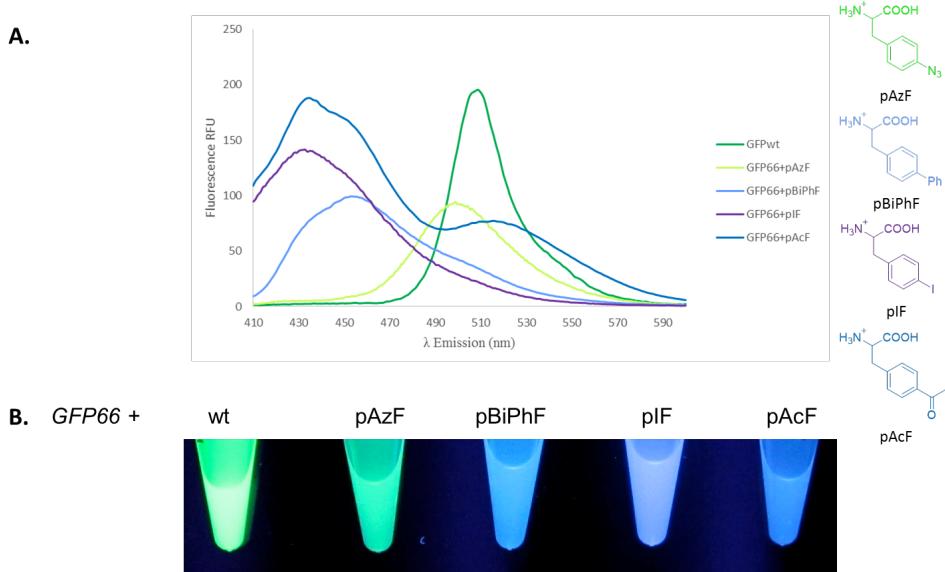
In the third week students present their results as well as a discussion of topics related to protein expression, purification, and fluorescence. Each student is assigned a specific topic to prepare a 5-10 min presentation. Examples of topics include: biological relevance of lysozyme, methods of cell lysis, DNA mutagenesis for the introduction of the TAG codon, aaRS evolution, mechanism of chemical transformations, antibiotics and antibiotic resistance, protein induction by IPTG/Arabinose, vectorology, the chemical basis of GFP fluorescence, and numerous other topics related to the laboratory. Additionally, the results of the laboratory are discussed and compiled fluorimetry data is presented to truly illustrate the effect of the different amino acids on GFP fluorescence

### **Results and Discussion**

Overall, the experiments described above translated well in a laboratory setting. While the traditional pGLO GFP lab has been employed in our Advanced Biochemistry Laboratory for the past several years, the unnatural amino acid component has been included in the past two years. Each year averages between 8-12 students in the lab, for a total of approximately 20 total student participants. The double transformations that the

students performed yielded approximately 5 colonies per lab group, which is to be expected of double transformations. Single transformations afforded much higher transformation efficiencies, to those referenced in the literature and ranged from 0 to  $10^9$  cfu. The efficiencies were typically student dependent based on previous experience and their ability to follow directions. These results promote discussions into antibiotic resistance and requisites for colony growth on antibiotic plates. Protein expression levels were robust, the GFP<sub>WT</sub> and the UAA-containing GFP<sub>66</sub> both yielding between 0.5 to 1 mg/mL of purified protein.

The UAA-mutagenesis of the GFP<sub>66</sub> was successful for all UAA used in this laboratory, as demonstrated by the successful shift in the UAA-containing GFP<sub>66</sub> relative to the GFP<sub>WT</sub>. Gratifyingly, spectral shifts of GFP fluorescence were observed for each mutant, demonstrating the relationship between protein structure and function, and the influence of a single amino acid mutation (see Figure 4). The most dramatic blue shift was observed for the pAcF mutant, most likely due to the resonance into the carboxylic oxygen. Overall, students were able to effectively transform bacteria with the plasmids, as well as express and purify functional protein, and analyze its spectral properties.



*Figure 31. Fluorescence results for the expression of GFP mutants. (A) Fluorescence spectra obtained on a fluorimeter exciting at 395 and scanning emission from 410 to 600 nm. (B) Visualization of mutant protein fluorescence on a transiluminator with an excitation wavelength of 365 nm.*

This experiment was piloted in an upper-level biochemistry lab targeted for juniors and seniors with a background in the subject. Compared to a standard protein expression/purification lab, in which students transform GFP<sub>WT</sub> into *E. coli* and subsequently purify the protein, this exercise provides students with exposure to considerably more biochemical and molecular biology techniques. In particular, the use of the Schultz methodology for UAA incorporation allowed students to engage with topics from lecture, such as translation and protein structure-function relationships, in a way that was not accessible via a more traditional lab (Table 1). Based on student evaluations of the laboratory, an appreciation of how the experiment concretely illustrates protein structure-function relationships was the most powerful learning outcome. Moreover, the use of the fluorimeter allowed students to visualize the impact of their UAA incorporation, thereby reinforcing how even a single amino acid alteration can drastically impact a protein's function. This concept is perfectly illustrated by the

integration of UAAs into the laboratory and can easily be achieved over the course of 2 lab meetings.

*Table 3. Comparison of Standard GFP Laboratories Versus the UAA Laboratory*

Learning Outcomes for Standard GFP Transformation Experiment	Learning Outcomes for UAA-Mutagenesis of GFP Experiment
<ul style="list-style-type: none"> <li>• Basic understanding of transformation techniques as well as protein purification (which obviously pairs well with protein properties).</li> <li>• Provides student with an opportunity to understand selection processes in conjunction with plasmid design (i.e. antibiotic resistance as a means to ensure plasmid maintenance).</li> <li>• Provides an illustration of molecular genetics by providing a means to discuss operators and promoters</li> </ul>	<ul style="list-style-type: none"> <li>• All those listed in the “Standard” GFP Transformation Experiment</li> <li>• Facilitates discussion of translational processes via the Schultz’ methodology for UAA incorporation</li> <li>• Facilitates discussion of protein structure-function relationship, as the incorporation of different UAAs alters the GFP-chromophore, affecting fluorescence profile</li> <li>• Provides students with the opportunity to practice use of a fluorimeter, a standard laboratory instrument in most chemical settings. Also facilitates a discussion of fluorescence, a topic typically relegated to physical chemistry</li> <li>• Provides students with exposure to current technologies (e.g. UAA incorporation) that are widely employed within the field of chemical biology.</li> </ul>

The experiment also provides practical introductions to bacterial transformations and illustrates the differences in transformation efficiency between various protocols. Additionally, the necessity for double-transformations facilitates discussions for mechanism of antibiotic resistance and the necessity of antibiotic resistance onto the plasmids to facilitate bacterial selections. Finally, key aspects of fluroimetry are also introduced in the experiment as differing functional groups result in spectral shifting of the fluorophore emission.

Upon completion of the laboratory during week 3, the learning outcomes were assessed both by presentation projects and by a laboratory evaluation form (see Experimental). Students were asked to distil the big picture ideas as well as assess the effectiveness of the laboratory. Overall, students were overwhelmingly positive in the assessment of the lab and all of the students were able to clearly and concisely communicate both the purpose of the lab and fundamental concepts about plasmids and protein translation. Finally, learning outcomes were assessed by the completion of a typical laboratory report summarizing the experiment and the overall results for each student.

## **Conclusion**

Unnatural amino acids are an excellent mechanism for the study of protein structure-function relationships. This developed laboratory for an upper division biochemistry course provides a simple means of demonstrating this feature of UAAs. Moreover, the mechanism of UAA incorporation facilitates substantial learning and solidifies several key concepts including: bacterial transformations, molecular cloning, and fluorescence.

## **Experimental**

Experimental Handout Adapted for Laboratory Exercise – Week 1

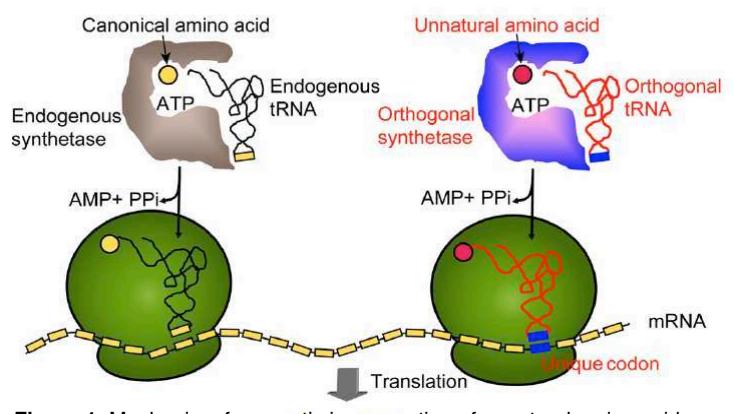
## Experiment 4 – Transformation of *E. coli* with the gene for Green Fluorescent Protein (GFP)

Week 1 – Transformation of *E. coli*, plating and calculations of transformation efficiency, and protein expression

Week 2 – Purification of GFP by column chromatography and fluorescence measurements.

### Background:

Unnatural amino acids (UAs) allow scientists to expand the rather limited biochemical functionality found in the 20



**Figure 1.** Mechanism for genetic incorporation of unnatural amino acids. An orthogonal synthetase must be introduced that does not cross-react

canonical amino acids and introduce novel reactivity not typically found in proteins<sup>1,2</sup>. Incorporated via suppression of a TAG stop codon in translation, which can be introduced into the DNA at a desired residue position of the target protein. In order to incorporate UAA's, the cell needs to be supplied with new orthogonal machinery (so that there is no cross-talk between the endogenous translation machinery of the cell and the machinery necessary to incorporate the UAA). A foreign aminoacyl-tRNA synthetase (aaRS) must be expressed in the cell that will recognize the UAA and "charge" it to an exogenous tRNA with the antisense codon for TAG. Now, as the cell's translation machinery reads our desired protein's mRNA, it will incorporate the UAA in any codons that code for TAG (thankfully the TAG, or Amber, stop codon is rarely used in *E. coli*, and therefore this new machinery will not impact cell viability/normal functioning). Scientists can readily manipulate the DNA to introduce TAG stop codons into desired residue positions on the protein's DNA. Overall, this process allows for the site-specific incorporation of UAA's in pre-defined positions. Figure 1 provides a great overview of the new cellular machinery described above. The alteration of a single amino acid can have significant impact on protein function. This week, we will introduce UAA's in place of Tyrosine 66 in Green Fluorescent Protein (GFP). Position 66 is a residue essential in the formation of the GFP fluorophore. Based on the properties of the incorporated UAA, the fluorescence of GFP will change. We will employ one of several commercially available UAAs, and each laboratory group will be assigned an amino acid. While typically, a unique aaRS must be evolved for an individual amino acid, recent work has found that the pCNF-aaRS is promiscuous and able to recognize multiple different UAAs, a property that we will exploit for UAA incorporation. We will be measuring these altered fluorescence spectra after purification of the protein in Week 2.

### Helpful definitions

Competent cells: Capable of taking up plasmid DNA. Typically, cells are treated with calcium chloride which alters their membrane in such a way that DNA can pass through.

Transformation: Once the bacteria have taken up endogenously added DNA, they are called "transformed."

LB plates – contain agar and a nutrient broth named after Luria and Bertani.

LB/amp plates – contain LB growth media, agar and ampicillin

LB/amp/ara plates – contain LB growth media, agar, ampicillin and a sugar inducer called arabinose

Transformed *E. coli* will grow on LB/amp plates; ampicillin is an antibiotic similar to penicillin. Transformed bacteria will express an enzyme that is capable of destroying the ampicillin present in the plate media. Thus, only those bacteria will grow on LB/amp plates. However, to express GFP, an inducer must be present as well. Thus, GFP should only be expressed on LB/amp/ara plates.

UAA – unnatural amino acid

aaRS- aminoacyl-tRNA synthetase, a protein responsible for charging tRNA molecules with specific amino acids

pEVOL-pCNF – a plasmid coding for the tRNA synthetase that recognizes our UAA, as well as the tRNA with an antisense codon for TAG. This plasmid gives the bacteria resistance to *chloramphenicol*.

pET-GFP-66-TAG – a plasmid coding for GFP with a TAG codon in residue 66. This plasmid gives the bacteria resistance to *ampicillin*.

#### Transformation procedure

This week you will be performing two transformations. One will be a single transformation and will produce the wild-type GFP. The second transformation will be a double transformation involving two plasmids, and will involve the electrical shock (electroporation) of competent cells to introduce both the GFP protein (with a mutated TAG codon at position 66) and the necessary translational machinery.

The procedure for transforming *E. coli* is very simple and is described in the “Transformation Kit-Quick Guide.” You will use 10 µl of the pET-GFP-WT plasmid solution provided. We will use Pipetmen to dispense solutions because they are more precise than the plastic pipets shown in the “Transformation Kit-Quick Guide”. There are a few changes to the kit procedure that I will announce in lab before we begin the procedure.

For the electric shock transformation, all of the following must be placed on ice: an aliquot of competent cells, some media, an electropuvette, pET-GFP-66-TAG plasmid, and pEVOL-pCNF. The process is as follows:

**NOTE: to allow you to move faster, it is useful to preset and place pipet tips on all the pipetmen that you will be using.**

- 1) Take up 0.5 µL of pEVOL-pCNF – inject this into the aliquot of cells
- 2) Take up 0.5 µL of pET-GFP-66-TAG – inject this into the aliquot of cells
- 3) Take up all 50 µL of the cells (set pipette to 55uL to ensure you get the full volume) –and then transfer the full volume to the electropuvette (NOTE: the opening for the electropuvette is very slim, so do your best to make sure that the very tip of the pipette-tip is in between the slot).
- 4) Place the now-full electropuvette into the electroporator, which should be preset to 1600V. Shock your cells!
- 5) RAPIDLY add 500uL of chilled media to the cells – this will help them recover from the shock.
- 6) Transfer the ~550uL of transformed cells to a labeled eppendorf tube and allow to recover for 1 hr at 37 °C.

**If your cells “spark/arc”:** Redo the procedure above, using 0.4uL of DNA each (for a total of 0.8uL of DNA) and shock the cells at 1400V.

After the 1hr recovery, plate a 100  $\mu$ L aliquot of the transformed cells onto an AMP/CHLOR LB Agar plate.

**YOU MUST CHECK YOUR PLATES TOMORROW AFTERNOON FOR COLONIES!**

Expression Protocol:

Additionally, you will set up a protein expression that you will purify next week. Protein expression is an extremely time dependent process, as expression must be induced at an optimal time in the cell growth cycle to maximize yield. Prior to lab, your instructor has inoculated a liquid culture of LB media (containing ampicillin and chloramphenicol) with the bacteria containing both the pEVOL-pCNF plasmid and the pET-GFP-66-TAG plasmid.

- 1) Monitor the growth of the cells by measuring the OD<sub>600</sub> of the bacterial culture on the spectrophotometer (this may require you to make a 1:10 dilution of the cells prior to measurement)
- 2) When an OD<sub>600</sub> of 0.8 has been achieved, obtain 10 mL of bacterial culture and induce GFP production by the addition of 10  $\mu$ L of 20% arabinose (to induce the production of the translational machinery), 10  $\mu$ L of 1M IPTG (to induce the production of GFP), and 100  $\mu$ L of an assigned unnatural amino acid (100 mM).
- 3) The culture will now be placed in an incubator at 30 °C for 16 h to allow for protein expression.
- 4) Your instructor will centrifuge the culture tomorrow at 4,000 rpm for 10 min, discard the supernatant, and store the cell pellet at -80 °C for you to use next week in the GFP purification.

Things to do:

1. Check for transformed *E.coli* on the thermal transformations of wild-type protein LB/amp and LB/amp/ara plates. Check all plates with the UV light to look for “glowing” GFP containing colonies.
2. You will count the total number of colonies on the LB/amp and LB/amp/ara plates. With these numbers, you can calculate transformation efficiency. We will go over this calculation in lab.
3. Observe your UAA GFP pellet for fluorescence using the transilluminator.

Experimental Handout Adapted for Laboratory Exercise – Week 2

## **Experiment 4 Week 2– Bacterial cell lysis and partial purification of GFP by hydrophobic interaction chromatography**

*Please follow the instructions below.* The GFP purification quick guide is posted but I have made significant changes. Use Pipetmen for all volume transfers. Green fluorescent protein (GFP) will bind to a hydrophobic interaction column(HIC) in the presence of 2 M ammonium sulfate (hydrophilic proteins will not bind). Then we elute GFP from the HIC with a buffer of lower salt concentration.

### **Solutions used:**

TE buffer – 10 mM Tris, 1 mM EDTA pH 8.0

Binding buffer – 4 M  $(\text{NH}_3)\text{SO}_4$  in TE;

TE Column Equilibration buffer - 2 M  $(\text{NH}_3)\text{SO}_4$  n TE

Column Wash buffer -1.3 M  $(\text{NH}_3)\text{SO}_4$  In TE

Lysozyme in TE buffer

### **Bacterial cell lysis**

1. Each student will receive a 2 ml portion of bacterial culture. The cells will have been grown (*name that verb tense!*) for 24-36 hours at 30 °C with shaking. Look at the cell suspension under UV light. Does it glow?

Observation: \_\_\_\_\_

2. Spin the cells in a microfuge tube for 5 minutes at maximum speed to obtain a pellet. Since there are 2 mL, I suggest you pipet 1 mL into a microfuge tube, spin for 5 minutes, decant the supernatant and then add the second ml of bacterial culture. Spin again for 5 minutes and decant the supernatant.

3. Resuspend the cell pellet in 250 µl TE buffer by pipetting up and down several times.

**Modified lysozyme step** – lysozyme degrades the cell wall. Then freeze/thaw cycles should “finish the job.”

4. Add 100 µl lysozyme solution to the resuspended cells. Incubate the sample at 37 °C for 30 minutes. Be sure to vortex the sample about every 5 minutes during the 30 minute period.

5. Freeze the sample in liquid nitrogen to rupture the cell walls. Thaw quickly in warm water. Repeat the freeze/thaw cycle.

6. Spin the sample for 10 minutes at maximum speed to pellet the insoluble bacterial debris.

***While the tubes are spinning, prepare the HIC column (see below)***

Page 2 of 2

After the spin, carefully pipet the supernatant into a new, clean microfuge tube. Look at the pellet and the supernatant under UV light.

Observations:

pellet \_\_\_\_\_

supernatant \_\_\_\_\_

The total volume is now 350 µl, so add 350 µl binding buffer, 4 M (in TE, before loading the column).

Final concentration of  $(\text{NH}_3)\text{SO}_4$  = \_\_\_\_\_

You will load **ALL** 700 µl onto the column.

## 7. Column prep and loading

You will need 3 tubes for collecting fractions from the column. Label them 1, 2 and 3.

Clamp the HIC column in an upright position. Remove the top and bottom caps and allow the column to drain (into a waste beaker). Prepare the column by adding 2 ml Equilibration buffer (2 M (NH<sub>3</sub>)SO<sub>4</sub> in TE) to the top of the column. After it has drained through, re-cap the column and place collection tube 1 under it.

8. Load your entire supernatant sample (700 µl) on the column bed carefully. Allow the supernatant to drip down the side of the column wall so as not to disturb the HIC column bed.

9. Uncap the column and allow the column to drain slowly. It is during the loading step that the column packing material (HI resin) and the GFP “find each other” so you do not want to rush this step.

Check fraction 1 under UV light. Observation \_\_\_\_\_

10. Place tube 2 under the column. Apply 700 µl Column Wash solution (1.3 M (NH<sub>3</sub>)SO<sub>4</sub> in TE) to the top of the column. Collect as fraction 2.

Check fraction 2 under UV light.

Observation \_\_\_\_\_

11. Place tube 3 under the column. To elute GFP, apply 700 µl TE buffer. Collect as fraction 3. Check fraction 3 under UV light.

Observation \_\_\_\_\_

Label your GFP column fractions with your initials and place them in the designated rack before you leave. Did you check them with the UV lamp to see if they GLOW?? Really?

*Who will have the highest yield of GFP??*

12. Record the emissions spectrum of both the WT GFP sample and the mutant GFP sample on the fluorimeter

- 1) Turn on the fluorimeter and let it warm up.
- 2) To a quartz cuvette, add **10 uL of GFP into 3 mL of (TRIS OR PBS)**
- 3) Place into holder in the fluorimeter
- 4) Operation of fluorimeter software
  - a. Open **FL WinLab**
  - b. Pick any scan method

- c. A window will pop up – select the heading titled “Emission”
  - i. Parameters: start (nm): 405; end (nm): 650; excitation (nm): 395; ex slit (nm): 10; em slit (nm): 10; scan speed (nm/min): 50
- d. Name the file a 4 letter name for your group (EX: GRPA)
- e. Press the “go” on the stop light to begin the read

Experimental Report Guidelines – End of Lab

## **Experiment 4 Report and oral presentation guidelines**

### **Written report**

1. Write a brief summary (one paragraph) of the goals of the two-week experiment and the outcome.
2. Note anything unusual that may have happened.
3. Week 1 – report the results of the transformation. In a table, summarize what you saw on the 4 plates. Also explain why some colonies expressed GFP and others didn't. Show calculations of transformation efficiency. Remember: Transformation efficiency is the number of colonies/amt of DNA used. If you saw nothing on the plates, explain why.
4. Week 2 – report the results of the purification of GFP from the bacterial culture. Why did the protein bind to the column? Do you think it was pure? How would you check? Discuss the observed fluorescence spectra. Was there a difference between the UAA protein and the WT protein?

The entire written report need not be more than 1-2 pages.

In addition to presenting the report information (especially #3 above), each of you will also answer a question. You should think of it as “teaching the rest of the class” about a particular topic related to the two-week GFP experiment. That means that you should prepare figures with essential chemical structures or be willing to draw them on the board. Total presentation time should be 7-8 minutes/student. Questions from the audience are encouraged. If you don’t ask them of each other, I definitely will. Heck, I probably will anyway.

**Student 1**

We used lysozyme and freeze/thaw cycles to rupture the bacteria. How does lysozyme work? What other methods could we have used to lyse the bacteria without destroying GFP?

**Student 2**

Explain how the “heat shock” procedure for transforming bacteria works (this was week 1). What is the purpose of the  $\text{CaCl}_2$  transformation solution? What are other common methods to get plasmid DNA into bacterial cells?

**Student 3**

How is the GFP fluorophore formed (see the J. Chem. Ed. article)? This may require you to look at the sequence/structure of GFP. Is there any reason why this protein evolved?

**Student 4**

In the J. Chem. Ed. article posted on Blackboard, the researchers explain the amino acid mutations in GFP that have been done and how those mutations affect the fluorescence/color of the resulting GFP proteins. Using the amino acid sequences, the article and what you know about chemistry, explain the color changes. (It is essential that you show the class the sequences in Table 1.)

**Student 5**

Explain how site-directed mutagenesis of GFP was performed to make the mutants described in the J. Chem. Ed article posted on Blackboard and used in the experiment.

**Student 6**

How does arabinose induce expression of GFP? (Be sure to include the structure of arabinose) Compare the arabinose induction system with the lac operon.

**Student 7**

What is the chemical mechanism of ampicillin resistance? Compare the structures of ampicillin and penicillin. This will require some explanation of the structure of the bacterial cell wall.

**Student 8**

What is the drug Augmentin? Provide some information about the structure and mechanism of the enzyme,  $\beta$ -lactamase that destroys ampicillin. What other inhibitors of  $\beta$ -lactamase are available? Or under development? Tell us about at least one more.

**Student 9**

Explain how hydrophobic interaction chromatography works. Be sure to explain the purpose of the buffers used (equilibration, binding, wash and TE). If you used ion exchange chromatography, under what conditions, if any, would you expect GFP to bind to the column? This may require you to look at the sequence/structure of GFP.

**Student 10**

Explain the process of incorporating unnatural amino acids into proteins. How are synthetases evolved to recognize unnatural amino acids?

**Student 11**

Explain the basis of polyspecificity of a synthetase. How else have unnatural amino acids been employed than to alter the fluorescence of GFP (pick 1-2 examples).

You may use PowerPoint or write on the board. If you have any questions about the written report or the presentation, please ask!

The complete pdf guidebooks (50-60 pages) for the two GFP experiments that you did are posted on Blackboard under course documents. You need not print out the whole thing but you will find the majority of the information you need to answer the questions there.

**References**

- <sup>1</sup>Liu, C.; Schultz, P. "Adding new chemistries to the genetic code" *Annu. Rev. Biochem.* **2010**, 79, 413-44.  
<sup>2</sup>Young, T.; Schultz, P. "Beyond the canonical 20 amino acids: Expanding the genetic lexicon" *J. Biol. Chem.* **2010**, 285, 11039-11044.

Evaluation form used to gather student feedback

**Lab Evaluation Form**  
Unnatural Amino Acids and GFP

The purpose of this evaluation is to gather input so that we can improve this lab and your learning outcomes. Please answer these questions without referring to the lab manual. Your response will be kept anonymous and are not graded. Thank you for participating!

Your Name \_\_\_\_\_

1. What is the purpose of having 2 plasmids? What proteins are specifically encoded on each?
  
2. Summarize the major findings of this lab.
  
3. What is the strongest feature of this lab? In other words, what contributed most to your learning?

4. Was bringing unnatural amino acids into the lab useful?

5. Do you have any additional comments about this lab?

List of Chemicals Used in Laboratory Exercise

Chemical	CAS	MW (g/mol)	Density (g/mL)	Hazard
4-iodo-L-phenylalanine	242 50-85-9	29 1.09	---	Potential irritant
4-acetyl-phenylalanine	22888-49-9	207.23	1.226	Potential irritant
L-4,4'-biphenylalanine	<a href="#">155760-02-4</a>	241.2896	---	Potential irritant
4-azido-L-phenylalanine	33173-53-4	206.2	---	Potential irritant

isopropylthio- $\beta$ -galactoside	1552 9-019	238.32	---	Potential irritant
L-(+)-arabinose	532 8-37-0	150.13	---	Potential irritant
ampicillin	<a href="#">7177-48-2</a>	403.45	---	allergies
chloramphenicol	56-75-7	323.13	---	irritant
acrylamide	<a href="#">79-06-1</a>	71.08	---	neurotoxin

### Instructor Prep Guide

Week 1 – Transformation of *E. coli*, plating and calculations of transformation efficiency, and protein expression

To prepare (beyond the items required by the pGLO kit):

- Electrocompetent cells
- 100 mM Stocks of the UAAs to be used
  - For 100 mM pIF dissolve 29 mgs of pIF into 1 mL of deionized water
  - For 100 mM pAcF dissolve 21 mgs of pAcF into 1 mL of deionized water

- For 100 mM pBiPhF dissolve 24 mgs of pBiPhF into 1 mL of deionized water
- For 100 mM pAzF dissolve 21 mgs of pAzF into 1 mL of deionized water
  - NB: sometimes the UAAs have issues going into solution. As such, we found it useful to both sonicate the UAAs as well as adding approximately 5  $\mu$ L increments of a 10 M NaOH solution to the UAA (prior to bringing to full volume).
- Pure plasmid stocks of PET-GFP<sub>TAG66</sub> and pEVOL-pCNFaaRS
- Prepare IPTG (200mg/mL)
  1. Dissolve 2g IPTG (Isopropylthio-B-D-galactoside) in 8 mL of water.
  2. Adjust final volume to 10 mL with more H<sub>2</sub>O if needed.
  3. Transfer mixture to 12 mL syringe with filter attached to end.
  4. Filter 1 mL aliquots into epis.
  5. Store in freezer.
- Prepare 20% (w/v) Arabinose
- Amp/Chlor supplemented agar plates for selection of cells cotransformed with GFP66+pCNFaaRS
- Expression cultures of cells (at an OD<sub>600</sub>  $\sim$  0.7) already harboring plasmids for GFP66 and pCNFaaRS
  - Thus you will already need to have done the cotransformation yourself, plated these transformed cells on Amp/Chlor agar, selected an isolated colony for starter culture growth, and begin the expression cultures  $\sim$  2hr prior to the start of lab.

Week 2 – Purification of GFP by column chromatography and fluorescence measurements.

To prepare (beyond the items required by the pGLO kit):

- GFPwt expression cultures must be prepared prior (according to direction in the pGLO kit) one day before the lab meets.

Common student pitfalls While electroporation is, for the most part, a fairly easy technique, the risk of “sparking” cells is high for the beginner. As such, we found it useful to have a more experienced person demonstrate a proper electroporation, and then had students attempt their own. However, because “sparking” is often common with beginners, we did not have students redo their transformations if their cells sparked. Instead, because we were already providing the students with a pre-transformed expression culture, we reasoned it was not necessary for all students to have successful transformations. Indeed, the pGLO kit transformants would provide the students with transformed cells that could then serve as a discussion point for selection methods.

We found that the pGLO kit did not always yield successfully transformed colonies for all students. As such, we prepared our own expressions of GFPwt (using a PET-GFPwt plasmid and electroporation) and pelleted the cells prior to start of lab on week 2. Then, if any students did not have good colonies, and hence GFP expression, they could use the pelleted-GFP expressing cells we had already prepared for them.

## References

- (1) Young, T. S.; Schultz, P. G. Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *J. Biol. Chem.* 2010, 285 (15), 11039– 11044.

- (2) Wang, L.; Xie, J.; Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* 2006, 35, 225–249.
- (3) Wang, L.; Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* 2001, 8 (9), 883–890.
- (4) Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* 2009, 38 (10), 2865–2875.
- (5) Hicks, B. W. Recombinant green fluorescent protein isoforms: Exercises to integrate molecular biology, biochemistry and biophysical chemistry. *J. Chem. Educ.* 1999, 76 (3), 409–415.
- (6) Girón, M. D.; Salto, R. From green to blue: Site-directed mutagenesis of the green fluorescent protein to teach protein structure-function relationships. *Biochem. Mol. Biol. Educ.* 2011, 39 (4), 309–315.
- (7) Bilgiçer, B.; Kumar, K. Products of Chemistry Protein Design Using Unnatural Amino Acids. *J. Chem. Educ.* 2003, 80 (11), 1275–1281.
- (8) Breslow, R. Bioorganic Chemistry: A Natural and Unnatural Science. *J. Chem. Educ.* 1998, 75 (6), 705–718.
- (9) Young, D. D.; Jockush, S.; Turro, N. J.; Schultz, P. G. Synthetase polyspecificity as a tool to modulate protein function. *Bioorg. Med. Chem. Lett.* 2011, 21 (24), 7502–7504.  
771

*Part II: Editing GFP Fluorescence using the Biological Glaser-Hay*

Protein engineering is a powerful tool for the development of new therapeutics, catalysts, and biosensors.<sup>1</sup> While many advances in the field have been made, designing novel protein functionality is still a challenge, as it requires an intricate understanding of the subtle interplay between protein structure and function. Current engineering

techniques often focus on using selection and screens to optimize or enhance existing protein functionality.<sup>1</sup> While this method has proved useful for optimizing existing function, generating new function where it does not exist is still a hurdle.

Unnatural amino acids (UAAs) represent a unique means to introduce new chemical functionality into proteins. Using the Schultz methodology, UAAs are co-translationally incorporated into a protein's primary sequence, typically in response to a stop codon introduced at the DNA level.<sup>2,3</sup> As such, UAA mutagenesis allows for the site-specific incorporation of a novel chemical functionality into proteins, which can then be exploited towards the development of novel protein activities. By inserting new chemical reactivity into a protein, UAAs allow for a condensed and facile workflow for engineering new protein functionality without the need to perform tedious and lengthy selections and screens. For example, recent reports from our group and others have site-specifically incorporated fluorescent UAAs into proteins, yielding novel fluorescent properties in a protein in a direct and highly controlled fashion.<sup>4,5</sup>

While UAAs have allowed for the development of unique protein function, the evolved protein are often limited to a single new function, depending on the UAA incorporated.<sup>1</sup> In addition, the functionality is limited to the UAA itself, which suffers from constraints such as the requisite for an aminoacyl tRNA synthetase capable of recognizing the UAA, the synthetic accessibility of the UAA, and the size of the UAA which may preclude its uptake by a biological system.<sup>1-3,6</sup> A more appealing strategy would allow for the generation of a UAA-containing protein "template" upon which researchers could synthetically introduce different chemical moieties that would in turn lead to altered protein function depending on the moiety employed.

Bioorthogonal chemistry, which employs reactions that proceed to completion under physiological conditions ( $\text{pH} \sim 7$ ,  $37^\circ\text{C}$ ), offers a unique mechanism to add new chemical functionality to proteins.<sup>6,7</sup> Indeed, a variety of reactions have been developed that can add new chemistry to living systems. In particular, the cycloaddition between azides and alkynes that is either copper(I) mediated or strain promoted, has become a widespread technique to introduce new chemistry to proteins. More recently, our group has developed a biorthogonal variant of the Glaser-Hay reaction, which brings together two terminal alkynes to form a diyne on a protein in the presence of copper(I) under physiological conditions.<sup>8</sup> The resulting stable diyne linkage has a well-defined conjugated linear geometry, and due to the abundance of commercial available terminal alkynes, a variety of chemical moieties can be reacted onto a protein using this technique. As such, our group sought to use the power of this new chemistry to generate new and different protein function dependent upon the alkyne reaction partner and not purely the UAA.

Specifically, we designed a proof-of-concept experiment to alter the function of green fluorescent protein (GFP) via reaction of different terminal alkynes, onto the GFP chromophore. GFP is a 27kDa protein isolated from *Aequorea victoria* with novel photochemical properties arising from an internal chromophore composed of Ser65-Tyr66-Gly67.<sup>9</sup> New chemical properties afforded simply by UAA introduction in place of Tyr66 have already been documented to alter GFPs fluorescence profile. All UAAs incorporated were found to blueshift the fluorescence profile of GFP, with more highly conjugated UAAs exhibiting a greater degree of blueshifting.<sup>5</sup> Based on these results, the ability to modulate the conjugation of GFPs fluorophore using UAA mutagenesis is

apparent, and represents a convenient means to rationally design new protein function. However, this mutagenesis approach is limited by the size and complexity of the UAA. A potentially better approach involves exploiting the chemical functionality in preexisting UAAs to serve as a handle for biorthogonal reactions, acting as a “template” for the chemical derivatization of new protein function.

This chemically directed protein evolution could be achieved via the genetic incorporation of *p*-propargyloxyphenylalanine (*p*PrF, **1**) or *p*-ethynylphenylalanine (*p*EtF, **2**) into residue 66 of GFP (Figure 5). This provides a terminal alkyne reaction handle for reaction with different chemical moieties via the biorthogonal Glaser-Hay reaction. The resulting diyne linkage is highly conjugated, and a prime candidate to introduce larger conjugation patterns into GFP’s fluorophore without the need to evolve a new aaRS or use large UAAs. Moreover, the *p*EtF (**2**) is directly conjugated with the aromatic ring of the UAA, allowing for a comparison of the conjugation between the different UAAs. We hypothesized that the altered conjugation and chemical properties around the fluorophore will lead to new photophysical properties, demonstrating the utility of terminal alkyne containing UAAs and biorthogonal chemistry in a chemically programmable protein engineering strategy. Herein we report our findings on utilizing the Glaser-Hay reaction on GFP’s fluorophore to introduce novel function.

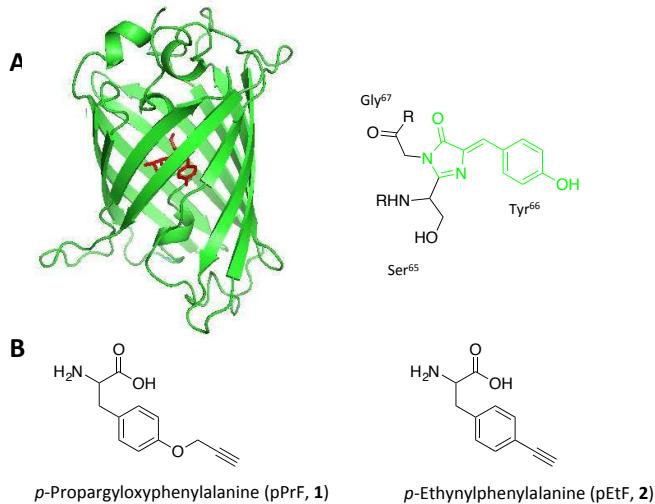


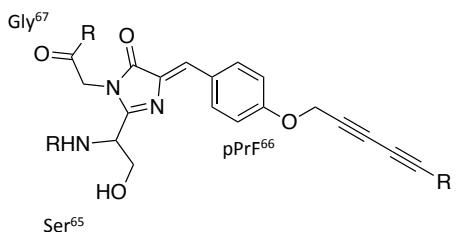
Figure 32. Modification of GFP function via UAA mutagenesis. A) GFP B-barrel (green) with the chromophore (red) and chromophore structure. B) Unnatural amino acids employed as bioconjugation handles.

In order to obtain protein possessing an alkynyl moiety, a GFP plasmid harboring a TAG mutation at position 66 was co-transformed with the polyspecific *p*CNF-aaRS/tRNA pair. Conveniently, this aaRS is capable of recognizing both **1** and **2** and for expression of an alkyne-containing GFP.<sup>5</sup> As the alkynyl UAA is incorporated into position 66 in GFP's fluorophore, the extended conjugation afforded by the UAA alters the spectral properties of GFP. To assess that *p*PrF-GFP<sub>TAG66</sub> was successfully produced, spectra for the GFP-variant were compared to the wild type. These spectra exhibited blue-shift, in agreement with the literature precedent, in the *p*PrF variant relative to the wild type. Incorporation was also confirmed by SDS-PAGE analysis of protein expression in the presence and absence of the UAA. A similar expression was performed using the *p*CNF aaRS/tRNA pair and *p*EtF to produce a separate GFP mutant with a bioconjugation handle. Conveniently, due to the modularity of this approach, only a single protein expression is necessary and all functional modification can be achieved synthetically.

This is in contrast to previous experiments, which required an individual protein expression for each UAA in order to modify function.

With a *p*PrF-GFP<sub>TAG66</sub> and *p*EtF-GFP<sub>TAG66</sub> in hand, we then sought to use our previously reported biorthogonal Glaser-Hay reaction to install new and varied chemical functionality into the chromophore of GFP (Figure 2). To investigate, we performed biorthogonal Glaser-Hay reactions on the chromophore's alkyne handle to couple aliphatic, aromatic, and alcohol containing alkynes. Briefly, Glaser-Hay reactions were performed by using a working concentration of 500 mM of CuI and TMEDA in the presence of alkyne-UAA bearing GFP<sub>TAG66</sub> and the cognate alkynyl partner (scheme 1). Reactions were run for 4 hr at 4°C and then 10 µL of the reacted GFP was placed in 3 mL of phosphate-buffered saline solution (pH ~7.2) for analysis using fluorescence spectroscopy. Gratifyingly, our initial attempts with aliphatic terminal alkynes worked well. Furthermore, the different characteristics of the alkyne moieties installed successfully shifted the fluorescence profile away from the parent alkynyl-GFP<sub>TAG66</sub> spectra, each in a unique way. While it might be expected that the requisite for the reaction to occur within the β-barrel of GFP may hinder this reaction from occurring, we hypothesize that the hydrophobic nature of interior of GFP actually promotes the soluble alkyne localization, thereby facilitating the reaction by increasing effective concentration. Additionally, SDS-PAGE analysis with Coomassie revealed that the Glaser-Hay reaction only minimally altered protein concentration suggesting only minimal protein degradation.

**A**



**B**

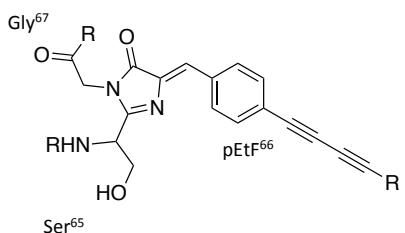
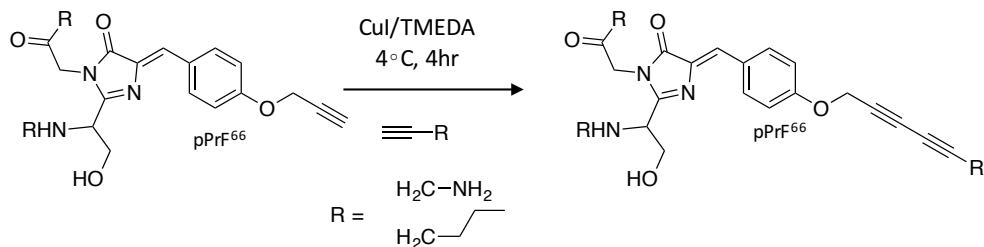


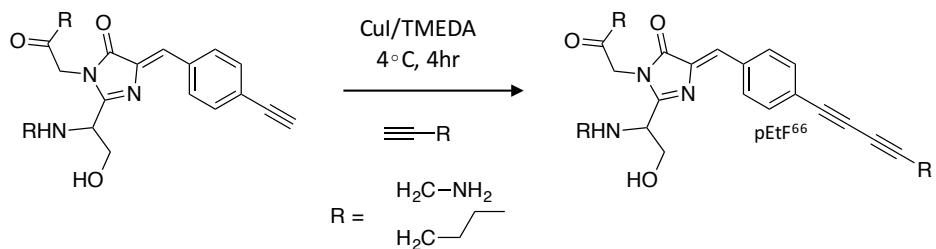
Figure 33. Structures of the fluorophores generated when performing a Glaser-Hay reaction with the two UAAs. A) Product from the coupling of a terminal alkyne to a GFP mutant containing **1**. B) Product from the coupling of a terminal alkyne to a GFP mutant containing **2**.

Scheme 9

**A**



**B**



We found that our initial Glaser-Hays on pPrF-GFP<sub>TAG66</sub> had different effects on the fluorescent profile of GFP (Figure 7A). Reacting 1-hexyne on the chromophore caused a general broadening and quasi-red shift of the fluorescent spectra. Reacting propargyl amine on the chromophore caused a slight band broadening, as well as potential increase in fluorescence intensity, likely due to the additional resonance available on the chromophore thanks to the lone pair on the amine group.

We next sought to explore the effects of reacting directly onto the phenyl ring of residue 66. This was feasible with the GFP mutant harboring **2**. Interestingly, the direct attachment of a terminal alkyne onto the ring resulted in an even greater blue shift of the pEtF-GFP<sub>TAG66</sub> compared to the both the *p*PrF and wild-type variants, likely due to the increased conjugation due to the direct attachment of that terminal alkyne (Figure 7B). Because of this shift, a different excitation wavelength was necessary, as 395nm was found to not excite the pEtF-containing chromophore. Based on absorption experiments, we selected 280 nm as the wavelength to excite the pEtF-GFP<sub>TAG66</sub> and all its Glaser-Hay derivatives. In the same fashion as the *p*PrF, the biorthogonal Glaser-Hay was performed on pEtF-GFP<sub>TAG66</sub> using the same reaction partners. Once again, the 1-hexyne was found to broaden the fluorescence spectra. Interestingly, the propargyl amine reaction had a drastic red shift relative to the pEtF parent chromophore. We believe this helps validate our initial speculation that resonance thanks to the lone pair on the amine has a drastic impact on the fluorescent properties of the chromophore, as in this instance the whole system is in direct conjugated.

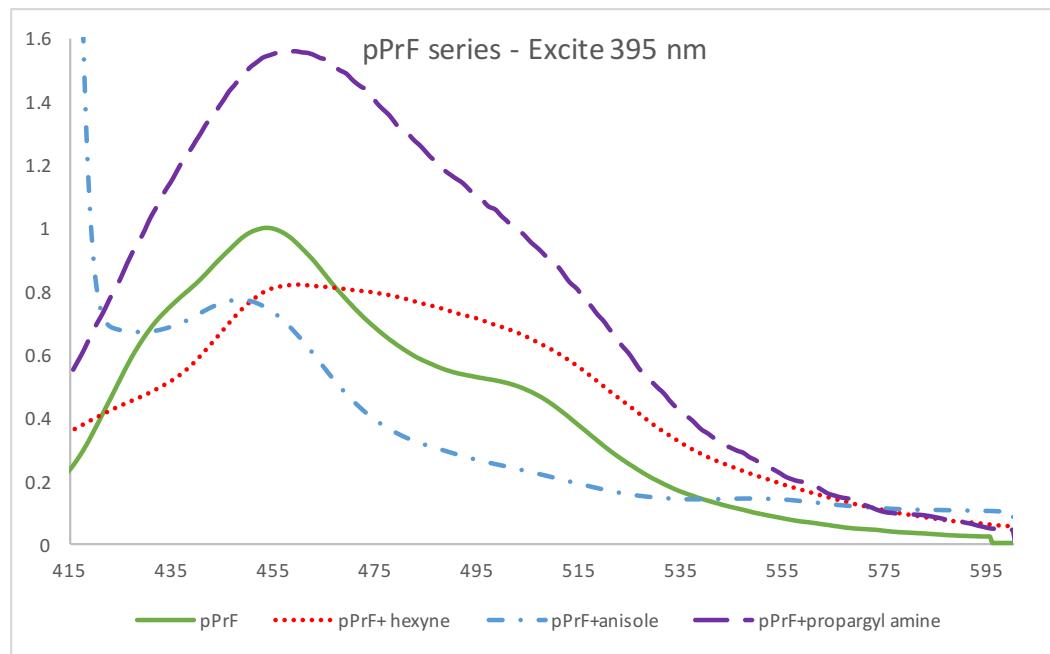
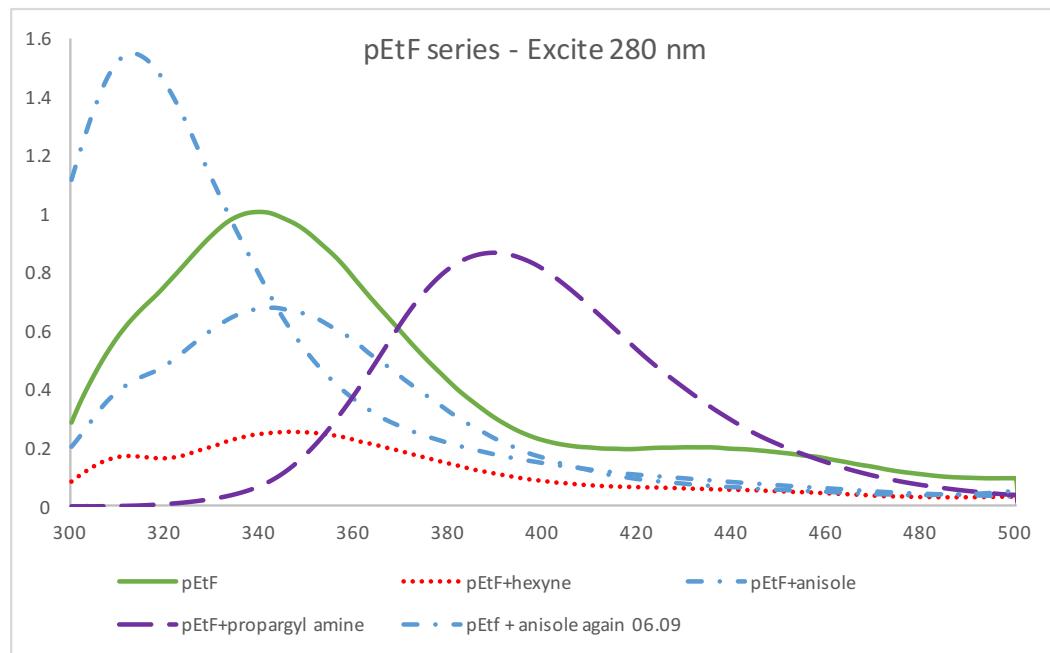
**A****B**

Figure 34. Fluorescent profile obtained after reacting various alkyne containing partners with pPrF or pEtF-containing GFP<sub>TAG66</sub>. A) Fluorescent spectra for Glaser-Hay modified pPrF-GFP chromophore. B) Fluorescent spectra for Glaser-Hay modified pEtF-GFP chromophore.

We have attempted Glaser-Hays with aromatic conjugation partners using the same reaction conditions as above. However, it was found that the signals obtained following these reactions were just for the free aromatic partner floating in solution. As such, we cleaned and concentrated these reactions using a 10 kDa molecular weight cut off spin column to remove any unreacted aromatic group for a test reaction between pEtF-GFP<sub>TAG66</sub> and 4-ethynylanisole. The data is still inconclusive, as the spectra does not appear to have shifted relative to the the parent pEtF chromophore (Figure 7B). We are worried that the aromatic alkynes may be shielded by solvent, perhaps preventing these alkynes from entering the hydrophobic region within GFP's B-barrel. We are continuing to attempt these Glaser-Hay couplings, and will be attempting a different reaction protocol with a greater concentration of protein to see if this enhances our yields, and thus signal, of the aromatic derivatized GFP chromophores.

## Conclusion

We have extended our work on the biological Glaser-Hay to utilize the biorthogonal chemistry to modulate protein function. Using two previously reported alkyne-containing UAAs within the chromophore of GFP (position 66), we have successfully performed the Glaser-Hay reaction on the chromophore of GFP. The resulting diyne linkage alters the fluorescence profile of GFP depending on the moiety attached to the terminal alkyne. Our future work seeks to extend the reaction to aromatic containing alkynes, which we hope will have a greater impact on GFP fluorescence due to the increased conjugation found in an aromatic system. Our findings highlight the

potential of biorthogonal chemistry, particularly diyne forming chemistries, to modulate protein function without the need for tedious selections and screens.

## Experimental

General Biological Glaser-Hay Protocol To 1.5 mL Eppendorf tube was added 3  $\mu$ L of a 500 mM CuI solution in water and 3  $\mu$ L 500 mM TMEDA solution in water. This mixture was then allowed to sit in the fridge for 10 mins. After waiting, 22  $\mu$ L of PBS was added, followed by 10  $\mu$ L of the alkyne containing GFP<sub>TAG66</sub> in PBS (~0.5mg/mL) and 4  $\mu$ L of a 40 mM solution of the cognate alkyne in DMSO. The mixture was allowed to react at 4°C for 4 hr.

Protocol for Fluorimetry Scans After reacting for 4 hr, 10  $\mu$ L of the reaction mixture was added to a quartz cuvette and diluted up to 3 mL with PBS. This was then excited at either 395 nm (for the pPrF-containing chromophore) or 280 nm (for the pEtF-containing chromophore) with a 10 nm slit width for the excitation and emission wavelengths. Slit widths were increased or decreased as necessary depending on the intensity of the reaction product's signal. The scan speed was set at 500 nm/min.

General Protocol for Biological Glaser-Hay with Aromatic Reaction Partners To 1.5 mL Eppendorf tube was added 3  $\mu$ L of a 500 mM CuI solution in water and 3  $\mu$ L 500 mM TMEDA solution in water. This mixture was then allowed to sit in the fridge for 10 mins. After waiting, 22  $\mu$ L of PBS was added, followed by 10  $\mu$ L of the alkyne containing GFP<sub>TAG66</sub> in PBS (~0.5mg/mL) and 4  $\mu$ L of a 40 mM solution of the cognate alkyne in DMSO. The mixture was allowed to react at 4°C for 4 hr. Following this, the unreacted

aromatic alkyne was washed away using a 10 MWCO spin column (Corning) and rinsing with 50  $\mu$ L portions of PBS 8 times. After the eight rinse, the solution was concentrated down to  $\sim$ 25  $\mu$ L, as indicated on the spin column. Of this cleaned solution, 10  $\mu$ L was placed into a quartz cuvette and diluted to 2 mL with PBS for fluorescence analysis.

## References

1. Brustad, E. M. & Arnold, F. H. Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **15**, 201–210 (2011).
2. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249 (2006).
3. Young, T. S. & Schultz, P. G. Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *J. Biol. Chem.* **285**, 11039–11044 (2010).
4. Maza, J. C., Villa, J. K., Landino, L. M. & Young, D. D. Utilizing Unnatural Amino Acids to Illustrate Protein Structure-Function Relationships: An Experiment Designed for an Undergraduate Biochemistry Laboratory. *J. Chem. Educ.* **93**, 767–771 (2016).
5. Young, D. D., Jockush, S., Turro, N. J. & Schultz, P. G. Synthetase polyspecificity as a tool to modulate protein function. *Bioorganic Med. Chem. Lett.* **21**, 7502–7504 (2011).
6. Maza, J. C., Jacobs, T. H., Uthappa, D. M. & Young, D. D. Employing Unnatural Amino Acids in the Preparation of Bioconjugates. *Synlett* **27**, 805–813 (2016).
7. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chemie - Int. Ed.* **48**, 6974–6998 (2009).
8. Lampkowski, J. S., Villa, J. K. & Young, D. D. Development and Optimization of Glaser-Hay Bioconjugations. *Angew. Chemie Int. Ed. English (pending Revis.)* (2015).
9. Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* **38**, 2865–2875 (2009).

## **CHAPTER 5: IMMOBILIZATION OF THE RNA-GUIDED ENGINEERED NUCLEASE CAS9**

RNA-guided engineered nucleases (RGEN) represent a novel form of targeted gene editing tools. In these systems, an RNA molecule provides targeting specificity to a protein endonuclease via RNA:DNA complementarity<sup>1,2</sup>. Synthetic oligonucleotides are commercially available, and the ease with which these systems can be programmed to cut DNA has allowed for rapid developments in the field of molecular biology. The facility with which DNA editing can be programmed has allowed for rapid new developments in transgenic models that were previously tedious or inaccessible due to embryonic lethality. For example, a recent technique employs Cas9 deleting adult-cardiac specific genes, developing new adult mouse models for cardiac disease that were previously inaccessible due to embryonic lethality.<sup>3</sup> In addition, the ability to selectively excise portions of DNA has made the technology an attractive new therapeutic. A recent report in *Nature: Scientific Reports* detailed the use of CRISPR/Cas9 to remove latently infected HIV-1 from T-lymphocytes, effectively providing a cure for HIV/AIDS.<sup>4</sup> The clustered regularly interspaced short palindromic repeats (CRISPR), which act in association with CRISPR-associated (Cas) genes, is a type of RGEN that has gained recent attention as a new tool for genetic manipulation.

Within the system, the Cas endonuclease, called Cas9, acts in conjunction with a CRISPR RNA (crRNA), which provides specificity by base pairing with the DNA target, in effect serving as a bacterial adaptive immunity. When a bacterium survives a viral

infection, it stores a short segment of the viral genome in the CRISPR region of the bacterial genome; subsequent infection activates transcription of this short viral segment which serves as a crRNA, directing the Cas9 nuclease to destroy the viral genome (Figure 1A). In addition to the crRNA, a trans-activating CRISPR RNA (tracrRNA) is required to allow proper interaction between the crRNA and the DNA (Figure 1A). Recently, the Cas9 from *Streptococcus pyogenes* has been made amenable to laboratory use (Figure 1B)<sup>2</sup>.

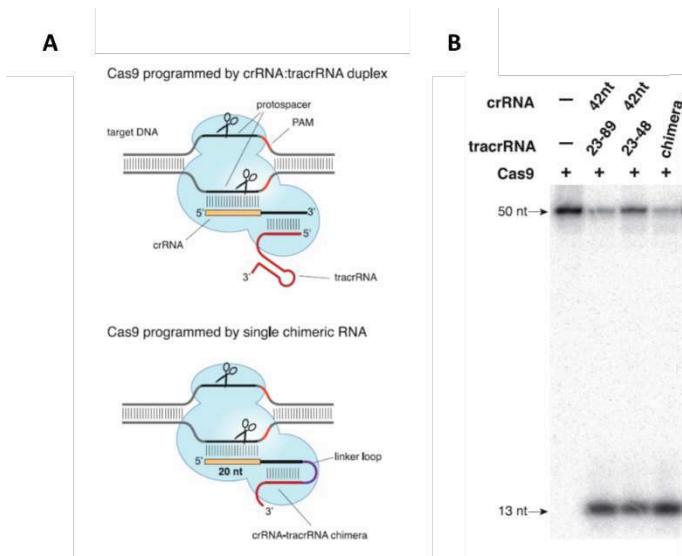
Oligonucleotides can be easily designed and synthesized to facilitate directed Cas9 targeting of specific DNA sequences in a range of organisms. Furthermore, the ease with which DNA binding can be programmed with Cas9 is also making the technology an attractive option as a new biomolecular tool. For example, CRISPR interference (CRISPRi) has become a valuable new technique for sequence specific gene transcriptional modulation. By using a catalytically inactive variant of Cas9 (also called a “dead” or dCas9), researchers were able to sterically hinder transcription of a gene targeted by a guide RNA.<sup>5</sup> The technique has even been extended to transcriptional activation (CRISPRa) by fusing dCas9 to viral transcriptional activators, such as VP64.<sup>6</sup> In this example, the protein serves as a platform for the docking of transcription activators, and is directed to the gene of interest by the sequence of the guide RNA. Even more recently, researchers were able to generate a variant of Cas9 containing the ligand-binding domain of human estrogen receptor- $\alpha$  (ER-LBD).<sup>7</sup> Upon binding of the antagonist 4-hydroxytamoxifen, the ER-LBD undergoes a conformational change bringing the N and C termini of the domain closer together, thereby rendering the Cas9 active once again. This mechanism was shown to serve as an allosteric switch for the exogenous

control of either Cas9 or dCas9, granting an even greater degree of user-control over genetic manipulation.

Cas9, therefore, represents a new tool to perform user defined genetic manipulations. An ideal requisite for genetic manipulation is the propensity of the RGEN to have maximal on-target cleavage with no off-target cleavage. Given that Cas9 is directed by RNA:DNA complementarity, and that there are often redundant sequences within the entire genome or between homologous chromosomes, there is growing concern that Cas9 mediated cleavage may have off-target effects.<sup>8</sup> However, recent high-throughput studies have provided new insight into Cas9 specificity that will lead to optimal guide RNA design.

### **Guide RNA mismatch is linked to off-target cleavage**

Early studies on the CRISPR/Cas9 system sought to elucidate the role of the crRNA and tracrRNA in mediating directed DNA cleavage. By using an electrophoretic mobility shift assay, researchers were able to demonstrate that the constant tracrRNA is necessary for crRNA binding to target DNA, presumably by helping to orient the crRNA via formation of a hairpin loop structure. By sequentially truncating the tracrRNA, it was then possible to determine the minimum tracrRNA sequence required to maintain DNA cleavage. This was then fused to the last 20 nucleotides of the crRNA 3' end, which were shown to provide the system's specificity to target DNA (see Fig. 1). The resulting chimeric single guide RNA (sgRNA) produced targeted DNA cleavage at levels comparable to the naturally occurring system (see Fig. 1)<sup>2</sup>.



**Figure 35. (A)** Chimeric single guide RNA design mimics critical RNA structures in the crRNA:tracrRNA duplex. The yellow bar denotes the 20 nucleotide targeting sequence in the sgRNA, which can be programmed by the user. Presence of the PAM 5'NGG3' sequence is a restriction on sgRNA design, as only sequences 5' of the PAM can be targeted. **(B)** Chimeric sgRNA+Cas9 cuts a DNA template as efficiently as the crRNA:tracrRNA+Cas9 system. This is observed via the presence of a cleaved 13 nt sequence after gel electrophoresis. Adapted from Jinek et al., 2012.

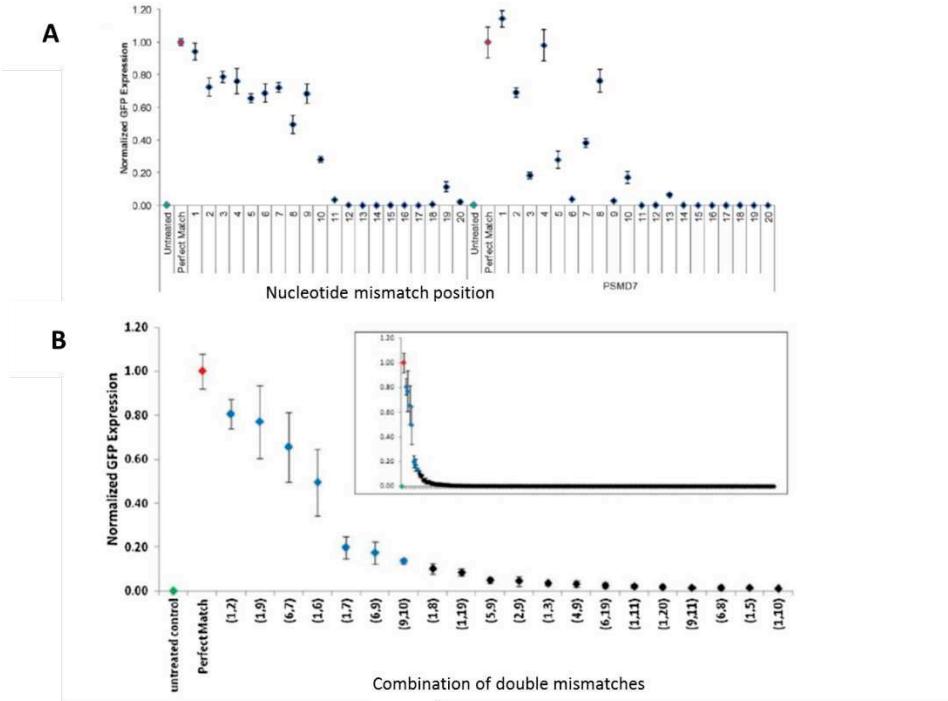
In theory, the sgRNA would allow programmed Cas9 cleavage to any DNA target by base pair complementarity. However, additional studies revealed the requisite of a protospacer adjacent motif (PAM) on the 3' end of the target DNA sequence in the form of 5'-NGG-3' (see Fig. 1). Even a single mutation in the GG sequence abolishes Cas9 cleavage<sup>2</sup>.

While these studies demonstrated programmable Cas9 cleavage, it was still unclear if there would be any off-target effects that would minimize the specificity and thus applicability of the enzyme. However, this requires scientific ingenuity, as exploring every positional mismatch within the sgRNA and the whole genome is incredibly challenging. Early studies focused on examining mismatches within the sgRNA targeting sequence towards a target gene. This served as a proxy for off-target cleavage within the genome. Researchers designed sgRNAs targeting the human hemoglobin  $\beta$  (*HBB*) gene

with different mismatches. Using the T7E1 assay (see Fig. 2), the authors demonstrated that the mismatched sgRNAs could lead to indel formation in the *HBB* gene, indicating non-perfect complementarity is permissible for Cas9 cleavage (Cradick et al., 2013). These findings were corroborated in a functional assay that relied on Cas9 disruption of GFP via indel formation. Relative to the perfectly matched sgRNA, the authors observed that the sgRNA tolerated single and double mismatches within the 20 nucleotide specificity sequence, especially if they were in the 5' end most distal from the PAM<sup>8</sup>.

### **Current findings reveal less promiscuity in cleavage**

Unlike previous studies, researchers generated guide RNAs with single and double mismatches at every position and combination in the RNA's 20 nucleotide targeting sequence using automated oligonucleotide synthesis. Using a functional assay in which fluorescence serves as a metric for off-target cleavage, the researchers performed a comprehensive analysis of guide RNA positional mismatch effects. Once again, their findings revealed that the most PAM distal nucleotides were tolerated for off-target cleavage. However, upon exploring every possible double mismatch combination, they found that only sequential double mismatches at this distal site had any real effect on off-target cleavage (see Fig. 2). Triple mismatches abolished off-target cleavage in this study. Their data further validate that mismatches between the sgRNA and off-target sites are only tolerated at the most PAM distal nucleotides within the target sequence. Furthermore, the authors argued that off-target effects would be minimized by the requisite combination of favorable mismatches adjacent to a PAM sequence<sup>10</sup>. This positional data can be employed in the future to design less promiscuous sgRNAs.

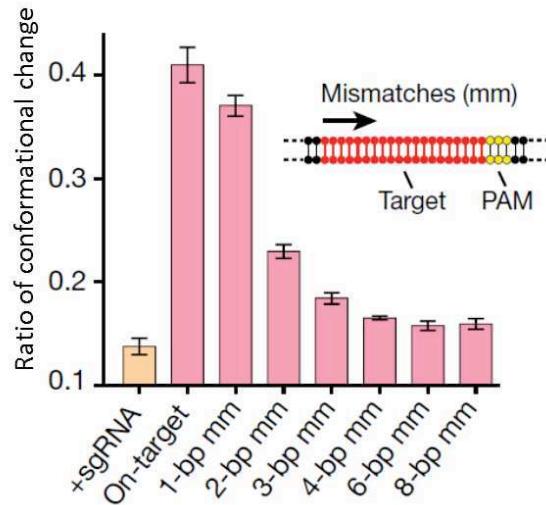


*Figure 36. Positional effect of mismatches between the guide RNA and the DNA template as detected via a functional Ub-GFP assay. Numbers denote position of the mismatch, with 1 corresponding to the nucleotide most distal from the PAM and 20 corresponding to the nucleotide directly adjacent to the PAM. All GFP values are normalized to the perfect match levels. (A) As previously observed, nucleotides more distal from the PAM sequence tolerate mismatches more readily, as observed by an increase in GFP fluorescence. (B) Double mismatches are less permissible than single mismatches, with a sequential mismatch in nucleotides 1,2 tolerated the most. The inset demonstrates all possible combinations of double mismatches within the guide RNA, demonstrating that vast majority of double mismatches are not tolerated. Adapted from Anderson et al., 2015.*

These authors' conclusions were further validated by a powerful new assay capable of capturing Cas9 off-target cleavage sites within the entire genome. This technique, termed Digenome-seq, allowed the authors to probe the entire genome for off-target cleavage of a sgRNA targeting the *HBB* gene. Using Digenome-seq, the authors were able to capture the on-target site as well as two previously validated off-target sites. Another 122 other potential off-target sites were also captured. However, whole genome sequencing (WGS) of Cas9 transfected cells revealed that these 122 other sites were not accompanied by indels, suggesting that mutations at these off-target sites are rare. Based

on their findings, off-target cleavage would not be expected to exert an effect on the cell<sup>11</sup>. In comparison to previous techniques, this method was the first to allow researchers to probe the genome for off-target sites of a sgRNA they wish to employ in the future.

Recently, the Doudna group revealed insight into the mechanism of Cas9:DNA binding that negates off target cleavage. By using Forster resonance energy transfer (FRET) experiments to explore conformational changes in Cas9 upon binding DNA, the group revealed that Cas9 adopts a unique conformation with perfect sgRNA:DNA complementarity (see Fig. 3). Mismatches, even in the distal portion of the sgRNA targeting sequence, have distinct conformations<sup>S</sup> that presumably disfavor enzymatic cleavage. This provides structural evidence for a conformation-dependent proofreading mechanism for Cas9 based on the complementarity of the sgRNA and the DNA. As such, it is thought that this proofreading mechanism helps minimize off-target cleavage (Sternberg et al, 2015).



*Figure 37. FRET experiments involve attaching two fluorophores to a protein. As the protein undergoes conformational changes, the distance between the fluorophores changes. This alters their fluorescence intensity which serves as a metric for structural changes within the protein. FRET data from the Doudna group demonstrates a conformational control of DNA target cleavage in Cas9. In comparison to the perfect complementarity between the sgRNA and the*

*target DNA (on-target), mismatches demonstrate a different degree of conformational change (y-axis). Adapted from Sternberg et al., 2015.*

## DNA-Fishing using Cas9

The ease with which Cas9 can be programmed to target certain genetic loci has made it an increasingly attractive tool, especially towards molecular genetics. Recent work has established a methodology to use an inactive form of Cas9 (“dead” or dCas9), which can still bind specified DNA sequences using a gRNA, to pulldown specific genetic loci. These loci are pulled down with their associated proteins, allowing researchers to efficiently isolate certain genetic regions and their associated proteins, expanding our limited understanding of the varied roles proteins play in regulating expression.<sup>13</sup> The technique, termed engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enCHIP), was applied to the human interferon (IFN) regulatory factor-1 (hIRF-1) and was demonstrated to pull down new and validated protein targets.<sup>14</sup> While the technique works and exhibits minimal off-target effects, it requires binding of the dCas9 complex to the DNA *in vivo*, perhaps precluding certain protein factors that may interact at that region or altering the favorability for other regulators to bind.

Due to the increasing widespread applicability in a variety of molecular biology endeavours, we sought to develop a UAA-containing Cas9. Our hope was that by incorporating UAAs into Cas9, we could generate either a photoactivatable Cas9, using photocaged UAAs, or immobilize Cas9 onto a solid support for DNA-pulldown applications. It is our hope that the by installing new biochemistries into Cas9 via UAAs that will further increase the applicability of this technology for a more diverse array of applications.

## Results and Discussion

Our group has already established a protein immobilization protocol using UAA technologies. Immobilization of Cas9 on a bead may allow specific targeting of certain genetic loci, producing a more complete picture of the role of protein function in regulating DNA, as well as affording a new means to purify genetic material of interest. Cas9 immobilization to a resin or bead may also allow for the regeneration of the system, allowing researchers to purify defined DNA sequences multiple times.

Initially, sites within the amino acid sequence of the Cas9 enzyme that would amenable to UAA-mutagenesis were identified. As a consequence of most bacterial aaRSs having been evolved for tyrosine derivatives, we relegated our investigation to tyrosine residues within the protein. In addition, the search was further constrained by the necessity to utilize a surface exposed tyrosine residue that did not perturb canonical Cas9 DNA cleavage. It was our hope that this exposed residue could serve as a handle for biorthogonal functionalization.

Our initial search identified Tyr1265 as a prime candidate for immobilization, as it is distal to Cas9's catalytic domains and is surface exposed (see Fig 4). Our choice was further validated by a recent study that sought to identify engineering hotspots within Cas9.<sup>7</sup> It was found that amino acid residues in the CTD of Cas9 were particularly amenable to alteration without drastically impairing protein function. While Y1265 was not directly explored, residues 1262 and 1267 were, and they both continued to exhibit functionality, albeit a bit decreased. Moreover, because the screen was performed by insertion of a large protein domain, we hypothesize that the small size of the UAA insertion will have a minimal effect on Cas9 function.

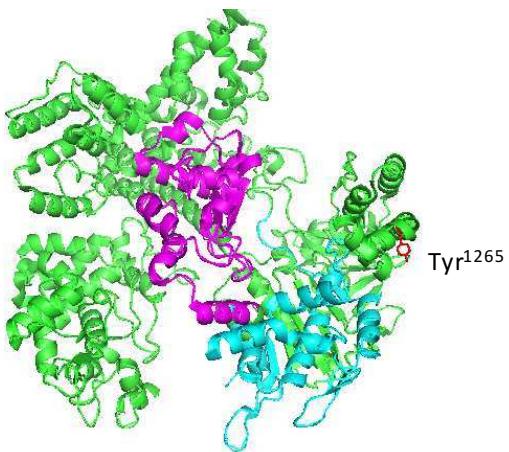


Figure 38. Cas9 with HNH (magenta) and RuvC (cyan) highlighted. Proposed site of UAA mutagenesis (Y1265) is shown in red.

Having selected Tyr1265 as the ideal site for UAA incorporation, we then designed primers for PCR mutagenesis, converting Y1265 into a TAG for UAA-mediated amber suppression. PCR mutagenesis using Kappa Hi-Fidelity polymerase was performed according to the Kappa protocol. Subsequent sequencing and alignment of the plasmid obtained from this mutagenesis showed that we had indeed inserted a TAG into position 1265, generating a Cas9Y1265X variant.

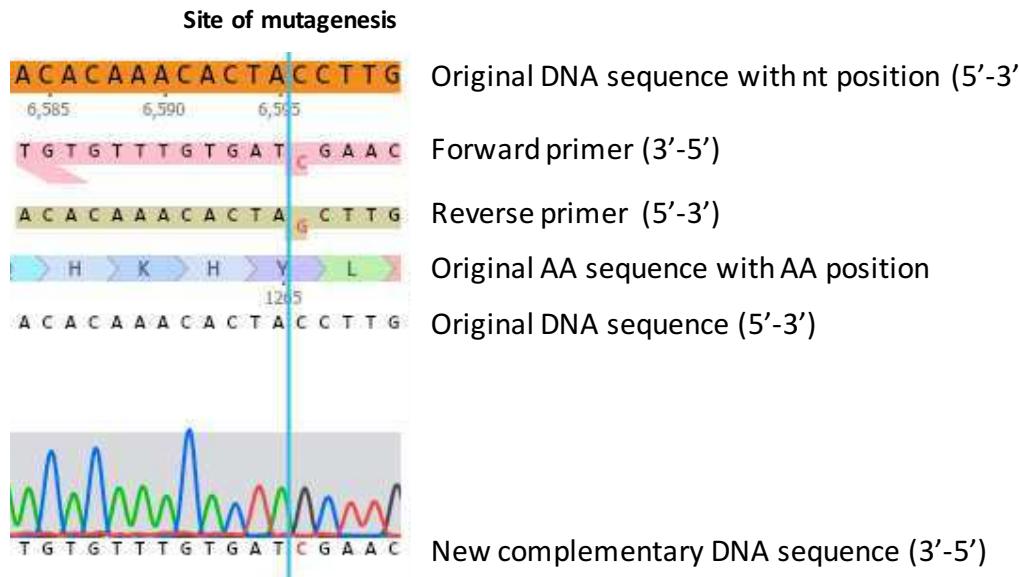


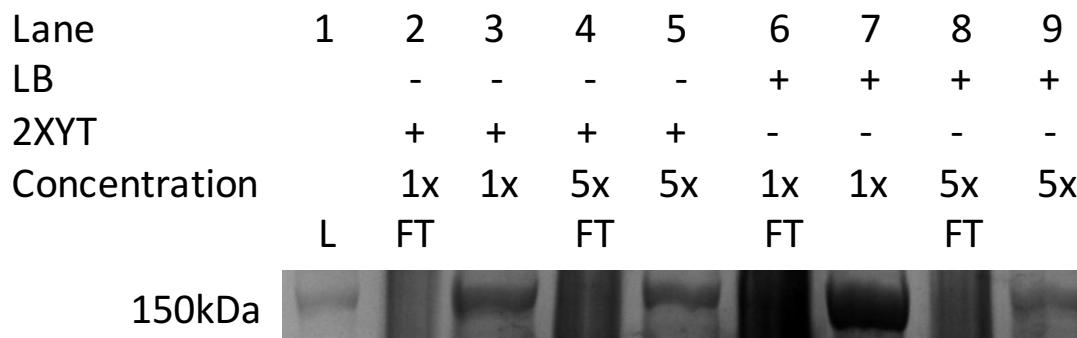
Figure 39. Results from sequencing and alignment for our PCR mutagenesis. Flow chromatogram on the bottom highlights in the red the incorporation of our TAG mutation on the complementary DNA sequence (3'-5').

Following insertion of the TAG mutation, it was then necessary to optimize our expression protocol. Cas9 is an approximately 160kDa protein, and thus expression represents a serious energetic toll for the cell. Combining the expression of this large protein with the added layer of UAA incorporation increases the complexity of the process. To optimize UAA incorporation into Cas9Y1265X we performed an expression experiment with the pET-Cas9Y1265X and the pEVOL-WT plasmids. pEVOL-WT is the wild type aaRS from *M. jannaschii*, and will charge tyrosine to the cognate tRNA with an antisense codon for the TAG stop codon. Thus, after this process the protein produced is the wild type Cas9, but amber suppression is required. This method, therefore, serves as a good proxy for our capability to incorporate a UAA without the added complexity of a UAA in the system.

To optimize Cas9Y1265X expression, we performed expression experiments using 100mL expression cultures of *E. coli* BL21(DE3). These cultures were either

grown in Luria-Burtani (LB) broth or 2x Yeast tryptone (2XYT), a more nutritionally rich broth that is often advantageous for the expression of large proteins. In addition, we also explored the effects of concentration our expression cultures prior to induction, as our previous work with GFP has shown that a 5x concentration enhances protein yields while minimizing the amount of UAA required.

Our results from the expression experiment show that we were indeed able to generate wild type Cas9 using the amber suppression method. To our surprise, the 1x LB expression protocol performed best. We hypothesize that the large demands imposed by the expression of such a large protein makes our concentration protocol very demanding on the cell, as it now finds itself in a nutritionally diminished environment while expressing Cas9. In addition, the 2XYT expressions did not outcompete the 1x LB version, perhaps because the increased nutritional supplies of the 2XYT broth allowed the cells to focus energetic resources towards continued growth and division instead of increasing cellular size and protein expression.



*Figure 40. Protein expression experiment for pET-Cas9Y1265X and pEVOL-WT. Our data indicates that the 1x LB protocol performs best, as evidenced by denatured SDS-PAGE following staining. FT in the figure refers to flow-through, which demonstrates that we have captured all the Cas9 produced by the cells.*

With an optimized expression protocol for the incorporation of a UAA into Cas9Y1265X, we next sought to produce a UAA-containing Cas9. A pET-Cas9Y1265X

was co-transformed with the polyubiquitous *pCNF*-synthetase. The 1x LB protocol was then used with 100 mL expression cultures either in the presence or absence of the UAA *pAzF*, introducing an azide functionality into the protein. Gratifyingly, our data show that the UAA is incorporated (see Fig 7, Lanes 3-5) in the presence of UAA. Furthermore, in the absence of UAA, no protein is produced, as the cell encounters an early stop codon in protein translation (see Fig 7, lanes 7-9). Furthermore, no truncated protein products can be detected since our Cas9 contains a C-terminal hexaHis tag, meaning only completely successful read-throughs produce our purification tag.

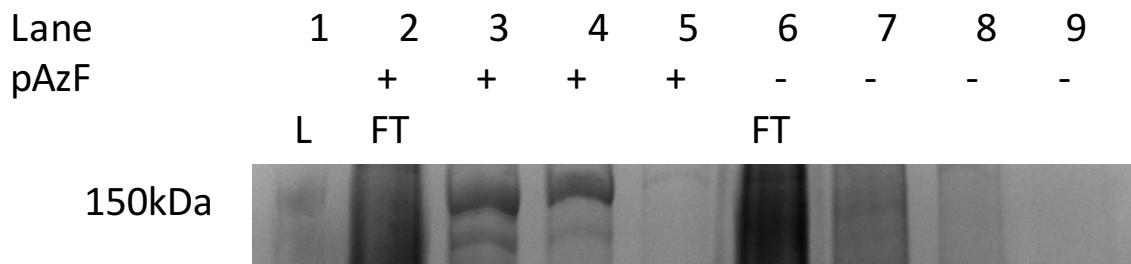


Figure 41. Expression of a *pAzF*-containing Cas9Y1265X.

## Experimental

General Protocol for PCR Mutagenesis PCR mutagenesis was performed in a Bio-Rad iCycler 96 well reaction module thermocycler. The protocol was adapted from the Kappa Polymerase HiFi PCR mutagenesis protocol. Two, 1x master mix for the PCR reaction were created by adding 5  $\mu$ L Kappa HiFi Buffer, 0.75  $\mu$ L 10 mM dNTPs, 0.75  $\mu$ L, 0.75  $\mu$ L of 10 mM of forward primer dissolved in sterile deionized water, 0.75  $\mu$ L of 10 mM of reverse primer also dissolved in sterile deionized water, and 12.25  $\mu$ L of sterile deionized water. To the reaction master mix was added 0.5  $\mu$ L of kappa polymerase at the very end, prior to starting the reaction in the thermocycler. To the control master mix

was added 0.5  $\mu$ L of sterile deionized water in place of the kappa polymerase. This recipe for the 1x master mix was increased by 2.5x or 5x depending on the number of reactions that were being run. For example, for the PCR mutagenesis converting Tyr1265 to TAG1265, four experimental PCR tubes were prepared, two with 5 ng of pET28b-Cas9-His (5  $\mu$ L of 1ng/ $\mu$ L of template DNA) and two with 10 ng of pET28b-Cas9-His (5  $\mu$ L of 2ng/ $\mu$ L of template DNA). To these PCR tubes was then added the corresponding reaction or control master mix, which was then loaded into the thermocycler using the following protocol: 1) initial denaturation at 95°C for 2 mins; 2) denaturation at 98°C for 2 mins followed by annealing at 65°C or 15 secs [note, 2°C were added to whatever the IDT reported annealing temperature was because the Kappa HiFi buffer is a high salt buffer that increases annealing temperatures] followed by an extension at 72°C for 4.5 mins (30 sec/kbp), this step was cycled for a total of 20 cycles; 3) a final extension was performed at 72°C for 5 mins; 4) the thermocycler completed the run by holding at 4°C. Following the PCR mutagenesis, the samples then underwent a DpnI digestion at 37°C for 2 hours to remove any methylated template DNA. The DpnI was deactivated at 80°C for 15 mins, and the samples were then PCR cleaned and concentrated (Zymo Research DNA clean and Concentrator Kit) and eluted in 10  $\mu$ L of sterile deionized water. Two microliters of each sample were then transformed into competent *E. coli* (BL21(DE3)), recovered in 500  $\mu$ L of 2XYT media for 1 hour, and then this full volume was plated on agar plates supplemented with kanamycin. After allowing to grow overnight, plates were compared for differential growth between the reaction and control. If differential growth was observed, four colonies were picked at random in four different quadrants of the plate. These were grown to confluence overnight at 37°C, a glycerol stock was prepared

using 1 mL of culture, while the rest was used to extract DNA for sequencing and expression tests.

## Primers Designed for PCR Mutagenesis of pET28b-Cas9-His

Primer ID	Mutation	Explanation	Template Plasmid	IDT Calculated Tm	Verified Tm	Sequence
A9	D10TAG F	to convert Cas9 to dCas9 via inactivation of catalytic D(restor with YOBNY?)	pET28b-Cas9-His	67.3		aagttactccattggctctatgcggcacaacaggcgt
A10	D10TAG R		pET28b-Cas9-His	67.3		caaggcttttgtccqatcgaacccataaaggtaactt
A11	Y72TAG F	Tyr interacting with DNA base (pi-pi stacking)	pET28b-Cas9-His	68.0	68.0	aaacacgacggcgcagatagaccggcagaa
A12	Y72TAG R		pET28b-Cas9-His	68.0	68.0	aaacacgacggcgcagatagaccggcagaa
A13	Y450TAG F	Tyr inactivated in dCas9???	pET28b-Cas9-His	64.5		cattttggataccatgtatgttagccccctcg
A14	Y450TAG R		pET28b-Cas9-His	64.5		ccggggggcclacatatacggtatccqaaatg
A15	Y1131TAG F	Tyr near DNA	pET28b-Cas9-His	66.2	68.2	atggatgggaccccaagaataggggatgttcgtca
A16	Y1131TAG R		pET28b-Cas9-His	66.2	68.2	tccatgttcgcgccttcttgggtccaaatct
A17	Y1265TAG F	Tyr surface exposed and away from DNA/HNH and RuvC catalytic domains	pET28b-Cas9-His	63.2	65.0	ACACAAACACTAGCTTGTAGATGAGATCATCGAGCAAATAAGC
A18	Y1265TAG R		pET28b-Cas9-His	63.2	65.0	GCTTATTGCTGTCATGATCTCAAGCTAGTTGGTGT

## Sequencing of PCR Mutagens

Samples were sent off to sequencing to GeneWiz using their “premixed” DNA Sequencing Services following their sample submission guidelines. For the Cas9Y1265X mutagen the universal T7 terminator primer was selected for sequencing. This primer will be utilized to sequence Cas9Y1131X as well. For mutagens Cas9D10X and Cas9Y450X the universal T7 promoter primer will be used for sequencing. The Cas9Y450X mutant is located too far into the DNA sequence of the Cas9 gene for reliable sequencing using either of the universal primers above, as such we will design a sequencing primer nearer to this mutation site.

Optimization Cas9TAG Cotranslational UAA Incorporation A pET-Cas9Y1265X  
plasmid (0.5  $\mu$ L) was co-transformed with a pEVOL-WT plasmid (0.5  $\mu$ L) into Escherichia coli BL21(DE3) cells using an Eppendorfeporator. The cells were then plated and grown on LB agar in the presence of chloramphenicol (34 mg/mL) and kanamycin (50 mg/mL) at 37°C overnight. One colony was then used to inoculate LB media (5 mL) containing both kanamycin and chloramphenicol. The culture was incubated at 37°C overnight and used to inoculate an expression culture. Expression cultures were either grown in LB media (100 mL 50 mg/mL Kan, 34 mg/mL Chlr) or 2XYT media (100 mL

50 mg/mL Kan, 34 mg/mL Chlr) at an OD600 0.1. The cultures were incubated at 37 °C to an OD600 between 0.6 and 0.8. Protein expression was either induced immediately via addition 20 % arabinose (100 µL) and 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; 100 µL) or the cultures were concentrated 5x to a volume of 20 mL and expressed via addition of 20% arabinose (20 µL) and IPTG (20 µL). The cultures were incubated at room temperature for 24 h then centrifuged at 5,000 rpm for 20 minutes, as the larger expression cultures required longer centrifuge times to ensure all the cells were pelleted. The pelleted cells were then stored at -80 °C for 20 mins. The cell pellet was resuspended using 500 µL of Bugbuster (Novagen) containing lysozyme, and incubated at room for 30 minutes. The solution was then centrifuged at 5,000 rpm for 20 mins. The supernatant was added to an equilibrated His- pur Ni-NTA spin (Qiagen) column and Cas9 was purified according to manufacturer's protocol. Purified GFP was analyzed by SDS-PAGE (BioRad 6% precast gels, 150V, 1.5h), and employed without further purification.

Expression of UAA-Containing Cas9Y1265X A pET28b-Cas9Y1265X-His plasmid (0.5 µL) was co-transformed with a pEVOL-pCNF plasmid (0.5 µL) into Escherichia coli BL21(DE3) cells using an Eppendorf eporator. The cells were then plated and grown on LB agar in the presence of chloramphenicol (34 mg/mL) and kanamycin (50 mg/mL) at 37 °C overnight. One colony was then used to inoculate LB media (10 mL) containing both kanamycin and chloramphenicol. The culture was incubated at 37 °C overnight and used to inoculate an expression culture (100 mL LB media, 50 mg/mL Kan, 34 mg/mL Chlr) at an OD600 0.1. The cultures were incubated at 37 °C to an OD600 between 0.6 and 0.8, and protein expression was induced by addition of UAA (1 mL, 100 mM) and 20 % arabinose (100 µL) and 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; 100

$\mu$ L). The cultures were incubated at room temperature for 16–20 h then centrifuged at 5,000 rpm for 10 minutes and stored at -80 °C. Purification was performed according to the protocol above using a Ni-NTA resin.

## References

1. Savić, N. & Schwank, G. Advances in therapeutic CRISPR/Cas9 genome editing. *Transl. Res.* (2015). doi:10.1016/j.trsl.2015.09.008
2. Jinek, M. *et al.* A Programmable Dual-RNA – Guided. **337**, 816–822 (2012).
3. Carroll, K. J. *et al.* A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. *Proc. Natl. Acad. Sci.* **113**, 201523918 (2015).
4. Kaminski, R. *et al.* Elimination of HIV-1 Genomes from Human T-lymphoid Cells by CRISPR/Cas9 Gene Editing. *Sci. Rep.* **6**, 22555 (2016).
5. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
6. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat Meth* **10**, 977–979 (2013).
7. Oakes, B. L. *et al.* Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat. Biotechnol.* (2016). doi:10.1038/nbt.3528
8. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–6 (2013).
9. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* **41**, 9584–9592 (2013).
10. Anderson, E. M. *et al.* Systematic Analysis of CRISPR-Cas9 Mismatch Tolerance Reveals Low Levels of Off-Target Activity. *J. Biotechnol.* **211**, 56–65 (2015).

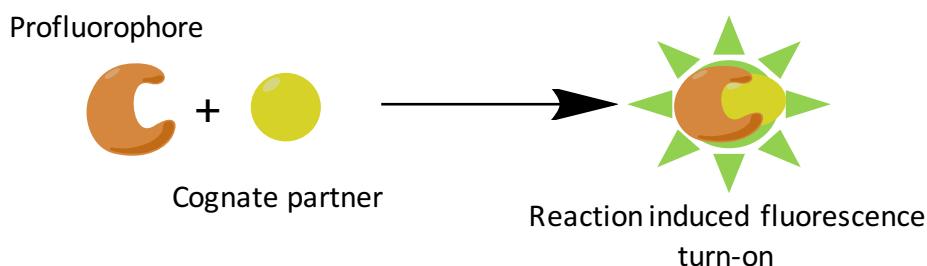
11. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* **12**, 237–43, 1 p following 243 (2015).
12. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. a. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110–113 (2015).
13. Fujita, T. & Fujii, H. Isolation of specific genomic regions and identification of associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Chromatin Protoc. Third Ed.* **439**, 43–52 (2015).
14. Fujita, T. & Fujii, H. Identification of proteins associated with an IFN??-responsive promoter by a retroviral expression system for enChIP using CRISPR. *PLoS One* **9**, 1–9 (2014).

## CHAPTER 6: TOWARDS THE DEVELOPMENT OF GLASER-HAY “TURN-ON” PROBES

In order for a fluorescent probe to prove beneficial, it must possess certain key features. The first constraint is size, as the probe must be small enough to enter the cell, yet large enough to provide enough information for specificity and a prolonged signal.<sup>1</sup> Additionally, the signal must occur selectively where the probe is meant to bind, with minimal interfering background.<sup>1</sup> Most importantly, the fluorescent probe must bind exclusively to the substrate of interest.<sup>1</sup>

Turn-on fluorescent dyes, or profluorophores, represent a unique means of developing chemical probes. These probes are chemically designed to produce a signal after undergoing a particular reaction, which is dictated via the installation of a unique handle onto the dye (Scheme 1).<sup>2,3</sup> These types of probes have gained increased attention, as they produce low background signal in the absence of their cognate partner, improving overall signal-to-noise ratios. Furthermore, their cognate partner can be readily installed into certain regions of the cells (via UAA mutagenesis, for example), allowing the scientist to better control the localization of their fluorescent signal.

*Scheme 10*



Based on the fluorescent imaging advantages intrinsic to profluorophores, we proposed a strategy that utilizes UAAs, in conjunction with biorthogonal chemistry, to selectively bind profluorophores to genetically modified proteins that were previously described to incorporate UAAs. Such an approach allows for the generation of site-specificity within the reaction of the protein and the label. This affords a high level of control over the process, and the prevention of protein inactivation and loss of function. Additionally, this strategy eliminates the background signal, as fluorescence cannot be observed unless conjugation has occurred. Finally, the literature contains a myriad of profluorophores that have already been synthesized and can be employed. This permits for the discovery of dyes of the correct size to provide *in vivo* diffusion, as well as different excitation spectra that can be used for simultaneous labeling of multiple proteins within a single organism.

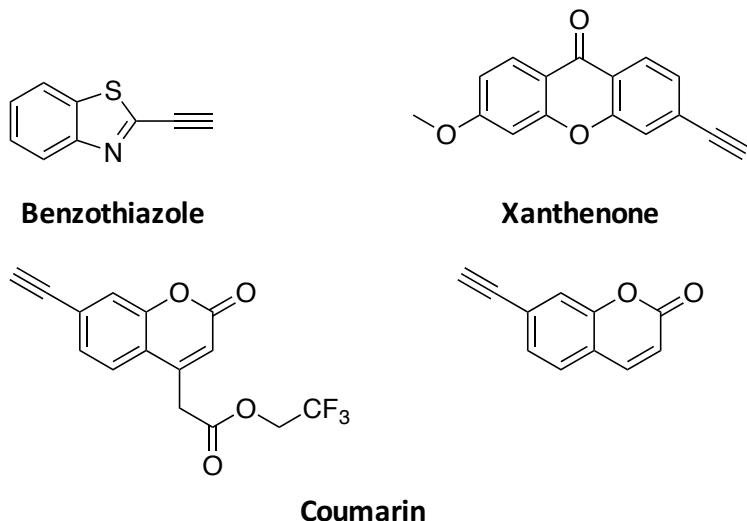
The unique functional handles that UAAs can introduce into proteins offer a means to tackle these complexities, as well as reduces the challenges associated with protein labeling. This strategy is has primarily relied upon upon click chemistry, a reaction that proceeds to completion readily and in relatively mild conditions.<sup>8</sup> Of particular utility is click chemistry involving azides and alkynes that readily undergo a Huisgen 1,3-dipolar cycloaddition which produces a stable triazole moiety optimal for *in vivo* conjugation.<sup>9</sup> Additionally, the scarcity of azides and alkynes in biologic systems makes these two functional groups optimal for use in developing a click chemistry strategy of fluorescent labeling, as the chance of cross-reactions are minimal. Moreover, research has identified profluorophores, or “turn-on” probes, that become fluorescent only

after a successful click reaction has formed the triazole moiety. Indeed, there are a variety of click-chemistry compatible turn-on probes in the literature.<sup>2–6</sup>

Recently, our group expanded the role of terminal alkyne in biorthogonal chemistry by developing biological variants of the Glaser-Hay and Cadiot-Chodkiewicz couplings.<sup>7,8</sup> Both reactions produce a covalent diyne linkage. The Glaser-Hay reaction involves the coupling of two terminal alkynes in the presence of copper(I). The Cadiot-Chodkiewicz coupling utilizes a bromoalkyne and a terminal alkyne in the presence of copper(I) to produce the diyne linkage, adding a degree of chemoselectivity while also minimizing the requisite amounts of copper(I). In order to further expand the utility of these chemistries in chemical biology, we sought to explore the potential of these reactions for “turn-on” probe labeling of proteins via UAAs. We hypothesized that the added conjugation present in the diyne linkage would mirror that found in the triazole ring, restoring electron density to the profluorophores and eliciting fluorescence following successful conjugation.

## Results and Discussion

In order to probe the potential of the biorthogonal Glaser-Hay and Cadiot-Chodkiewicz for “turn-on” probe development, we first attempted to synthesize a variety of terminal-alkyne containing profluorophores. Thankfully, a myriad of such compounds have already been reported in the literature, possessing a variety of well-defined fluorophore cores.<sup>3,4,6</sup> As such, we sought to synthesize a benzothiazole-, coumarin-, and xanthenone-alkyne containing profluorophroes.

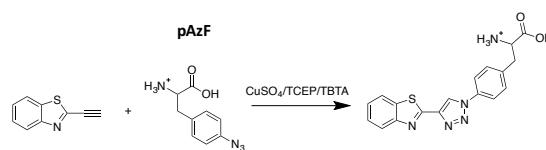


*Figure 42. Alkyne-containing profluorophores to be synthesized to probe the application of the Glaser-Hay towards turn-on probes.*

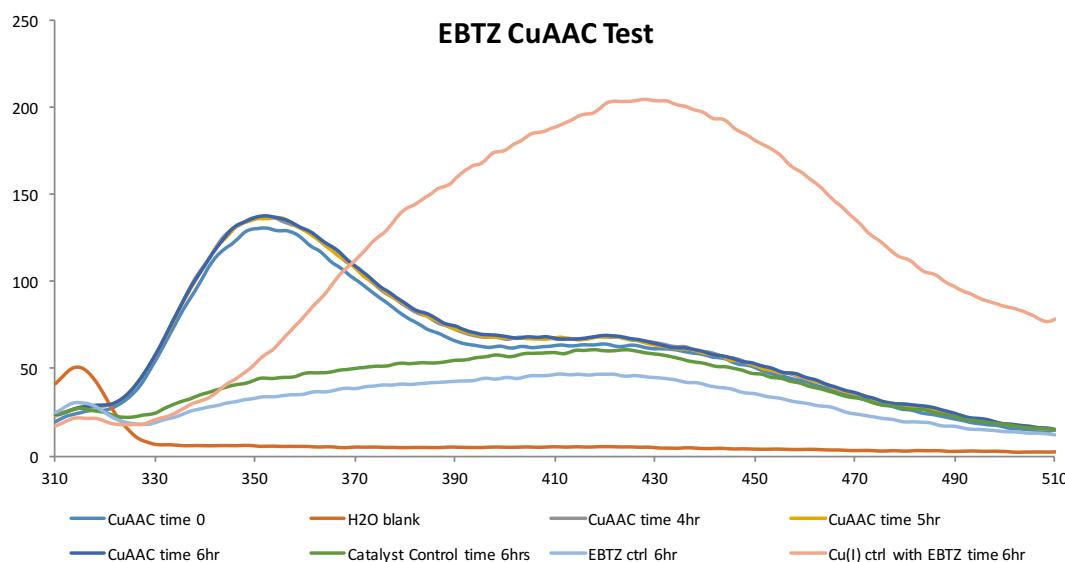
The first fluorophore synthesized was ethynylbenzothiazole (**1**, EBTZ, see Experimental for synthetic scheme). Starting from 2-aminobenzothiazole, an iodo group was installed 2-iodobenzothiazole, which could then subsequently undergo a Sonogashira cross-coupling with trimethylsilylacetylene. Following TMS deprotection using potassium carbonate in methanol, column chromatography in 6:1 hexanes:ethyl acetate afforded the 2-ethynylbenzothiazole as a brown solid (7% yield). While the compound was not initially soluble in the eluent system, a small volume a dichloromethane was used to help solubilize the reaction mixture for loading onto the silica gel column. Solubility issues were a major roadblock for the synthesis of all profluorophores, likely due the highly aromatic nature of many of these fluorophore cores. We hypothesize that tight pi-stacking between the aromatic rings prevents the solvation of the product. Typically the addition of a polar solvent, such as dichloromethane or methanol, to the loading solution as well as sonication of the mixture helped solubilize the bulk of the product for flash chromatography.

To verify that the synthesized EBTZ was indeed a profluorophore, a click reaction was performed using p-azidophenylalanine (pAzF) in water, which mirrored the same fluorogenic click reaction reported in the literature. The same “turn on” peak as reported in the literature was observed at ~354nm in the click reaction when excited at 285nm. In the absence of the copper catalyst system, no real peak was observed, and in the presence of solely CuSO<sub>4</sub>, a different peak at ~430nm was observed, likely due to turn on of the fluorophore as it bound the copper salt. Furthermore, unlike the literature, our fluorogenic click reaction was run in the presence of the nitrogenous ligand TBTA, which is reported to drastically increase the rate of the copper(I)-catalyzed cycloaddition of azides and alkynes. As such, based on the fluorimetry data it appears that our reaction reached completion within 4 hr, unlike the literature which required ~24 hr to achieve maximum signal.

**A.**



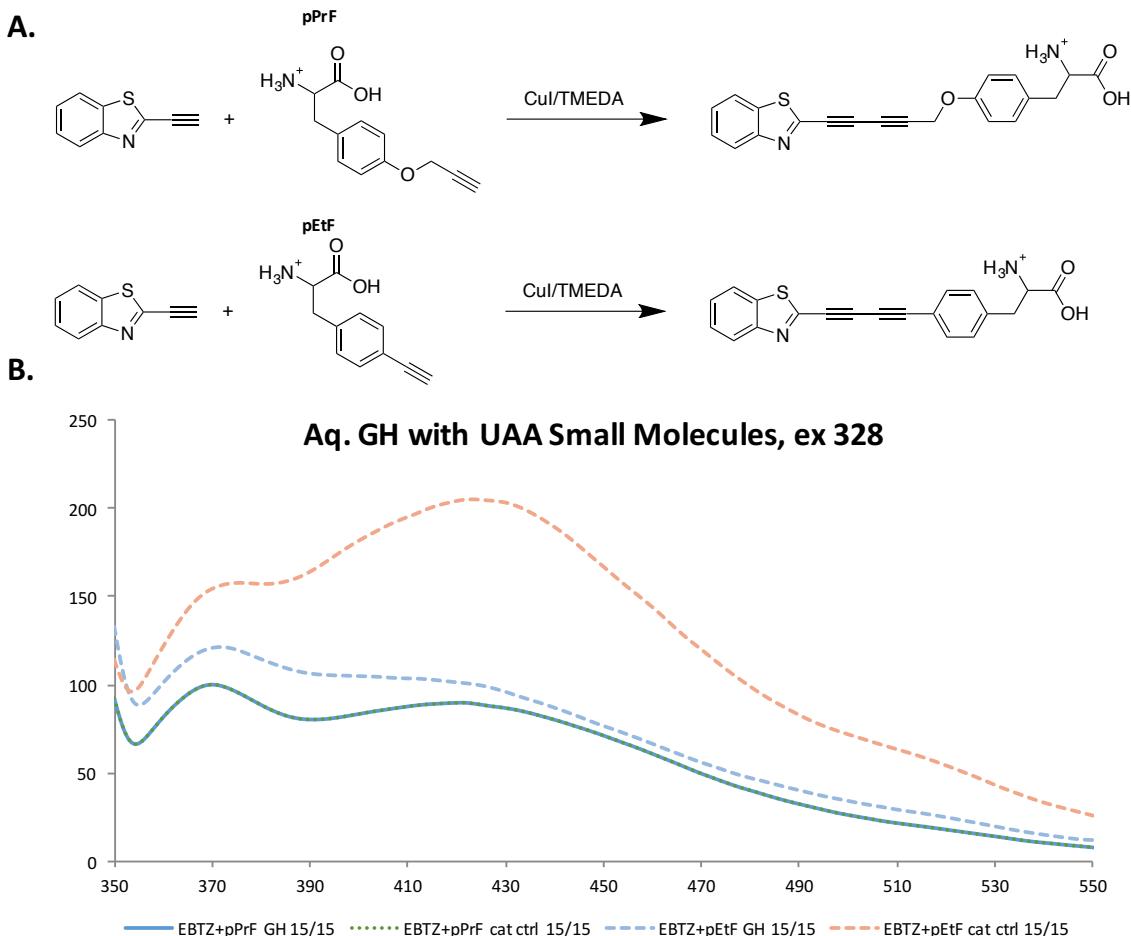
**B.**



Taken together, the above data verified that the synthesized molecule was indeed an alkyne-bearing profluorophore. As such, we next sought to investigate whether our recently developed biological Glaser-Hay reaction could also produce a turn-on response in the probe. We hypothesized that the added electron density found in the Glaser-Hay's diyne linkage would mimic that found in the copper click's triazole ring, serving to promote the relaxation of excited state electrons through fluorescence.

To begin, we performed an organic Glaser-Hay with EBTZ in the presence of propargyl alcohol and phenylacetylene, which mimicked the functional groups found in our alkyne-bearing UAAs (p-propargyloxyphenylalanine, pPrF, and p-ethynylphenylalanine, pEtF). Absorption spectra taken on the UV-Vis revealed an absorption maxima at ~328 nm for the Glaser-Hay between phenylacetylene and EBTZ, with no real absorption peaks found for the reaction with propargyl alcohol. As such, this wavelength was chosen to excite the alkyne-bearing UAAs conjugated to EBTZ.

Interestingly, the Glaser-Hay control reaction, in the absence of the copper catalyst system, with the pEtF had a higher fluorescence response than the Glaser-Hay reaction between pEtF and EBTZ, with a maximum emission at ~420nm, which corresponded to the same emission profile of the free pEtF in water. These results were contradictory to what we expected, as we thought directly conjugating EBTZ to an aromatic ring would increase fluorescence response. As such, it appears that the aromatic variant of the Glaser-Hay onto the EBTZ profluorophore serves to quench the fluorescence intrinsic to the aromatic UAA. There was no fluorescence change observed for the reaction between pPrF and EBTZ.



We attempted the biological Glaser-Hay reaction in a protein context using p-propargyloxyphenylalanine-GFP<sub>TAG151</sub> in the presence of EBTZ. However, our results were inconclusive, as even after extended denaturation the GFP signal swamped our fluorescence spectra. Future work will attempt the Glaser-Hay using a p-propargyloxyphenylalanine-Ubiquitin<sub>TAG47</sub>, which is a non-fluorescent small protein that will not convolute our fluorescence spectra with its own fluorescent signal.

In order to verify the potential of the Glaser-Hay reaction to either quench UAAs or profluorophores, we sought to synthesize other UAAs to help corroborate our initial findings with EBTZ. These syntheses are ongoing, with varying degrees of success.

The next profluorophore we attempted to synthesize from the literature was a 3-methoxy-6-ethynyl-9H-xanthen-9-one, which possesses a xanthone core.<sup>5</sup> To begin, 2,2',4,4'-tetrahydroxybenzophenone was subjected to a microwave-mediated condensation reaction in the presence of 10% NaOH to afford 3,6-dihydroxy-9H-xanthen-9-one. This product was then asymmetrically methylated using methyl iodide in a biphasic solvent system of toluene and 50% NaOH. Unfortunately, we were never able to isolate the asymmetrically methylated product, largely due to solubility issues in all our available organic and aqueous solvents. If we had, we would have installed a triflate group on the free hydroxyl group, which could then undergo a Sonogashira cross-coupling with trimethylsilylacetylene to yield 3-methoxy-6-(2-(trimethylsilyl)ethynyl)-9H-xanthen-9-one. This TMS-protected variant could then be deprotected to afford the terminal alkyne containing **2**.

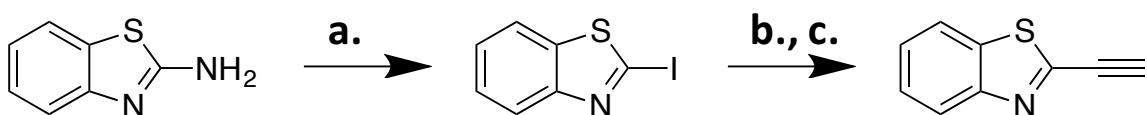
The last proposed profluorophore moiety we are preparing involve ethynyl-containing coumarin profluorophores.

## Conclusion

We have begun the synthesis of a variety of profluorophores to probe the ability of the Glaser-Hay to serve as a “turn-on” mechanism for these probes. Our initial studies appear to reveal an actual “turn-off” effect when utilizing the Glaser-Hay in conjunction with alkyne bearing profluorophores.

## Experimental

### Synthesis of 2-ethynylbenzothiazole (**1**)<sup>3</sup>

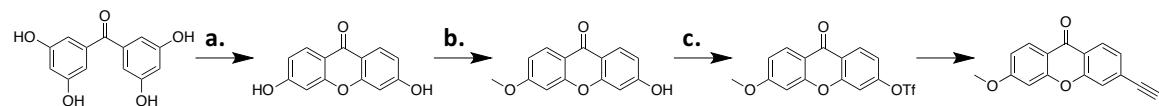


Synthesis of 2-iodobenzothiazole (**1a**) via **a.** To a flame dried round bottom and stir bar was added 1.875 g (8.085 mmol) of 2-aminobenzothiazole and 7.5 g of p-toluenesulfonic acid (39.428 mmol). Meanwhile, a salt solution of 5.25 g of potassium iodide (31.627 mmol) and 1.75 g of sodium nitrite (25.362 mmol) in 7.5 mL of deionized water was prepared. To the round bottom containing the 2-aminobenzothiazole was added 10 mL of acetonitrile, and to this was promptly added the salt solution. The mixture was stirred at 0°C for 2 hr and then stirred at room temperature overnight. The next day 175 mL of deionized water were added to the reaction round bottom, and the pH of the mixture was adjusted to ~8-9 (as indicated via pH strips) using 1M sodium bicarbonate. A spatula tip of sodium thiosulfate was added to this, which caused a brownish precipitate to crash out. Any additional sodium thiosulfate was found to hinder the workup procedure. The whole mixture was filtered on a Buchner funnel and washed with cold water to give 2.106 g of **1a** as a brown powder (8.066 mmol, 99% yield) as confirmed by <sup>1</sup>H NMR in CDCl<sub>3</sub> according to the literature.

Synthesis of 2-ethynylbenzothiazole (**1**) via **b. & c.** To a flame dried round bottom and stir bar, flushed with argon beforehand, was added 0.870 g of **1a** (3.334 mmol). The round bottom was then capped with a septum and flushed with argon. To this round bottom was added 7 mL of dry dichloromethane. This solution was degassed using argon for 20 minutes, and then 942 μL of trimethylsilylacetylene (0.65 g, 6.618 mmol) and 3.3 mL of triethylamine (2.409 g, 23.807 mmol) was added via syringe through the cap, maintaining the integrity of the argon atmosphere. This mixture was stirred under argon,

and the septum was briefly opened to add 47 mg of bis(triphenylphosphine)palladium(II) dichloro (0.067 mmol) and 14 mg of copper(I) iodide (0.074 mmol). The reaction was then flushed with argon and was allowed to stir at 65°C for 4 hr. After 4 hr, 10 mL of a brine solution was added to the mixture. The reaction was then extracted with dichloromethane (10 mL x 3) and this organic layer was back extracted using bring (40 mL x 4). The organic layer was then dried with magnesium sulfate, filtered, and concentrated *in vacuo* to afford 0.463g of 2-trimethylsilyl-ethynylbenzothiazole (**1b**) (2.000 mmol, 60% yield) as a brown powder. To the round bottom containing **1b** was added a stir bar and 10 mL of methanol to check if this solvent would solubilize the compound. If it did, an additional 10 mL portion of methanol was added to **1b**, if not, an additional 10 mL of dichloromethane was added instead. To the solution was added 1.244 g of potassium carbonate (9.000 mmol) and the reaction was stirred at room temperature for 2-3 hrs. Next, the reaction was extracted in a 1:2 volume of dichloromethane to brine (20 mL of dichloromethane x 3, 40 mL of brine total). The organic layer was then back extracted one time with 60 mL of brine. The organic layer was then dried with magnesium sulfate, filtered, and concentrated *in vacuo*. Initial attempts to purify the reaction with silica gel chromatography in a MeOH:DCM eluent failed. As such the reaction mixture was dissolved up in a small volume of DCM and purified via silica gel chromatography using a 6:1 hexanes:ethyl acetate as the solvent system. Twenty-two mgs of 2-ethynylbenzothiazole (0.138 mmol, 7% yield) were obtained from the column as a brown residue that was not fluorescent via TLC. <sup>1</sup>H NMR

#### Synthesis of 3-methoxy-6-ethynyl-9H-xanthen-9-one (**2**)<sup>5</sup>



Synthesis of 3,6-dihydroxy-xanthen-9-one (**2a**) via a. To a microwave vial containing a stir bar was added 800 mgs of 2,2',4,4'-tetrahydroxybenzophenone (3.250 mmol). This was dissolved in 4 mL of a 10% NaOH solution (w/v), and then subjected to microwave irradiation in dynamic setting (10 mins at 120°C; microwave protocol on Young Lab CEM as Profluoro st 1). Next, the reaction was subjected to fixed power setting at 150 W, with a max temperature of 180°C for twelve pulses. After the twelve pulses, the reaction was diluted with 4 mL of deionized water and acidified with 6 M HCl to give a white, viscous gel like substance. This was filtered via Buchner filtration to give **2a** as a white to off-white powder.

Synthesis of 3-hydroxy-6-methoxy-9H-xanthen-9-one (**2b**) via b. To 1.25g of **2a** (5.5 mmol) was added 62.5 mL of toluene, which did not solubilize the material. To this mixture was added 1.25 g (5.5 mmol) of benzyl triethyl ammonium chloride followed by 375  $\mu$ L (0.781 g, 5.5 mmol) of methyl iodide. Next, 31 mL of a 50% aqueous NaOH (w/w) solution was added and the reaction was capped and stirred at 110°C for 6 hr. After the 6 hr the separate phases were separated via a separatory funnel, and the aqueous layer was collected. The aqueous layer was then acidified with ~12 mL of 12N HCl, which produced a color change from cloudy yellow, to clear yellow, to a white/milky precipitate formed once a pH = 4 was reached. The acidified layer then sat on ice for 2 minutes and filtered via Buchner filtration. The collected crystals sat overnight and were then recrystallized in 10 mL of boiling toluene. This gave 725 mgs of an off white product which we thought may be the asymmetrically methylated **2b**.

Synthesis of 6-methoxy-9-oxo-9H-xanthen-3-yl-trifluoromethansulfonate (**2c**) via c. To a flame dried round bottom and stir bar was added 0.400g (1.65 mmol) of **2b** dissolved in

20 mL of DCM. To this was added 665  $\mu$ L of dry pyridine (0.653 g, 8.25 mmol) and the mixture was cooled over ice to 0°C. Everything was flushed with argon while the reaction mixture cooled. Then, using a needle and syringe, 416  $\mu$ L of trifluoromethanesulfonic anhydride (which may or may not have been anhydrous; 0.698 g, 2.475 mmol) dropwise, to the best of my ability. The temperature was gradually raised to room temperature and the reaction was stirred for an additional 2 hrs. After the 2 hrs, 15 mL of deionized water was to quench the reaction. This mixture was stirred for an additional 5 mins, and the organic and aqueous phases were then separated. The organic layer was washed with a 1N HCl solution in water (15 mL x 2), and it was then washed once more with a 15 mL portion of deionized water. The organic layer was then dried with magnesium sulfate (which we now believe may be interacting with our profluorophores and fluorophores, preventing adequate recovery after this step), filtered, and concentrated *in vacuo*. Silica gel chromatography in 3:1 hexanes:ethyl acetate yielded 5 mgs of something that was a single spot on TLC. We were concerned that the silica gel may be interacting with the triflate group, binding our product to the column during the purification process.

Cu(I)-Catalyzed Azido-Alkyne Cycloaddition with EBTZ, 1 To a 5 mL SnapCap tube was added 2  $\mu$ L of a 50 mM CuSO<sub>4</sub> solution in water, followed by 2  $\mu$ L of a 50 mM TCEP solution in water, then 2  $\mu$ L of a 5 mM TBTA solution in DMSO, then 2  $\mu$ L of a 100 mM solution of p-azidophenylalanine (pAzF) in water, and finally 40  $\mu$ L of a 1 mM solution of **1** in DMSO. All of this was diluted to 3 mL in deionized water. As a control, a reaction between pAzF and **1** in the absence on the copper click catalyst system was performed. A **1** fluorescent control was prepared by dissolving 40  $\mu$ L of th solution of **1**

in 3 mL of deionized water. Finally, the effects of CuSO<sub>4</sub> on inducing a turn on response in **1** was performed by mixing 2  $\mu$ L of the 50 mM CuSO<sub>4</sub> solution with 40  $\mu$ L of the solution of **1**. Fluorescence profiles of all the 3 mL samples were taken in a quartz cuvette by exciting at 285 nm and scanning from 310 to 510 nm with 8 nm excitation and emission wavelength slits at a scan speed of 200 nm/min.

Organic Small Molecule Glaser-Hay with EBTZ, **1** To a flame dried vial and stir bar was added ~10-20 mg of CuI, which was subsequently dissolved in 1 mL of pseudo-dry tetrahydrofuran (THF). To this mix was added 60  $\mu$ L of TMEDA and the catalyst system was allowed to stir. Meanwhile, 100  $\mu$ L of a 10 mM solution of **1** in DMSO (16  $\mu$ g, 0.100  $\mu$ mol) was added to the catalyst solution followed by 2  $\mu$ L of the small molecule partner (either propargyl alcohol or phenylacetylene at a >10 molar equivalency). This reaction was allowed to stir at 60°C overnight. Following this, both reactions were dissolved in an extra 1 mL of THF. Absorbancies were taken for both reactions using a UV/Vis to verify the maximum absorption wavelength, which corresponded to the maximum excitation wavelength.

Glaser-Hay with UAAs and EBTZ, **1** To a 3 mL SnapCap tube was added 2  $\mu$ L of a 500 mM solution of CuI in water followed by 2  $\mu$ L of a 500 mM solution of TMEDA. Next was added 2  $\mu$ L of an alkyne containing UAA (either p-propargyloxyphenylalanine or p-ethynylphenylalanine) and 40  $\mu$ L of a 1 mM solution of **1** in DMSO. This mix was then diluted to 3 mL using deionized water. Catalyst control reactions were set up by mixing the UAAs and **1** as above without adding the CuI or TMEDA. These were allowed to react overnight, and the 3 mL sample was then added to a quartz cuvette for fluorescence analysis. Samples were excited at 328 nm and emission profiles were scanned for from

350 to 550 nm using 8 nm excitation and emission wavelengths slits and a scan speed of 200 nm/min.

## References

1. Urano, Y. *et al.* Selective molecular imaging of viable cancer cells with pH-activatable fluorescence probes. *Nat Med* **15**, 104–109 (2009).
2. Sivakumar, K. *et al.* A Fluorogenic 1 , 3-Dipolar Cycloaddition Reaction of 3-Azidocoumarins and. *Org. Lett.* **6**, 4603–4606 (2010).
3. Qi, J. & Tung, C. H. Development of benzothiazole ‘click-on’ fluorogenic dyes. *Bioorganic Med. Chem. Lett.* **21**, 320–323 (2011).
4. Rong, L. *et al.* A coumarin derivative as a fluorogenic glycoproteomic probe for biological imaging. *Chem. Commun. (Camb).* **50**, 667–9 (2014).
5. Li, J. Q., Hu, M. Y. & Yao, S. Q. Rapid Synthesis, Screening, and Identification of Xanthone- and Xanthene-Based Fluorophores Using Click Chemistry. *Org. Lett.* **11**, 3008–3011 (2009).
6. Liu, J. *et al.* 7-Ethynylcoumarins: selective inhibitors of human cytochrome P450s 1A1 and 1A2. *Chem Res Toxicol* **25**, 1047–1057 (2012).
7. Maza, J. C., Nimmo, Z. M. & Young, D. D. Expanding the scope of alkyne-mediated bioconjugations utilizing unnatural amino acids. *Chem. Commun.* (2015). doi:10.1039/C5CC08287K
8. Lampkowski, J. S., Villa, J. K. & Young, D. D. Development and Optimization of Glaser-Hay Bioconjugations. *Angew. Chemie Int. Ed. English (pending Revis.* (2015).