5-2016

Investigation of Unnatural Amino Acids as a Means to Modulate Protein Function

Taylor Harrison Jacobs
College of William and Mary

Investigation of Unnatural Amino Acids as a Means to Modulate Protein Function

A thesis submitted in partial fulfillment of the requirement

for the degree of Bachelor of Science in Chemistry from

The College of William and Mary

by

Taylor Harrison Jacobs

Accepted for Honors

Dr. Douglas Young, Committee Chair

Dr. Robert Hinkle

Dr. William McNamara

Dr. Margaret Saha

Williamsburg, VA
May 5, 2016

Investigation of Unnatural Amino Acids as a Means to Modulate Protein Function

Taylor H. Jacobs

Leesburg, VA

A Thesis Presented at the College of William and Mary in Candidacy of Honors

Chemistry Department

The College of William and Mary
May 3, 2016

ABSTRACT

In order to elucidate the biological processes that occur in everyday life, chemical biologists have developed technologies that allow the study of a various biological systems. Bioorthogonal chemistry is an ever-growing technology that involves performing chemical reactions with biological systems that do not rely on existing biological chemistries. In this work we attempt to develop and characterize novel bioorthognal chemistries that further expand the utility of this field. Additionally, we explore the utility of fluorescent probes in labeling applications. Finally, we attempt to create novel methods of control for a gene-editing protein using light as a mechanism of regulation.

TABLE OF CONTENTS

Chapter 1: Introduction to Unnatural Amino Acids and Bioconjugations
1.I. Unnatrual Amino Acids and Incorporation 10
1.II. Bioorthogonal Chemistry 14
1.II.A. Click Chemistry 15
1.II.B. Oxime Chemistry 16
1.III. Applications of Bioconjugation Reactions 17
Chapter 2: Development of Alkoxyamine Unnatural Amino Acids for Bioconjugation Applications
2.I. Introduction 20
2.II. Results and Discussion 23
2.II.A. Synthesis of Unnatural Amino Acids 23
2.II.B. Incorporation of Unnatural Amino Acids 26
2.II.C. Bioconjugation Experiments 32
2.III. Experimental 37
2.IV. Future Directions 42
2.V. Conclusion 43
Chapter 3: Synthesis of Profluorophore 2-ethynylbenzo[d]thiazole and Bioconjugation in Protein
3.I. Introduction 44
3.II. Results and Discussion 45
3.II.A. Synthesis of 2-ethynylbenzo[d]thiozole 46
3.II.B. Cu(I)-1,3-Cycloaddition Reactions 48
3.III. Experimental 49
3.IV. Future Directions 50
3.V. Conclusion 51
Chapter 4: Mutagenesis of Cas9 for Incorporation of Unnatural Amino Acids
4.I. Introduction 52
4.II. Results and Discussion 55
4.II.A. Site Directed Mutagenesis of Cas9 Plasmid 57
4.II.B. Incorporation of Unnatural Amino Acid Into Cas9 59
4.III. Experimental 60
4.IV. Future Directions 60
4.V. Conclusion 61
Appendix A: NMR Spectra 62
Appendix B: Mass Spectra 73
Appendix C: Primer Sequences 76
Acknowledgements 77
References 78

TABLE OF FIGURES

Figure 1. Central Dogma 8
Figure 2. Translation of natural and unnatural amino acids 9
Figure 3. Unnatural amino acids 10
Figure 4. General procedure for double sieve selection of aaRS 13
Figure 5. Unspecific antibody conjugation 20
Figure 6. Specific antibody conjugation 20
Figure 7. Diagram of genetic circuit 21
Figure 8. Ketone UAAs 22
Figure 9. Site-specific immune-PCR 24
Figure 10. Synthetase screen data 29
Figure 11. SDS-Page gel of alkoxyamine UAA incorporation into GFP-151 31
Figure 12. Graph of incorporation effiecncy of alkoxyamine UAAs 32
Figure 13. SDS-PAGE gel oxime reaction of GFP-151 mutants with prodan 33
Figure 14. Oxime ligation of GFP-151 UAA mutants reacted with ubiquitin pAcF mutants 36
Figure 15. Diagram of profluorophore linkage alkyne and azide functionality 45
Figure 16. Fluorescence spectrum for a Cu(I)-1,3-dipolar cycloaddition with pAzF mutated GFP-151and 2-ethynylbenzo[d]thizole (**3)**. 48
Figure 17. The overall mechanisms of the three types of CRISPR Cas9 systems 53
Figure 18. Cas9 protein and target DNA complex. 54
Figure 19. Interactions of amino acid Y450 in Cas9 56

TABLE OF SCHEMES

Scheme 1- Cu[I]-1,3-dipolar cycloaddition reaction mechanism 16
Scheme 2- Oxime condensation reaction 18
Scheme 3- Oxime condensation reaction mechanism 21
Scheme 4- Overall alkoxyamine UAA synthesis mechanism 26
Scheme 5- Elimination by-product of alkoxyamine synthesis 27
Scheme 6- Oxime ligation with GFP-151 and prodan 33
Scheme 7- Non-biological oxime ligation reaction 35
Scheme 8- GFP-151 alkoxyamine UAA mutants and ubiquitin pAcF mutant 37
Scheme 9- Overall reaction protocol for 2-ethynylbenzothiazole 46
Scheme 10- Diazotization-iodination reaction to form 2-iodobenzo[d]thiozole. 47
Scheme 11- Sonogashira coupling reaction. 48
Scheme 12- Deprotection routes to create (**3**). 48
Scheme 13- ONBY photocleavage 56

TABLE OF TABLES

Table 1- Overall yields of alkoxyamine UAAs 27

CHAPTER 1: INTRODUCTION TO UNNATURAL AMINO ACIDS AND BIOCONJUGATIONS

Proteins are a class of essential macromolecules that perform a multitude of functions necessary within the cells of all living organisms.[1, 2] These functions include catalyzing chemical reactions, transporting various molecules, providing a means of communication throughout the body, and as structural support for cells. Proteins are composed of various combinations of twenty naturally occurring amino acids. The information to form the sequence of amino acids of a protein is stored within the DNA sequence within the organism's genome. This DNA is first transcribed into mRNA, which is then translated into protein (Figure 1).

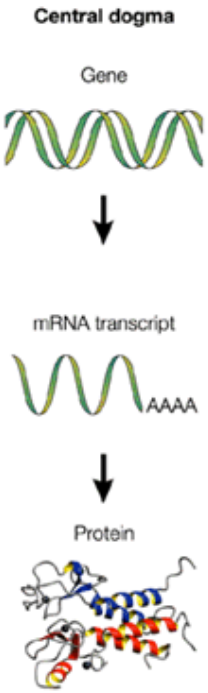


Figure 1. The central dogma, the process of converting genetic information in to functional protein. Adapted from *Nat. Rev. Genet.* **2002** 3, 43-52.[3]

Translation occurs on the ribosome, where three base pair segments of the mRNA, called codons, are used to encode a single amino acid. Codons on the mRNA are recognized by anticodons on tRNA molecules. Each tRNA has a corresponding amino acid that is bound to it through aminoacylation, by an aminoacyl tRNA synthetase (aaRS) enzyme, refer to figure 2.[1] The tRNA charged with an amino acid is then used by the ribosome to form new peptide bonds between amino acids. This process repeats itself multiple times to biosynthetically prepare proteins. The primary sequence of amino acids affords intramolecular interactions to form secondary and tertiary structures, as well as quaternary structures with other polypeptide chains. Once properly assembled, these proteins are then able to perform various tasks within the cell.

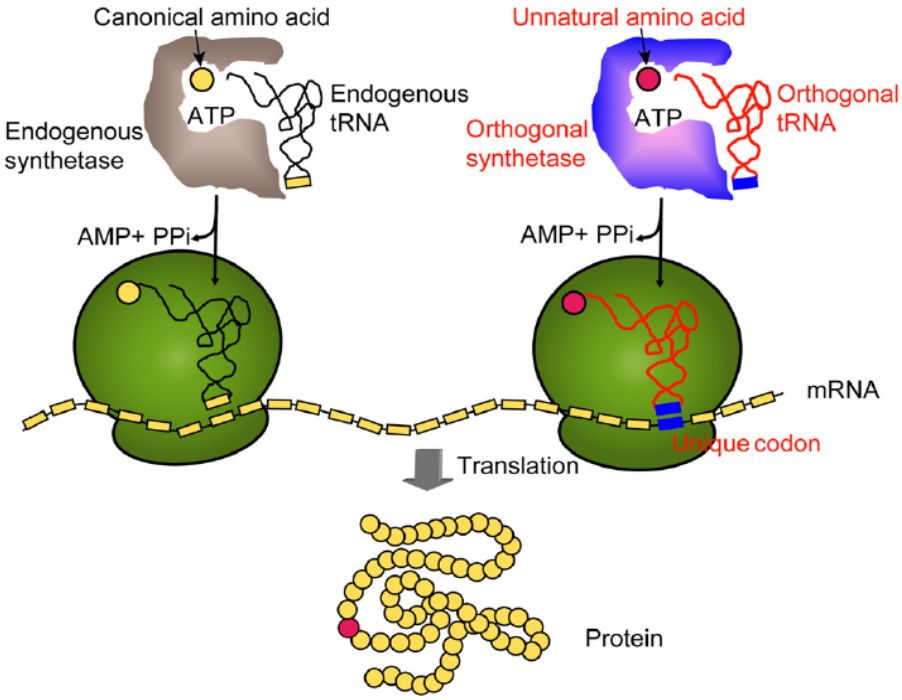


Figure 2. Comparison of natural translation and translation using unnatural amino acids and a mutated tRNA synthetase. Adapted from *Chem. Biol.* **2009** 16, 323-336.[4]

The diversity of the polypeptide sequences and structures allows for various functions; however, there are limitations due to the limited chemical functionality of the natural 20 amino acids. The canonical amino acids are made of only five elements that create acidic, basic, polar, and hydrophobic residues. Incorporation of unnatural amino

acids (UAAs) into a protein can introduce a multitude of unique chemical functionalities that may expand the function of a protein. These amino acids can also be inserted into proteins and manipulated to facilitate the study of the protein structure and function.[1, 2]

UAAs also have many clinical applications including: labeling HER2 cancer cells with fluorescent markers[5], selectively binding cytotoxic cancer drugs to antibodies to provide better specificity[6], and artificially stimulating immune systems.[2] With these diseases affecting such a large portion of our population, working to perfect and expand this methodology represents a mechanism to introduce new therapeutics and diagnostics. Moreover, there has been recent work that employed UAAs for work in synthetic biology.[7] Synthetic biology is a field of biology that creates new tools to help solve biological problems or create new ways to better study biological phenomena. The use of UAAs in these capacities indicates a bright future for advancement with this technology.

1.I. UNNATURAL AMINO ACIDS AND INCORPORATION

There are approximately 70 UAAs that have been synthesized and incorporated into proteins. The diversity of the chemical functionalities of the UAAs affords a myriad of novel protein functions (Figure 1).[1, 2] These amino acids have functionalities that are, for the most part, non-disruptive to the necessary functions of the cell, but confer new chemical properties to the protein to facilitate unique functions.

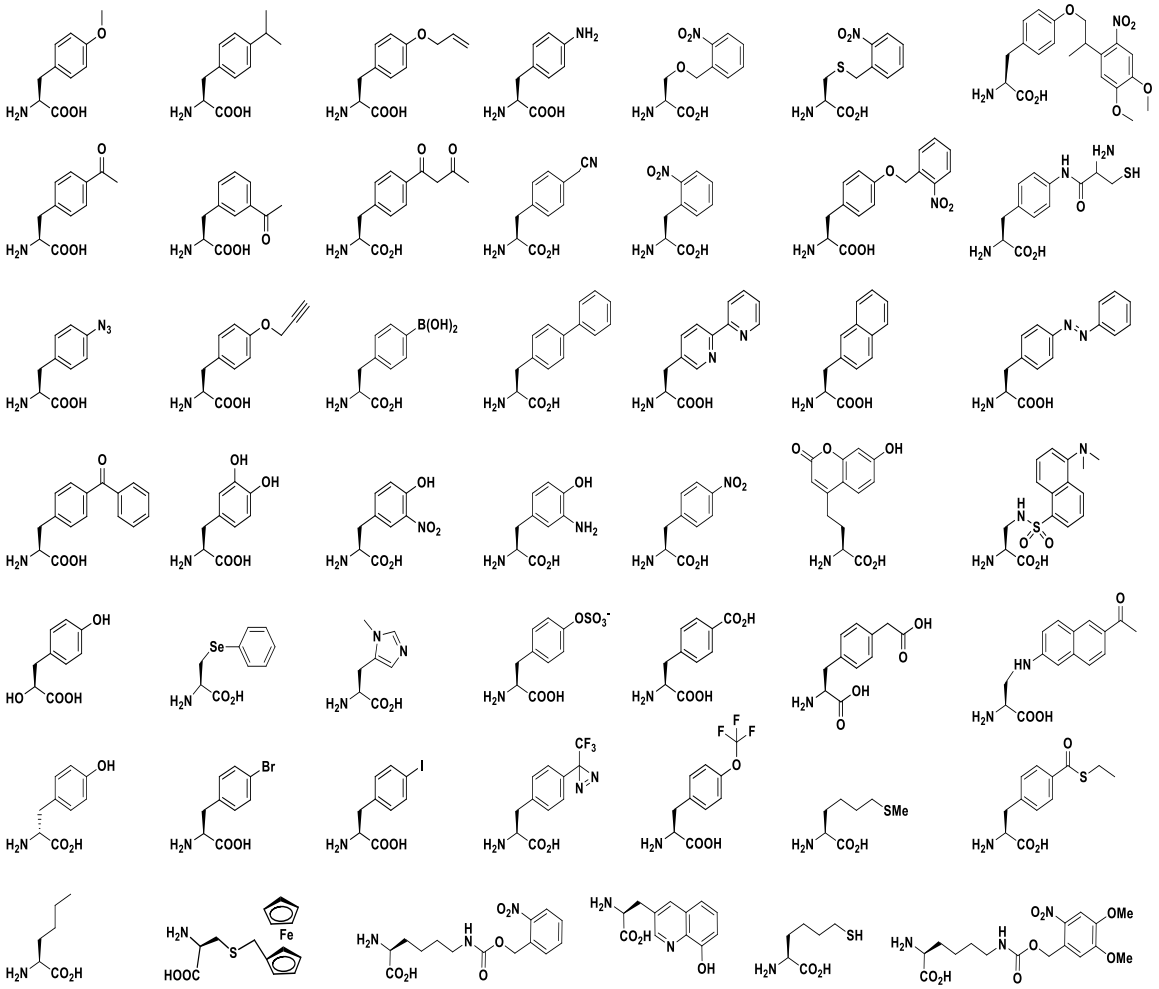


Figure 3. Structures of unnatural amino acids previously synthesized and incorporated.

To incorporate these UAAs into protein an orthogonal aminoacyl tRNA synthetase (aaRS) is required to aminoacylate the unnatural amino acid to a nonsense suppressor tRNA.[1, 2, 8] This represents a bioorthogonal aaRS/tRNA pair that does not cross-react with any of endogenous translational system. The incorporation of aaRS/tRNA from an archaebacteria species into E. *coli* though allows for adequate functionality for translation within the organism without negatively interacting with endogenous aaRS/tRNA. Additionally, the amber stop codon is typically employed, as it is the least commonly used stop codon in *E. coli*.[1] Often, a tyrosine aaRS and mutated amber tRNA from the archaea *Methanocaldococcus jannaschii* is used to allow *E. coli* incorporate UAAs. Other archaebacteria aaRS/tRNA have also been used with efficiency similar to that of *Methanocaldococcus jannaschii*, for example, a pyrrolysine pair derived

from the achaebacterium *Pyrococcus horikoshii*, has also been developed by Santoro et al.[9] The ability to exploit different amino acid precursors provides a greater diversity in UAAs that can be manipulated, and ultimately incorporated into proteins.

Liu et al. have also worked to improve the efficiency of the aminoacylation and translation of the mutant tRNAs.[10] In addition to mutagenizing the anti-codon, modifications to other residues of the tRNA were also conducted. Mutations within the acceptor arm, the D-loop, and the anti-codon binding loop were performed to find the optimal translation with UAAs.

To allow the tyrosyl-*M. jannaschii* tRNA (*Mj*tRNA^{Tyr}) recognize the TAG codon, the tRNA anti-codon is mutated to CUA.[1, 2, 8] To selectively alter the *M. jannaschii* aaRS (*Mj*TyrRS) to aminoacylate a UAA instead of the natural tyrosine, a library of mutant *Mj*TyrRS are created and are evolved through a double-sieve selection procedure. Due to the inherent promiscuity of the active site within the aaRS selective pressure can be applied to the aaRS library to find specific aaRSs that recognize the desired UAA.[11] The selection begins with a positive round as the aaRS library and tRNA are introduced into cells that contain a plasmid encoding a chloramphenicol resistance protein harboring a TAG mutation. The cells are then grown on media that has chloramphenicol in addition to the desired UAA. Consequently, only the surviving colonies will have been able to charge the UAA or an endogenous amino acid to provide antibiotic resistance. Cells possessing aaRSs that do not recognize the UAA result in a terminated chloramphenicol resistance protein, and consequently cannot survive. The plasmids from the positive hits are then isolated and subjected to a negative round of selection. The plasmids are transformed into cells containing a plasmid with a barnase gene posessing three amber stop codons. The barnase protein is cytotoxic as it is a bacterial ribonuclease, which will destroy bacterial RNA and will lead to the death of the cell if it is transcribed. These cells are grown in the absence of the UAA. Thus, if the cell is able to express barnase, the aaRS does incorporate amino acids at the stop codon, but it is not specific to the unnatural and can interact with the canonical amino acids. Since the barnase is present in the cells, the undesirable aaRSs are eliminated from the selection via the death of the cell. Therefore, the only cells that survive both rounds of selection possess an aaRS that only charges the UAA and not any of the 20 natural amino acids. The double selection process

is repeated multiple times to ensure the aaRS will charge the tRNA with the UAA with high specificity (Figure 4).

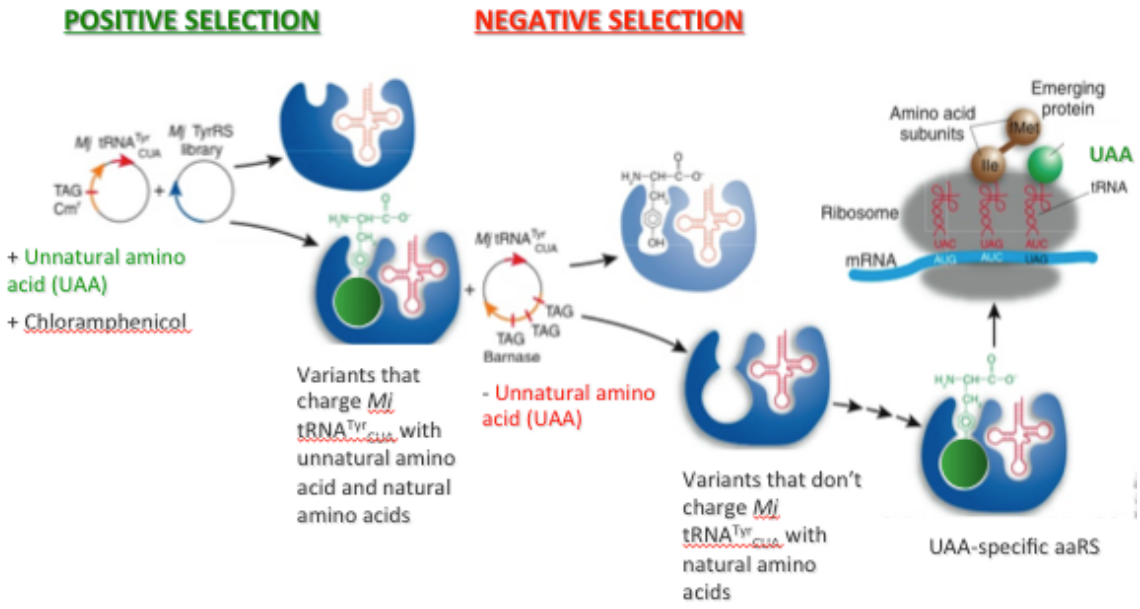


Figure 4. The general procedure for double sieve selection for a functioning orthogonal mutant aaRS that will incorporate a specific Unnatural Amino Acid.

These selective pressures promote the charging of an unnatural amino acid and not the charging of endogenous amino acids. However, there is no selective pressure for different UAAs to be charged to the tRNA. As a result, it remains possible that the evolved aaRS may be able to recognize and charge other UAAs that were not involved in the selection. This promiscuity can afford the incorporation of multiple UAAs by a single aaRS.

In addition to being recognize by an aaRS and charged to the tRNA, the UAA must meet other requirements to afford its incorporation into a protein.[2] For instance, the UAA must be permeable to the cell or readily synthesized by the cell. The UAA must also be compatible with all of the ribosomal machinery of the cell. These conditions are typically met when using prokaryotic organisms but issues have arisen when translating the technology to eukaryotic organisms. Consequently, yeast (a eukaryotic organism) have been exploited to incorporate UAAs with great success using similar methods as

employed with E. *coli* (Figure 2). More substantial problems exist when UAAs are integrated into other eukaryotic cell lines, like mammalian cell lines, due to complications with cell growth, the two-stage selection process and cell divisions.[2, 12] There has been some success though, by first manipulating yeast cells to incorporate UAAs and then transfecting mammalian cell lines with yeast DNA. Further improvement of eukaryotic expression still required making this approach more viable.

Additionally, current research involves finding new methods to incorporate UAAs and to continue to improve the current methods. New archaebacterial aaRS/tRNA are being used in *E. coli* and new codons, other suppressor codons and four base codons, are being exploited to express multiple different UAAs in proteins.[13] In order to increase protein yields, expression methods are being optimized as well.[14] For instance, the plasmid vectors that are used to incorporate the genes to express the aaRS/tRNA were optimized with to be highly efficient, with the best promoters, ribosomal binding sites, and terminators. A plasmid vector called pEVOL was engineered, which provided the greatest protein yields and greatest stability for incorporation into *E. coli*. With these improvements UAA incorporation is continually developing into a much more manageable and useful technology.

1. II. BIOORTHOGONAL CHEMISTRY

One of the greatest advantages of employing unnatural amino acids is that it facilitates the incorporation of bioorthogonal functional groups. For an amino acid to be completely orthogonal it must not disturb the function of the protein, nor impede the other functions that occur within the cell.[2] This is not to say that the residues that are incorporated are unreactive; however, they are only reactive with other introduced substrates. The introduced functionalites that undergo non-specific cross-reactions with the natural cellular processes are not completely orthogonal and can be detrimental to the organism. Interaction should be localized to the introduced functionalities and the reaction conditions that occur between them must also occur under physiological conditions, as deviation will also lead to biological damage.

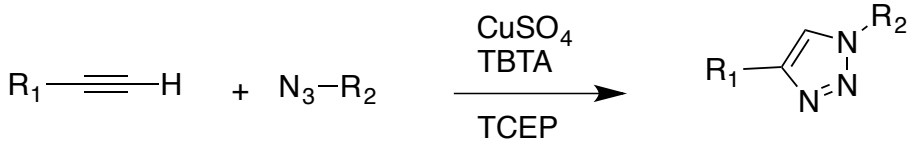
This research focuses on the synthesis of UAAs possessing bioorthogonal handles that will facilitate novel chemistries within the cell. The functionalities can then react with tags or other molecules in a site-specific manner in a bioconjugation reaction to further modify the protein and its function.

With theses moieties (ketones/aldehydes, azides, alkynes, and alkenes) present in only specified proteins containing UAAs, reactions will only occur with the protein of interest.[2] The unnatural functionalities must fulfill two essential criteria to function as efficient tools. First, the functional groups should not be able to interact with normal functions of the organism; only with other introduced functionalities. Second, the reactions must also occur in physiological conditions, as to minimize effects on the organism. Several reactions satisfy the first condition, but many reactions with UAAs do not meet the second condition, occurring outside of standard physiological conditions. Therefore, the development of new functionalities and reaction conditions is important to maximize the utility of bioconjugations in biological systems. Using these bioorthogonal reactions we can study processes in the cell like DNA replication, protein function, and glycosylation without disruption of the natural processes

1.II.A. CLICK CHEMISTRY

A commonly used bioconjugation reaction in our lab, was developed by Chan et al., involves a copper (I) catalyzed-1,3-dipolar cycloaddition, and is known as a "click reaction" (Scheme 1).

Scheme 1.



While this reaction is truly bioothogonal, as no naturally occurring alkynes or azides exist within proteins, it is challenging to perform in a truly *in vivo* setting.[15] The reaction has a high reaction rate and high specificity, but high concentrations of copper within an organism are toxic.[8, 15] Thus, a modified reaction has been developed using a strained cyclooctyne with fluorine substituents which performs the same dipolar

cycloaddtion with azides, but does not require the copper catalyst.[16] This copper-free mechanism has comparable kinetics to the copper catalyzed reaction but is not cytotoxic. More reactions such as this one are needed to continue to allow for the cost of orthogonally.

1.II.B. OXIME CHEMISTRY

Oxime ligations are another form of bioconjugation involving the reaction of ketones or aldehydes with alkoxyamine groups (Scheme 2). This reaction is both bioorthogonal and can proceed in nearly physiological conditions. The reaction employs ketones and aldehydes to react with alkoxyamine compounds to form covalent oxime linkages (Figure 5).[8, 17, 18]

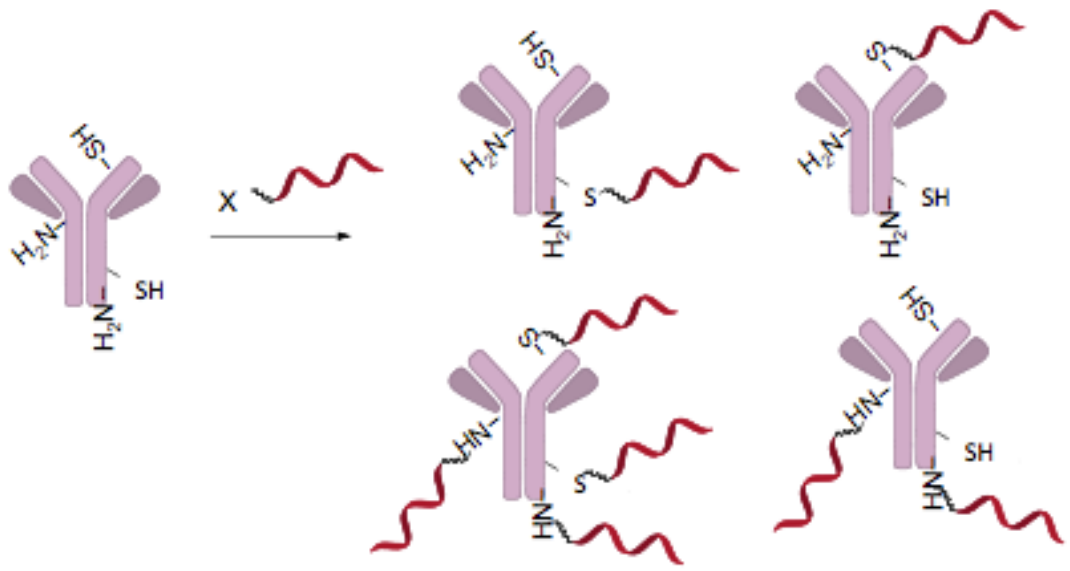
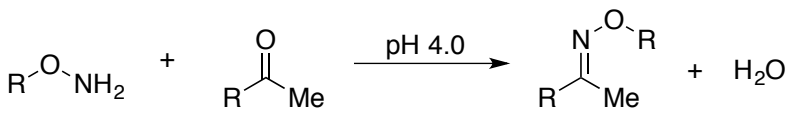


Figure 5. The current non-specific antibody conjugation method.

These alkoxyamines have enhanced reactivity due to the α-effect, where substituents on the α-position of amines have a significant role on the nucleophilicity of the compound.[19] Substituents contribute significantly to the alkoxyamine's reactivity with an electrophile as they possess lone pairs of electrons that can contribute to a stabilized positive transition state. This reaction using an alkoxyamine functional group generates a stable covalent bond and can be useful in forming linkages with various molecular tools through interaction with aldehyde/ketone functionalities. For instance, recent work has

used oxime ligations to fluorescently label glycosylated cell surfaces that harboring exposed aldehyde functionalities.[17] Cell surface glycans with sialic acid residues were modified to have aldehyde functionalities by periodate oxidation. Alkoxyamine functionalities on fluorescent probes could then interact via an oxime condensation. Disadvantages of this reaction are that this reaction was optimal in acidic conditions, which is not ideal for biological systems, and the reaction has slower reaction rates than other common bioconjugation reactions. However, Dirksen et al. elucidated that aniline, or aniline-like molecules, have a nucleophilic catalytic effect.[18] This greatly accelerated the reaction and allowed it to occur at near physiological conditions, making it highly efficient in labeling cells. Ultimately, the incorporation of this functionality into biological macromolecules will enable stable and orthogonal conjugation reactions with fluorescent probes, cancer drugs, or other molecular tools that contain ketone/aldehyde functionalities.

Scheme 2.



1. III. APPLICATIONS OF BIOCONJUGATION REACTIONS

Incorporation of UAAs and corresponding bioconjugations employing the residues creates various unique opportunities for proteins. For instance, profluorophores are a class small biomolecules that have fluorescent structures, like conjugated rings and conjugated alkyl chains, which are quenched by electron donating or electron withdrawing functionalities.[20, 21] In the case of azide and alkyne functionalized profluorophore molecules, these two chemical handles can undergo a 1,3 dipolar cycloaddition to form a triazole ring structure, which delocalizes the electrons in the structures and the product subsequently exhibits fluorescence. This represents a very useful tool to fluorescently label cells, with very little background noise due to unspecific labeling. Since it is already possible to incorporate an alkyne or azide UAA reaction partner into proteins, this technology can readily be translated to a protein context. While

several fluorescent UAAs already exist and have been incorporated into proteins, they tend to be larger residues and can possibly disrupt a protein's functionality. Moreover, the native fluorescence necessitates the washing away of unincorporated UAA, leading to substantial issues as this is difficult to do in a biological system. Utilization of pro-fluorogenic residues can therefore be used to specifically label cancer cells with fluorescent tag without undesired background fluorescence. This can ultimately aid surgeons to locate and excise malignant tumors during operations.[5]

Other novel applications of biocongugation technologies include the synthesis of specific antibody-drug conjugates.[6] Common cancer therapeutics are produced by reaction of either cysteine or lysine residues on an antibody with a toxic small molecule. Once the antibody is bound to cancer cells it is endocytosed into the cell, degraded, thus releasing the cytotoxic cancer drug into the cell. The synthesis of antibodies with UAAs allows for greater specificity in the generation of this conjugate, compared to the unspecific conjugation methods currently used (Figure 5).

Antibody conjugates were synthesized by Axup et al. with a UAA incorporated in specific residues of the antibody (Figure 5). This UAA possessed a keto group that was reacted with a cancer drug that with an alkoxyamine functional group to generate an oxime linkage. This methodology allows for specific drug application, as the antibodies are only endocytosed in cancer cells, and therefore the drug will only be released into cancer cells, decreasing systemic toxicity.

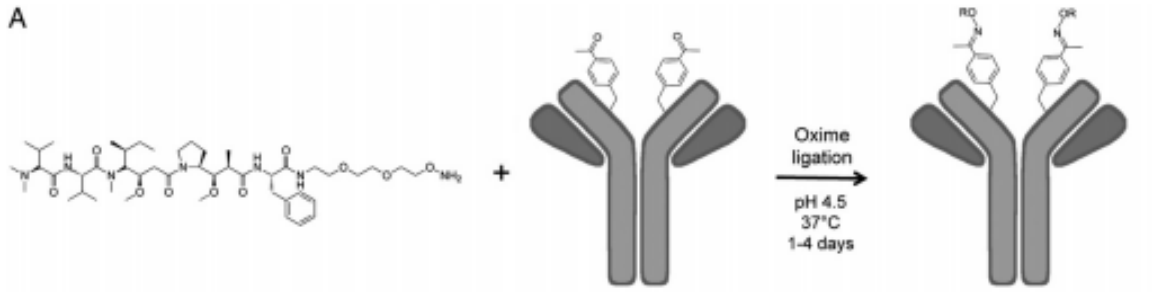


Figure 6. Keto group of incorporated UAA that has an alkoxyamine interaction with it to form an oxime linkage, from *Proc. Natl. Acad. Sci. USA* **2012**, 109 (40), 16101-6.[6]

Within the emerging field of synthetic biology UAA technologies have proven to be extremely valuable. This subset of biology looks to develop new biological tools that will aid in our attempts to begin to understand various life processes and create new

functions for organisms. UAAs can play a significant role in the development of these tools. For instance, Minaba et al. used UAA introduction into bacterial cells as a signal for a translational switch known as an AND switch.[7] This switch, when activated by the binding of the UAA, will act as a promoter on the gene transcribing the aaRS/tRNA pair that will incorporate the UAA (3-Iodo-L-tyrosine) into a protein. Without the presence of UAAs in the cell, transcription will cease and eventually the protein will be degraded. This was used as a proof of concept that primative translational switches can be designed to induce transcription and translation of aaRS/tRNA pairs by the presence of UAAs, which allows more specific control of when UAAs will be integrated into the protein of more complex organisms. This could lead to improved study of cellular processes in the future, as a researcher could expand upon this work and possibly design a control mechanism that would not only allow the cellular uptake of a UAA, but also the subsequent incorporation into a protein (Figure 7).[7]

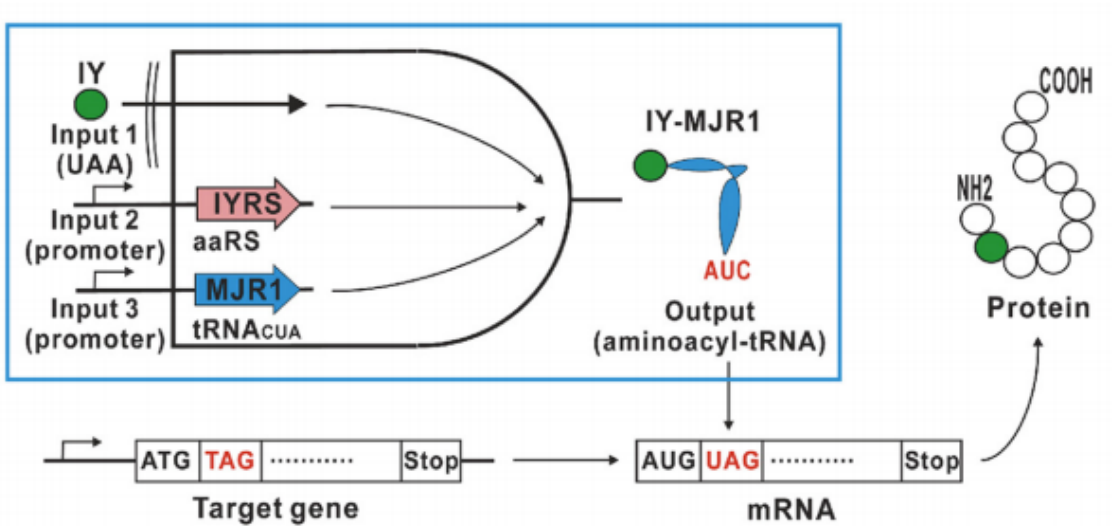


Figure 7- Diagram of the genetic circuit using UAA from *Appl. Environ. Microbiol.* **2014**, 80 (5), 1718-25.[7]

These functions emphasize the vast utility of this technology. With its improvement and expansion there is significant of promise in what can continue to come in the future.

CHAPTER 2: DEVELOPMENT OF ALKOXYAMINE UNNATURAL AMINO ACIDS FOR BIOCONJUGATION APPLICATIONS

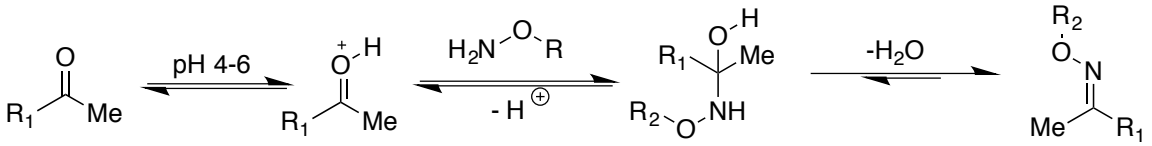
2. I. INTRODUCTION

A commonly exploited bioconjugation reaction involves the reaction of an alkoxyamine with a carbonyl moiety to generate a stable oxime linkage. This reaction is typically initiated by protonation of the aldehyde/ketone functionality, often requiring a slightly acidic environment (pH ranging from 4-6).[22-24] Unfortunately, due to the innate reactivity of aldehyde UAAs containing these functionalities are difficult to genetically incorporate. But in recent work, Tuley et al. incorporated an aldehyde UAA, *meta*-formyl-phenylalanine, into protein and labeled it using a hydrazine dye.[25] This labeling reaction was rapid as compared to reactions with ketones (182 M^{-1} s^{-1}) at pH 7.[25, 26] Ketone functionalized UAAs are more commonly incorporated into protein and are used frequently in studies of oxime ligations. Functionalities like *p*-acetyl-L-phenylalanine (pAcF) possess a ketone functionality, and can readily be incorporated into proteins using the *p*AcFRS to introduce a bioorthogonal handle for these reactions (Figure 8).[11] The alkoxyamine functionality has yet to be incorporated into protein as an amino acid functionality, it is usually a functionality of a non-biological associating molecule.

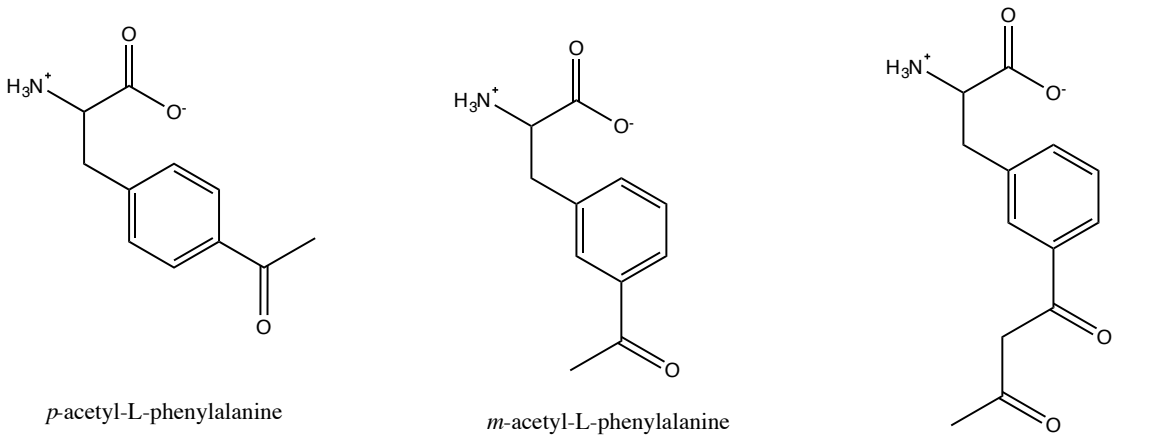
There are various advantages and disadvantages to using oxime ligations, creating a unique niche of utility within the study of biological conjugation. The primary advantage of these reactions includes their overall ability to occur under physiological conditions in the presence of special catalysts.[22, 27] Moreover, the reaction yields a stable, geometrically well-defined, covalent oxime linkage, and unlike many other bioconjugations does not rely on cytotoxic transition metal catalysts.[28] However, these reactions also have disadvantages relative to other bioorthogonal chemistries. The fact that the reaction is acid catalyzed, requiring a moderately acidic environment, is deleterious for living cells, as protein degradation occurs and other cellular functions may be hindered.[22] Perhaps even more disadvantageous is the slow reaction rate of the ligation.[22, 28] With an approximate reaction rate of less than 10^{-2} $M^{-1}s^{-1}$ these reactions are sluggish and highly pH dependent.[23] Using higher concentration of the reactants within

the cell has been used to accelerate the reaction, but can lead to toxicity and non-specific reactions with molecules in the cell that also have ketone/aldehyde moieties (e.g. pyruvate, triglycerides, nucleic acids, acetyl-CoA).[22] In order to remedy these slow reaction rates, Zeng et al. have shown that the addition of strong nucleophiles, like aniline, will effectively accelerate the reaction rate.[17] Aniline and similar compounds have been demonstrated to act as catalysts that enhance the electrophilicity of the carbon, using its conjugated ring system to withdraw electrons, and therefore rendering it more susceptible to addition of an alkoxyamine. These additives have been shown to increase the reaction rate to 8.2 $M^{-1}s^{-1}$ in a neutral environment (Scheme 3).

Scheme 3.



Kim et al. investigated tagging cells using this reaction by reacting *p*-acetyl-L-phenylalanine and *m*-acetyl-L-phenylalanine residues incorporated in the outer membrane protein LamB (Figure 8).[29] These UAAs were then reacted with fluorescein and biotin hydrazides in a site-specific manner. In a separate study, Zeng et al. incorporated a diketone containing derivative of phenylalanine in the Lys7 Z-protein domain and labeled the protein with an Alexa Fluor 488 hydroxylamine derivative or a hydroxylamine biotin derivative.[30] The results of these studies indicated that this method of protein tagging is highly efficient and site specific.



p-acetyl-L-phenylalanine

m-acetyl-L-phenylalanine

m-β-diketone-L-phenylalanine

Figure 8. Ketone functionalized UAAs

Additionally, these reactions have been applied to label or kill HER2 expressing cancer cells[31]. HER2 oncogenes express membrane receptor proteins that are found in breast cancer cells. Bispecific antibodies that allow for the linkage of cancer cells and T-cells which express membrane receptor protein CD3 using UAAs with biocojugation accessable reisdues. The anti-HER2/anti-CD3 bispecific-antibody linkage was effectively prepared through the incorporation of *p*AcF, which interacted with a bifunctional ethylene linker. This linker is functionalized with an alkoxyamine residue and an azide or alkyne residue to allow for conjugation to a HER2 cell or a T-cell and then linkage of the two by Cu(I)-1,3-cycloaddition, allowing for a more directed immune response and increased cytotoxicity for HER2 cancer cells with picomolar concentrations of the antibody.[32] This technology was also employed to create antibody drug conjugates (ADC), generating immunoconjugates with cytotoxic molecules to specifically target and kill HER2 expressing cancer cells, while significantly reducing off-target cytotoxicity.[6] Finally, oxime methodologies have been employed to detect cancer cells by immuno-PCR using similar antibody bioconjugates[31]. These antibodies utilize an oxime bioconjugation of an oligonucleotide to the structure of an antibody, which has many sites for PCR amplification. These sights are amplified by rolling circular amplification using a circular oligonucleotide. Amplification produces short (~20 nt) sequences that can then be fluorescently labeled using fluorescently tagged oligonucleotide

complements, allowing fluorescent labeling of antibody bound Her2 expressing cancer cells (Figure 9).[31]

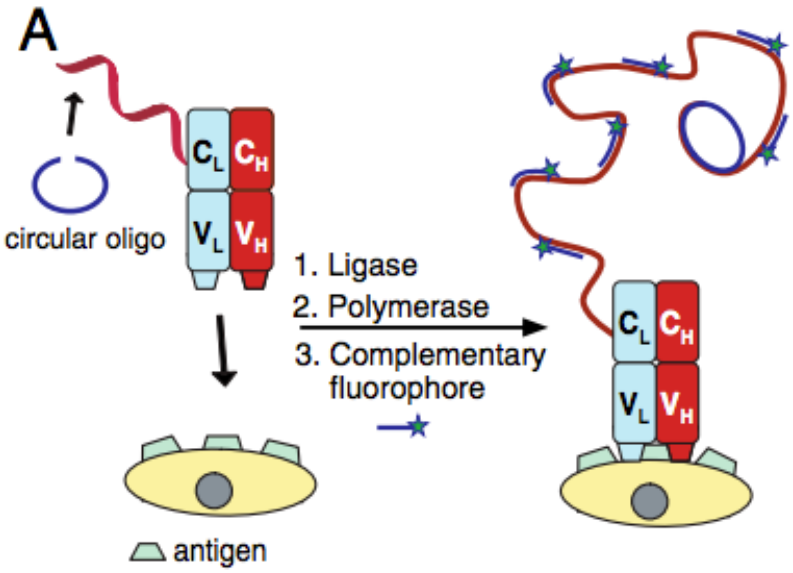


Figure 9: Site-specific immuno-PCR as a method of fluorescently labeling Her2 cancer cells, from *Proc. Natl. Acad. Sci. USA* **2012,** 109 (10), 3731-6.

Ultimately, in all of these examples, the specificity and bioorthogonality of the UAAs have substantially advanced the field and provided novel treatment and detection methodologies for cancerous cells. Consequently, further development of the oxime ligation to increase physiological compatibility, accelerate reaction rates, and incorporate aminoxy functionality into biological systems will only increase the utility of this reaction.

In all of the previously described applications molecules with alkoxyamine functionalities have been reacted with protein containing a ketone. An alkoxyamine residue has not yet been incorporated into a biological system via unnatural amino acid, technologies. This UAA would possess substantial advantages, especially as the reactivity can be more highly controlled than the ketone analogs previously employed. This moiety can ultimately be used in a bioconjugation reaction with either a ketone or aldehyde partner. The incorporation of an aminooxy unnatural amino acid into proteins will allow site-specific oxime ligation reactions to occur without the concern of cross-

reactivity that may occur with more highly reactive aldehyde UAAs. This also facilitates bioconjugation reactions with previously synthesized ketone containing amino acids. These linkages will result in stable covalent bonds and can be employed in the development of new reaction methodologies that are more useful and applicable in biological systems.

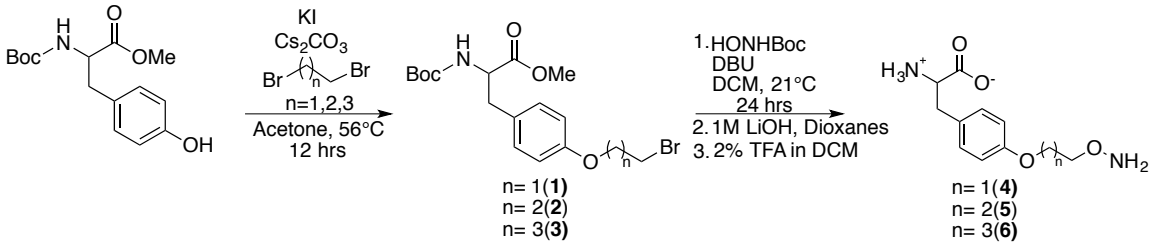
In this project, three tyrosine derivatives will be synthesized with alkoxyamine functionalities and subsequently incorporated into green fluorescent protein. These different unnatural amino acids will possess different numbers of carbon tether lengths (2-4 methylene units), which will separate the orthogonal handle from the core of the protein. This alteration in reaction distance to the steric hindrance of the protein may facilitate more optimal reactions. This has previously been demonstrated with alkynyl and azido tethered UAAs, demonstrating a correlation between tether length and reactivity.[33] This will allow us to optimize reaction conditions to best facilitate its interaction with ketones/aldehydes. Through study of these variants we will elucidate which structure is optimal for synthesis, incorporation into protein, and formation of oxime linkage as an amino acid within a protein.

2.II. RESULTS AND DISCUSSION

2.II.A. SYNTHESIS OF UNNATURAL AMINO ACIDS

To optimize bioconjugation using an alkoxyamine functionalized tyrosine derivative, three unnatural amino acids were synthesized with 2, 3, and 4 alkyl chain length tethers. These methylene units, extend the functionality from the steric bulk of the protein, and may allow for better interaction with a ketone/aldehyde reaction partner. These tyrosine derivatives can be synthesized in a two-step process utilizing S_N2 nuecleiophilc substitution reactions (Scheme 4).

Scheme 4.

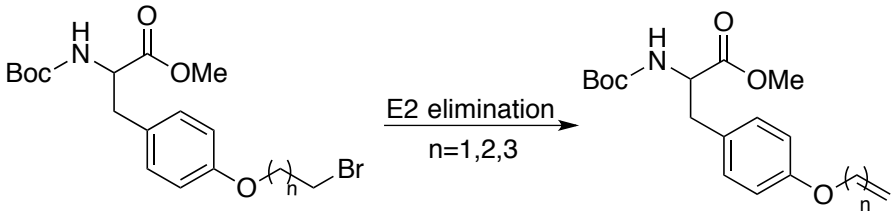


The synthesis is intitiated with the addition of a brominated alkyl chains to a common *N*-tert-butyloxy (Boc)/OMe protected tyrosine precursor via a substitution reaction with dibromoethane, dibromopropane or dibromobutane. This reaction afforded tyrosines possessing brominated alkyl chains extending from the *para*-substituted oxygen (**1-3**). These bromines served as suitable leaving groups for the next step of the synthesis involving a substitution reaction of a Boc- protected aminooxy nucleophile.

The initial alkyl bromine substitution reactions were based on previous work in the Young Lab. A primary issue associated with this synthesis was the propensity to under go an elimination reaction instead of one of the S_N2 reactions as desired. This was especially prevalent in the reactions involving the 2 and 3 carbon tether bromo-tyrosine reactions. To prevent elimination and facilitate the desired substitution reaction, catalytic amounts of potassium iodide were added to the reaction, which would ideally substitute with the bromine, and act as a better leaving group for the desired substitution reaction. This modification was not successful as it only yielded unreacted starting materials. In another attempt to obtain the desired reaction, a literature search revealed that substitution reactions work more efficiently in acetone as a solvent rather than the original DMF solvent.[34] This modification favored the substitution reaction and facilitated the synthesis of all three bromo-tyrosine derivatives. The effectiveness of the acetone solvent system was most likely due to the altered polarity of the solution as compared to DMF. DMF is more polar providing greater solvent dipole-dipole interactions with the bromine, allowing an elimination reaction to proceed.

Based on this previous work, the initial substitution reactions to yield the tethered bromides were performed using both potassium iodide as a nucleophilic catalyst, as it can replace the bromine and function as a better leaving group for substitution reactions, and acetone as a solvent. In these reactions, the majority of the starting material underwent the desired S_N2 reaction, but some elimination side product was still observed (Scheme 5). As the alkyl chain length decreased there was an increase in the amount of elimination byproducts observed, due to increased stability of the elimination intermediate. We hypothesized the elimination was still occurring because the reaction vessel and solvents were not anhydrous and polar water molecules were still facilitating an elimination reaction. After thorough flame dying of all vials and stir bars the amount of byproduct was reduced. The byproduct was then removed via silica flash chromatography and the desired bromo-tyrosine (**1-3**) products were obtained in good to moderate yields and high purities.

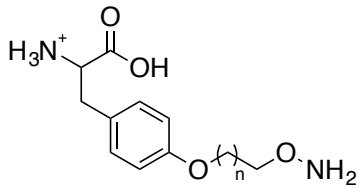
Scheme 5.



The bromotyrosines (**1-3**) were then subjected to a subsequent S_N2 reaction in which the bromine is replaced by a Boc protected aminooxy functionality. Specifically, N-Boc Hydoxylamine is dissolved in dichloromethane in the presence of a non-nucleophilic base 1,8-diazabicycclo-7-undecene (DBU). DBU was employed, as it is a weak base that will not remove any protecting groups but will help facilitate the hydroxyamic acid to undergo an S_N2 reaction. This reaction proceeded as expected and the desired products were obtained, with only very small amounts of elimination by-products. Elimination most likely occurred due to the reduction in chain lengths or use of DBU, even though it is a weaker base, the steric interactions of the base may have facilitated the elimination. The reaction was then purified using silica flash chromatography and the desired products were isolated.

While the desired aminooxy functionality was installed, both it and the amino acid backbone still harbored protecting groups to prevent undesired chemical reactions in the previous steps. A Boc protecting group was used to protect the amino group on the backbone as well as the aminooxy functionality and is acid labile. A methyl ester was used to protect the carboxylic acid group of the amino acid and is base labile. To remove these groups a successive addition of 1M aqueous lithium hydroxide followed by neutralization and subsequent treatment with 2% TFA in DCM was performed to yield **4-6** . The deprotected products (**4-6**) were then concentrated under vacuum and analyzed using ^{1}H and ^{13}C NMR spectroscopy. The overall yields of the final products of this protocol are summarized in Table 1.

Table 1. Overall Yields for the Syntheses of Alkoxyamine UAAs

UAA	n	% Yield
	1 (**4**)	24.6% (**p2AOY)**
	2 (**5**)	30.4% (**p3AOY)**
	3 (**6**)	28.2% (**p4AOY)**

2.II.B. INCORPORATION OF UNNATURAL AMINO ACIDS

The newly synthesized alkoxyamine UAAs could then be investigated for incorporation into green fluorescence protein (GFP) using the Schultz methodology previously discussed (see Chapter 1). Based on previous experiments the surface-exposed residue 151 was selected as the site of incorporation and expression of the protein was attempted utilizing *E. coli*. This surface-exposed residue of GFP is an ideal site for incorporation as allows for presentation and accessibility of the amino acid residue to permit subsequent bioconjugation reactions. It also serves as a useful screening mechanism for successful incorporation of the UAA into the protein. This is due to the pET-GFP plasmid harboring a stop codon (TAG) at site 151. Thus, the only way

expression will occur is if the tRNA can be charged with the UAA and utilized to suppress the stop signal. If the bacteria do not have the mutated synthetase or do not properly charge the UAA then translation will halt at the codon, preventing proper GFP expression and resulting in an absence of fluorescence

In order to charge the UAAs with to the tRNA, an aaRS that at recognizes this novel UAA must be identified. An aaRS can either be evolved in a double sieve selection (Chapter 1), or the polyspecific nature of certain mutant aaRSs can be exploited. For our purposes we will be doing the latter due to the time consuming nature of the selection process, and the most effective aaRSs were identified using a fluorescence screen of six aaRSs. These 6 aaRSs were selected due to either their previously described promiscuity or due to their ability to incorporate structurally similar UAAs to the aminooxy UAAs.[11] Using a microplate reader, *E. coli* transformed with a pET-GFP-TAG_{151} and one of the pEVOL-aaRS plasmids were expressed with each of the 3 tethered UAAs in triplicate. These were then compared to control expressions, which had not received the UAA to see which aaRS was the most efficient and able to incorporate each UAA (**4-6**) (Figure 10 a-f)

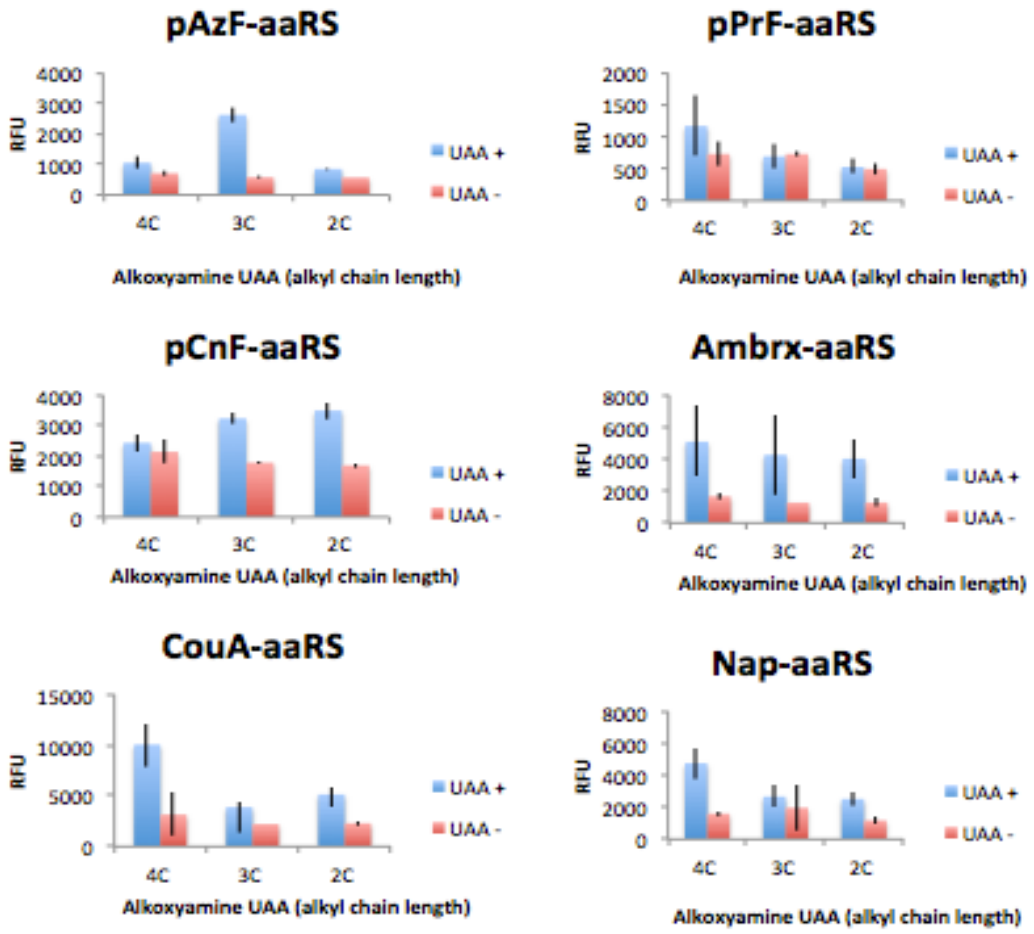


Figure 10. Synthetase screen of **4,5,** and **6** using the a) pAzF-aaRS, b) pPrF-aaRS, c) pCNF-aaRS, d) CouA-aaRS, e) Ambryx-aaRS, and f) Nap-aaRS.

From this screen we identified that the aaRS previously used to incorporate a napthalene UAA (Nap) and the aaRS evolved to incorporated a coumarin UAA (CouA) were most the most effective at incorporating the four carbon tether alkoxyamine UAA (**6**). We hypothesize these to be effective aaRSs due to their larger binding pocket relative to other aaRSs, facilitating the more substantial space requirements of the 4-methylene tethered aminooxy UAA. The *para*-azidophenylalanine (*p*AzF) synthetase was the most effective at incorporating the three carbon tethered alkoxyamine UAA (**5).** This is most likely due to the structural similarity of the two UAAs and the similar polarities of the azide and aminooxy functionalities. Finally, the Ambrx aaRS can incorporate the two carbon tether UAA (**4**), as well as the other two tether length UAAs (**5** and **6**), with high

efficiency. This was expected as this aaRS has been demonstrated to be one of the most promiscuious aaRSs known, and has been previously used to incorporate UAAs with methylene tethers. These represent exciting results as an aaRS selection is not necessary, and as they demonstrate that our newly synthesized aminooxy UAAs can readily be incorporated into a protein context. Due to the universality of the Ambryx aaRS for all three UAAs, we attempted to utilize it for all future expressions.

Having identified viable aaRSs for incorporation in the screen, we next sought to confirm these results in more practical expressions that facilitated the isolation and purification of the aminooxy-GFP mutatnts. Expressions were performed using the Ambryx-aaRS, and the 3 methylene aminooxy UAA (**5)** was efficiently incorporated into GFP. The 2 carbon UAA derivative (**4**) was incorporated into GFP using the Ambrx – aaRS effectively as well, but the fluorescence intensity was less than the fluorescence of (**5**), signifying a lower efficiency of incorporation (Figures 11 and 12). Finally, the 4 carbon UAA (**6**) did not incorporate as readily as it appeared by the data obtained from the synthetase screen. This result prompted some troubleshooting of the expression to improve incorporation efficiency. Early attempts to optimize incorporation of (**6**) attempted to vary several different aspects of the expression. These variables included the optical density at induction and the scale of the expressions. Changing these variables did not improve the expression yields, so the next variable altered was the aliquot of pEVOL-aaRS plasmid utilized in the transformations, as previously it had been observed that some DNA aliquots had been contaminated. After multiple attempts to transform and express the plasmids, expression of (**6**) in GFP was finally achieved, albeit at low expression levels. This is most likely due to the fact that the long tether creates steric hindrance in the active site binding pocket of the aaRS. All of the UAA incorporated protein was then purified using Ni-NTA affinity chromatography. This afforded the immobilization the mutant GFP due to the presence of a 6X histidine tag, which could be subsequently eluted into solution with increasing concentration of imidazole.

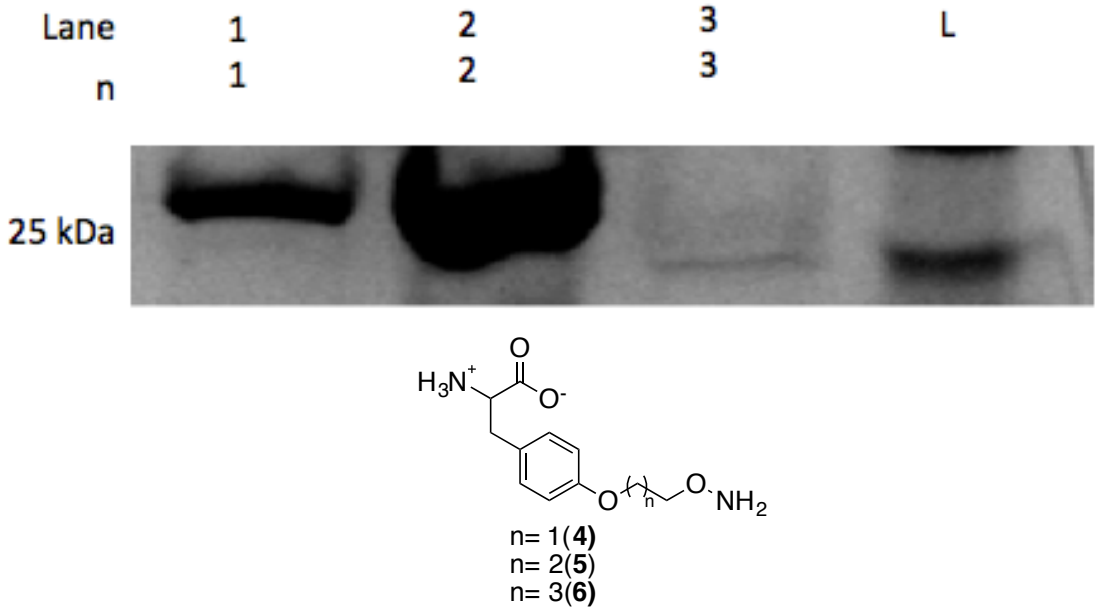


Figure 11. SDS-PAGE gel of (**4,5, and 6**) mutant GFP-151 protein

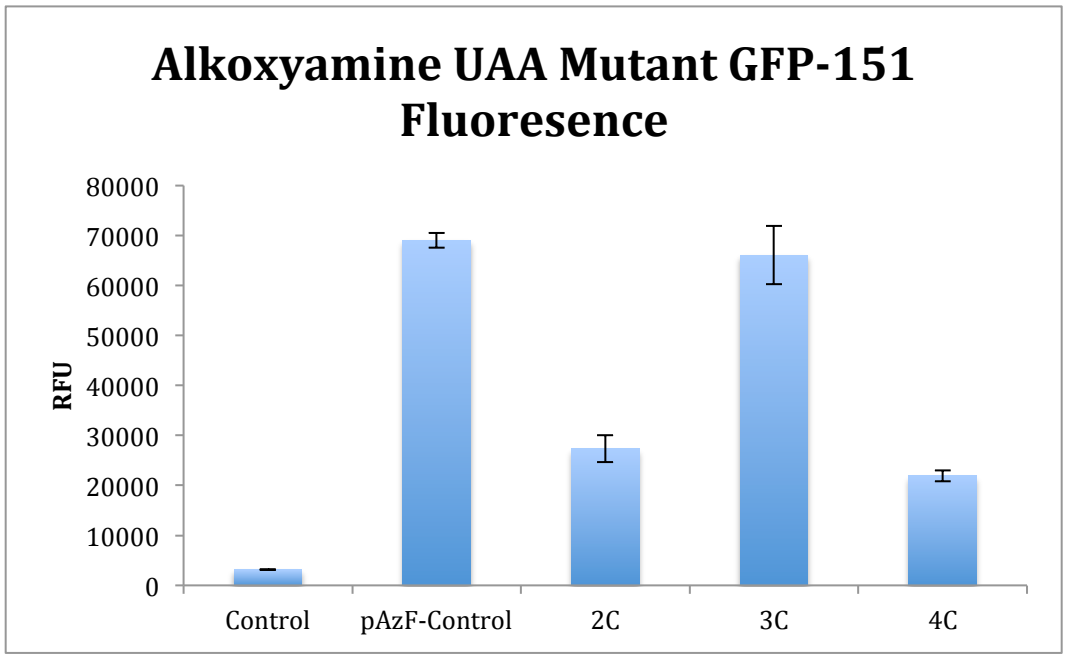


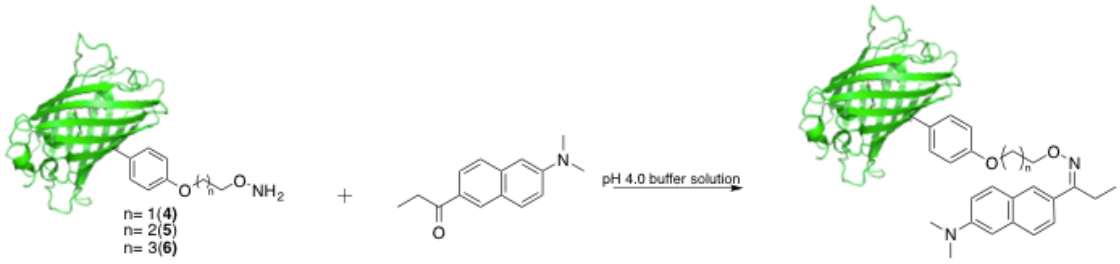
Figure 12. Incorporation efficiency of (**4, 5, and 6**) into GFP-151 using Ambrx-aaRS compared to incorporation of pAzF in GFP-151 using Ambrx-aaRS and a negative control.

2.II.C. BIOCONJUGATION EXPERIMENTS

Prodan Assay

Once the alkoxyamine UAAs (**4-6**) were successfully incorporated into GFP, we wanted to assess the ability of each UAA to undergo an oxime ligation with an aldehyde or ketone, and ascertain if the variable tether could be utilized to optimize the reaction. Our first attempt to create an assay involved the reaction of the UAAs with a prodan dye, provided by Dr. Abelt's lab, which contains a ketone and is a fluorescent molecule. The molecule was reacted with the UAA containing GFP and would yield a fluorescent product even after the GFP was denatured (Scheme 6). The mutant GFP (100μM) and prodan (10 mM) were reacted overnight at 37°C, following a literature protocol from Brustad et al., we then purified by size exclusion chromatography, to eliminate excess prodan fluorophore.[35] The purified protein was then analyzed by SDS-PAGE. During this process, the GFP was denatured by the SDS and lost its inherent fluorescence on the gel, resulting in the only observable fluorescence arising from the GFP-prodan bioconjugate.

Scheme 6.



Unfortunately, this assay proved to be unsuccessful at assessing the success of the oxime ligation. The fluorescence was not a concentrated band were the GFP was found but smearing of the fluorescent molecule (Figure 13). We first suspected that this might be due the reaction conditions (pH, concentrations, etc.) not being ideal for this type of bioconjugation.

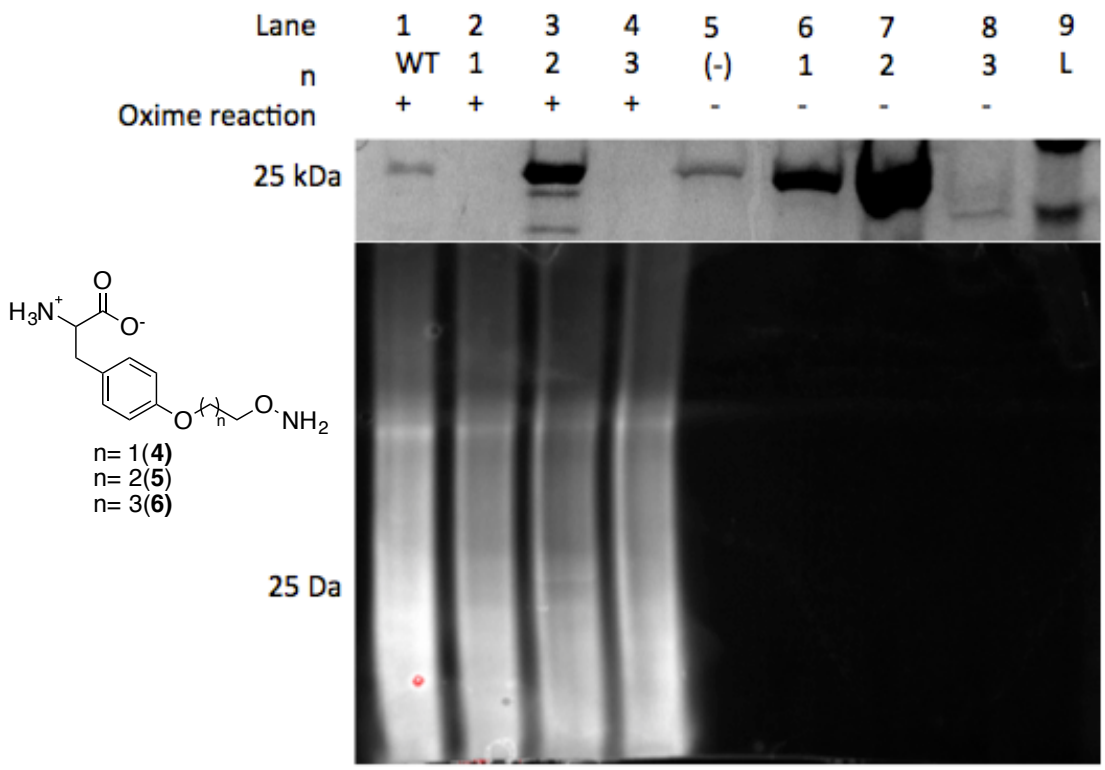
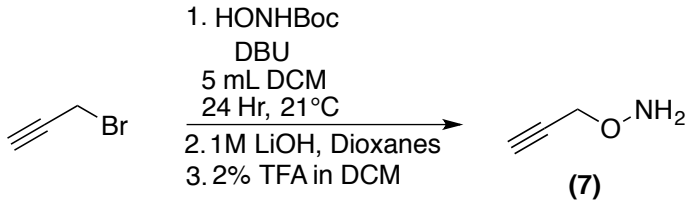


Figure 13. SDS-PAGE gel of oxime ligation reactions between (**4, 5, and 6**) GFP-151 mutants and prodan. Lanes 6-8 demonstrate successful incorporation of the UAAs into GFP-151. Lanes 2-4 show protein was present after the reaction but the reaction did not occur as prodan is smeared and unbound as seen on the fluorescence image.

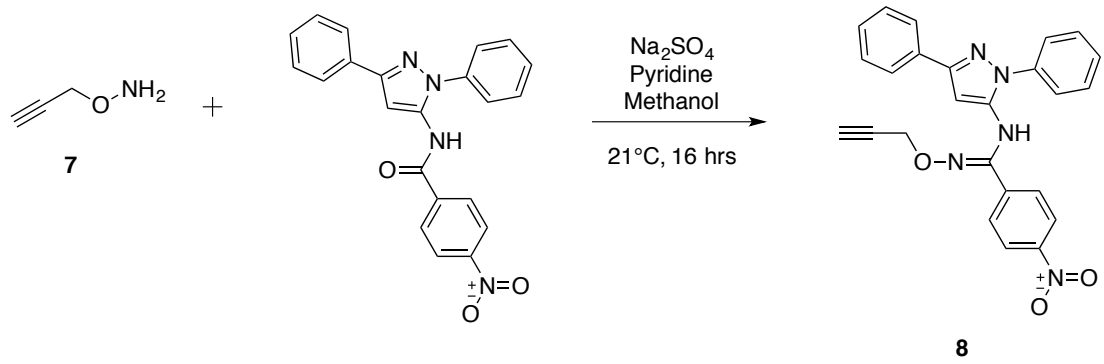
In the literature, reaction conditions that were not suitable for biological systems were identified; however, these conditions could be modified to potentially minimize the detriment to the protein by reducing the amount of pyridine added, and conduct the reaction in an aqueous environment.[36] To test this reaction we first used propargyl aminooxy and 4-nitro-*N*-(1,3-diphenyl-1H-pyrazol-5-yl)benzamide (DPAP) which has ketone functionalities and reacted it with a propargyl aminooxy derivative (Schemes 7a and 7b).

Scheme 7a and 7b:

a)



b)



This reaction was successful, as confirmed by ^{1}H NMR spectroscopy, and we therefore adapted this reaction protocol to the UAA incorporated protein and prodan. Unfortunately, it still did not result in an oxime ligation between the GFP and the prodan. No fluorescence was observed by SDS-PAGE analysis despite the presence of protein when stained with coomassie blue. It was then hypothesized that the prodan would not react with the UAA because the molecule tends to aggregate with other prodan molecules in an aqueous environment. This tendency to aggregate would not allow the alkoxyamine and the ketone functionalities to be in close proximity to react, and the unreacted prodan was then eliminated during the purification of the protein. Thus, a new assay was necessary to assess the effectiveness of the oxime ligation.

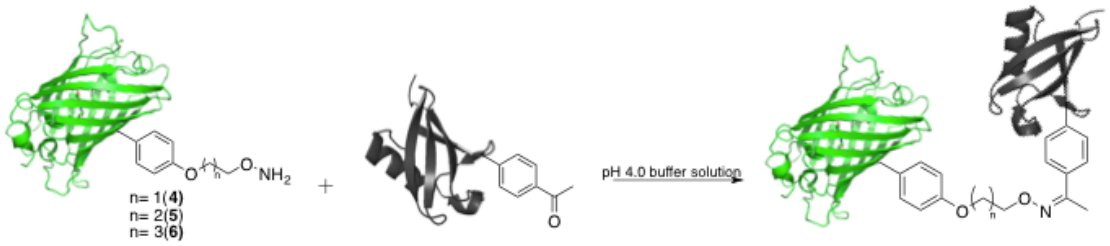
Ubiquitin Assay

Instead of using fluorescence to determine the success of the oxime ligation we began to look for methods where we could perform the reaction with a biological macromolecule, like protein or DNA, which would result in a gel shift in a SDS-PAGE

due to the addition of a compound of a significant molecular weight. Initially, we hoped to utilize a DNA oligonucleotide with a ketone/aldehyde modification, but these are no longer commercially available, and would be synthetically challenging to access. The next logical strategy involves the coupling of two proteins, while this is challenging due to the lack of both ketone/aldehydes and aminooxy groups, we were able to introduce them into the proteins using the UAA mutagenesis technology, and already had the aminooxy-GFP in hand. Previous attempts in the lab to dimerize GFP have been limited due to the size of the protein, so an alternative protein bionconjugation partner was investigated. Specifically we aimed couple the UAAs (**4-6**) that are incorporated into GFP with *para*-acetylphenylalanine (*p*AcF) incorporated into a small protein called ubiquitin. Polymeric chains of ubiquitin perform various signaling processes, and thus are very exciting for future experiments in more biologically relevant settings. For instance, ubiquitin chains can promote protein degradation, cell cycle regulation, transcriptional regulation, vesicular transport of protein through the cell, and are involved in the processing of antigens.[37] A ubiquitin plasmid with a TAG mutation at residue 48 was provided by Dr. Ashton Cropp from Virginia Commonwealth University. Ubiquitin is a 8.5 KDa protein that will show corresponding shift in the size of a GFP band, once bound, on a SDS-PAGE gel.

After obtaining the plasmid from VCU, initial issues arose in the expression of the protein as we were having trouble transforming the plasmid and obtaining little yield on when finally expressed. Several troubleshooting experiments were attempted, including the transformation of the plasmid, the expression protocol, and altering the percentage of acrylamide gel (12%). Finally, we were able to express and confirm that we had ubiquitin protein containing the *p*AcF UAA. This mutant ubiquitin was then ready for the optimization of the oxime ligation reactions with the previously prepared alkoxyamine UAA incorporated GFP (Scheme 8).

Scheme 8.



Unfortunately, the ligations using this protocol also did not work, as there was no shift in the SDS-PAGE gel (Figure 14). This most likely was due to low protein concentration minimizing the propensity of the two to encounter each other in the proper orientation to facilitate a reaction. It could also be due to unfavorable ratios of the two proteins for the ligation reaction to occur. Typically in the literature employs a 5 equivalents excess of one reactant to the other. Therefore, we were unable to demonstrate a successful ligation using the alkoxyamine UAA and future troubleshooting of these assays needs to be conducted to ascertain if the coupling is indeed feasible within a protein context.

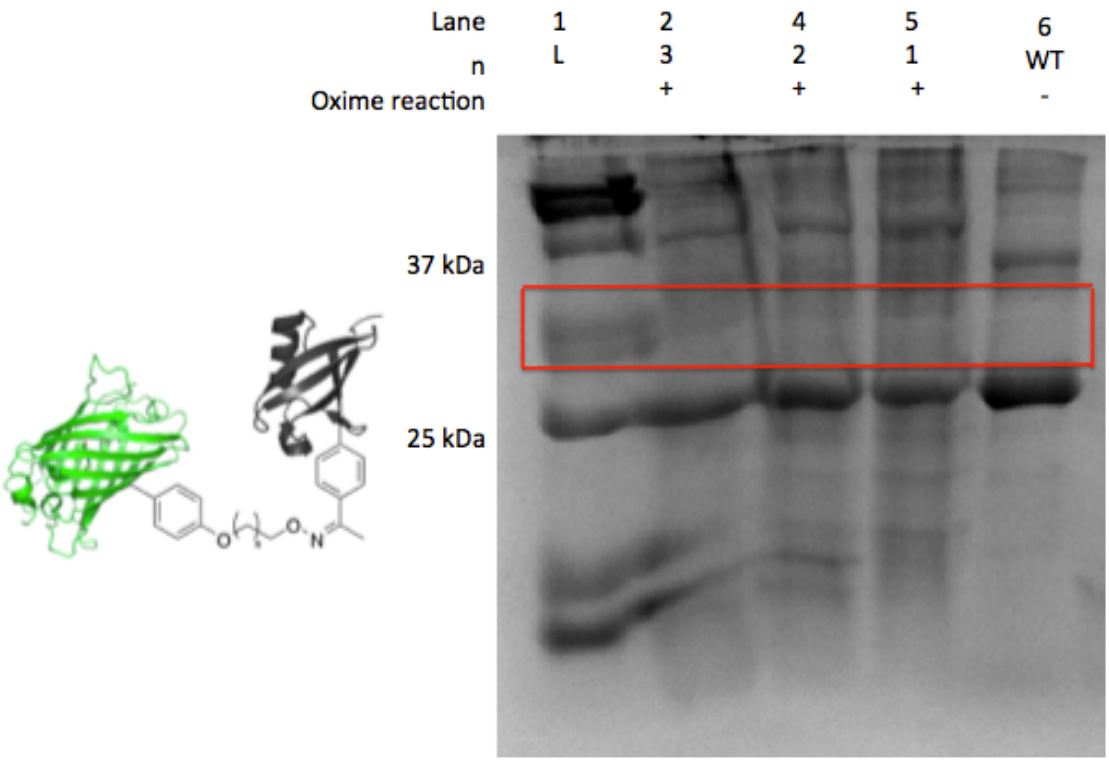


Figure 14-Oxime ligation of (**4, 5, and 6**) mutated GFP-151 and pAcF mutated Ubiquitin-48. The bands at 25 kDa are GFP. If the ligation were successful the band would shift, as it the ligated proteins would be about 33.5 kDa, would be within the red box.

2.III. EXPERIMENTAL

Synthesis of *p*-Bromo-ethyl-Boc-OMe-Tyrosine (1). A solution of Boc-Tyrosine-OMe (0.500 g, 1 eq., 1.69 mmol) was prepared in acetone (7 mL) in a flame-dried vial. To this solution, cesium carbonate (1.655 g, 3 eq., 5.08 mmol) was added, followed by the addition of 1,2-dibromoethane (292 µL, 2 eq., 3.39 mmol). The reaction was stirred overnight at 56 °C, cooled to room temperature, filtered into a round bottom and washed with EtOAc. The filtrate was then concentrated in vacuum and further purified using flash chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate for 10 fractions, 5:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions, 1:1

hexanes/ ethyl acetate for 10 fractions) to afford a clear oil (0.184 g, 0.46 mmol, 27% yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.00 (d, J=8 Hz, 2H), 6.79 (d, J= 8 Hz, 2H), 4.07 (q, J=8 Hz, 1H), 3.41 (t, J= 4 Hz, 2H), 3.36 (t, J=8 Hz, 3H)

Synthesis of *p*-Bromo-propyl-Boc-OMe-Tyrosine (2). A solution of Boc-Tyrosine-OMe (0.500 g, 1 eq., 1.69 mmol) was prepared in acetone (7 mL) in a flame-dried vial. To this solution, cesium carbonate (1.655 g, 3 eq., 5.08 mmol) was added, followed by the addition of 1,3-dibromopropane (345 μL, 2 eq., 3.38 mmol). The reaction was stirred overnight at 56°C. It was then cooled to room temperature, filtered into a roundbottom and washed with ethyl acetate. The filtrate was then concentrated in vacuum and further purified using flash chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate for 10 fractions, 5:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions) to afford a clear oil (0.504 g, 1.20 mmol, 71% yield). ^{1}H NMR (400 MHz, $CDCl_3$): 6.95 (d, J=8 Hz, 2H), 6.73 (d, 8 Hz, 2H), 3.90 (t, J=8 Hz, 2H), 3.60 (s, 3H), 2.16 (m, 2H), 1.93 (s, 9H)

Synthesis of *p*-Bromo-butyl-Boc-OMe-Tyrosine (3). A solution of Boc-Tyrosine-OMe (0.500 g, 1 eq., 1.69 mmol) was prepared in acetone (7 mL) in a flame-dried vial. To this solution, cesium carbonate (1.655 g, 3 eq., 5.08 mmol) was added, followed by the addition of 1,4-dibromobutane (404 μL, 2 eq., 3.39 mmol). The reaction was stirred overnight at 56°C. It was then cooled to room temperature, filtered into a roundbottom and washed with ethyl acetate. The filtrate was then concentrated in vacuum and was further purified using flash chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate for 10 fractions, 5:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions) to afford a clear oil that, once cooled, crashed out as white crystal (0.494 g, 1.148 mmol, 68% yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.99 (d, J=8 Hz, 2H), 6.76 (d, 8 Hz, 2H), 3.66 (s, 3H), 3.43 (t, 8 Hz, 2H), 1.87 (m, 2H)

Synthesis of *p*-Alkoxyamine-ethyl-Tyrosine (4). DCM (5 mL) was used to transfer (**1**) (0.184 g, 1 eq., 0.46 mmol) into a flame-dried vial. To this solution, *N*-Boc-hydroxylamine (0.091 g, 1.5 eq., 0.69 mmol) was added, followed by the addition of 1,8-Diazabicycle-7-Undecene (68 μL, 1 eq., 0.46 mmol). The reaction was stirred for 24 hours at room temperature, then extracted using DCM and brine, 10 mL of brine and three 5 mL washes with DCM. The organic layer was dried magnesium sulfate, filtered, and concentrated in vacuum. The product was further purified using flash chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate for 10 fractions, 5:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions, 1:1 hexanes/ ethyl acetate for 10 fractions) and identified using TLC. This afforded a clear oil which was subsequently deprotected. Dioxane (500 μL) and 1 M Lithium hydroxide (500 μL) were combined and added to a vial containing **4**. The reaction was stirred at room temperature for 2 hours followed by the removal of dioxane via rotatory evaporation. An additional 1 mL of water was added, followed by the dropwise addition of 6 M hydrochloric acid until a pH of 4 was obtained. The product was then extracted with ethyl acetate into a new vial. The organic layer was then dried with magnesium sulfate and then filtered and concentrated in vacuum. Trifluoroacetic acid (2%, 1 mL) was added to

the vial on ice. The vial was then stirred at room temperature for 1 hour, and the DCM was then removed by rotatory evaporation affording a slightly yellow crystal (0.100 g, 0.42 mmol, 24.6% yield). MS (ESI) Expected: 241.11, Actual: 240.11 m/z ($M^+ + H^+$), ^{1}H NMR (400 MHz, CD_3OD): δ 6.05 (s, 3H), 7.00 (d, 8 Hz, 2H), 6.69 (d, 8 Hz, 2H), 1.30 (s, 2H). ^{13}C NMR (400 MHz, CD_3OD): δ 172.34, 128.40, 112.67, 66.28, 34.88

Synthesis of *p*-Alkoxyamine-propyl-Tyrosine (5). DCM (5 mL) was used to transfer (**2**) (0.504 g, 1 eq., 1.20 mmol) into a flame-dried vial. To this solution, *N*-Boc-hydroxylamine (0.240 g, 1.5 eq., 1.80 mmol) was added, followed by the addition of 1,8-Diazabicycle-7-Undecene (180 µL, 1 eq., 1.20 mmol). The reaction was stirred for 24 hours at room temperature, then extracted using DCM and brine, 10 mL of brine and three 5 mL washes with DCM. The organic layer was dried using magnesium sulfate, filtered, and concentrated in vacuum. The product was further purified using flash chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate for 10 fractions, 5:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions, 1:1 hexanes/ ethyl acetate for 10 fractions) and identified using TLC. This afforded a clear oil which was subsequently deprotected. Dioxane (500 µL) and 1 M Lithium hydroxide (500 µL) were combined and added to a vial containing **5**. The reaction was stirred at room temperature for 2 hours, followed by the removal of dioxane via rotatory evaporation. An additional 1 mL of water was added, followed by the dropwise addition of 6 M hydrochloric acid until a pH of 4 was obtained. The product was then extracted with ethyl acetate into a new vial. The organic layer was then dried with magnesium sulfate and then filtered and concentrated in vacuum. Trifluoroacetic acid (2%, 1 mL) was added to the vial on ice. The vial was then stirred at room temperature for 1 hour, and the DCM was then removed by rotatory evaporation affording a clear oil that formed white crystal when cooled. (0.131 g, 0.52 mmol, 30.4% yield). MS (ESI) Expected: 255.13 m/z ($M^+ + H^+$) Actual: 254.13 m/z ($M^+ + H^+$), ^{1}H NMR (400 MHz, CD_3OD): δ 7.00 (d, 8 Hz, 2H), 6.72 (d, 8 Hz, 2H), 3.94 (t, 8Hz, 2H), 1.87 (s, 2H),^{13}C NMR (400 MHz, CD_3OD): δ 172.30, 131.85, 129.40, 114.35, 66.66, 27.60

Synthesis of *p*-Alkoxyamine-butyl-Tyrosine (6). DCM (5 mL) was used to transfer (**3**) (0.494 g, 1 eq., 1.15 mmol) into a flame-dried vial. To this solution, *N*-Boc-hydroxylamine (0.229 g, 1.5 eq., 1.72 mmol) was added, followed by the addition of 1,8-Diazabicycle-7-Undecene (171 µL, 1 eq., 1.15 mmol). The reaction was stirred for 24 hours at room temperature, then extracted using DCM and brine, 10 mL of brine and three 5 mL washes with DCM. The organic layer was dried using magnesium sulfate, filtered, and concentrated in vacuum. The product was further purified using flash chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate for 10 fractions, 5:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions, 1:1 hexanes/ ethyl acetate for 10 fractions) and identified using TLC. This afforded a clear oil which was subsequently deprotected. Dioxane (500 µL) and 1 M Lithium hydroxide (500 µL) were combined and added to a vial containing **5**. The reaction was stirred at room temperature for 2 hours, followed by the removal of dioxane via rotatory evaporation. An additional 1 mL of water was added, followed by the dropwise addition of 6M hydrochloric acid until a pH of 4 was obtained. The product was then extracted with ethyl acetate into a new vial. The organic layer was then dried with magnesium sulfate and

then filtered and concentrated in vacuum. Trifluoroacetic acid (2%, 1 mL) was added to the vial on ice. The vial was then stirred at room temperature for 1 hour, and the DCM was then removed by rotatory evaporation affording a slightly yellow crystal (0.128 g, 0.477 mmol, 28.2% yield). MS (ESI) Expected: 269.14, actual: 268.14 m/z ($M^+ + H^+$), ^{1}H NMR (400 MHz, CD_3OD): 1) δ 7.00 (s, 3H), 6.79 (d, 8 Hz, 2H), 3.85 (m, 2H), 3.64 (m, 2H), ^{13}C NMR (400 MHz, CD_3OD): δ 172.31, 156.11, 128.17, 112.32, 64.90, 27.65

Synthesis of *O*-(prop-2-yn-1-yl)hydroxylamine (7): DCM (5 mL) was added to a flame-dried vial. To this solution, propargyl bromide (0.375 µL, 1.0 eq., 4.20 mmol) was added, followed by the addition of *N*-Boc-hydroxylamine (0.840 g, 1.5 eq., 6.31 mmol) and 1,8-Diazabicycle-7-Undecene (627 µL, 1 eq., 4.20 mmol). The reaction was stirred for 24 hours at room temperature. It was then extracted using DCM and brine, 10 mL of brine and three 5 mL washes with DCM. The organic layer was dried using magnesium sulfate, filtered, and concentrated in vacuum. The product was further purified using flash chromatography (silica gel, gradient of 7:1 hexanes/ethyl acetate for 10 fractions, 5:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions, 1:1 hexanes/ ethyl acetate for 10 fractions) and identified using TLC. The identified fractions were then combined and concentrated down in vacuum, affording a clear oil and was subsequently deprotected. Trifluoroacetic acid (2%, 1 mL) was added to the vial on ice. The vial was then stirred at room temperature for 1 hour, and the DCM was then removed by rotatory evaporation affording a clear oil product (0.207 g, 2.91 mmol, 69.3% yield). ^{1}H NMR (400 MHz, CD_3OD): δ 3.33 (s, 2H), 2.73 (t, 8 Hz, 1H) ^{13}C NMR (400 MHz, CD_3OD): δ 76.27, 73.86, 61.36

General Alexa488-alkoxyamine Reaction, Adapted for alkoxyamine UAA with Prodan[38]: GFP incorporated with **(1, 2, or 3)** was concentrated down using concentrator columns (10k MWCO, Corning Spin-X) and switched to alkoxyamine reaction buffer (50mM sodium acetate, 150 mM NaCl, at pH 4.0). The protein was added to a 1.5 mL Eppendorf ® tube (20 µL). The solution containing prodan was added to the tube (10 µL) and the tube was shaken overnight at room temperature and the conjugation was assessed using an SDS-PAGE gel.

General Alexa488-alkoxyamine Reaction, Adapted for alkoxyamine UAA with Ubiquitin[38]: GFP incorporated with **(1, 2, or 3)** was concentrated down using concentrator columns (10k MWCO, Corning Spin-X) and switched to sterile deionized water. The protein was concentrated down to about 5 µL of solution in a 1.5 mL Eppendorf ® tube under vacuum. The Ubiquitin incorporated with *para*-acetylphenylalanine was concentrated with concentrator columns and the solution was switched to the alkoxyamine reaction buffer (50 mM sodium acetate, 150 mM NaCl, at pH 4.0). The Ubiquitin protein in alkoxyamine reaction buffer was added to the 1.5 mL Eppendorf ® tube with concentrated GFP protein. The tube was then shaken overnight at room temperature and the conjugation was assessed using an SDS-PAGE gel.

General Non-Biological Oxime Ligation Reaction[36]: First, methanol (1 mL) and pyridine (50 µL) were added to a flame dried vial. To the vial the akynone, specifically 4-

nitro-*N*-(1,3-diphenyl-1H-pyrazol-5-yl)benzamide (DPAP), (0.5 equivalents), the alkoxyamine, specifically O-(prop-2-yn-1-yl)hydroxylamine, (1.0 equivalent), and sodium sulfate (1.0 equivalent) are added then stirred for overnight at room temperature. The reaction was run on a TLC. The reaction was diluted with water (5 mL) and then extracted with ethyl-acetate (1 mL, 5x). The organic layer was dried with magnesium sulfate, filtered, and concentrated in vacuum. The product was purified using flash chromatography (silica gel, gradient of 3:1 hexanes/ethyl acetate for 10 fractions, 3:1 hexanes/ethyl acetate for 10 fractions, 1:3 hexanes/ethyl acetate for 10 fractions), and identified using TLC. The identified fractions were combined and concentrated down in vacuum. To determine if the reaction using DPAP and *O*-(prop-2-yn-1-yl)hydroxylamine was successful ^{1}H NMR spectroscopy was used to determine if **8** was synthesized. ^{1}H NMR (400 MHz, CD_3OD): δ 8.04 (d, 4 Hz, 2H), 8.02 (d, 4 Hz, 2H), 7.61 (d, 4 Hz, 2H), 7.29 (m, 8H), 3.26 (s, 2H), 2.47 (t, 4 Hz, 1H)

General Non-Biological Oxime Ligation Reaction, Adapted for Biological Macromolecules[36]: To a 1.5 mL Eppendorf ® tube, protein with an incorporated alkoxyamine UAA, specifically GFP with **1**, **2**, and **3**, (1.0 equivalent), Sodium sulfate (1.0 equivalent), and an alkynone, specifically prodan, (0.5 equivalent) are added. To the solution 2 μL of methanol and 2 μL of pyridine were added and shook at room temperature overnight. To determine if the reaction was a success, the reaction was then concentrated using concentrator spin columns (10k MWCO, Corning Spin-X) and switched to PBS buffer, then run on a SDS-PAGE gel. The gel was then imaged for fluorescence and then coomassie stain.

General GFP Expression: A pET-GFP-TAG variant plasmid (0.5 μL) was cotransformed with a pEVOL-ambrx plasmid (0.5 μL) into *Escherichia coli* BL21(DE3) cells using an Eppendorf Eporator™ electroporator. The cells were plated onto Luria-Bertani (LB) agar plates with chloramphenicol (34 mg/mL) and Ampicillin (100 mg/mL) resistance at 37°C overnight. Single colonies that had grown on the plate were then inoculated into LB media (2-4 mL) containing both ampicillin and chloramphenicol resistance. This culture was grown to turbidity overnight at 37°C and subsequently used to inoculate expression cultures (20 mL of LB media, 34 mg/mL of chloramphenicol, and 50 mg/mL of ampicillin resistance) to an OD_{600} of 0.1. The cultures are then incubated at 37°C to grow to an OD_{600} between 0.6-0.8 and expressions were then centrifuged at 5,000 rpm for 10 minutes and the supernatant was discarded. The pelleted cells were then resuspended in 4 mL of LB media (with 34 mg/mL of chloramphenicol and 100 mg/mL of ampicillin resistance). The resuspended cells were subsequently induced with the addition of **1**, **2**, or **3** (40 μL, 100 mM), 20% arabinose (4 μL) and 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; 4 μL). The cultures were then shook for 16-20 hours at 30°C. The cultures were then centrifuged at 5,000 rpm for 10 minutes, the supernatant was discarded, and the cell pellet was stored at -80°C for at least 20 minutes. The 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, the supernatant was then collected and transferred into an equilibrated His-pur Ni-NTA spin column (Qiagen) with nickel resin

(200 μL) and GFP was purified according to manufacturer's protocol. Purified GFP was analyzed by SDS-PAGE (BioRad 10% gels, 150 V, 1.5 hours), and employed with further purification depending on the solution required for specific reaction conditions. The concentrations were determined using fluorescence measurements.

General Ubiquitin Expression: A pET-ubiqutin-TAG variant plasmid (0.5 μL) was cotransformed with a pEVOL-ambrx plasmid (0.5 μL) into *Escherichia coli* BL21(DE3) cells using an Eppendorf Eporator electroporator. The cells were plated onto Luria-Bertani (LB) agar plates with chloramphenicol (34 mg/mL) and Ampicillin (100 mg/mL) resistance at 37°C overnight. Single colonies that had grown on the plate were inoculated into LB media (2-4 mL) containing both ampicillin and chloramphenicol resistance. The culture was grown to turbidity overnight at 37°C and subsequently used to inoculate expression cultures (20 mL of LB media, 34 mg/mL of chloramphenicol, and 50 mg/mL of ampicillin resistance) to an OD_{600} of 0.1. The cultures are then incubated at 37°C to grow to an OD_{600} between 0.6-0.8 and expressions were then centrifuged at 5,000 rpm for 10 minutes and the supernatant was discarded. The pelleted cells were resuspended in 4 mL of LB media (34 mg/mL of chloramphenicol and 100 mg/mL of ampicillin resistance). The resuspended cells were subsequently induced through with the addition of pAcF (40 μL, 100 mM), 20% arabinose (4 μL) and 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; 4 μL). The cultures shook for 16-20 hours at 30°C. The cultures were then centrifuged at 5,000 rpm for 10 minutes, the supernatant was discarded, and the cell pellet was stored at -80°C for at least 20 minutes. The 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, the supernatant was then collected and transferred into an equilibrated His-pur Ni-NTA spin column (Qiagen) with nickel resin (200 μL) and GFP was purified according to manufacturer's protocol. Purified GFP was analyzed by SDS-PAGE (BioRad 10% gels, 150 V, 1.5 hours), and employed with further purification depending on the solution required for a specific reaction.

2.IV. FUTURE DIRECTIONS

The oxime ligation is a useful tool for bioconjugation chemistry due to its hydrolytic stability and relatively physiological reaction conditions. Future work is needed to optimize reaction conditions for protein conjugation. This could include manipulation of protein concentrations or the addition of strong nucleophilic catalysts such as aniline.[17] Once optimized, these bioconjugations can be employed in more biologically relevant proteins toward the development of new imaging or chemotherapeutic reagents."

Currently, there are other UAAs that contain ketone functionalities. These carbonyls react well with alkoxyamines, but aldehydes are more efficient coupling partners for oxime ligations due to their increased reactivity. However, due to this

increased reactivity, aaRS evolution experiments have not found a significant degree of success. In order to more efficiently incorporate this moiety and prevent interaction with cellular components, work will be done to synthesize and incorporate a photocaged aldehyde UAA. Ideally, upon excitation by a specific wavelength of light, the caging group be removed to expose the aldehyde functionality. This amino acid will then be able to undergo the oxime reaction (without substantial undesired side reactions) and will represent an ideal reaction partner for the already developed alkoxyamine UAA.

2.V. CONCLUSION

I was able to synthesize three novel alkoxyamine functionalized tyrosine derivatives. These unnatural amino acids have different alkyl chain lengths to optimize alkyl tether length for bioconjugation reactions. Each of the UAAs were also successfully incorporated into green fluorescent protein using a polyspecific aaRS. The incorporation of this alkoxyamine functionality allows for new bioorthogonal chemistry reactions that can be used for bioconjugation or fluorescent imaging with fluorescent molecules. With the synthesis of the alkoxyamine UAAs and their incorporation into protein already accomplished, only the optimization of reaction conditions remains as a hurdle to utilizing these novel amino acids in a biological setting.

CHAPTER 3: SYNTHESIS OF PROFLUOROPHORE 2-ETHYNYLBENZO[d]THIAZOLE AND BIOCONJUGATION IN PROTEIN

3.I. INTRODUCTION

Fluorescent dyes are a class of molecules that are of extremely useful in the labeling of biological structures and macromolecules. This biological labeling provides a means to study the synthesis and movement of certain biological molecules like protein, nucleic acids, carbohydrates or whole cells.[21] Fluorescent dyes can also be used as a mechanism of studying various biological processes. For instance, fluorescent probes have been used to study glycosylation in vivo or specific types of cell using labels of cell surface glycoproteins.[39] Shieh et al. have also used these probes to label newly synthesized DNA and RNA to measure cell proliferation.[40] Unfortunately, the fact that these molecules are already fluorescent, and are difficult to remove from the system of study once introduced, leads to a high amount of background fluorescence when applied to a biological system. This background fluorescence of unreacted dyes can sometimes significantly complicate results and invalidate experiments. To remedy this, many fluorescent probes produce a shift in their emission wavelength once reacted or bound to a biomolecule, decreasing undesired signal. Moreover, even though some fluorophores experience a shift in the emission wavelength once bound, the use of fluorescent molecules that attach to common chemical functionalities found in biological system can also facilitate unwanted background labeling.[41] These complications result in over-labeling and mislabeling of biological molecules and additional research is needed to address these issues.

One solution to these complications with biological imaging involves the development of small molecules that are non-fluorescent due to quenching effects of attached moieties in their unreacted state, but upon reaction become fluorescent.[39] These functionalities can be either electron donating or electron withdrawing and prevent the electronic transitions necessary for fluorescence.[20] Therefore, in the unbound/unreacted state these probes are non-fluorescent. To restore the fluorogenic nature of these molecules they must react with another functionality that results in delocalization of electrons allowing the molecules to undergo the electronic transitions and fluoresce. This allows for application of these probes without the concern of washing the system of unbound probes, as they will not afford background fluorescence.

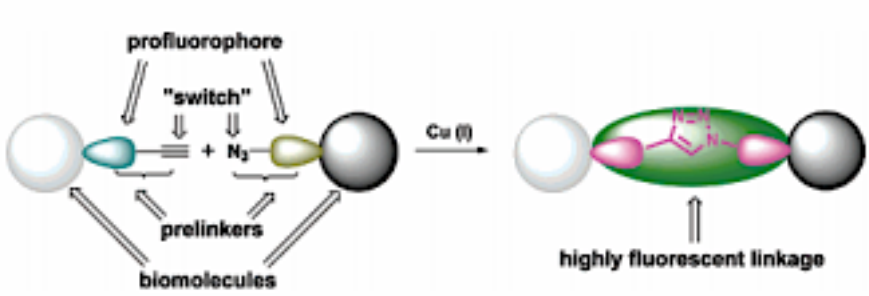


Figure 15- Diagram of profluorophore linkage of an alkyne and azide functionality from *Org. Lett.* **2004,** 6 (24), 4603-6.[20]

The fact that the reactions can be directed via the use of non-endogenous chemical functionalities, labeling of specific biomolecules can have a greater degree of specificity and efficiency.[40] For instance, alkyne and azide functionalities undergo a Cu(I)-1,3-cycloaddition reaction that forms a triazole ring structure, effectively delocalizing their electrons back into conjugated systems (Figure 15).[41] These two functionalities are not naturally found in biological systems and are thus considered orthogonal. This is ideal as they can be incorporated into biological systems using mutated amino acids, nucleic acids, and carbohydrates then can be site specifically labeled with a dye possessing a complementary functionality to that in the biomolecule. Other then the reaction potentially requiring catalytic amounts of copper that can be cytotoxic to the cellular environment, the reaction can readily occur in physiological conditions. Thus, the incorporation of unnatural amino acids that contain these complementary functionalities to profluorogenic molecules allow for a more efficient means of tracking proteins within a biological system.

3.II. RESULTS AND DISCUSSION

2-ethynyl[d]benzothiazole is an example of an alkyne functionalized profluorophore molecule that can undergo a Cu(I)-1,3-dipolar cycloaddtion with an azide functionality. The alkyne functionality is electron withdrawing and diminishes the fluorescence of the conjugated ring structure. Upon reaction with an electron donating azide group the electrons are delocalized and a fluorescent molecule is generated. Qi et al. synthesized this molecule and demonstrated its ability to undergo this reaction with a

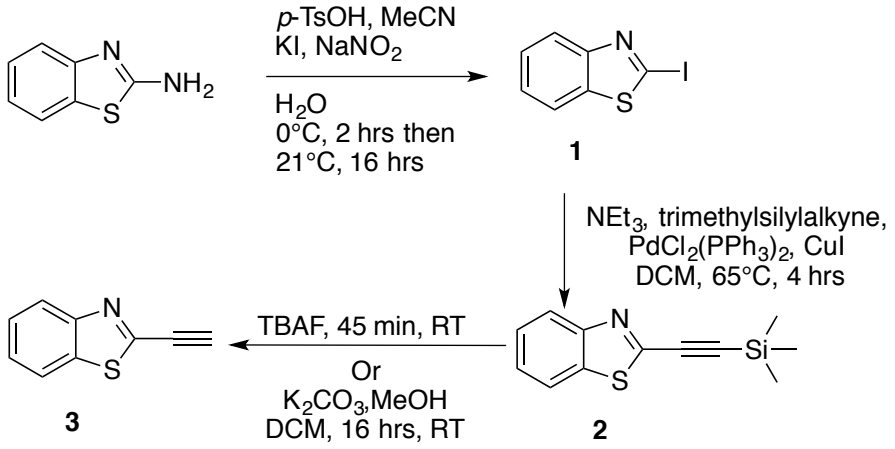
single azido functionalized amino acid in solution, not within the structure of a protein, demonstrating the profluorphore's utility in binding with amino acid derivatives.

However, the reaction in a protein context creates a greater challenge for profluorophores. Proteins are composed of many amino acid residues that interact with each other to form secondary, tertiary, and quaternary structures. These structures are vital for the proper function of the protein in the cell, but add a level of complexity to bioconjugation reactions. So the work by Qi et al. is important in demonstrating the fluorogenic ability of the profluorophore molecule but it does not explore the ability of this molecule to react and fluoresce in a biological context. In this work we explore how the ability of this profluorophore to react with an incorporated azide UAA in the context of a biological macromolecule, specifically a green fluorescent protein.

3.II.A. SYNTHESIS OF 2-ETHYNYLBENZO[d]THIOZOLE

Using the synthetic method developed by Qi et al. the synthesis of 2-ethynylbenzothiozole was performed (Scheme 9).[41]

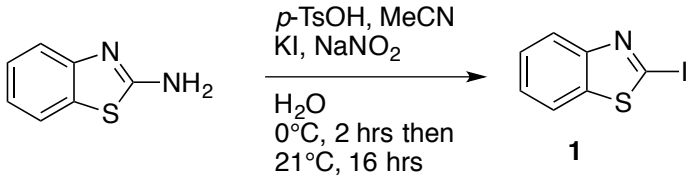
Scheme 9.



The first step of the synthesis begins with commercially available 2-aminobenzo[d]thiazole. This compound is subjected to a diazotization-iodination reaction in solution with *p*-toluenesulfonic acid in acetonitrile (Scheme 10). The first step of the reaction is stirred at 0°C for 2 hours, and yields the diaozotization intermediate that is now susceptible to substitution by halides, specifically iodine. This is accomplished by

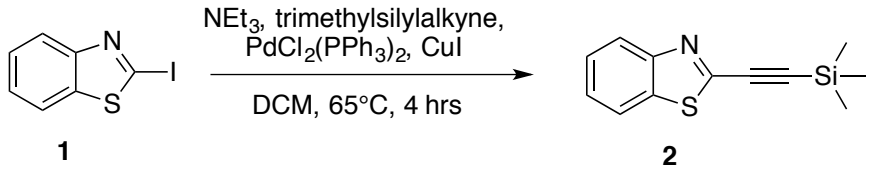
the reaction with KI, yielding a brown solid after filtration and purification. This step of the overall reaction was worked well and resulted in good product yields of 2-iodobenzo[d]thiozole (**1**) (1.40 mmol, 69.9% yield).

Scheme 10.



The 2-iodobenzothiozole product (**1**) is then subjected to a Sonogashira coupling reaction. This cross-coupling reaction uses catalytic amounts of copper iodide and a palladium catalyst ($Pd(PPh_3)_2Cl_2$) to functionalize the benzothiozole with an trimethylsilyl protected alkyne (Scheme 11). The product (**2)** is then extracted and purified using silica gel chromatography. Due, to the non-polar nature of the molecule and its similarity in polarity of the starting material it was challenging to isolate the product using silica flash chromatography. Many of fractions that appeared to have product also contained significant impurities from starting material and would be subjected to a second column, resulting in significantly diminished yields.

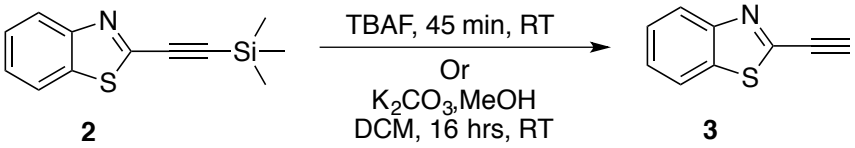
Scheme 11.



In addition, the literature indicates that the protecting group is removed during the silica gel chromatography purification, but this was not observed during my synthesis of the compound (Scheme 12). Consequently, tetra-*n*-butylammonium fluoride (TBAF) was

employed to remove the trimethylsilyl group. Unfortunately, the product was too challenging to isolate from excess TBAF contamination and other routes were examined for deprotection. The next method for removal of trimethylsilyl groups investigated involved using potassium carbonate and methanol as reagents. Gratifyingly, this method of deprotection seemed to be effective as assessed by NMR spectroscopic analysis. After the difficulty separating the product from the starting material and troubleshooting the deprotection of the molecule, the yield of (**3)** was very low (0.06 mmol, 5.46% yield) and actual application of the profluorphore was difficult to assess. It is hypothesized that the yield was very low because the protected molecule (**2**) and the deprotected product (**3)** have relatively similar polarities they were hard to separate. They both require very non-polar solvent systems to enter the mobile phase and it was hard to find a solution that would separately elute the two products. Therefore, products would be impure or would have reduced yield due to multiple column chromatography purifications in an attempt to isolate (**3**).

Scheme 12.



2 **3**

3.II.B. Cu(I)-1,3-CYCLOADDITION REACTIONS

With small amounts of 2-ethynylbenzothiazole obtained in the previously described synthesis, the product was diluted into a 1 mM solution that was then employed in a Cu(I)-1,3-cycloaddition reaction. The compound was reacted with a GFP that has been mutated to contain an azide derivatized tryrosine using the Shultz methodology previously described. The profluorophore was also reacted with a benzyl azide molecule to compare both the biological and chemical reaction. Both reactions were performed with their own set of controls to afford an effective assessment of the reaction and the generated fluorophore.. The reactions followed the general reaction protocol for Cu(I)-

1,3-cycloadditions and were reacted overnight at 4°C. The reaction solution of the protein was purified and then all of the reactions were diluted to the same concentrations using PBS, and their fluorescence was measured. Qi et al. reported excitation of the reacted groups at 285nm and emission at 363nm, so we excited at the same wavelength and measured the emission from 300nm-600nm. However, we observed no changes in fluorescence relative to the control. After, multiple attempts at rescanning and re-diluting there was still no change in the fluorescence (Figure 16).

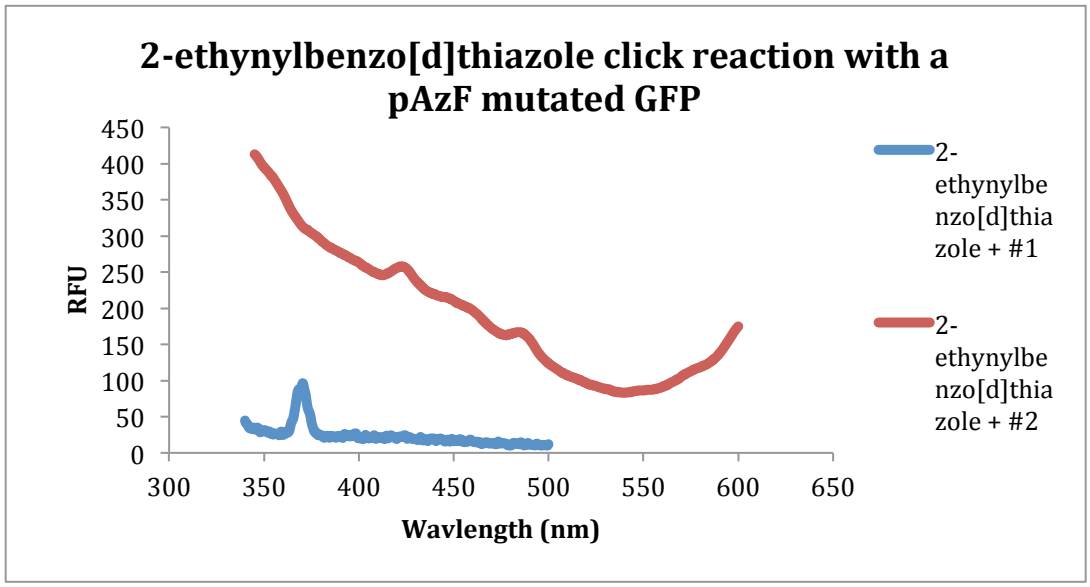


Figure 16. Fluorescence spectrum for a Cu(I)-1,3-dipolar cycloaddition with pAzF mutated GFP-151 and 2-ethynylbenzo[d]thizole (**3**).

3.III. EXPERIMENTAL

Synthesis of 2-Iodobenzo[d]thizole (1)[41]- To a flame dried vial was added 2-aminobenzo[d]thiazole (0.300 g, 1 eq., 2.00 mmol) and *para*-toluenesulfonic acid (*p*-TsOH) (1.14 g, 3 eq., 5.99 mmol) in 8.0 mL of acetonitrile. In a separate vial potassium iodide (0.862 g, 2.6 eq., 5.19 mmol), sodium nitrite (0.276 g, 2 eq., 3.99 mmol), and 1.19 mL of water were combined. The two solutions were then mixed into a third flame-dried vial with a stir bar at 0°C, and the reaction was stirred for 2 hours on ice. The reaction was then warmed to room temperature and stirred overnight. At this point the reaction was quenched using water (9 mL), and the pH of the solution was adjusted to a pH of 8-9 using sodium bicarbonate. Sodium thiosulfate (0.010 g, 0.06 mmol) was added and the solution was filtered and collected via büchner filtration. The precipitate was washed thoroughly with deionized water to yield a red-brown solid. (0.365 g, 1.40mmol, 69.9% yield) ^{1}H NMR (400 MHz, CD_3OD): δ 8.04 (1 H, dd, J1=8Hz, J2=1 Hz), 7.85 (1H, dd, J1=8Hz, J2=1Hz), 7.42 (2H, m)

Synthesis of 2-ethynylbenzo[d]thiazole (3)[41]- DCM (2.5 mL) was added to a flame-dried vial, followed by, 2-iodobenzothiazole (0.300 g, 1eq., 1.15mmol), triethylamine (1.12 mL, 7 eq., 8.04 mmol), trimethylsilylalkyne (327 µL, 2 eq., 2.30 mmol), $PdCl_2(PPh_3)_2$ (0.016 g, 0.020eq, 0.023 mmol), and Copper(I)Iodide (0.005 g, 0.0221eq., 0.0254 mmol). The reaction was then stirred under argon at 65°C for 4 hours. The reaction was diluted with EtOAc and then filtered into a round-bottom flask. The solution was then concentrated under vacuum to afford a dark brown oil. The oil was then purified by flash chromatography (silica gel, 100% hexanes for 10 fractions, 10:1 hexanes/ethyl acetate for 10 fractions, 7:1 hexanes/ethyl acetate for 10 fractions, and 5:1 hexanes/ethyl acetate for 10 fractions) the product was isolated as a dark brown solid. To remove the trimethylsilyl protecting group two pathways were investigated. In the first reaction attempted, the protected compound was treated with tetra-n-butylammonium fluoride (TBAF) (1 mL per 0.1 g of product) and stirred for 45 minutes, followed by extraction using DCM and brine in a separatory funnel. The organic layer was collected, dried using magnesium sulfate, filtered and then concentrated in vacuum. Unfortunately, this led to significant contamination of the product by TBAF so another deprotection method was implemented. This method used the protected alkyne benzothiazole (0.023 g, 1 eq., 0.1 mmol) and potassium carbonate (0.152 g, 4.5 eq., 1.10 mmol) in a (1 mL)1:1 DCM and methanol solution, that was stirred overnight at room temperature. The product purified by flash chromatography (silica gel, 100% hexanes for 10 fractions, 10:1 hexanes/ethyl acetate, 7:1 hexanes/ethyl acetate for 10 fractions) the product was a brown solid (0.010g, 0.06mmol, 5.46%) ^{1}H NMR (400 MHz, CD_3OD): δ 8.00 (1H, d, J=8 Hz), 7.86 (1H, d, J=8 Hz), 7.41 (2H, m), 3.67 (1H, s)

Biological Cu(I)-1,3-dipolar Cycloaddition Reaction: Click reactions were performed on GFP protein containing *para*-azidophenylalanine (pAzF) in reside 151 or benzyl azide in a 1.5 mL Eppendorf® tube. To the reaction tube, the following were added in the described order: $CuSO_4$ (2 µL, 50 mM), Tris[(1-benzyl-1-H-1,2,3-triazol-4-yl)methyl]amine (TBTA) (20 µL, 5 mM), 20µL of pAzF containing GFP in phosphate-buffered saline (PBS), 2-ethynylbenzo[d]thiazole (10 µL, 1 mM), which contains the alkyne functionality, and tris(2-carboxyethyl)phosphine (TCEP) (2 µl, 50 mM). Control reactions were performed with 20 µL of the pAzf mutated GFP in PBS or benzyl azide and of 2-ethynylbenzo[d]thiazole (10 µL, 1mM) in the absence the Cu-1,3-dipolar cycloaddition catalysts ($CuSO_4$, TBTA, and TCEP) to ensure that successful click reactions were not a result of non-specific interactions between the alkyne and azide functionalities. The reaction solution was then diluted with PBS to the same volume as the actual reactions. All of the reactions were incubated for 10 hours at 4°C. All of the reactions involving GFP were transferred to concentrator columns (10k MWCO, Corning Spin-X) pre-wet with PBS and spun at maximum speed in a tabletop centrifuge for 2 minutes. This was repeated with 7 washes (PBS, 500 µL) and the final volume was then adjusted with extra spins to bring samples to ~25 µL. The samples were then measured for fluorescence using a Perkin Elmer LS 55 luminescence spectrophotometer and compared with unreacted starting material. The samples were diluted in PBS then excited at 285 nm and emission was measured at 300-600 nm, according to literature.[41]

3.IV. FUTURE DIRECTIONS

In order to perform more effective Cu(I)-1,3-cycloaddition reactions with 2-ethynylbenzo[d]thiazole, methods must be optimized for isolation of the compound and for the removal of the alkyne protecting group. For compound isolation, solvent systems that are used to separate the products from the starting materials need to be investigated. Ideally, further fine-tuning to create a better non-polar gradient of solvents may allow for a more efficient separation. Another approach that can be implemented while trying to improve upon column separation is scaling up the reaction conditions. This would produce a greater amount of the product, and even if there was a need for further separation, a much larger quantity would facilitate conjugation studies. The contamination by the reagents and reduction in yield remove the protecting group was detrimental to the overall synthesis of this molecule. Thus, improving upon these methods or implementation of novel methods would increase the usable yield of this synthesis.

Once the small molecule synthesis has been optimized, more efficient, Cu (1)-1,3-cycloaddion reactions can be performed. In addition to measuring the fluorescence shift in the reacted product the reaction can also be assessed using SDS-PAGE gels, to determine if the reaction was a success. Increased concentrations of not only the profluorophore, but also the azide containing protein would aid in actually facilitating the analysis of the reaction. For the chemical molecule control using benzyl azide, increasing the reaction volume to then have the ability to isolate the product for NMR analysis would also aid in troubleshooting the reaction.

Ultimately, other conjugations can also be performed using this functionalized profluorophore, like Glaser-Hay reactions, which would also ideally yield fluorescent probes. Moreover, reaction with biologically relevant proteins will further demonstrate the utility of the reaction. With further improvement of the synthesis of this molecule and the optimization of reaction conditions will provide a better understanding of the effects of this fluorescent probe on biological macromolecules.

3.V. CONCLUSION

I was able to synthesize small amounts of an alkyne-functionalized benzothiozole. The reduction in the yield occurred during the final steps of the synthesis route due to an inability to isolate the product. Unfortunately, click reactions with the isolated molecule were unsuccessful as there was no significant change in the fluorescence spectrum in the reaction with the protein or with a small molecule, benzyl azide. Future work should be performed to improve the reaction yields of this profluorophore and more labeling experiments should be performed using Cu(I)-1,3-cycloadditions as well as other reactions like the Glaser-Hay coupling. Further characterization of this molecule in the context of a biological macromolecule will help elucidate this profluorophore's utility within a biological system.

CHAPTER 4: MUTAGENESIS OF Cas9 FOR INCORPORATION OF UNNATURAL AMINO ACIDS

4.I. INTRODUCTION

Genome engineering is a field of growing importance in biology and biochemistry. One of the most recent and most impactful, developments in genomic engineering is the discovery, utilization and modification of the CRISPR Cas9 system. The CRISPR Cas9 system is a mechanism of defense found bacteria and archaea against viral infection, conferring bacterial immunity.[42] Ever since it was discovered in 2002, researchers have been working with the system to exploit it as a tool for precise genomic modification. From the work that has been done CRISPR Cas9 technology has evolved for use in site-specific homologous recombination, DNA cleavage, transcriptional activation, and transcriptional repression.[42] New modifications are constantly being performed to this biological tool to even further increase its utility.

There are three types of CRISPR Cas9 systems that have been discovered in bacteria and archaea.[42] Each type has a similar mechanism for viral DNA cleavage (Figure 17). The bacteria or archaea has a genomic CRISPR locus that transcribes a single transcript containing a series of repeat sequences with variable sequences that correspond to foreign genetic material. This is a form microbial immune memory where the bacteria will incorporate sequence, a variable region, from a viral DNA into the CRISPR loci and can then be used in targeting. Following transcription of this CRISPR RNA, the RNA undergoes a processing pathway that cleaves the transcript into a functional trans-activating RNA (tracrRNA) and 20-nucleotide guide sequence, which constitute a CRISPR RNA (crRNA).[43] This RNA forms a complex with the Cas9 endonuclease and allows for complementary binding to the DNA of interest. In addition, in order for the Cas9/crRNA complex to complementarily bind to the DNA it must also recognize a species-specific 2-4 base pair protospacer adjacent motif (PAM) that is next to gene of interest in the target DNA. The dual lobed endonuclease then changes conformation once bound to DNA, to bring the two nuclease domains, HNH and RuvC, to the DNA.[44] These domains cleave both strands of the viral DNA rendering it non-functional. With the complementary base pairing and PAM recognition, this system is very site specific. A study by Bikard et al. found that the Cas9/crRNA complex binding is

affected by a few as 2 mismatches in the guide RNA sequence and the target DNA, significant reductions in effectiveness were witnessed in Cas9.[45] This site specificity and effective DNA cleavage make this a effective gene targeting and editing tool.

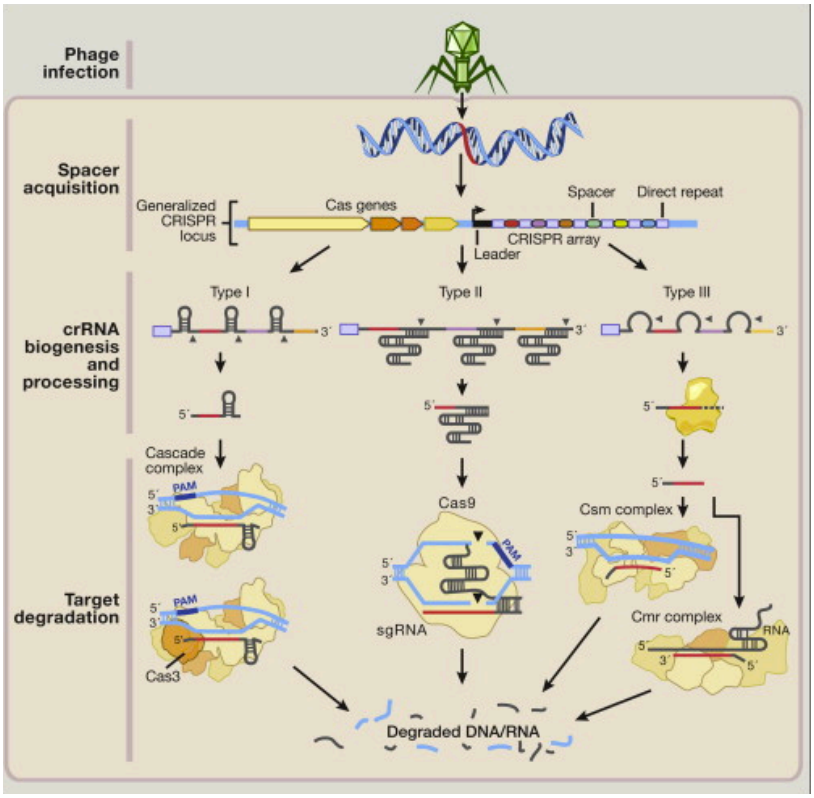


Figure 17. The overall mechanisms of the three types of CRISPR Cas9 systems, from *Cell* **2014,** 159 (3), 647-61.[42]

To take advantage of this system, Cas9 proteins and engineered single-guide RNAs (sgRNAs), which act as processed tracrRNAs and crRNAs, are incorporated into various organisms. A commonly used Cas9 endonuclease is a type I endonuclease from *S. pyogenes*.[42] These sgRNA guided Cas9's are used to induce double strand breaks in genomic DNA and subsequent homologous recombination with the addition of short double stranded DNA sequences with homologous end regions. Moreover, they have been employed to generate controlled DNA damage by inducing double strand breaks that are repaired using non-homologous end joining. Other groups like the Doudna lab have worked to expand CRISPR Cas9's genetic engineering utility through mutations of the two catalytic domains of the Cas9 protein to inhibit its DNA cleavage properties.[43] These mutations in the RuvC and HNH domains are D10A and H840A, respectively.

This mutant Cas9 protein is known as "dead" Cas9 or dCas9. It has been used to site specifically bind to double stranded DNA of interest and provides a steric inhibition of transcription (figure 18). Further modifications to the dCas9 protein have been made by groups like Gilbert et al. to combine transcriptional activator proteins and repressor proteins to confer greater site-specific control of expression or repression of genes.[46] Other modifications have also been made to bind green fluorescent protein to Cas9 as a mechanism to monitor its function in the cell, and mutations to nuclease domains have been done individually to create a Cas9 protein that will induce a single strand break in the DNA, called nicking.[42]

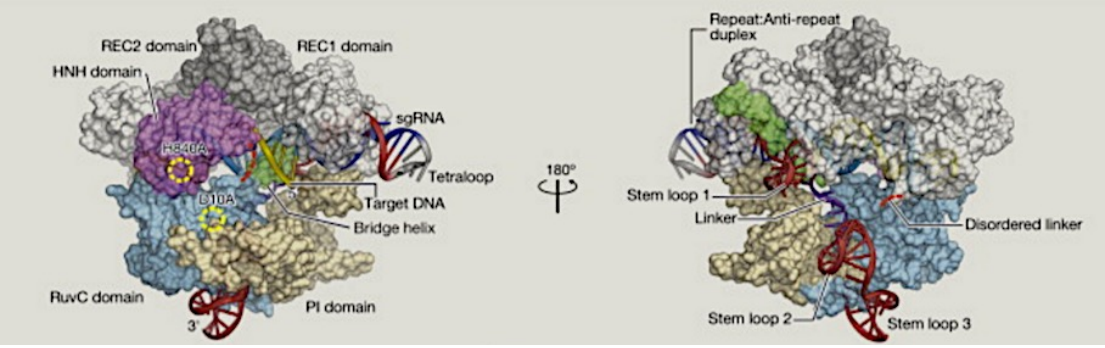


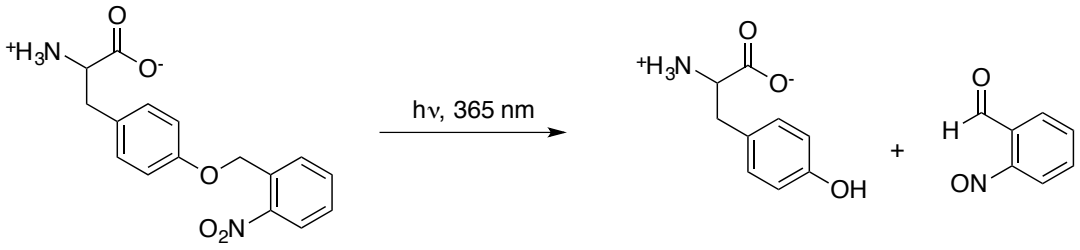
Figure 18. Cas9 protein and target DNA complex, from *Cell* **2014,** 159 (3), 647-61.[42]

In combination, the site specificity of the CRISPR Cas9 system and their broadened applications make this technology a groundbreaking advancement in the field of genetic engineering as a whole. With our work we intend to add to the utility of this technology through the creating of a novel posttranslational inducible control of the activity of the Cas9 protein in its DNA binding and cleavage processes. Work has previously been conducted in this area, introducing using small-molecule inducers and light induction to the Cas9's nuclease function.[42]

Though small-molecule induction of Cas9 here is a method of gene editing function, systemic control of the protein's function is not specific to a certain region of the body. Therefore, gene editing is not specific and occurs throughout the organism. With light induced control of the protein there is an additional level of spatial and temporal control. There exist unnatural amino acids that have protecting groups that block canonical amino acid residues, inhibiting their function, that can be cleaved by

irradiation at a defined wavelength. These amino acids are referred known as photocaged amino acids. Using photocaged UAAs we intend to create a Cas9 protein that can be switched on with irradiation of the residue with a specific wavelength of light. This will allow the researcher posttranslational control of genetic modifications. Moreover, it facilitates the photo-switching control of transcriptional activation or repression, DNA editing, or DNA damage. Adding an additional level of control to the various applications of the CRISPR Cas9 system has numerous advantages, which we hope to exploit.

Scheme 13.



Previous work by Deiters et al. and Arbely et al. have shown that a photocaged tyrosine derivative known as *o*-nitrobenzyl-O-tyrosine (ONBY) can be incorporated and used effectively in both bacterial and mammalian systems (Scheme 13).[47] Our lab has been working with this UAA in other experiments and are currently working on optimization of incorporation of these photocaged derivatives of tyrosine into protein. This work has been ongoing to optimize expression with this UAA using different polyspecific synthetases in order to successfully incorporate the UAA into the Cas9 protein to finally achieve the photo-controlled Cas9 protein.

4.II. RESULTS AND DISCUSSION

Many modifications have been engineered into the site-specific endonuclease Cas9 to optimize its utility in genetic engineering as a whole. One mutated version of Cas9 in particular is of great interest is from the work of Dr. Jennifer Doudna. Her lab created a version of Cas9 (dCas9) whose nuclease domains are both catalytically inactivated by amino acid substitution, D10A and H840A. This mutant Cas9 is therefore useful, in combination with gene regulatory proteins, in binding to specific sites on DNA

and either activating or repressing transcription. We hope to expand upon this work, as we are interested in substituting UAAs into these mutation sites. We hypothesize that we can incorporate a photocaged UAA that will act in a similar manner to the amino acid substitutions of the dCas9 mutant. However, when irradiated with a specific wavelength of light the protecting group will be removed, exposing an amino acid residue that will restore the Cas9 nuclease's catalytic functionality.

In addition to synthesizing a dCas9 that can be photo-induced into a catalase active Cas9, we are also investigating another amino acid site due to information we encountered in the literature.[43] Jiang et al. determined that there is a tyrosine residue (Y450) that plays an influential role in the binding of the target DNA with the crRNA and Cas9. The tyrosine residue induces a kink in the crRNA between the 15th and 16th base pair that causes the bases 11-15 to have a tilted orientation that allows for interaction of bases 16-20 of the crRNA with the target DNA. Once the target DNA strand is bound the tryrosine then shifts 120° into a non-interactive confirmation (Figure 19). To take advantage of this influential amino acid residue, we also have targeted this for photocaged UAA incorporation. This would give us light inducible control of any Cas9 mutant's binding to a target DNA region. Moreover, we anticipate a higher potential for success, as the de-caging event will unmask a tyrosine residue that is present in the wild-type protein.

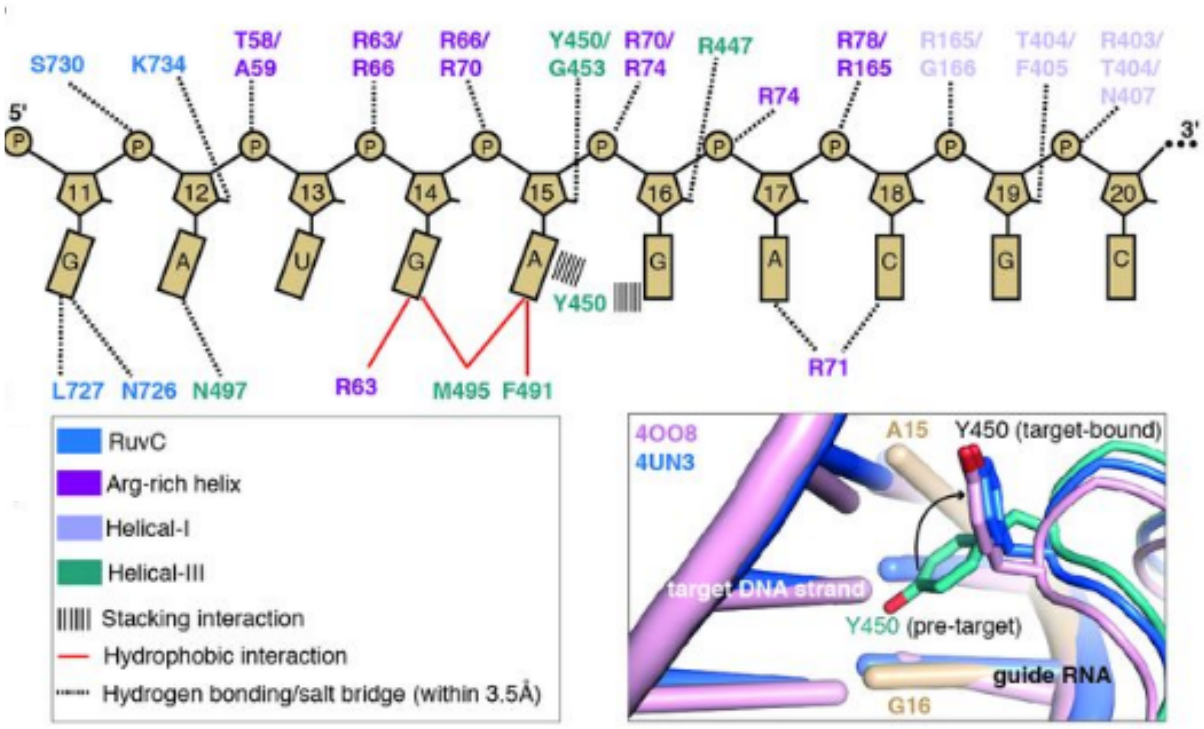


Figure 19. Amino acid Y450 interaction with crRNA and conformational transition between target bound and unbound state, from *Am Chem Soc* **2012,** 134 (29), 11912-5.[43]

UAA incorporation into these sites of the Cas9 protein will allow for light inducible control of both DNA binding interaction as well as the catalytic activity of the nuclease domains. This can afford the spatial control of the Cas9 mutants and lead to increased utility of this genetic modifying technology.

4.II.A SITE DIRECTED MUTAGENESIS OF Cas9 PLASMID

Using a Cas9 plasmid with a hexa-His tag (pET-Cas9-6XHis) from the Liu lab at MIT we designed primers using the QuickChange® primer design program. We specifically designed these sets of primers to introduce TAG mutations at amino acid codons on the parent plasmid for residues 10, 450, and 840. Using a high-fidelity polymerase specific thermocycling protocol, we conducted a mutagenic PCR based on the appropriate melting temperatures of each primer set. This process repeats 25-35 times, yielding low concentrations of mutated plasmid. The mutated plasmids at this stage are contaminated with original template plasmids that do not harbor the mutation.

In order to transform only the desired plasmids the template plasmid needed to be specifically degraded, this was accomplished using an endonuclease that recognizes and cleaves only methylated DNA (template strands). The mutated plasmids were then purified and buffer exchanged using a traditional PCR purification kit modified with a deionized water elution step to minimize ionic concentrations for electroporation. Following transformation, cells that were able to grow on antibiotic resistant LB plates were then inoculated into 2-4 mL of antibiotic containing LB media. The bacteria were then grown overnight at 37°C and the turbid cultures were then mini-prepped to obtain higher concentrations of mutated plasmids.

We only performed a site-directed mutagenesis on the amino acid residue 450 codon. Initially, we had complications getting the site-directed mutagenesis reactions to work due to the large size of the Cas9 plasmid, as it is 9,265 base pairs long. The large size of this plasmid required more specialized polymerases to minimize the introduction of undesired mutations during the lengthy polymerization. Additionally, the PCR protocol also required a degree of optimization in the concentration of template that required, the duration of each amplification step, and the temperature required for annealing the primers to the plasmid.

After, several attempts to optimize thermocycle protocols and utilizing different polymerases, the Kappa HiFi™ polymerase and an optimized version of their protocol was used to perform a successful SDM. The PCR product was purified, transformed into bacteria, and grown on LB plates. These colonies were then used to successfully inoculate a liquid LB culture, which was grown overnight and mini-prepped. We are currently awaiting DNA sequencing analysis of the isolated plasmids to truly confirm the presence of the TAG mutation; however, colony growth relative to a control (no PCR amplification), indicate that the mutagenesis was successful.

4.II.B. INCORPORATION OF UNNATURAL AMINO ACID INTO Cas9

Once the mutant plasmids were purified using a plasmid purification kit, they were co-transformed with the pEVOL-*p*CNF, the polyspecific tRNA synthetase plasmid. The cells were grown overnight on LB plates then inoculated into a 4 mL culture and grown to turbidity. This starter culture was then used to inoculate a larger expression

culture (20 mL). The culture was then induced at an OD_{600} of 0.6 with *para*-propargylphenylalanine (pPrF) as a test of whether we could incorporate UAAs into the protein. We selected this amino acid due to its availability in the laboratory and for its potential use in other experiments. Ultimately, once expression conditions are optimized, we aim to transition to the caged tyrosine UAA.

The success of the expression and protein purification was analyzed by SDS-PAGE. Unfortunately, the gels did not seem to have the correct protein band corresponding to full length Cas9. We hypothesized that this could be due to poor expression yield of the large protein by the bacteria or that the SDM was actually unsuccessful. Moreover, we noticed that the cultures had a distinct odor that we believe was due to contaminated stocks of the UAA *p*PrF. However, when we repeated the expression with newly synthesized pPrF we observed similar results, as no Cas9 protein was produced.

4.III. EXPERIMENTAL

General SDM protocol using Kappa HiFi™ - Into a Eppendorf tube (0.2 mL) 5x Kappa HiFi™ buffer (5 µL), 10 mM dNTPs (0.75 µL), 10 mM forward primer (5'-CATTTCGGATACCCTAGTATGTAGGCCCCCTCG-3') (0.75 µL), 10 mM reverse primer (5'-CGAGGGGGCCTACATACTAGGGTATCCGAAATG-3') 0.75µL, and plasmid template (5 µL;1 ng/µL), were added to sterile deionized water (12.25 µL). The reaction was initiated by the addition of Kappa polymerase (0.5 µL), for a final total volume of 25 µL. The reactions were placed in a Biorad iCycler 4.006 Thermocycler under the following program: 95°C for 2 min, followed by an amplification cycle of 98°C for 30 seconds, 66.5°C for 15 seconds, and 72°C for 4.5 minutes. This cycle was repeated 30 times followed by a final extension at 72°C for 5 minutes, and then incubation of the final product at 4°C. After completion of the PCR, the unmodified template DNA was digested with the addition of DpnI (1 µL) and incubated at 37°C for 2 hours, The enzyme was then denatured at 80°C for 15 minutes and the reaction was purified using a PCR purification kit (Qiagen) and eluted into a deionized water (10 µL) to allow for efficient electroporation. The plasmid (1.0 µL) was then transformed into *Escherichia coli* BL21(DE3) cells using an Eppendorf eporator® electroporator. The cells were then plated on Luria-Bertani (LB) agar plates with kanamycin (50 mg/mL) and incubated at 37°C overnight. Single colonies were then selected and inoculated into LB media (2-4 mL) containing kanamycin. This culture was grown to turbidity and then mini-prepped with a DNA purification kit (Qiagen).

General Cas9 Expression- A pET-Cas9-6xHis plasmid that has undergone the site-directed mutagenesis (0.5 μL) was cotransformed with a pEVOL-ambrx plasmid (0.5 μL) into *Escherichia coli* BL21(DE3) cells using an eppendorf eporator electroporator. The cells were then plated onto Luria-Bertani (LB) agar plates with chloramphenicol (34 mg/mL) and kanamycin (50 mg/mL) resistance at 37°C overnight. Single colonies were then inoculated into LB media (2-4 mL) containing both ampicillin and chloramphenicol resistance. This culture was grown to turbidity overnight at 37°C and subsequently used to inoculate expression cultures (20 mL of LB media, 34 mg/mL of chloramphenicol, and 50 mg/mL of kanamyocin resistance) to an OD_{600} of 0.1. The cultures are then incubated at 37°C to an OD_{600} between 0.6-0.8 and expressions were then centrifuged at 5,000 rpm for 10 minutes and the supernatant was discarded. The pelleted cells were then resuspended in 4 mL of LB media (with 34 mg/mL of chloramphenicol and 50 mg/mL of kanamyocin resistance). The resuspended cells were subsequently induced through the addition of *para*-propargyl-phenylalanine (pPrF) (40 μL, 100 mM), 20% arabinose (4 μL) and 0.8 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; 4 μL). The cultures were then incubated for 16-20 hours at 30°C, followed by centrifugation at 5,000 rpm for 10 minutes, and the cell pellet was stored at -80°C for at least 20 minutes. The pellet was then 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, the supernatant was then collected and transferred into an equilibrated His-pur Ni-NTA spin (Qiagen) column with nickel resin (200 μL) and GFP was purified according to manufacturer's protocol. Purified Cas9 was analyzed by SDS-PAGE (BioRad 10% gels, 150 V, 1.5 hours)

4.IV. FUTURE DIRECTIONS

Future work involves first confirming whether the mutagenesis of the amino acid residue 450 was successful in creating a TAG codon. This could include attempts to incorporate different UAAs or sending the plasmids to be sequenced to see if we obtained the desired mutation. Work will also be done to create new mutations at the other amino acid sites on the Cas9 plasmid. Other potential sites of UAA incorporation include the incorporation of UAAs that can undergo bioconjugation reactions by incorporations of bioorthogonal UAAs on outer residue positions of the Cas9 protein.

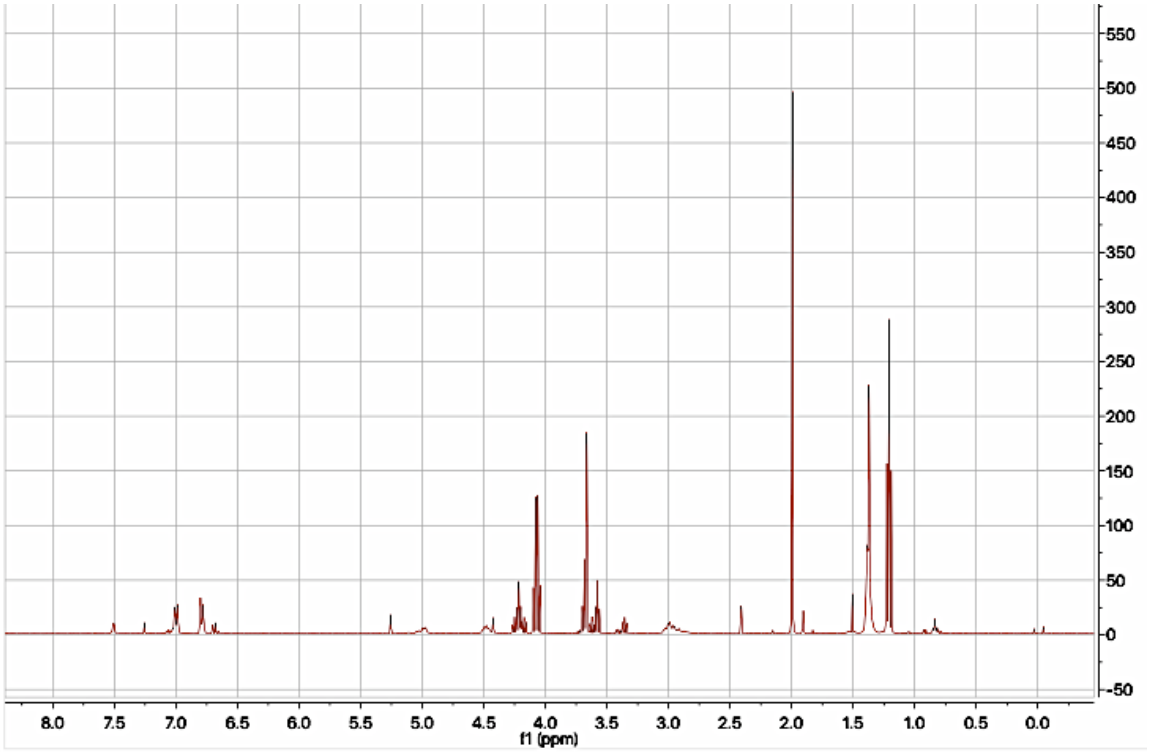
Once the mutagenesis and Cas9 expression with a photocaged tyrosine derivative (ONBY) has been optimized, experiments to determine if functionality of the protein can be modulated will be done. There are many different types of assays that can be performed to test the function of the protein before and after irradiation. For instance, a simple assay protocol is available from New England BioLabs. This is an *in vitro* assay where Cas9 proteins, with a single guide RNA, cleave complementary DNA sequences. The DNA fragments can then be analyzed by gel electrophoresis to see if DNA cleavage

occurs. More complex assays measuring transcriptional control, possibly of fluorescent proteins, can be developed to further characterize new sites of UAA incorporation into Cas9 proteins.

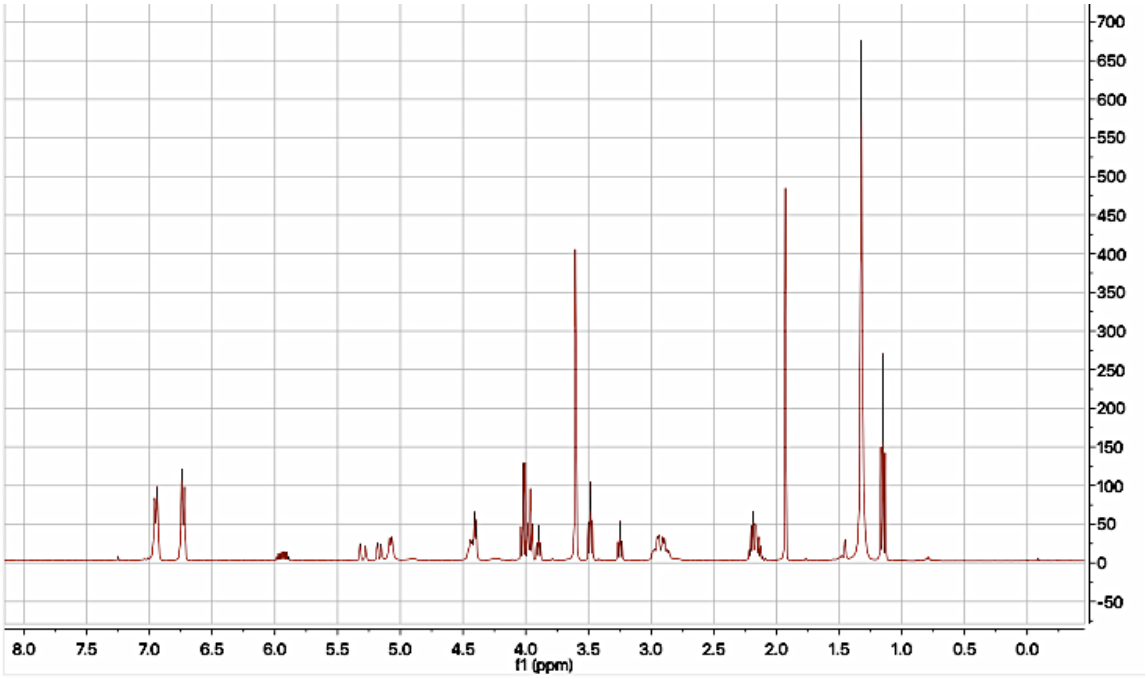
4.V. CONCLUSION

In conclusion, I was able to design primers that will create TAG mutations into a Cas9 plasmid in order to introduce UAAs that will control Cas9 function. These primers were then employed to attempt to mutagenize a Cas9 plasmid. This mutation seemed to be successful as transformation with the plasmid conferred kanamycin resistance. However, expression of the protein with UAAs was unsuccessful. Therefore further characterization of the mutagenesis by sequence analysis is necessary to determine if the mutagenesis was successful or if there are certain expression conditions that need to be optimized for proper protein generation. Future work is also being done to optimize the incorporation of photocaged UAAs into protein and to explore other sites in the Cas9 protein for the development of bioconjugation uses such as labeling with fluorescent molecules or immobilizing a functional protein onto a solid support. These advances will contribute novel applications to the expanding utility of this technology.

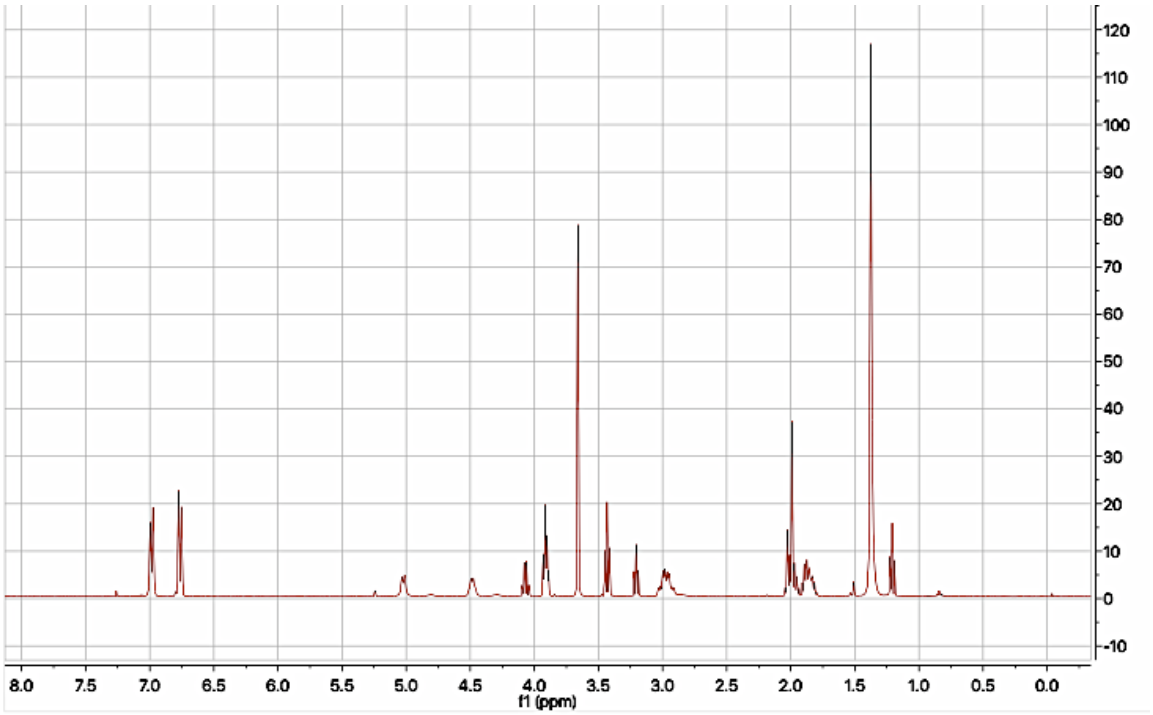
APPENDIX A: NMR SPECTRA
CHAPTER 2:



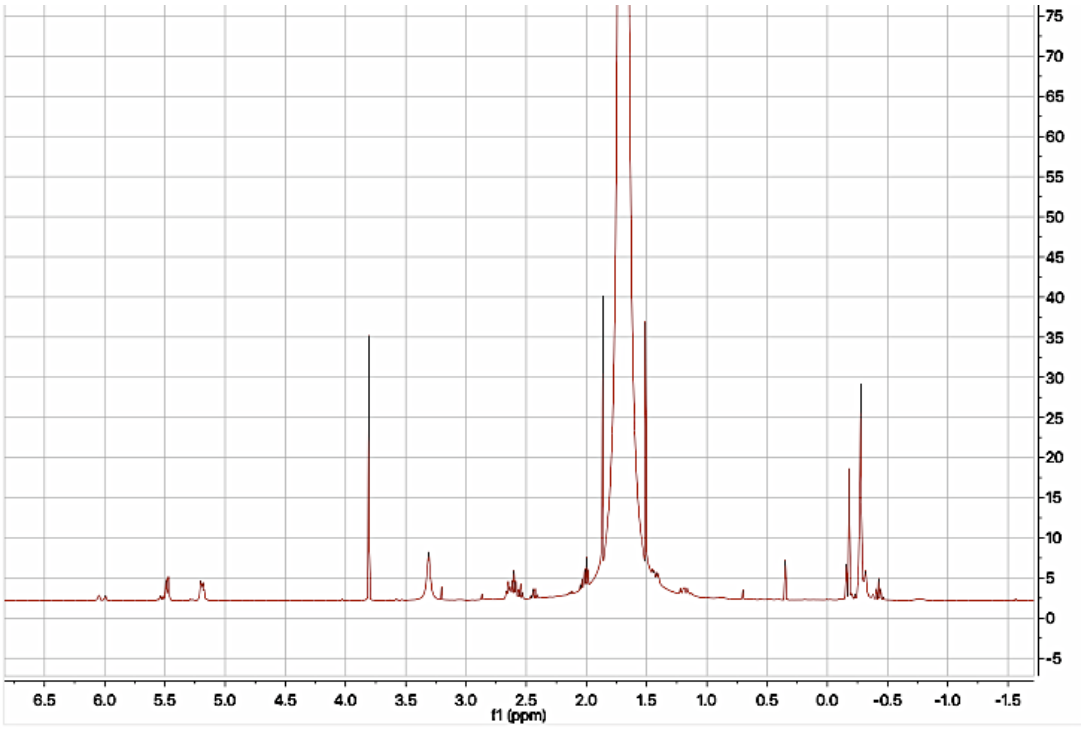
A1. ^{1}H NMR of ***p*-Bromo-ethyl-Boc-OMe-Tyrosine (1)-** NMR (400 MHz, $CDCl_3$)



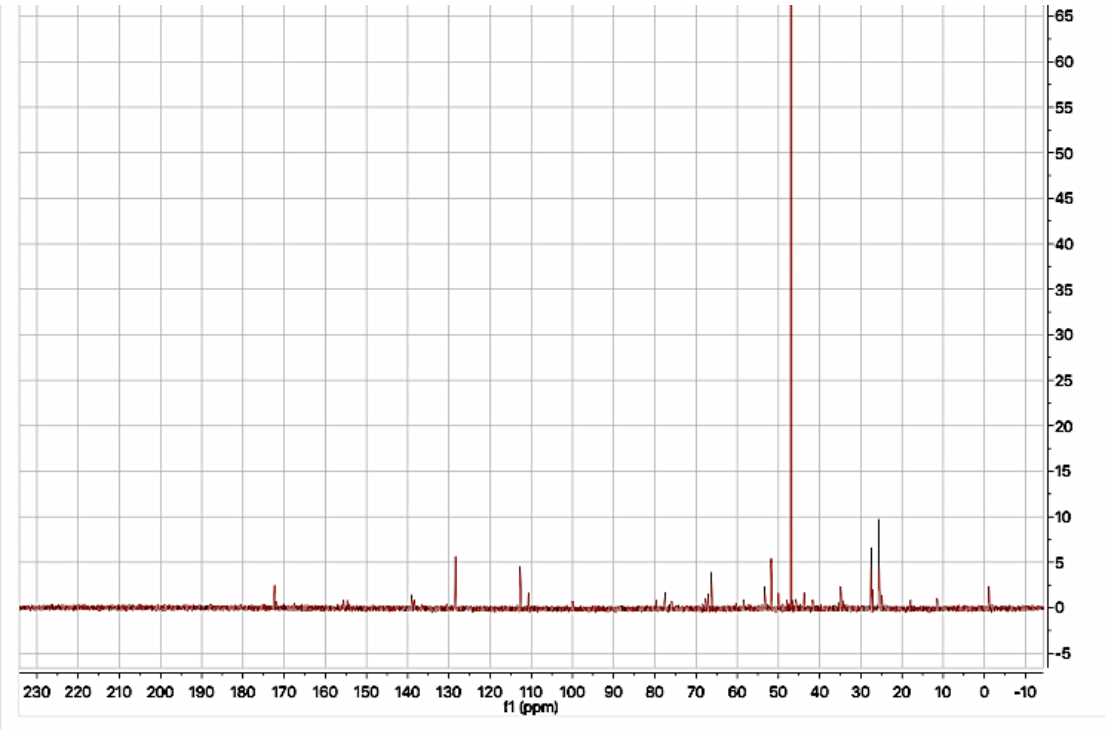
A2. ^{1}H NMR of ***p*-Bromo-propyl-Boc-OMe-Tyrosine (2)-** NMR (400 MHz, $CDCl_3$)



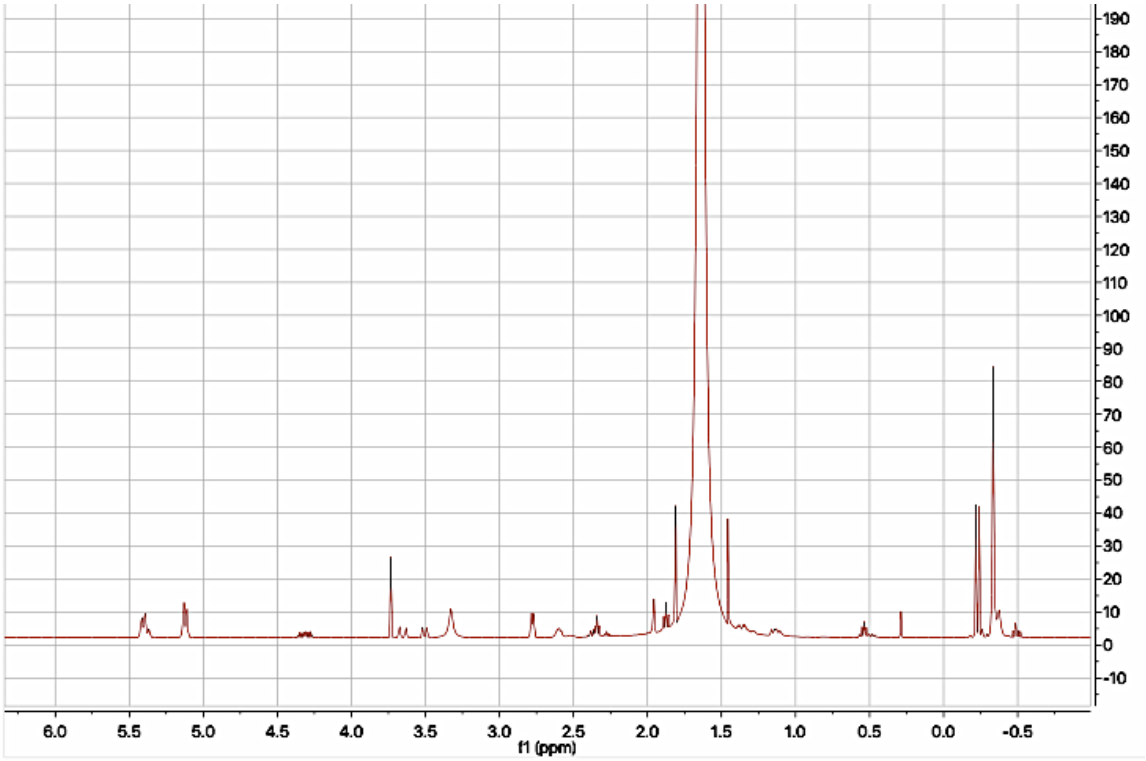
A3. ^{1}H NMR of ***p*-Bromo-butyl-Boc-OMe-Tyrosine (3)-** NMR (400 MHz, $CDCl_3$)



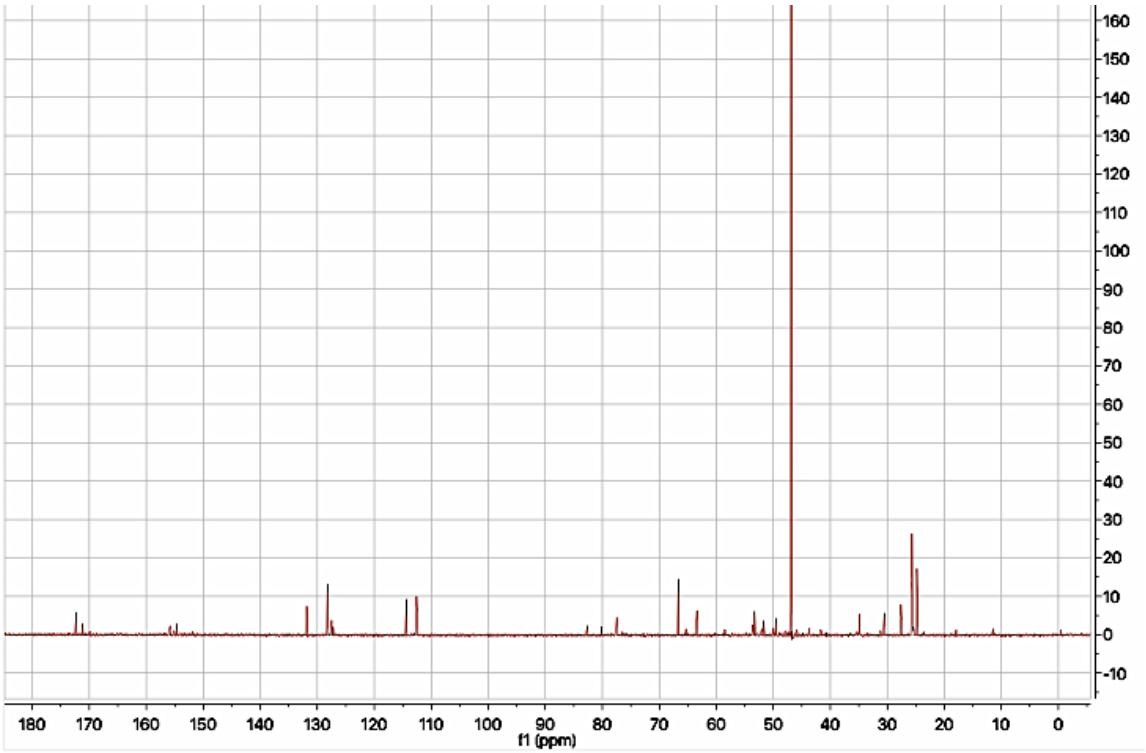
A4. ^{1}H NMR of ***p*-Alkoxyamine-ethyl-Tyrosine (4)-**NMR (400 MHz, CD_3OD)



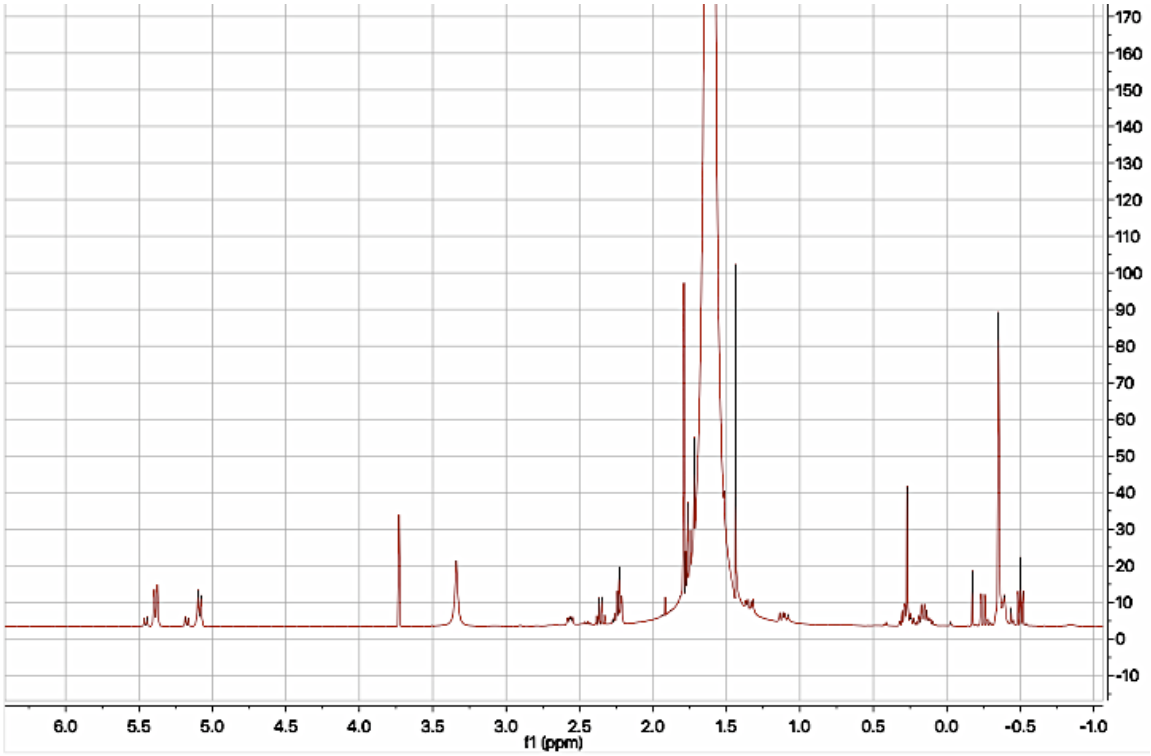
A5. ^{13}C NMR of ***p*-Alkoxyamine-ethyl-Tyrosine (4)-** NMR (400 MHz, CD_3OD)



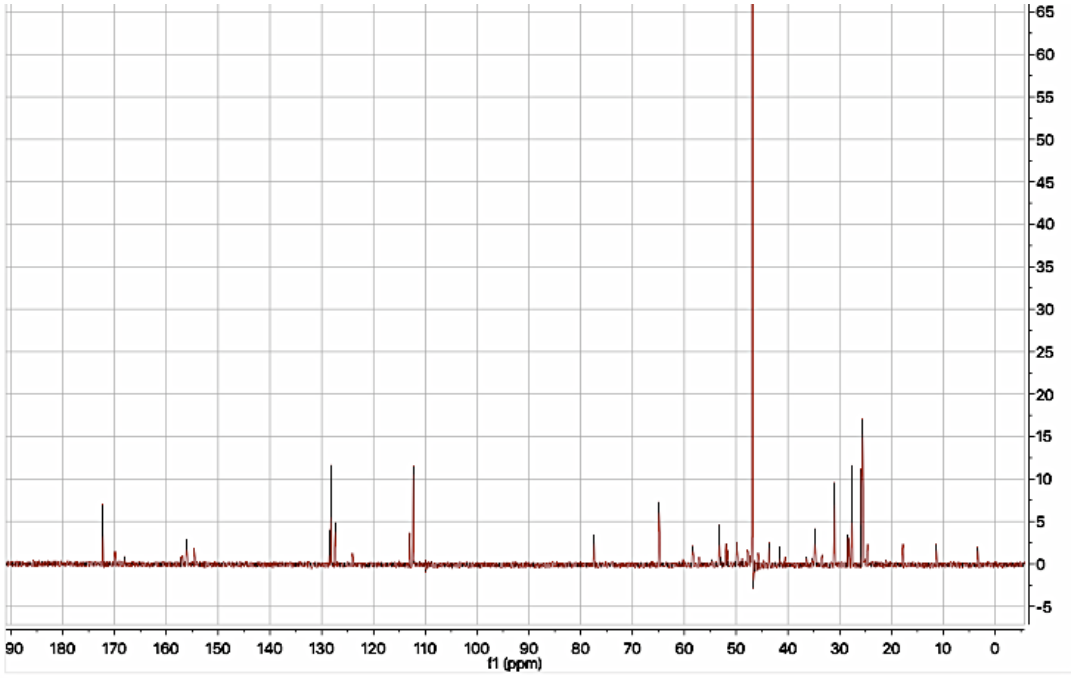
A6. ^{1}H NMR of ***p*-Alkoxyamine-propyl-Tyrosine (5)-** NMR (400 MHz, CD_3OD)



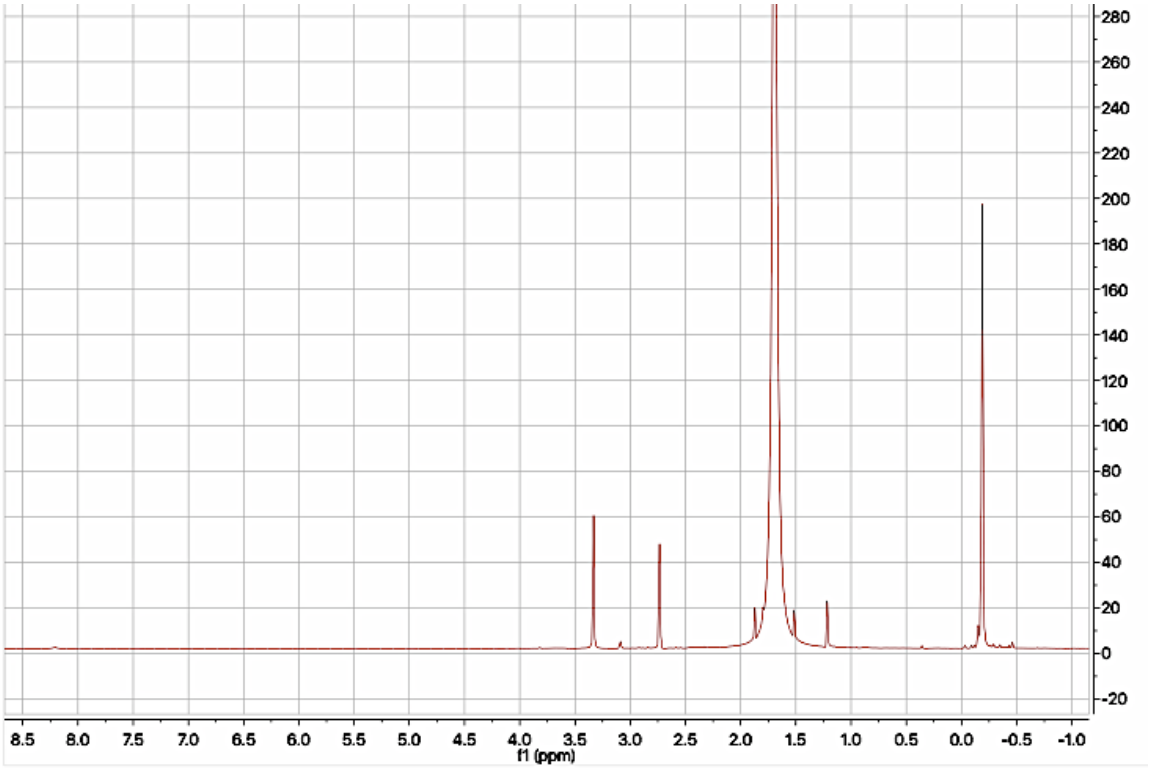
A7. ^{13}C NMR of ***p*-Alkoxyamine-propyl-Tyrosine (5)-** NMR (400 MHz, CD_3OD)



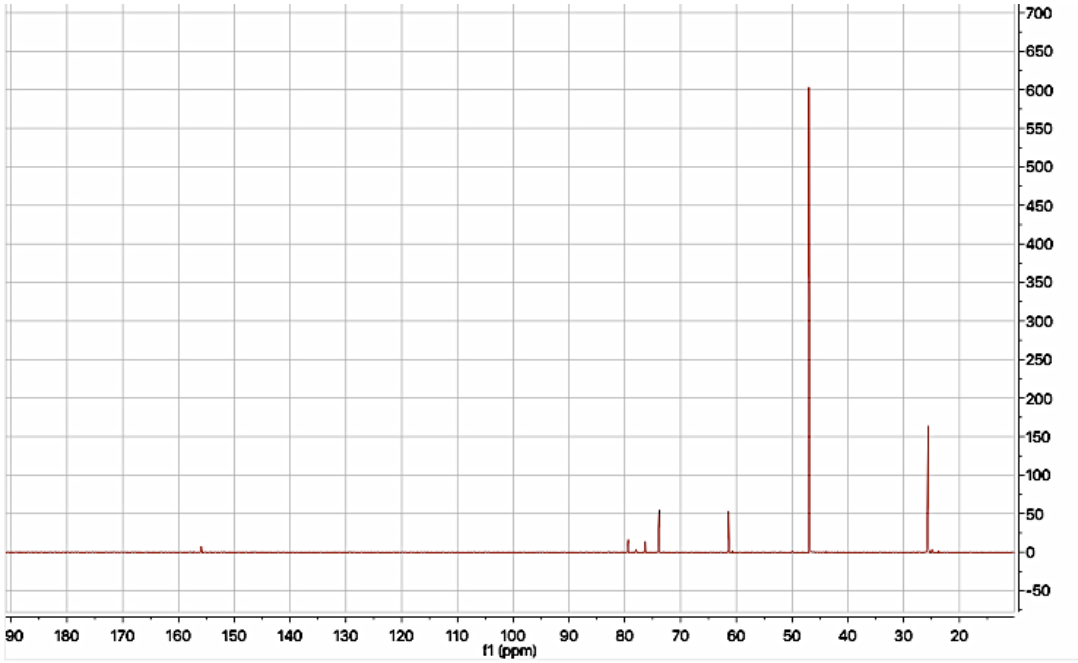
A8. ^{1}H NMR of ***p*-Alkoxyamine-butyl-Tyrosine (6)-** NMR (400 MHz, CD_3OD)



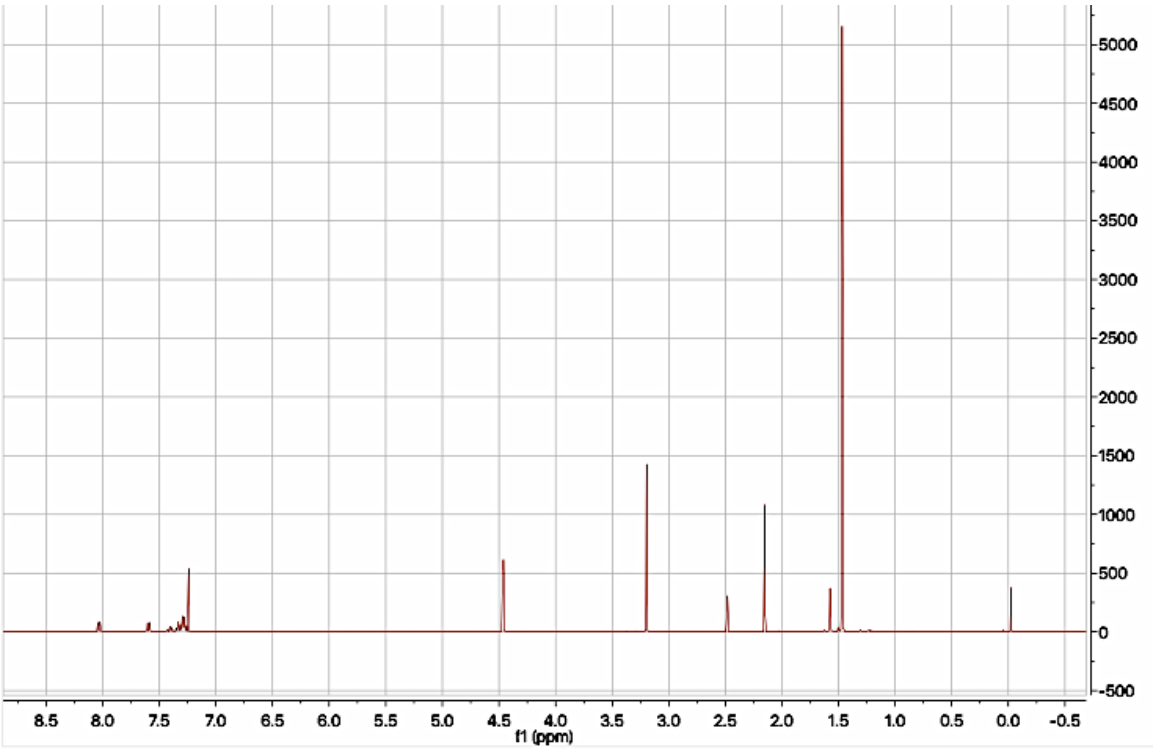
A9. ^{13}C NMR of ***p*-Alkoxyamine-butyl-Tyrosine (6)-** NMR (400 MHz, CD_3OD)



A10. ^{1}H NMR of ***O*-(prop-2-yn-1-yl)hydroxylamine (7)**-NMR (400 MHz, CD_3OD)



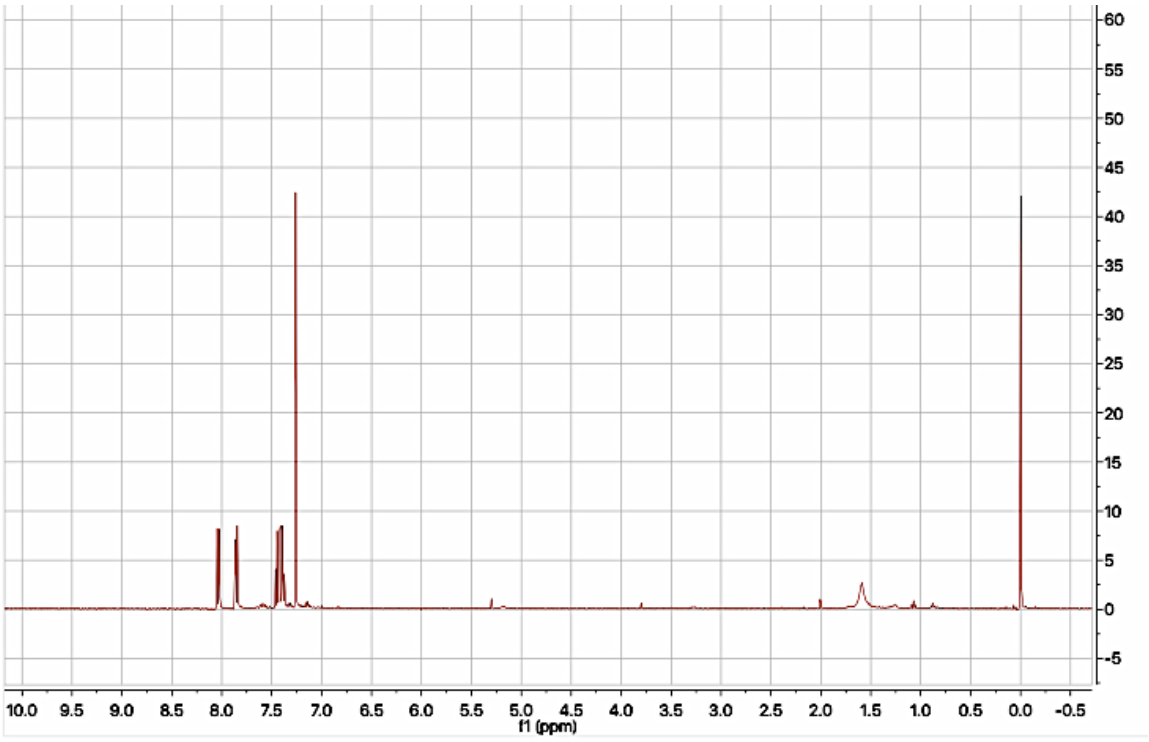
A11. ^{13}C NMR of ***O*-(prop-2-yn-1-yl)hydroxylamine (7)-** NMR (400 MHz, CD_3OD)



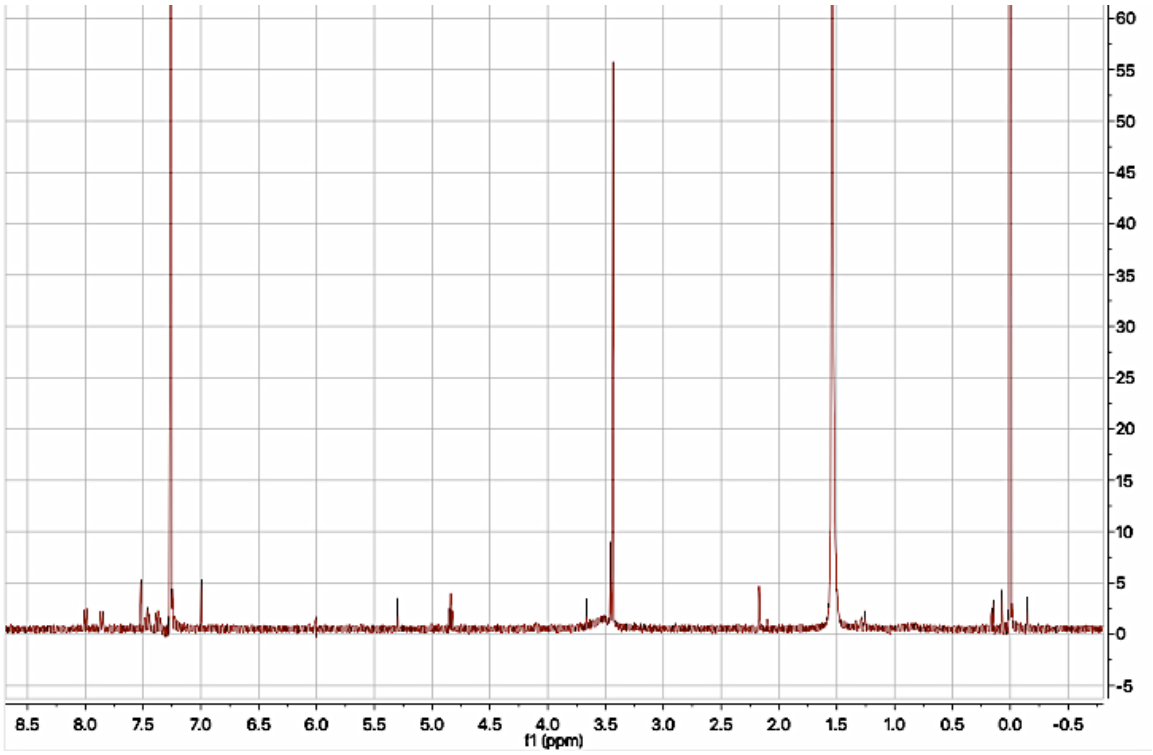
A12. ^{1}H NMR of **N-(1,3-diphenyl-1H-pyrazol-5-yl)-4-nitro-N'-(prop-2-yn-1-yloxy)benzimidamide (8)-**NMR (400 MHz, CD_3OD)

CHAPTER 3:

A13. ^{1}H NMR of **2-Iodobenzo[d]thizole (1)-** NMR (400 MHz, $CDCl_3$)

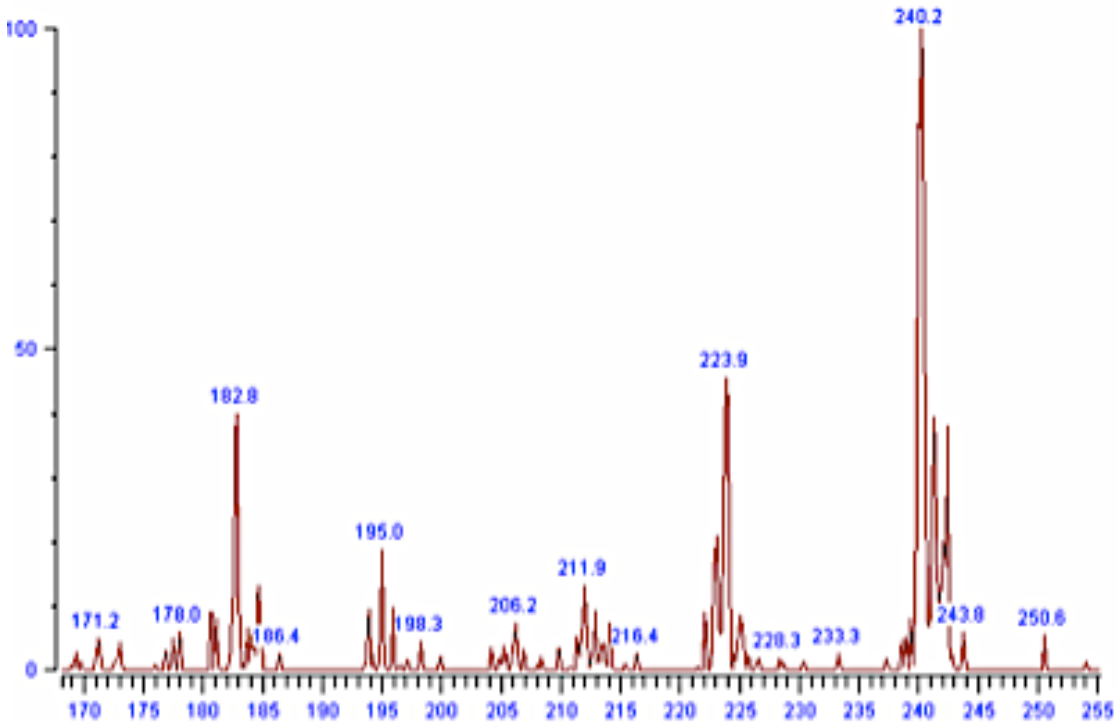


A14. ^{1}H NMR of **2-ethynylbenzo[d]thiazole (3)-**NMR (400 MHz, $CDCl_3$)

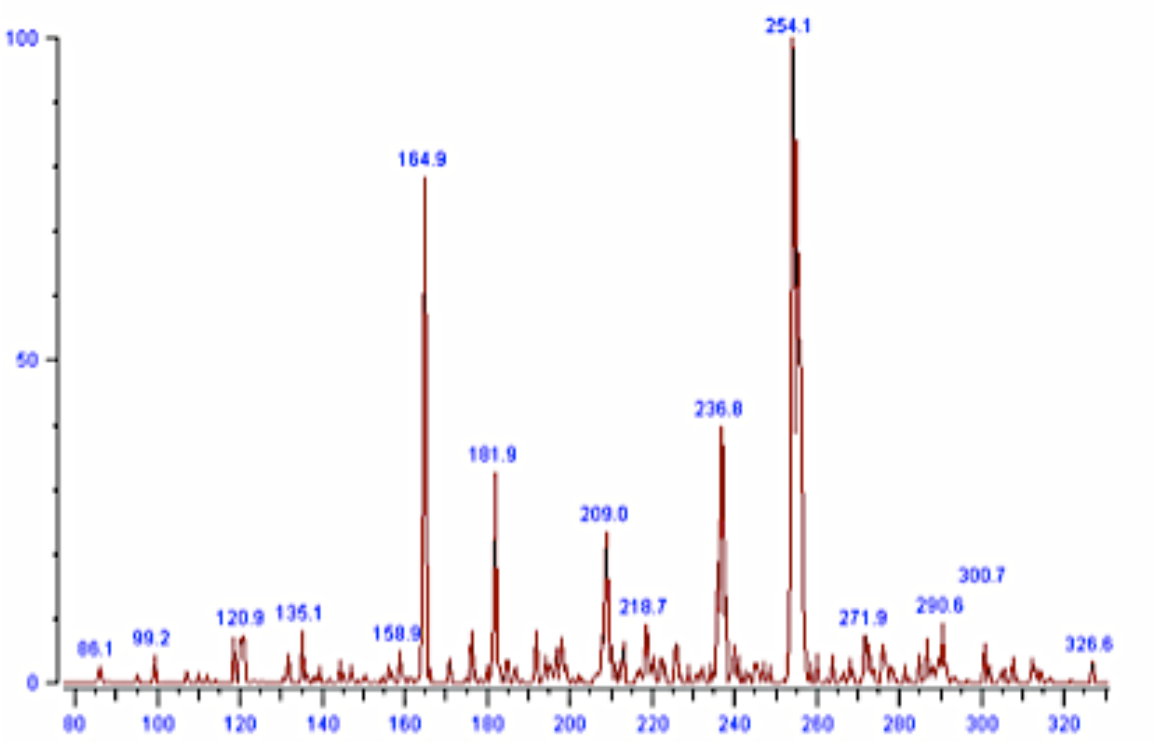


APPENDIX B: MASS SPECTRA
CHAPTER 2:

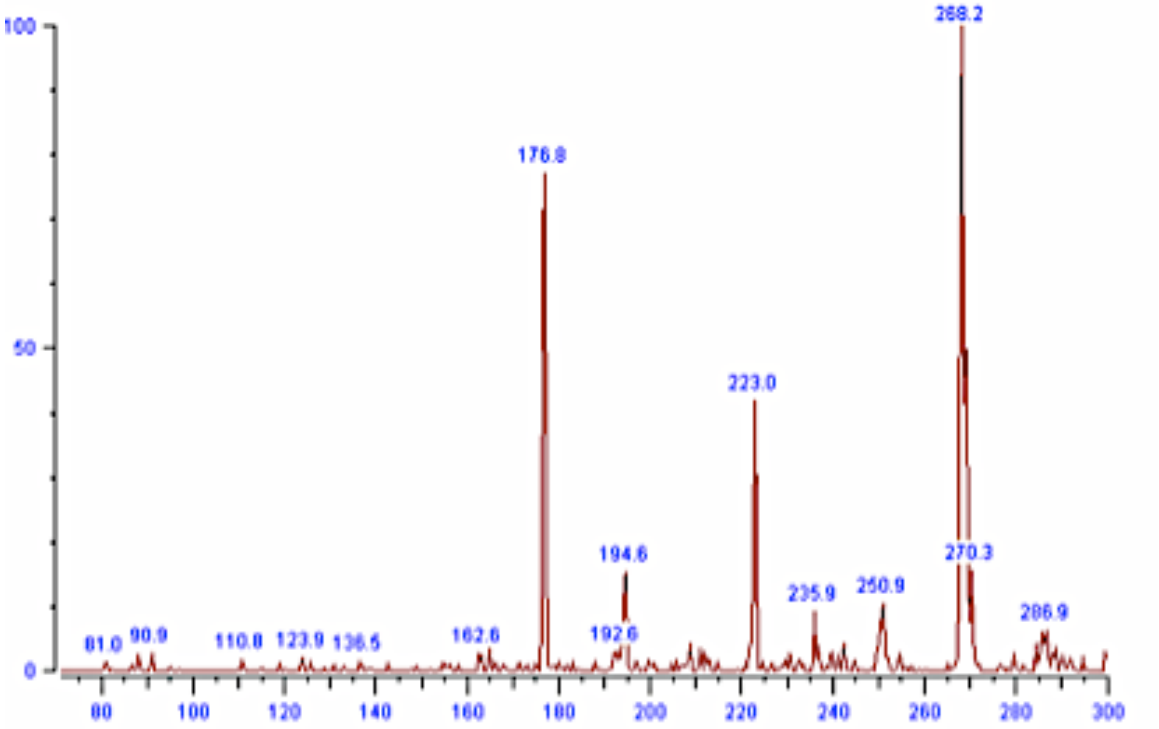
B1. Mass Spectrum of ***p*-Alkoxyamine-ethyl-Tyrosine (4)-** (ESI) m/z ($M^+ + H^+$)



B2. Mass Spectrum of ***p*-Alkoxyamine-propyl-Tyrosine (5)-** (ESI) m/z ($M^+ + H^+$)



B3. Mass Spectrum of ***p*-Alkoxyamine-butyl-Tyrosine (6)-**(ESI) m/z ($M^+ + H^+$)



APPENDIX C: PRIMER SEQUENCES
CHAPTER 4:

Cas9 D_10TAG_F
5'-AAGTACTCCATTGGGCTCTAGATCGGCACAAACAGCGTC-3'

Cas9 D_10TAG_R
5'-GACGCTGTTTGTGCCGATCTAGAGCCCAATGGAGTACTT-3'

Cas9 Y_450TAG_F
5'-CATTTCGGATACCCTAGTATGTAGGCCCCCTCG-3'

Cas9 Y_450TAG_R
5'-CGAGGGGGCCTACATACTAGGGTATCCGAAATG-3'

Cas9 Y_1265TAG_F
5'-ACACAAACACTAGCTTGAGATCATCGAGCAAATAAGC-3'

Cas9 Y_1265TAG_R
5'-GCTTATTTGCTCGATGATCTCATCAAGCTAGTGTTTGTGT-3'

dCas9 D_10TAG_F
5'-CAACTGAATTGGTGCCGATCTACAGGCCGATACTGTATTTTT-3'

dCas9 D_10TAG_R
5'-AAAAATACAGTATCGGCCTGTAGATCGGCACCAATTCAGTTG-3'

dCas9 Y_72TAG_F
5'-CTTACGACGTGTCTAGCGACGGCGGGC-3'

dCas9 Y_72TAG_R
5'-GCCCGCCGTCGCTAGACACGTCGTAAG-3'

dCas9 Y_450TAG_F
5'-ACATTTCGTATCCCTTAGTATGTCGGTCCTCTG-3'

dCas9 Y_450TAG_R
5'-CCAGAGGACCGACATACTAAGGGATACGAAATGT-3'

dCas9 Y_1131TAG_F
5'-GAATCGAAGCCACCCTATTTTTTCGAATCCCAATCTTTTTTG-3'

dCas9 Y_1131TAG_R
5'-CAAAAAAGATTGGGATCCGAAAAAATAGGGTGGCTTCGATTC-3'

ACKNOWLEDGEMENTS

I'd like to thank my family, who has supported me in all of my endeavors.

I'd like to thank the members of the Young lab for helping me throughout these projects and being there to always put a smile on my face.

I'd like to thank Johnathan Maza for teaching me how to work in the lab, for always giving me guidance, and for being a friend.

I'd like to thank Dr. Margaret Saha and members of IGEM for teaching me so much and supporting me through my work.

I'd like to thank my honors committee; Dr. Robert Hinkle, Dr. William McNamara, Dr. Margaret Saha, and Dr. Douglas Young for helping me to better understand science.

I'd like to thank Dr. Douglas Young for allowing me to work under him, giving me guidance, and having patience with me during my growth as a scientist.

REFERENCES

1. Young, T. S.; Schultz, P. G., Beyond the canonical 20 amino acids: expanding the genetic lexicon. *J Biol Chem* **2010,** *285* (15), 11039-44.
2. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010,** *79*, 413-44.
3. Doerge, R. W., Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* **2002,** *3* (1), 43-52.
4. Wang, Q.; Parrish, A. R.; Wang, L., Expanding the genetic code for biological studies. *Chem Biol* **2009,** *16* (3), 323-36.
5. Urano, Y.; Asanuma, D.; Hama, Y.; Koyama, Y.; Barrett, T.; Kamiya, M.; Nagano, T.; Watanabe, T.; Hasegawa, A.; Choyke, P. L.; Kobayashi, H., Selective molecular imaging of viable cancer cells with pH-activatable fluorescence probes. *Nat Med* **2009,** *15* (1), 104-9.
6. Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; Lu, Y.; Tran, H.; Seller, A. J.; Biroc, S. L.; Szydlik, A.; Pinkstaff, J. K.; Tian, F.; Sinha, S. C.; Felding-Habermann, B.; Smider, V. V.; Schultz, P. G., Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc Natl Acad Sci U S A* **2012,** *109* (40), 16101-6.
7. Minaba, M.; Kato, Y., High-yield, zero-leakage expression system with a translational switch using site-specific unnatural amino Acid incorporation. *Appl Environ Microbiol* **2014,** *80* (5), 1718-25.
8. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew Chem Int Ed Engl* **2009,** *48* (38), 6974-98.
9. Santoro, S. W.; Anderson, J. C.; Lakshman, V.; Schultz, P. G., An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli. *Nucleic Acids Res* **2003,** *31* (23), 6700-9; Cropp, T. A.; Schultz, P. G., An expanding genetic code. *Trends Genet* **2004,** *20* (12), 625-30.
10. 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* **1997,** *94* (19), 10092-7; Wang, L.; Schultz, P. G., A general approach for the generation of orthogonal tRNAs. *Chem Biol* **2001,** *8* (9), 883-90; Cload, S. T.; Liu, D. R.; Froland, W. A.; Schultz, P. G., Development of improved tRNAs for in vitro biosynthesis of proteins containing unnatural amino acids. *Chem Biol* **1996,** *3* (12), 1033-8.
11. Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G., An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **2011,** *50* (11), 1894-900.
12. Zhang, Z.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G., Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc Natl Acad Sci U S A* **2004,** *101* (24), 8882-7.

13. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.
14. Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G., An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol* **2010,** *395* (2), 361-74.
15. Tornoe, C.; 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. *Journal of Organic Chemistry* **2002,** *67* (9), 3057-3064; Chan, T.; Hilgraf, R.; Sharpless, K.; Fokin, V., Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Organic Letters* **2004,** *6* (17), 2853-2855.
16. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R., Copper-free click chemistry for dynamic in vivo imaging. *Proc Natl Acad Sci U S A* **2007,** *104* (43), 16793-7.
17. Zeng, Y.; Ramya, T. N.; Dirksen, A.; Dawson, P. E.; Paulson, J. C., High-efficiency labeling of sialylated glycoproteins on living cells. *Nat Methods* **2009,** *6* (3), 207-9.
18. Dirksen, A.; Dawson, P. E., Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjug Chem* **2008,** *19* (12), 2543-8.
19. Nigst, T. A.; Antipova, A.; Mayr, H., Nucleophilic reactivities of hydrazines and amines: the futile search for the α-effect in hydrazine reactivities. *J Org Chem* **2012,** *77* (18), 8142-55.
20. Sivakumar, K.; Xie, F.; Cash, B. M.; Long, S.; Barnhill, H. N.; Wang, Q., A fluorogenic 1,3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes. *Org Lett* **2004,** *6* (24), 4603-6.
21. Qi, J.; Han, M. S.; Chang, Y. C.; Tung, C. H., Developing visible fluorogenic 'click-on' dyes for cellular imaging. *Bioconjug Chem* **2011,** *22* (9), 1758-62.
22. Lang, K.; Chin, J. W., Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem Rev* **2014,** *114* (9), 4764-806.
23. Jencks, W. P., Studies on the Mechanism of Oxime and Semicarbazone Formation. *Journal of the American Chemical Society* **1958,** *81* (2), 475-481.
24. SAYER, J.; PESKIN, M.; JENCKS, W., IMINE-FORMING ELIMINATION-REACTIONS .1. GENERAL BASE AND ACID CATALYSIS AND INFLUENCE OF NITROGEN SUBSTITUENT ON RATES AND EQUILIBRIA FOR CARBINOLAMINE DEHYDRATION. *Journal of the American Chemical Society* **1973,** *95* (13), 4277-4287.
25. Tuley, A.; Lee, Y. J.; Wu, B.; Wang, Z. U.; Liu, W. R., A genetically encoded aldehyde for rapid protein labelling. *Chem Commun (Camb)* **2014,** *50* (56), 7424-6.
26. Maza, J. C.; Jacobs, T. H.; Uthappa, D. M.; Young, D. D., Employing Unnatural Amino Acids in the Preparation of Bioconjugates. *Synlett* **2015**.
27. Kim, C.; Axup, J.; Schultz, P., Protein conjugation with genetically encoded unnatural amino acids. *Current Opinion in Chemical Biology* **2013,** *17* (3), 412-419.
28. Kalia, J.; Raines, R. T., Hydrolytic stability of hydrazones and oximes. *Angew Chem Int Ed Engl* **2008,** *47* (39), 7523-6.
29. Zhang, Z.; Smith, B. A.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G., A new strategy for the site-specific modification of proteins in vivo. *Biochemistry* **2003,** *42* (22), 6735-46.
30. Zeng, H.; Xie, J.; Schultz, P. G., Genetic introduction of a diketone-containing amino acid into proteins. *Bioorg Med Chem Lett* **2006,** *16* (20), 5356-9.

31. Kazane, S. A.; Sok, D.; Cho, E. H.; Uson, M. L.; Kuhn, P.; Schultz, P. G.; Smider, V. V., Site-specific DNA-antibody conjugates for specific and sensitive immuno-PCR. *Proc Natl Acad Sci U S A* **2012,** *109* (10), 3731-6.
32. Kim, C. H.; Axup, J. Y.; Dubrovska, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G., Synthesis of bispecific antibodies using genetically encoded unnatural amino acids. *J Am Chem Soc* **2012,** *134* (24), 9918-21.
33. Maza, J. C.; McKenna, J. R.; Raliski, B. K.; Freedman, M. T.; Young, D. D., Synthesis and Incorporation of Unnatural Amino Acids To Probe and Optimize Protein Bioconjugations. *Bioconjug Chem* **2015,** *26* (9), 1884-9.
34. Chen, X. H.; Xiang, Z.; Hu, Y. S.; Lacey, V. K.; Cang, H.; Wang, L., Genetically encoding an electrophilic amino acid for protein stapling and covalent binding to native receptors. *ACS Chem Biol* **2014,** *9* (9), 1956-61; Xiang, Z.; Lacey, V. K.; Ren, H.; Xu, J.; Burban, D. J.; Jennings, P. A.; Wang, L., Proximity-enabled protein crosslinking through genetically encoding haloalkane unnatural amino acids. *Angew Chem Int Ed Engl* **2014,** *53* (8), 2190-3.
35. Brustad, E. M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A., A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J Am Chem Soc* **2008,** *130* (52), 17664-5.
36. Waldo, J. P.; Larock, R. C., Synthesis of isoxazoles via electrophilic cyclization. *Org Lett* **2005,** *7* (23), 5203-5.
37. Castañeda, C.; Liu, J.; Chaturvedi, A.; Nowicka, U.; Cropp, T. A.; Fushman, D., Nonenzymatic assembly of natural polyubiquitin chains of any linkage composition and isotopic labeling scheme. *J Am Chem Soc* **2011,** *133* (44), 17855-68.
38. Kim, J.; Seo, M. H.; Lee, S.; Cho, K.; Yang, A.; Woo, K.; Kim, H. S.; Park, H. S., Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis. *Anal Chem* **2013,** *85* (3), 1468-74.
39. Shie, J. J.; Liu, Y. C.; Lee, Y. M.; Lim, C.; Fang, J. M.; Wong, C. H., An azido-BODIPY probe for glycosylation: initiation of strong fluorescence upon triazole formation. *J Am Chem Soc* **2014,** *136* (28), 9953-61.
40. Shieh, P.; Dien, V. T.; Beahm, B. J.; Castellano, J. M.; Wyss-Coray, T.; Bertozzi, C. R., CalFluors: A Universal Motif for Fluorogenic Azide Probes across the Visible Spectrum. *J Am Chem Soc* **2015,** *137* (22), 7145-51.
41. Qi, J.; Tung, C. H., Development of benzothiazole 'click-on' fluorogenic dyes. *Bioorg Med Chem Lett* **2011,** *21* (1), 320-3.
42. Hsu, P. D.; Lander, E. S.; Zhang, F., Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **2014,** *157* (6), 1262-78.
43. Jiang, F.; Zhou, K.; Ma, L.; Gressel, S.; Doudna, J. A., STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **2015,** *348* (6242), 1477-81.
44. Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A., Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014,** *343* (6176), 1247997.

45. Bikard, D.; Jiang, W.; Samai, P.; Hochschild, A.; Zhang, F.; Marraffini, L. A., Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **2013,** *41* (15), 7429-37.
46. Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; Villalta, J. E.; Chen, Y.; Whitehead, E. H.; Guimaraes, C.; Panning, B.; Ploegh, H. L.; Bassik, M. C.; Qi, L. S.; Kampmann, M.; Weissman, J. S., Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **2014,** *159* (3), 647-61.
47. Arbely, E.; Torres-Kolbus, J.; Deiters, A.; Chin, J. W., Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J Am Chem Soc* **2012,** *134* (29), 11912-5; Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G., A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* **2006,** *45* (17), 2728-31.