Undergraduate Honors Theses   Theses, Dissertations, & Master Projects

5-2015

Unnatural Amino Acids in Proteins for Development of Novel Biochemical Tools

Jordan Villa
College of William and Mary

Recommended Citation

Villa, Jordan, "Unnatural Amino Acids in Proteins for Development of Novel Biochemical Tools" (2015). *Undergraduate Honors Theses.* Paper 152.
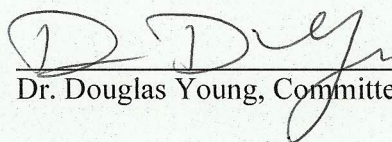https://scholarworks.wm.edu/honorstheses/152

Unnatural Amino Acids in Proteins for Development of Novel Biochemical Tools

A thesis submitted in partial fulfillment of the requirement
For the degree of Bachelor of Science in Chemistry from
The College of William & Mary
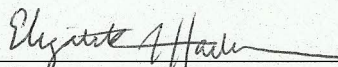
by

Jordan Kelly Villa
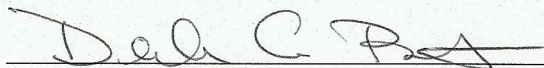
Accepted for Honors

Dr. Douglas Young, Committee Chair

Dr. Elizabeth Harbron

Dr. Deborah Bebout

Dr. Diane Shakes

Williamsburg, VA
May 6, 2015

Unnatural Amino Acids in Proteins for Development of Novel Biochemical Tools

Jordan Kelly Villa
Forest, Virginia

A Thesis presented to the
College of William and Mary in Candidacy for the Degree of
Bachelor of Science

Chemistry Department

The College of William and Mary
May, 2015

Table of Contents

Acknowledgements

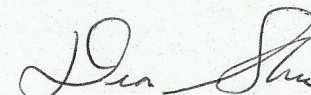I would like to express my gratitude and appreciation to Dr. Douglas Young for all of his guidance, patience, and constant jokes over the past four years in his lab. Thank you for always answering the millions of questions I have asked, even if the first response is a joking "no." I am so grateful to have had you as my advisor who made the lab a fun and inviting space. I hope that the results in this thesis are indeed "promising."

Thanks to the Harbron Lab for their collaboration on the fluorescence project. I look forward to seeing the final results of this project that you have worked so hard on.

I am also thankful to all of my professors here at William and Mary, especially Dr. Bebout, Dr. Harbron, and Dr. Shakes. I want to thank you not only for taking time to serve on my thesis committee, but for being such valuable professors. Thanks also to the William and Mary Chemistry Department as a whole. I had no intention of being a Chemistry major when I joined William and Mary in 2011, but my experiences with the professors of this department changed my mind and I am so thankful for that.

To the Young lab: I really don't have enough space to tell you how much you all mean to me. It has been a fantastic ride this whole way through with you all. It has been a joy to come in everyday to laugh and joke with you. You have been such great friends and you have helped to make lab be a second home.
Extremely special thanks to the two graduate students of Young Lab, Val Tripp and Jess Lampkowski. To Val: thank you so much for teaching me so much in the beginning, especially when I realize how late in the day I always showed up to lab. You were such a great mentor and friend, and I am so grateful that I got to know you. To Jess: You have put a smile on my face every day. Thank you for your constant positive energy, for always listening, and for being a great friend. Thank you for all of your hard work in the Glaser-Hay project—I am so glad we had a project together.

And finally, special thanks to my family and to CRA who have helped me the whole way through. Thank you for listening to me vent, telling me when I am being ridiculous and over stressing about unimportant details, and for always being there. I would have never been able to make it through without your unconditional love and support.

Abstract

Unnatural amino acids (UAAs) permit the incorporation of novel biochemical functionalities into proteins. This expansion of the genetic code has enabled enhanced spatial and temporal control of protein activity and conferred novel protein reactivity. This study examines the incorporation of three UAAs: fluoro-tyrosine, ortho-nitrobenzyl-tyrosine, and propargyloxy-phenylalanine towards various applications. Each UAA was successfully incorporated into a protein of interest (GFP or PRMT1) to facilitate the desired manipulation of protein function. The resulting alterations to GFP fluorescence, PRMT1 activity, or immobilization using Glaser-Hay bioconjugation demonstrate the success and practicality of the utilization of UAAs in the development of novel biochemical tools.

I.   Introduction to Unnatural Amino Acids

A. What are Unnatural Amino Acids (UAAs)

As apparent by the wide amount of diversity within living things, the twenty amino acids that compose all proteins are quite functional and adaptable due to their diverse range of biochemical functions. Proteins are the main machinery of a cell and provide cell structure, regulation of cell activities, and the overall catalysis of processes vital to cells.[1] Regulation of a cell's activities occurs through the regulation of protein concentration and protein activity. The genetic code, contained within DNA, consists of trinucleotide sequences called codons that encode one amino acid. As there are 64 possible codons from the four different nucleotides (4^3), there exists a certain amount of degeneracy in the code for amino acid incorporation. Regulation of transcription (the formation of mRNA from DNA) or translation (the formation of protein from mRNA) determines the amount of protein in the cell. Once protein is made from the twenty available amino acids, a number of post-translational modifications can occur. These modifications involve the additions of functional groups to the protein structure such as phosphorylation, methylation, acetylation, and glycosylation, which broaden the range of applications of proteins. However, despite these additions, the chemical composition of proteins is rather limited as the canonical twenty amino acids contain no alkynes, azides, halogens, etc. (Figure 1.1).[2]
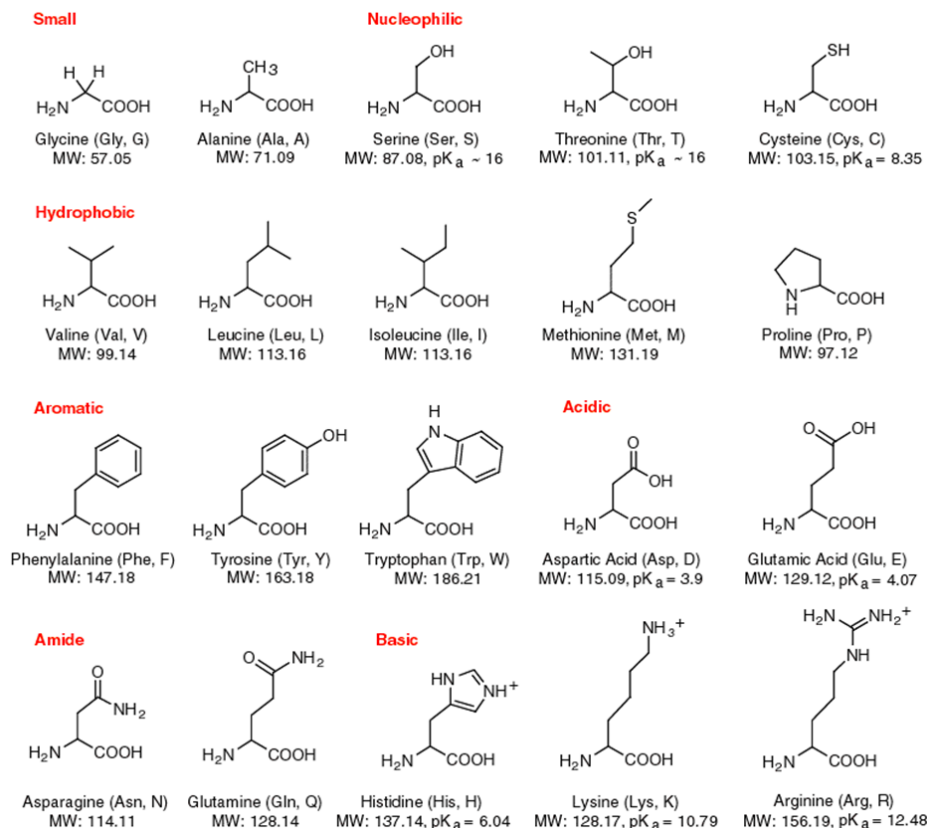
Figure 1.1: The structure of the canonical twenty amino acids. Each amino acid has a carboxylic and an amino functionality that allows them to be joined together in peptide bonds to form proteins. Each amino acid contains a slightly different R group that can be altered to produce UAAs.

To address these limitations, "unnatural" amino acids (UAAs) that contain novel functionalities can be synthesized. Some naturally occurring UAAs exist, such as the 21st and 22nd amino acids, selenocysteine and pyrrolysine, which are found in several bacteria and archaea.[3–5] Selenocysteine is modified with selenium instead of sulfur on a cysteine and pyrrolysine has a pyrroline ring on a lysine (Figure 1.2). The possibilities for the unnatural amino acids are only limited by our knowledge of chemistry, our synthetic capabilities, and our ability to incorporate the UAA into protein (due to constraints imposed by recognition of the UAA (e.g. size limits)).
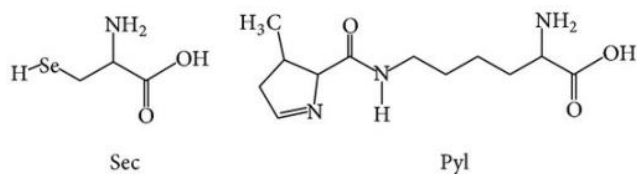
Figure 1.2: The structures of selenocysteine (Sec) and pyrrolysine (Pyl).

The amino acid modifications translate to structure and function alterations of the protein of interest. Addition of specific elements (such as metals, fluorine, reactive chemical functionalities, etc.) can be used for x-ray crystallography, fluorescent probes, or for reactive chemistry such as click reactions (Figure 1.3).[4] The development of UAA technologies is not only a quest for an expanded genetic library, but also a useful tool in therapeutics, engineering, and agriculture (fungicides) among other fields.[2,6] The manipulability of the UAAs allows for specificity that cannot be obtained when confined to using the original twenty amino acids.

This paper will explore the synthesis, incorporation into proteins, and applications of three unnatural amino acids: fluoro-tyrosine, ortho-nitrobenzyl-tyrosine, and propargyloxy-phenylalanine (Figure 1.4). As the range and utility of UAAa is a vast and expanding field, this paper presents only a small sampling of all that can be done with an expanded amino acid library.
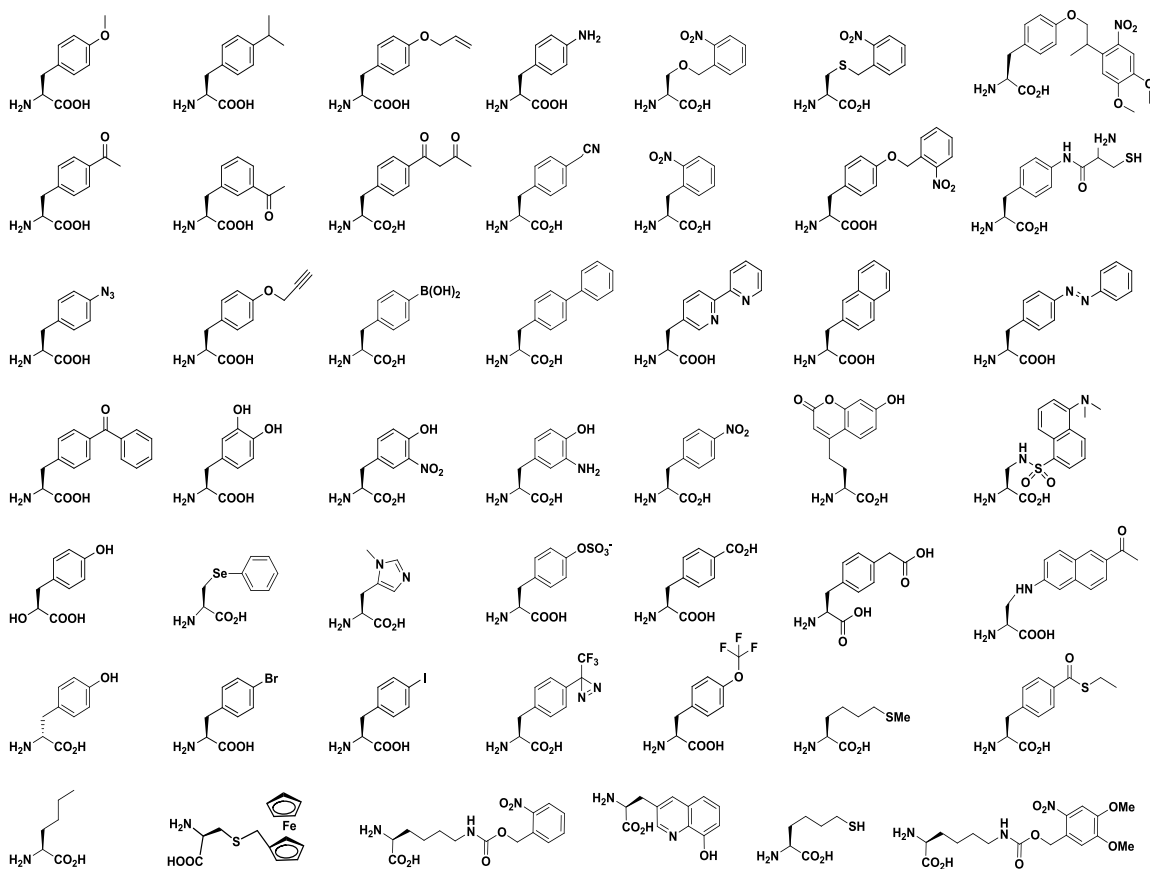
Figure 1.3: Examples of currently incorporated UAAs.

3-fluoro-tyrosine

ortho-nitrobenzyl-tyrosine

propargyloxy-phenylalanine
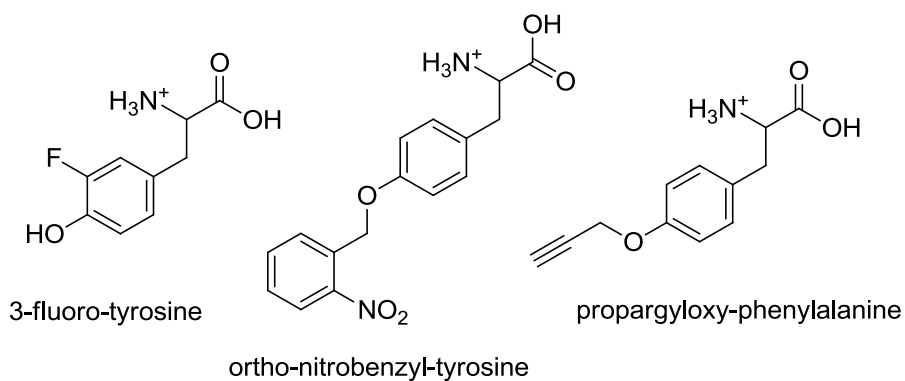
Figure 1.4: The structures 3-fluoro-tyrosine, ortho-nitrobenzyl-tyrosine, and propargyloxy-phenylalanine

B. UAA Protein Incorporation

There are a number of methods that can be used to incorporate the UAAs into proteins, each with their own advantages and drawbacks. New methods are continually being developed to perfect the speed and applications of the protein synthesis. Early methods began on a purely chemical basis (solid-phase synthesis), yet newer methods utilize cellular mechanisms in *E. coli*, yeast, and mammalian cells, to express proteins containing UAAs *in vivo.* This section focuses exclusively on the incorporation of the UAA into protein. The synthesis itself of the UAA is commonly done via common organic synthesis in the laboratory, and will thus not be covered.

B.1 Solid Phase Protein Synthesis (SPPS)

Solid Phase Protein Synthesis (SPPS) is primarily a chemical approach to the total synthesis of a protein. Each amino acid is individually linked onto the growing peptide chain like placing beads on a chain. A resin serves as the solid base/protecting group for the amino acids to attach to a linker which can easily be cleaved off at the end of peptide synthesis. To eliminate cross-reactions, the employed amino acids are modified with protecting groups like tert-butyloxycarbony (BOC) or fluorenylmethyloxycarbonyl (FMOC) groups which can easily be cleaved using a weak base or acid once the amino acid is attached to the chain.[7] The general process of SPPS is a cycle of deprotections, washes and couplings until the peptide chain is complete (Figure 1.5).[7] Once the polypeptide chain is complete, it can be cleaved off the resin and then purified. The duration of this process (purification included) varies with the length of the desired peptide: a peptide under ten amino acids takes approximately a day, while a forty amino

acid peptide takes a week (~80 hours).[8] Peptide synthesizers are also available to use, and

give comparable yield to those synthesized by hand.
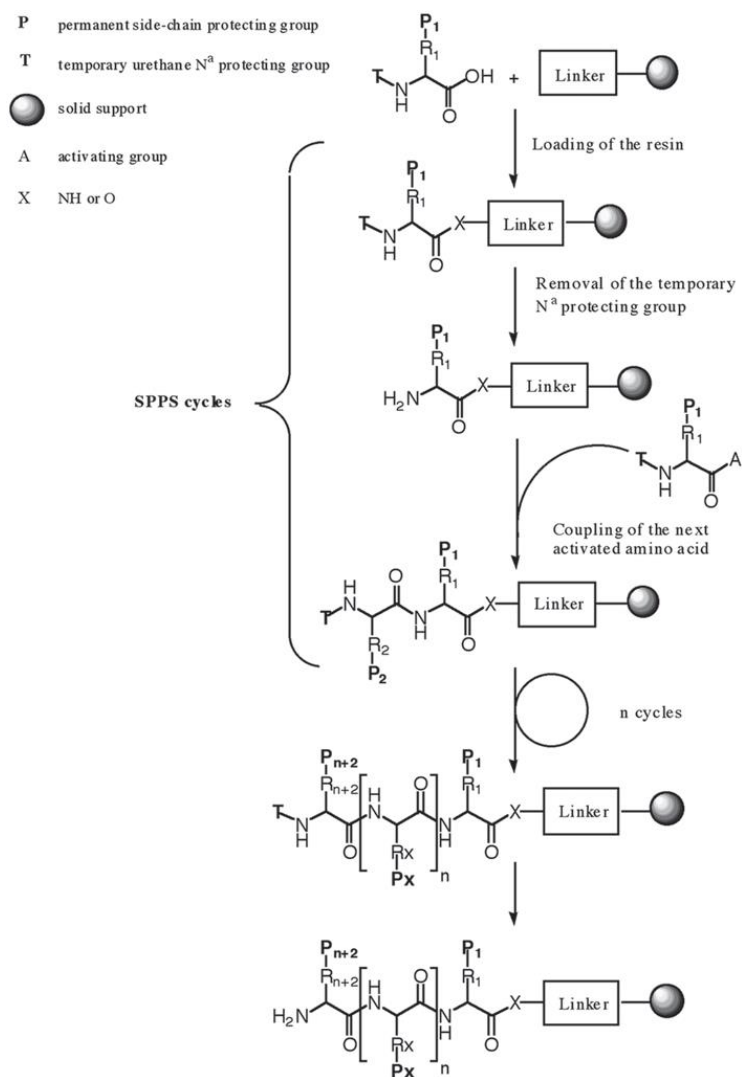
Figure 1.5: Overall scheme of SPPS.[7]

SPPS has been used to commercially produce small polypeptides such as insulin-like growth factors and human immunodeficiency virus 1 (HIV-1) protease which can then be used to determine the crystal structure of these proteins.[9] Due to the bottoms-up approach of this synthesis, UAAs can easily be added in either as a single unit of the chain, or as an entire synthetic strand. However, the major flaw of this method is the

length of the peptide that can be produced is limited to under 50-100 amino acids due to poor yields at increased length resulting from protein aggregation.[10] As many proteins extend beyond 100 amino acids (the largest protein Titin is 3–3.7 MDa), this method is not adequate to construct all proteins of interest.[11]

B.2 Global Incorporation

For larger proteins than what can be synthesized through SPPS, global incorporation of UAAs is a viable option. Global incorporation of a UAA involves the replacement of the canonical amino acid analog with UAAs throughout the protein of interest. This is done through starving the *E. coli* of the amino acid that most resembles the desired unnatural, while providing the desired amino acid analog (such as tyrosine for a fluorotyrosine incorporation).[12] This accomplishes a global incorporation of the UAA in multiple positions in the protein. As the van der Waals radius of fluorine is only 0.15Å larger than hydrogen, this is easily done for fluorinated analogues.[12] A case where multiple additions are favorable is ^{19}F NMR which has high sensitivity with minimal background signal.[2] However, global incorporation is not conducive to exponential cell growth as often times the analog is toxic to cells.[3,9] An additional disadvantage is that this method does not afford site-specific incorporation (in which the UAA is added at specific locations in the protein). Site-specific incorporation is important in precise adjustment of protein function, such as modifications made to the active site that would otherwise not be possible with global incorporation.

B.3 Orthogonal Aminoacyl tRNA Synthetases (aaRS)

Transcription is the biological process that transcribes DNA (the genetic material which encodes the information for proteins) to mRNA (the messenger material of this

code). The mRNA is then translated into protein through a process known as translation. Each set of three nucleotides (called a codon) encodes a specific amino acid. The tRNA contains an anti-codon and delivers the associated amino acid to the ribosome where it is attached to the growing polypeptide chain. The tRNA is charged with the respective amino acids by an aminoacylation catalyzed by aminoacyl-tRNA synthetases (aaRS) (Figure 1.6).
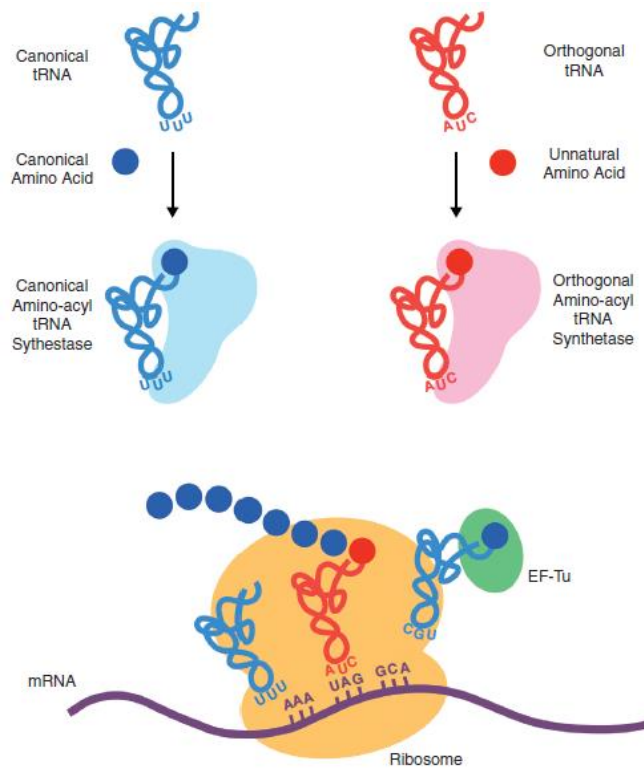
**Figure 1.6:** Normal translation with a canonical aaRS/tRNA pair is shown on the left where the tRNA is charged by the coordinate aaRS with the appropriate amino acid. The anticodon of the tRNA base-pairs with the codon to allow the insertion of the amino acid into the growing polypeptide chain.   On the right, the orthogonal aaRS/tRNA pair responds to the mutated codon to allow site-specific insertion of the UAA.[3]

Translation provides the main mechanism for site-specific UAA insertion into proteins through the modification of the aaRS and tRNA.[9,13] As the UAAs do not naturally have a tRNA that correlates for a codon, or an aaRS to attach it to the tRNA, a method of introducing these translational tools must be accomplished. There are three stop codons which do not have a specific tRNA, but rather encode a translational stop. One of these is the amber nonsense codon which is encoded by the trinucleotide sequence UAG. This codon can be re-purposed to incorporate UAAs in response to its presence in the mRNA transcript. One method to incorporate a UAA is to utilize chemically pre-aminoacylated tRNAs that recognize the UAG codon; however this mechanism is not a sustainable system due to low yields, difficult purification, and continual need for pre-charged tRNAs during protein synthesis.[14]

An orthogonal aaRS that can attach UAAs to the suppressor tRNA is desirable and more common approach to UAA incorporation. Orthogonal refers to the fact that the new aaRS does not cross-react with any of the other endogenous tRNAs and the tRNA does not react with any endogenous aaRSs.[9] An orthogonal aaRS and suppressor tRNA can be derived from phylogenetically different organisms from the host (like the archaea *Methanocaldococcus jannaschii*) to recognize this codon and the UAA of interest, while not recognizing any of the endogenous set.[3] The development of the aaRS/tRNA pair is accomplished via directed evolution utilizing two phases of selection: (1) positive selection to select mutants that incorporate any amino acid, and (2) negative selections to select against mutants that incorporate endogenous host amino acids (Figure 1.7).[3] For this process in each phase, a library of aaRS mutants is transformed into *E. coli* with the orthogonal tRNA with a TAG mutation in a selected gene. For the positive selection, the

9

gene containing the TAG encodes chloramphenicol resistance, such that when the cells are grown on a plate containing chloramphenicol and the UAA of interest, only those that have successfully suppressed the amber codon through insertion of the UAA will be able to survive. These colonies are then picked and the successful aaRS plasmids are subjected to the negative step as the colonies could have incorporated a canonical amino acid to survive. In the negative selection, the selection plasmid contains stop codons in a gene for barnase, a bacterial toxin, and the cells are grown without the UAA. If the amber codon is now suppressed, through incorporation of a canonical amino acid, the cell will die, whereas if they only recognize the UAA they do not produce barnase since no UAA is present in the selection. This selection process goes for 2.5 rounds and ultimately ends on a positive step to provide an aaRS/tRNA pair capable of incorporating the UAA.
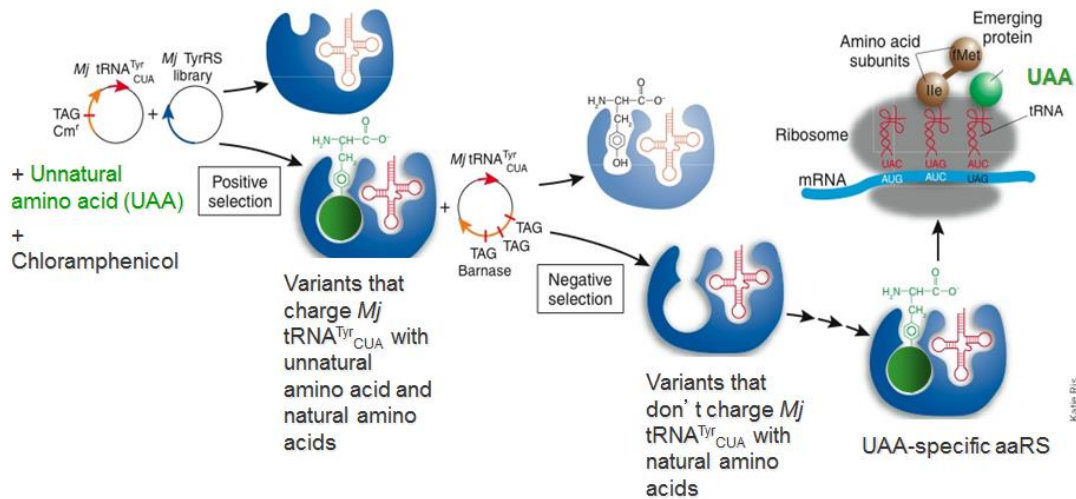
Figure 1.7: The selection process of the *Mj*aaRS/*Mj*tRNA pair (derived from *M. jannaschii*) to develop a aaRS that will recognize the UAA of interest. This cycle of selection will be completed 2.5 times, ultimately ending on a positive step, to develop a orthogonal synthetase for the UAA.

In this way, wherever the amber codon is present it is suppressed and the UAA will be inserted in its place. This also presents a mechanism for quality control, as the proteins that do not successfully incorporate the UAA will terminate in response to the stop codon. This implies that proteins that do not have the UAA will be significantly shorter and lack the functionality of the wild type protein, providing an opportunity for purification techniques to select only the proteins with the UAA correctly inserted if the terminated peptide was not naturally degraded.

The overall method of using nonsense codons is as follows: (1) mutation of the point of interest in DNA to a nonsense/stop codon; (2) isolation of a tRNA that recognizes the nonsense codon; and (3) selection aaRS that can attach the tRNA with the desired UAA.[5] While typically each UAA requires its unique aaRS, an advantageous aspect that can develop in the selection of the orthogonal aaRS is polyspecificity. Polyspecificity is the ability to recognize multiple UAAs, which results from the absence of selective pressure for other UAAs than the one employed in the selection.[15] The development of these promiscuous synthetases is critical to the continued use and development of this technology. One such example is *p*-cyanophenylalanine specific aminoacyl-tRNA synthetase (*p*CNF-RS) which is able to successfully incorporate 18 UAAs into protein.[15] This polyspecificity is developed coincidentally in the aaRS selection process through the use of one UAAS similar in structure to many others. Additionally, incorrect selection can produce a polyspecific synthetase through a decreased number of selection rounds. This leads to more flexibility in the recognition of UAAs by the aaRS, proving desirable synthetase polyspecifity for attaching more than one UAA to the tRNA.

Through the use of a cell's own translational machinery, polyspecific orthogonal aaRS and its cognate suppressor tRNA are able to selectively insert a number of UAAs into the protein of interest for future study.[3]

II. Fluorescent Biosensors: Fluorotyrosines

A. Introduction to Fluorescent Biosensors

Fluorescent proteins occur naturally and can be utilized as biomarkers/biosensors to monitor cell activity in real-time through the interactions of the analyte of interest with the fluorescent protein.[16,17] The addition of UAAs allow an opportunity for enhanced specificity and efficiency of these fluorescent biosensors in comparison with the wild type. The chemical/physical property of the UAA is selected to generate a fluorescence change (activation or quenching) upon contact of the analyte. Fluorescent protein biosensors have been used to study intercellular pH, redox potentials, protein-protein interactions, concentrations of specific small molecules (like metals), and enzyme activity.[16] For example, an unnatural fluorescence biosensor was used to observe the phosphorylation of Crk-II (an overexpressed protein in cancer cells).[18]

There are three possible designs of a natural fluorescent protein biosensor to which a UAA can be incorporated: (1) multiple fluorescent proteins are linked by a scaffold such that once a stimuli is sensed, there is a conformational change so that the distance between the two fluorescent proteins is altered to result in a shift in fluorescence resonance energy transfer (FRET); (2) the fluorescent protein with its associated sensing domain is split into two but in the presence of the analyte the two come back together to the fused native structure and fluorescence; (3) a single fluorescent protein undergoes a conformational change as the sensing domain recognizes the analyte which leads to shift in fluorescence intensity or hue (Figure 2.1).[16]
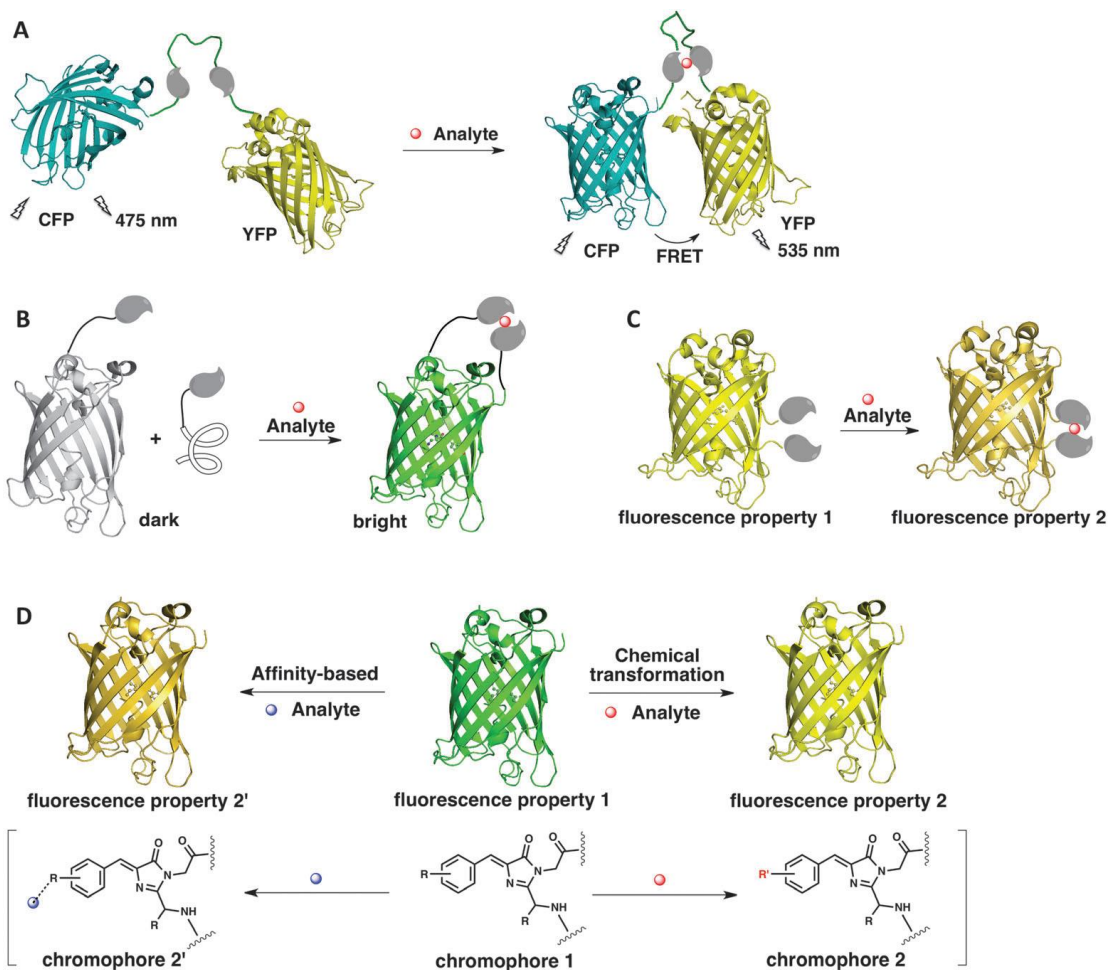
Figure 2.1: Possible designs of fluorescent biosensors. A) FRET biosensor B) bimolecular complementation biosensor C) single fluorescent protein and D) Single fluorescent protein with UAA.[16]

Green fluorescent protein (GFP) is a natural fluorescent protein that was found in deep sea jellyfish, *Aequorea victoria*. Its structure is a β-barrel surrounding an α-helix that contains the fluorophore responsible for GFP's fluorescence (Figure 2.2). GFP's fluorescence is determined by the fluorophore which is composed of a serine, dehydro-tyrosine and glycine at residues 65-67 that have cyclized to form an imidazolidone ring (Figure 2.3).[19] Gly67 is highly conserved in all fluorescent GFP mutants as it functions as

an important nucleophile with limited steric hindrance for the cyclization.[20] The fluorescence of GFP has been well characterized, which makes it a useful fluorescent probe.
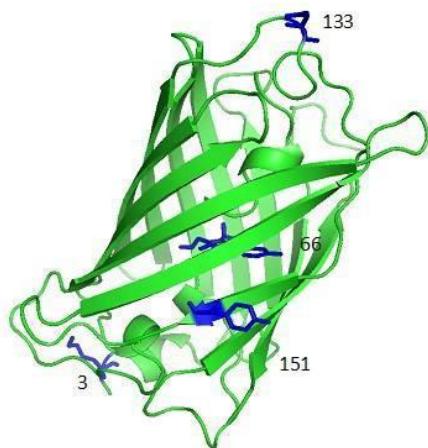
Figure 2.2: The structure of GFP is a β-barrel with an inner α-helix. The positions labeled are tyrosine residues of interest for UAA incorporation. Y66 is in the fluorophore while Y151 is on the rigid β-barrel
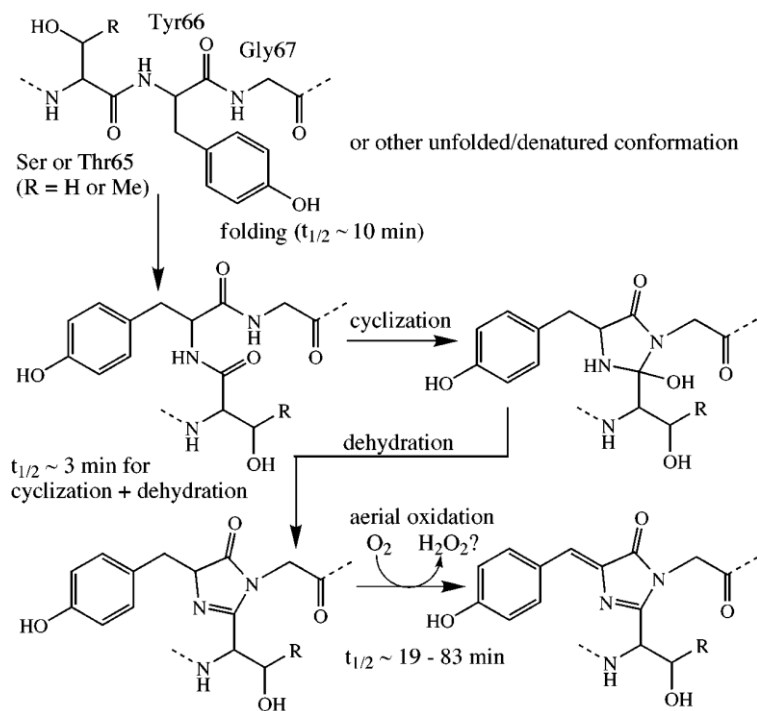
Figure 2.3: The formation of the fluorophore from the tripeptide region of GFP (Ser65, Tyr66, Gly67). The cyclization follows a series of steps including the initial folding and cyclization, enolization, and the dehydration and oxidation that finally results in the functional fluorophore.[20]

Due to the importance of the tyrosine phenol in the fluorophore, changes in protonation states result in two separate exited states of GFP. The protonated tyrosine is excited at 395 nm (emission at 512 nm) while the deprotonated state results in excitation at 475 nm (emission at 503 nm) (Figure 2.4).[19,20] A proposed mechanism for this switch between the protonation states of the buried fluorophore involves proton transfer using the hydrogen bonds of a buried water molecule and Ser205 to Glu222 while Thr203 stabilizes the phenolate oxyanion (Figure 2.5).[20] In a non-irradiated state, wild-type GFP exists in a 6:1 ratio of the neutral-to-anionic states, however this ratio shifts in favor of the anionic states after excitation due to the phenol becoming more acidic in its excited

16

state.[20] Because of this dependence on protonation of the phenol for excitation, GFP naturally displays pH sensitivity that could be modified for an enhanced range of sensitivity.
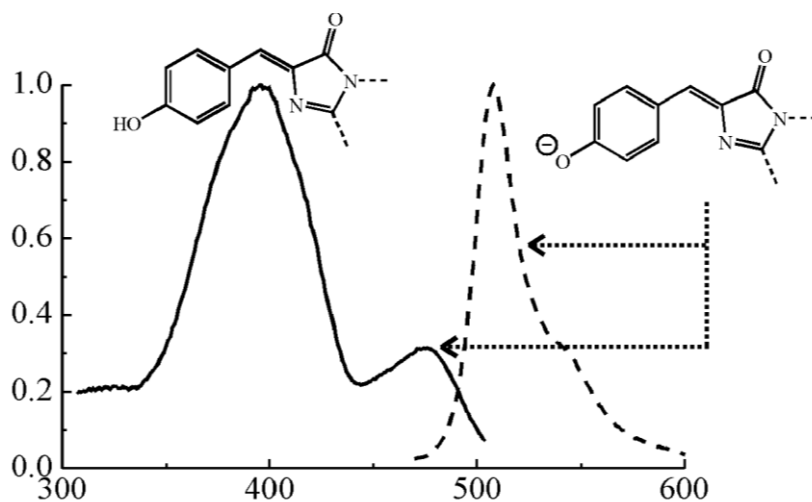
Figure 2.4: The excitation (solid line) and emission (dashed) of the two protonation states of GFP fluorophore.[20]
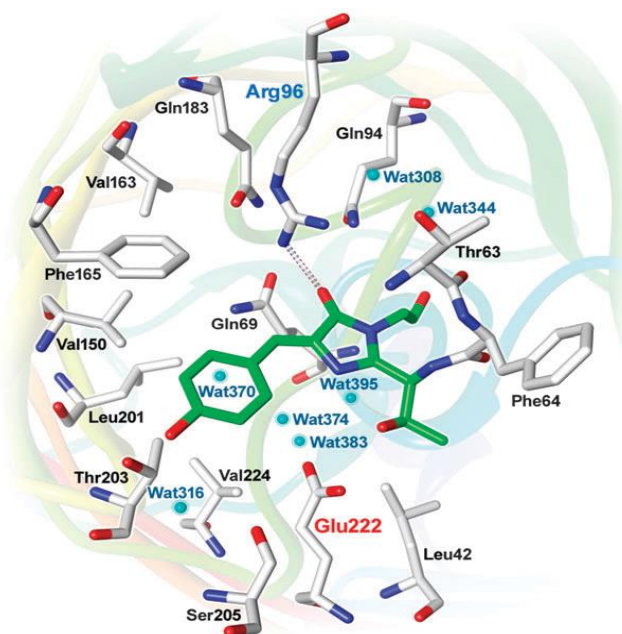
Figure 2.5: The chromophore and its environment in GFP.[19]

GFP is very commonly used as a modified fluorescent protein once it has been site-specifically engineered with UAAs. For a system to be biologically useful as a probe it is desirable that it contain the following characteristics: (1) specific to analyte, (2) display signal enhancement upon activation, (3) be orthogonal to cellular events, and (4) be genetically encoded by cells of interest.[16] All of these characteristics can easily be obtained using UAAs and GFP to study specific cellular effects with a noticeable visual correlation. This study seeks to determine the alteration in pH sensitivity of GFP as a result of UAA incorporation at multiple residues, including Y66 in the fluorophore.

B. UAA: Fluorotyrosines

The amino acid tyrosine (Y) is polar and contains a phenol ring. Fluorotyrosine is obtained through the addition of fluorine at one or multiple of the four possible positions on the phenol ring (Figure 2.6). The fluorine substitution is considered to be isosteric in that it does not perturb the spatial environment that it is placed in; however, it does have a very high electronegativity.[12] This electronegativity has the potential to alter the pK_a of the tyrosine phenol which can be useful in biological probes.
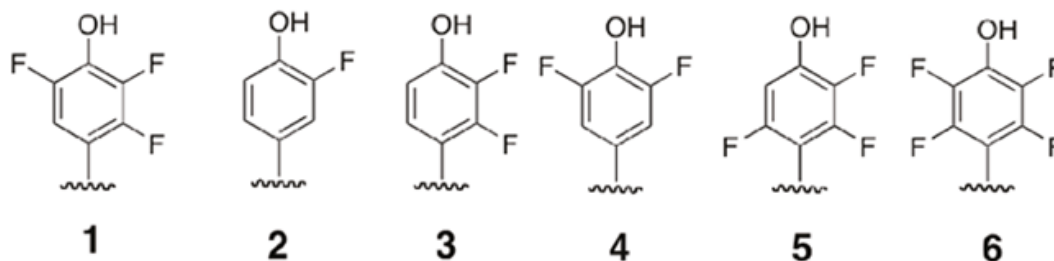
Figure 2.6: The structure of the six fluorotyrosine variants that were studied. (1) 2,3,5-F_3Y ; (2) 3-F_1Y; (3) 2,3-F_2Y; (4) 3,5-F_2Y; (5) 2,3,6-F_3Y; (6) 2,3,5,6-F_4Y.

The UAA 3-fluorotyrosine (**2**) has been used as a method to determine the structure and kinetics of proteins such as hemoglobin[21], organophosphate hydrolase[22] or human manganese superoxide dismutase[23]. Due to the electronic properties of the fluorine, it can also be used to study electron transport systems such as in photosystem II.[24] The ability for amino acids like tyrosine to form radicals is especially important in biological catalysis; fluorotyrosines provide a mechanism of studying the reactivity of tyrosine radicals.[25]

Fluorotyrosine may find application in biological probes such as GFP. As discussed previously, the tyrosine at position 66 is an important determinant in the fluorescence of the protein. Through the addition of a single fluorotyrosine residue into GFP at Tyr66, we may be able to alter the properties of the chromophore to change the protein function.[12] We hypothesize that the greater the number of fluorines that are added to Tyr66, the greater the electronegativity, and the greater the acidity, of the tyrosine phenol which will change the spectrophysical characteristics of the GFP. The changes in protonation state of tyrosine changes the chromophore which will allow it to be used as a pH sensor. The addition of fluorine to tyrosine changes the pK_a from the typical 10 of a tyrosine to 9.0-5.2 (depending on the number of fluorine additions) (Table 2.1).[12] This study explores the incorporation of a variety of fluorotyrosine into GFP to study this process.[26]

Tyrosine Derivative	pKa
Y	10
3-F$_1$Y	8.4
3,5-F$_2$Y	6.8
2,3-F$_2$Y	7.6
2,3,5-F$_3$Y	6.1
2,3,6-F$_3$Y	6.6
2,3,5,6-F$_4$Y	5.2

Table 2.1: The pKa of each fluorotyrosine derivative.

C. Methods

General

Solvents and reagents were obtained from either Sigma-Aldrich or Fisher Scientific and used without further purification. All GFP proteins were purified according to manufacturer's protocols using a Qiagen Ni-NTA Quik Spin Kit. Fluorescence was analyzed using a Perkin Elmer LS 55 Luminescence Spectrometer at excitations 300 nm, 320 nm, and 395 nm. Fluorotyrosines were custom ordered from a commercial source.

Expression of GFP-Fluorotyrosines

Escherichia coli BL21(DE3) cells were co-transformed with a pET-GFP-TAG-66 or pET-sfGFP-TAG-151 plasmid (0.5 μL) and pEVOL-3FY plasmid (0.5 μL) using an Eppendorf electroporator. Cells were then plated on LB-agar plates supplemented with ampicillin (50 mg/mL) and chloramphenicol (34 mg/mL) and grown at 37°C. After 16 h, a single colony was selected and used to inoculate LB media (4 mL) supplemented with ampicillin and chloramphenicol. The culture was grown at 37°C for 12 h. The culture was used to begin an expression culture of LB media (10 mL) at OD$_{600}$ 0.1, then incubated at 37°C, to an OD$_{600}$ of ~0.6, at which point cells were induced with 1 M IPTG (10 μL), 20% arabinose (10 μL) and 100 mM (100 μL) of respective fluorotyrosine (3FY;

2,3F$_2$Y; 3,5F$_2$Y; 2,3,5F$_3$Y; 2,3,6F$_3$Y). Cultures were grown for an additional 16 h at 37°C, then harvested by centrifugation (10 min at 10,000 rpm). The media was decanted and the cell pellet placed in the -80°C freezer for at 20 min. Purification was accomplished using commercially available Ni-NTA spin columns and according to manufacturer's protocol. Protein yield and purity was assessed by SDS-PAGE, and spectrophotometrically using a Nanodrop spectrophotometer.

Measurement of Fluorescence

Samples for fluorescence of both GFP-TAG-66 mutants and sfGFP-TAG-151 (concentration ~0.2-0.5 mg/mL) were prepared by a 1:300 dilution in PBS buffer or 1X Tris Buffer. Sample fluorescence was measured using a Perkin Elmer LS 55 Luminescence Spectrometer. Excitation was at 395nm with excitation and emission slit widths of 10 nm. Emission was recorded between 410 nm and 600 nm. Peak shifts were recorded for each fluorotyrosine variant. To explore the changes in pH sensitivity, aliquots of 2 μL 1M NaOH or 0.6M HCl were added. Acid was added if the intensity of the 512 nm peak was larger than the 450 nm peak (and the reverse for adding base). Qualitative observations of the fluorescence shifts were made after graphing emission peaks for each variant.

To more quantitatively determine the pH sensitivity, samples were given to the Harbron lab to analyze with a Varian Cary Eclipse Fluorescence Spectrophotometer. Samples were prepared by a 1:100 dilution with PBS buffer or 1X Tris Buffer with aliquots of 1M NaOH or 0.6M HCl. At each increment, the pH was measured using a Mettler Toledo microelectrode pH probe. Data analysis to test for titration curves was done using IGOR software program.

Synthesis of 2,3,5,6 F$_4$Y-A

The tetra-substituted fluorotyrosine (2,3,5,6 F$_4$Y (**6**)) was unable to be taken up by the cell and incorporated into GFP. To increase its uptake, the synthesis of a dipeptide of **6** and alanine was attempted. Initial trials to make the dipeptide were unsuccessful, so the reaction conditions were attempted with the synthesis of Tyr-Ala. Two set of dipeptide coupling conditions were attempted. One utilized the coupling agent N,N'-dicyclohexylcarbodiimide (DCC) (58.7 mgs, 0.8 eq, 0.285 mmol), however these conditions resulted in no recoverable product. A different coupling agent, 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI), was then attempted for subsequent reactions. To be able to react with the free carboxylic end of tyrosine, the carboxylic end of Boc-Ala-OH (0.25 g, 1eq, 1.32 mmol) was first methylated using NaHCO$_3$ (0.33 g, 3eq, 3.96 mmol) and CH$_3$I (0.071 mL, 1.1eq, 1.45 mmol) in 4 mL dimethylformamide (DMF). The reaction was stirred at 60ºC for 4 days (Scheme 2.1). The reaction was then extracted into ethyl acetate and water. Column chromatography was performed with 1:3 (hexanes: ethyl acetate) for 10 fractions, then 1:1 for 15 more fractions. The isolated Boc-Ala-Me was then reacted with 1% trifluoroacetic acid (1 mL) on ice for 5 min followed by stirring at room temperature for 1 hr to cleave the Boc protecting group (Scheme 2.1).

Scheme 2.1
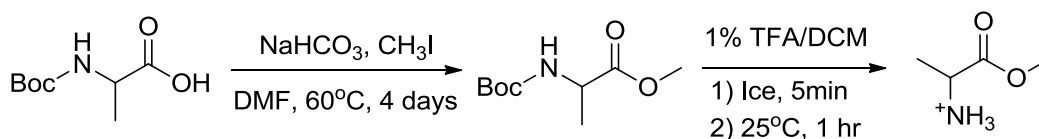
The tyrosine to react with NH$_2$-Ala-Me was obtained commercially and the methyl protecting group of Boc-Tyr-Me was cleaved with 1 M LiOH (0.5 mL) and 1M dioxane

(0.5 mL) on ice for 5 min before stirring for 1 hr at room temperature (Scheme 2.2). 1M

HCl was added dropwise until the solution reached a pH of 4. Boc-Tyr-OH was extracted

with ethyl acetate in a vial, and the organic extracted was dried with MgSO4 and

concentrated.

Scheme 2.2
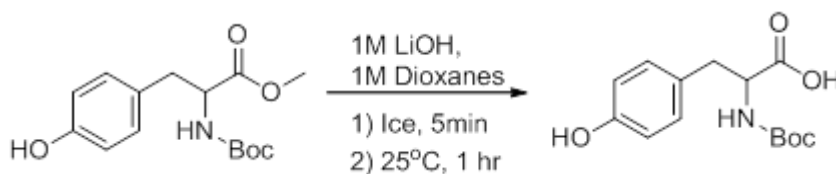
The dipeptide coupling was done between NH$_2$-Ala-Me (0.013 g, 1 eq, 0.126 mmol) and

Boc-Tyr-OH (0.035 g, 1 eq, 0.126 mmol) in a vial with hydroxybenzotriazole (HOBt)

(0.017 g, 1 eq, 0.126 mmol) in 2 mL of 5% DMF/DCM and trimethylamine (TEA)

(0.0176 mL, 1 eq, 0.126 mmol). Carefully EDCI (0.026 g, 1.05 eq, 0.132 mmol) was

added into vial and the reaction was stirred overnight at room temperature (Scheme 2.3).

The product was extracted with DCM and a column was run using 3:1 (hexanes: ethyl

acetate) for 20 fractions and 1:1 for 10 fractions. 7 mg of product was obtained in a yield

of 15%.

Scheme 2.3
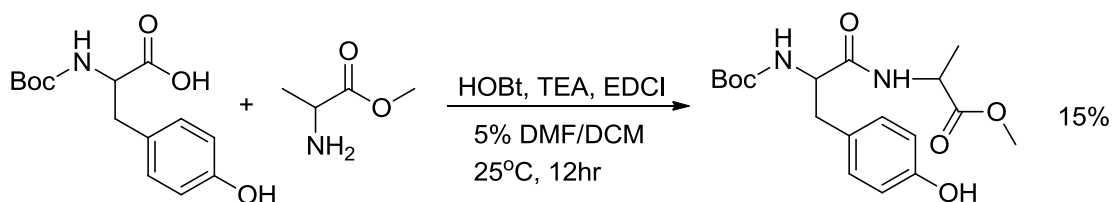
Using the procedure from above, more Boc-Ala-OH was methylated to react with the tetra-fluorinated tyrosine (Fmoc-4FTyr-OH (**6**)). Similar to the synthesis of Tyr-Ala dipeptide, **6** (0.02 g, 1 eq, 0.0603 mmol) and NH_2-Ala-Me (0.0124 g, 2 eq, 0.121 mmol) were added to a vial with HOBt (0.0081 g, 1 eq, 0.0603 mmol) and TEA (0.0084 mL, 1 eq, 0.0603 mmol) in 5% DMF/DCM. EDCI (0.0178 g, 1.05 eq, 0.0905 mmol) was added carefully to the vial which was stirred for two days at room temperature (Scheme 2.4). The product was extracted into DCM and brine. Column chromatography with 1:1 (hexanes: ethyl acetate) for 25 fractions and ethyl acetate for the final 5 fractions yielded product. A methanol flush was done after the column was performed to ensure that the product did not remain on the column. NMR data revealed residual starting material so an extraction in ethyl acetate was performed.

Deprotection of the Fmoc group was done using 10% piperidine/DCM (2 mL) for at least 12 hr, followed by a pipette column to remove the cleaved Fmoc-group (Scheme 2.4). 1% TEA was added to the column in addition to 1:3 (hexanes: ethyl acetate) for 13 fractions, followed by 1:5 for 10 fractions and 1:7 for an additional 5 fractions before a final methanol flush of the column. Final deprotection of the product was done as previously described to remove the methyl group; however, the product was in the water layer so product was concentrated rather than extracted (Scheme 2.4). 4 mgs of product was obtained in a 20% yield. The product was used later in expressions of pET-GFP-TAG66/pEvol-3FY, however protein purification did not yield successfully incorporated tetra-fluorotyrosine-alanine.

Scheme 2.4
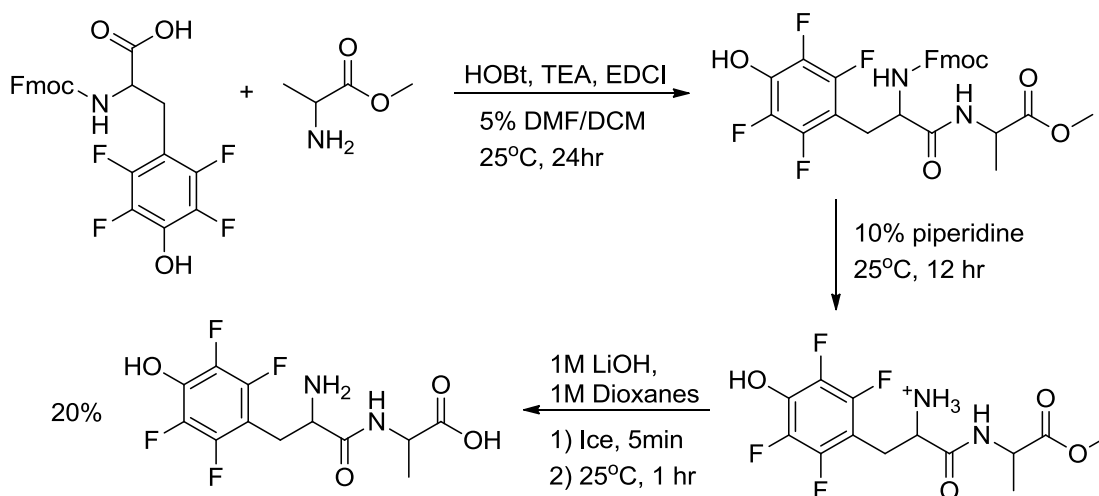
D. Results

To incorporate GFP with fluorotyrosine, *E. coli* cells were first transformed with pEVOL-3FY and either *p*ET-GFP-TAG66 or *p*ET-sfGFP-TAG151 plasmids to afford site-specific incorporation of fluorotyrosine at the two positions on GFP (Y66 in the fluorophore and Y151 on the rigid β-barrel). Superfolder GFP (sfGFP) is a GFP variant with optimized fluorescent properties (such as increased fluorescence intensity and robustness).[27] Five different fluorotyrosine UAAs (**1-5**) were incorporated into each of the positions of GFP to study. A tetra-substituted fluorotyrosine (**6**) also was attempted, however it was found to be too cell impermeable to be taken up by the cells and incorporated into protein. To attempt to resolve this, a dipeptide with alanine was unsuccessfully attempted. Future studies include the synthesis and analysis of this fluorotyrosine-alanine variant.

With the GFP-FY mutants in hand, fluorescence was measured to observe the relationship between the fluorotyrosine derivative and the spectral shifts on the GFP

fluorophore. Preliminary results using Perkin Elmer LS 55 Luminescence Spectrometer demonstrated unique spectra for each GFP-FY mutant with incorporation of fluorotyrosine at both positions Y66 and Y151. For each GFP-FY variant, fluorescence was measured at excitation wavelengths 300 nm, 320 nm, and 395 nm as determined by initial analysis based on the results of an excitation scan on the GFP. As the degree of spectral alteration from incorporation of the fluorotyrosines into GFP was unknown, these additional peaks were chosen. Additionally, higher excitations were not examined due to the possible interference of excitation with the emission of GFP-WT, which would make comparison challenging.

For the GFP-TAG66-FY mutants, the varying number of fluorines attached to the tyrosine shifted the 512 nm emission peak at all excitation wavelengths (Figures 2.7-2.9 and Table 2.2). The relationship between the incorporation of fluorotyrosines in the fluorophore, and the resulting shift in the emission peaks remains to be elucidated.
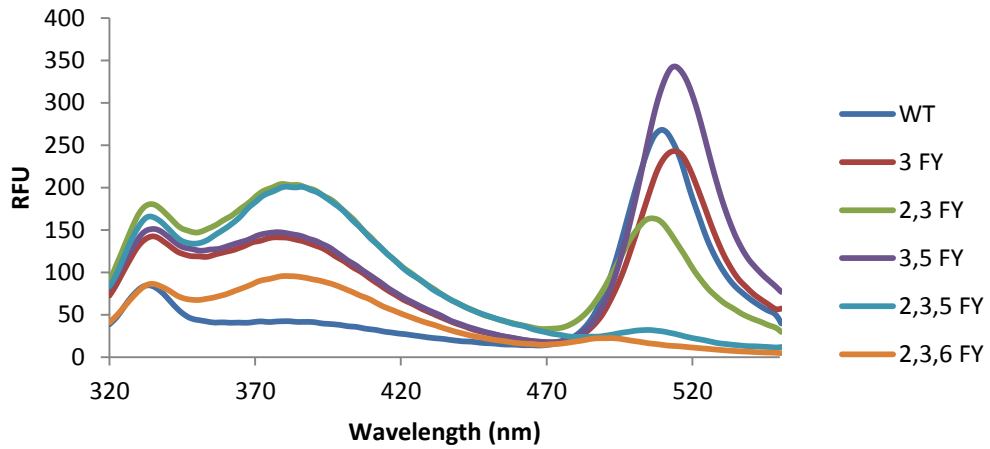
Figure 2.7: Emission spectra of GFP-TAG66-FY incorporated with each fluorotyrosine variant. Protein was excited at 300 nm.

Figure 2.8: Emission spectra of GFP-TAG66-FY incorporated with each fluorotyrosine variant. Protein was excited at 320 nm.

GFP-TAG66-FY at 395nm
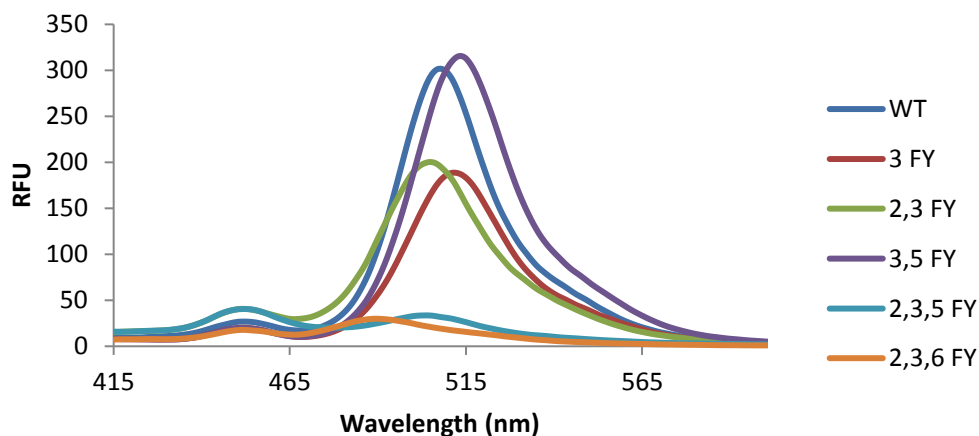
Figure 2.9: Emission spectra of GFP-TAG66-FY incorporated with each fluorotyrosine variant. Protein was excited at 395 nm.

GFP-TAG66-FY Variant	Emission Wavelength (nm)
WT	507.5
3 FY	511.5
2,3 FY	505
3,5 FY	513.5
2,3,5 FY	503.5
2,3,6 FY	489.5

Table 2.2: The emission wavelengths of each GFP-TAG66-FY variant at excitation 395 nm. Similar trends were seen at both excitations of 300 nm and 320 nm.

In contrast, the sfGFP-TAG151-FY mutants displayed no significant shifts of the 512 nm emission peak at all excitation wavelengths (Figures 2.10-2.12). The decrease in fluorescence intensity present in the spectra was potentially due to differences in protein concentration, rather than any meaningful trend. However, it could also be hypothesized

28

that the excitation at 395 nm is not the ideal excitation wavelength for the deprotonated form (475 nm). This could explain why the 450 nm emission peak for the deprotonated tyrosine is much lower than that of the 512 nm emission of the protonated tyrosine where it is being excited at the maximum excitation (395 nm). In addition the protonated and deprotonated fluorophores have different molar extinction coefficients (ε) that can also affect these differences in intensity; the protonated 395 nm excitation has ε equal to 30,000 $M^{-1}cm^{-1}$ while the deprotonated 475 nm excitation has ε of 7,000 $M^{-1}cm^{-1}$.[17] The lack of emission peak shifts in the sfGFP-TAG151-FY mutants as compared to the GFP-TAG66-FY variants could be due to the presence of the fluorotyrosine on the rigid outside of β-barrel that does not contribute directly to the fluorophore's fluorescence.

Figure 2.10: Emission spectra of sfGFP-TAG151-FY incorporated with each fluorotyrosine variant. Protein was excited at 300 nm.
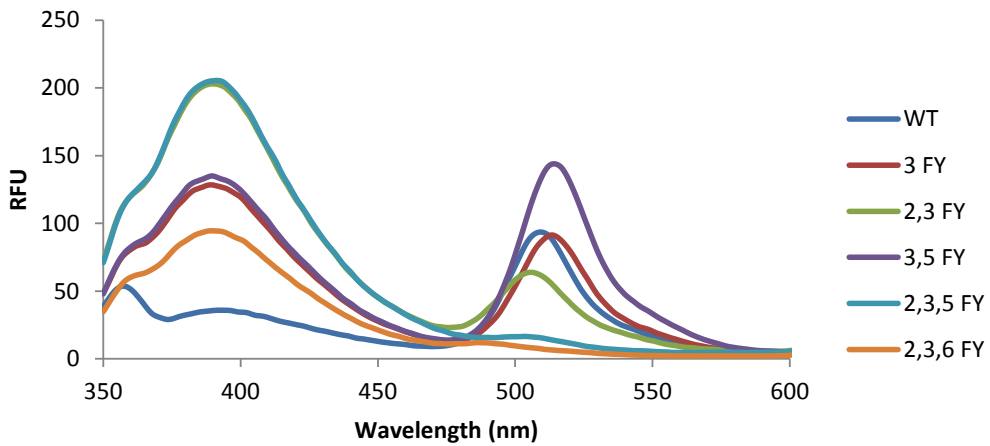
sfGFP-TAG151-FY at 320nm
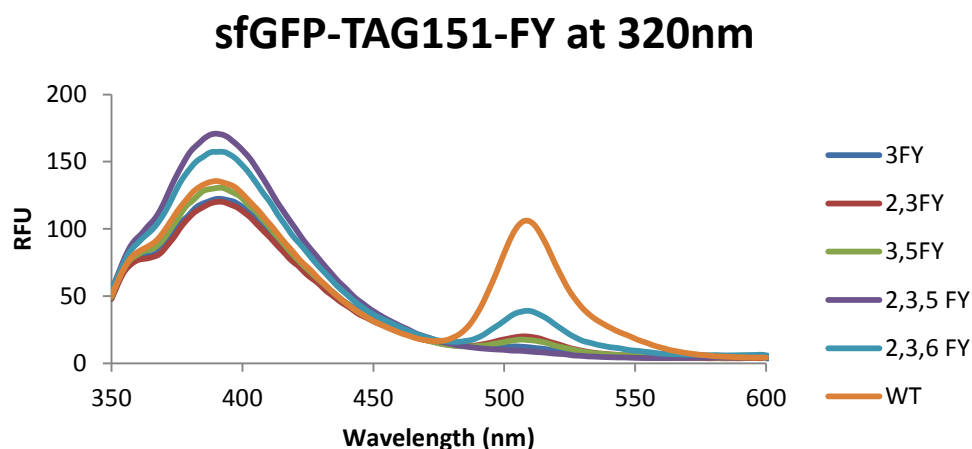
Figure 2.11: Emission spectra of sfGFP-TAG151-FY incorporated with each fluorotyrosine variant. Protein was excited at 320 nm.
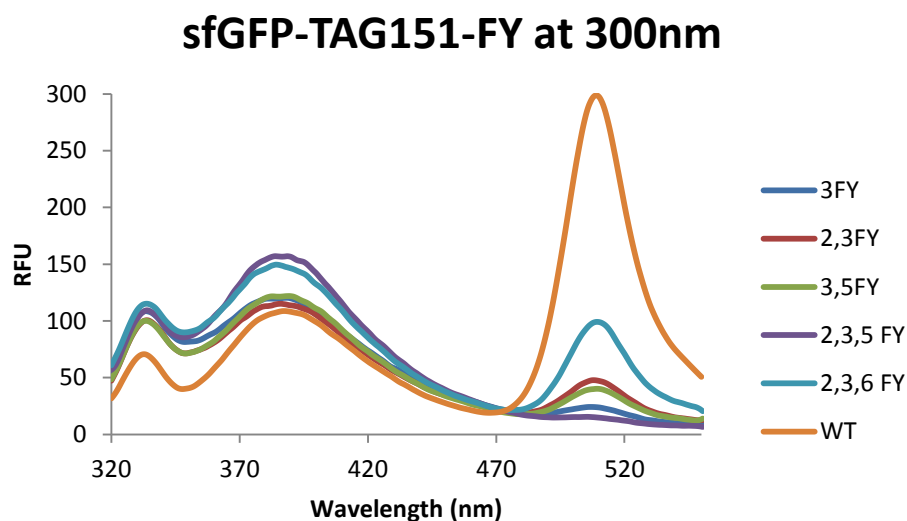
sfGFP-TAG151-FY at 395nm
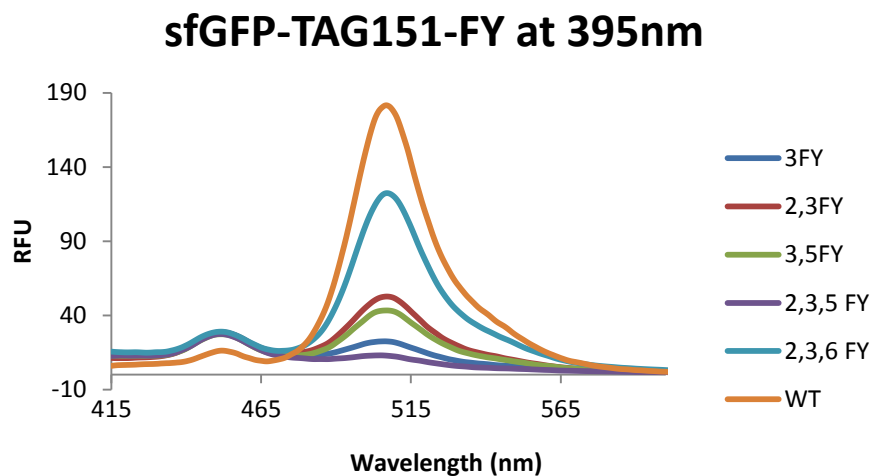
Figure 2.12: Emission spectra of sfGFP-TAG151-FY incorporated with each fluorotyrosine variant. Protein was excited at 395 nm.

The two peaks present in each spectrum at excitation of 395 nm correlate with the protonated (512 nm) and deprotonated (450 nm) forms of the tyrosine. GFP displays specific pH sensitivity (typically around pH 5-6) which can be demonstrated

spectroscopically through a shift in relative intensities of the 512 nm and 450 nm peaks. At higher pH's, the deprotonated form (emission at 450 nm) is favored, while at lower pH's the protonated is favored (emission at 512 nm).[19,20] Increasing the number of fluorine atoms, which results in a change in pKa, should result in shifted pH sensitivity for each GFP-FY variant. To determine the differences in pH sensitivity of each variant, aliquots of 1M NaOH or 0.6M HCl were added to the 1:300 dilution of protein in PBS buffer to determine the effect that the pH had on the ratio of the two peaks. While quantitative measurement of the changes in pH was not accomplished, these results demonstrated a shift in the two emission peaks in response to changes in pH.

While these results demonstrated specific shifts, the exact range of the pH sensitivity has not yet been determined. To amend this, samples were given to the Harbron lab where the experiments were repeated for each sample with the pH measured after each aliquot of acid or base. This research is still ongoing, but has showed promising correlation of different pH sensitivity with different sfGFP-TAG151-FY variants.

E. Conclusions

GFP is a highly utilized biosensor in many biological studies.[17,19] This study could result in an entire new set of GFP variants that could be used in more physiological conditions through shifting the pH sensitivity of GFP. These GFP variants would essentially become pH sensors to potentially detect changes in cellular pH. Two main applications of GFP-FY variants appear to be directly apparent for future studies. The first utilizes the differences in pKa's of each variant that are in a more physiological range. To determine changes in acidity of cells (such as cancer cells), the ratio of the two

fluorescence peaks (450 nm/ 512 nm) or the presence/absence of the 450 nm peak could be used to determine acidic conditions. A second method could utilize the shifts in emission wavelengths of the GFP-TAG66-FY variants to present a set of fluorescent probes that have the same excitation wavelength, but different emissions that could be used in a variety of studies. Overall this study offers an expanded selection of fluorescent probes to choose from for biological studies.

While research is still ongoing, preliminary results have returned evidence of distinct ranges of pH sensitivity for each GFP-FY variant. GFP-TAG66-FY variants contain unique emission spectra as compared to GFP-WT, while sfGFP-TAG151-FY variants present unique shifts in pH sensitivity. It appears surprising that the sfGFP-TAG151-FY variants respond more readily to pH changes than the GFP-TAG66-FY variants, especially considering that Y151 is on the rigid external β-barrel where shifts in tyrosine pKa should have a limited effect. It is hypothesized that the protonation state of the external β-barrel results in a conformational change to alter the environment of the internal fluorophore. However, additional research needs to be done to confirm this hypothesis.

Future studies also involve optimizing the fluorescence spectra peaks by calculating for the differences in molar absorptivity and absorption maxima of the two protonation forms. Because the deprotonated form's absorbance maximum is at 475 nm, excitation at 395 nm may not be detecting the full shifts in intensity between the two peaks. Measurement of the fluorescence spectra at excitation 475 nm could result in a more optimized fluorescence spectrum of the deprotonated tyrosine form. Moreover, a more in depth exploration of the pH sensitivity of both sfGFP-TAG151-FY and GFP-TAG66-FY

variants is needed. Eventual future studies will also explore the use of these fluorescent

probes in a biological setting to demonstrate their practical applications.

III. Caging of PRMT1 (Protein Arginine Methyltransferase 1)

A. Introduction to PRMT1

A.1 Arginine Methylation

Post-translational modifications allow an expansion of the chemical capabilities of proteins. Due to the additional properties of the appended group (electronegativity, hydrophobicity, steric hindrance, removal of hydrogen-binding etc.) these post-translational modifications can result in conformational changes of the protein, which can alter protein function. The two most common modifications are phosphorylation—the addition of a phosphate group to a protein—and methylation—the addition of a methyl group. Both modifications are important in the regulation of protein activity and signal transduction pathways. Arginine methylation has a number of physiological roles including signal transduction, mRNA splicing, transcriptional control, DNA repair, and protein translocation.[28]

Methylation is also important in the regulation of genes through the modification of histones that control the expression of DNA.[1] The DNA strands wrap around histones to form nucleosomes such that the DNA strand and histones resemble beads on a string (Figure 3.1). Where the histone binds the DNA more tightly, the DNA is unable to be transcribed. The tightness of binding is controlled through methylation, in which a methylated histone tail binds more tightly.
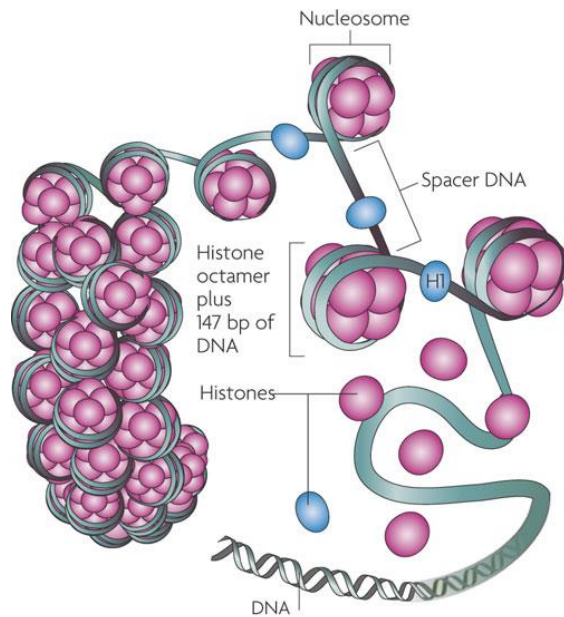
Nature Reviews | Microbiology

Figure 3.1: DNA wraps around an octamer of histones (two copies of H2A, H2B, H3 and H4) to compose a nucleosome. Histone H1 binds to the DNA in between two nucleosomes. The nucleosomes compose the basic units of chromatin.[44]

A.2 PRMT Mechanism

A particular enzyme involved in methylations is Protein Arginine Methyltransferase (PRMT), which transfers a methyl group from a methyl donor, S-Adenosyl methionine (AdoMet), to an arginine residue on the acceptor protein (Figure 3.2). Due to the chemical structure of arginine, methyltransferases can attach one or two methyl groups to the terminal guanidine nitrogen of arginine in a symmetric or asymmetric manner (Figure 3.3). There is no known enzyme that can catalyze both asymmetric and symmetric methylation.[28] As arginine's guanidine group contains five potential hydrogen bond donors, each methylation results in a removal of a hydrogen bond donor that alters the protein's shape, function, and binding interactions.[28]

Reaction catalyzed by a protein methyltransferase (PMT):
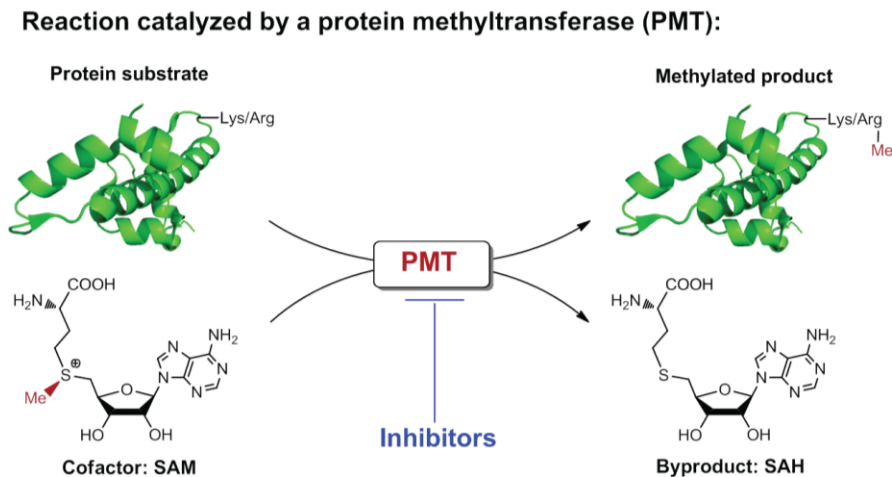
Figure 3.2: Protein methyltransferases (PMTs) can either methylate a lysine (PKMT) or arginine (PRMT) residue. Both PMTs utilize S-Adenosylmethionine (SAM) as the methyl donor. Inhibitors of this catalysis include the byproduct S-Adenisylhomocysteine (SAH).[31]

Methyltransferases that utilize AdoMet as the methyl donator have been classified into three categories based on structure and function.[29] The largest is class I that contain the methyltransferases with a common β-sheet structure to methylate DNA, RNA and proteins (this class includes PRMT).[29] Class II contain the SET lysine methyltransferases with a common SET domain that modulate gene activity; and Class III are the methyltransferases that associate with the membrane.[29] Of eleven members of the mammalian PRMT family, there are eight in humans that contain known enzymatic activity and are broken into two types based on their catalysis: Type I catalyze asymmetric methylation of arginine (including PRMT1,-2, -3, -4, -6 and -8) while Type II catalyze the symmetric methylation (including PRMT5, -7, -9 and the F-box proteins).[29] The only Type II PRMT that has been identified with certainty in humans is PRMT5.[30] Additionally the activity of PRMT2, -7, and -9 still remains to be elucidated.[28]

All of these PRMTs are extremely conserved among eukaryotes, especially in the 310 amino acid catalytic domain.[29]
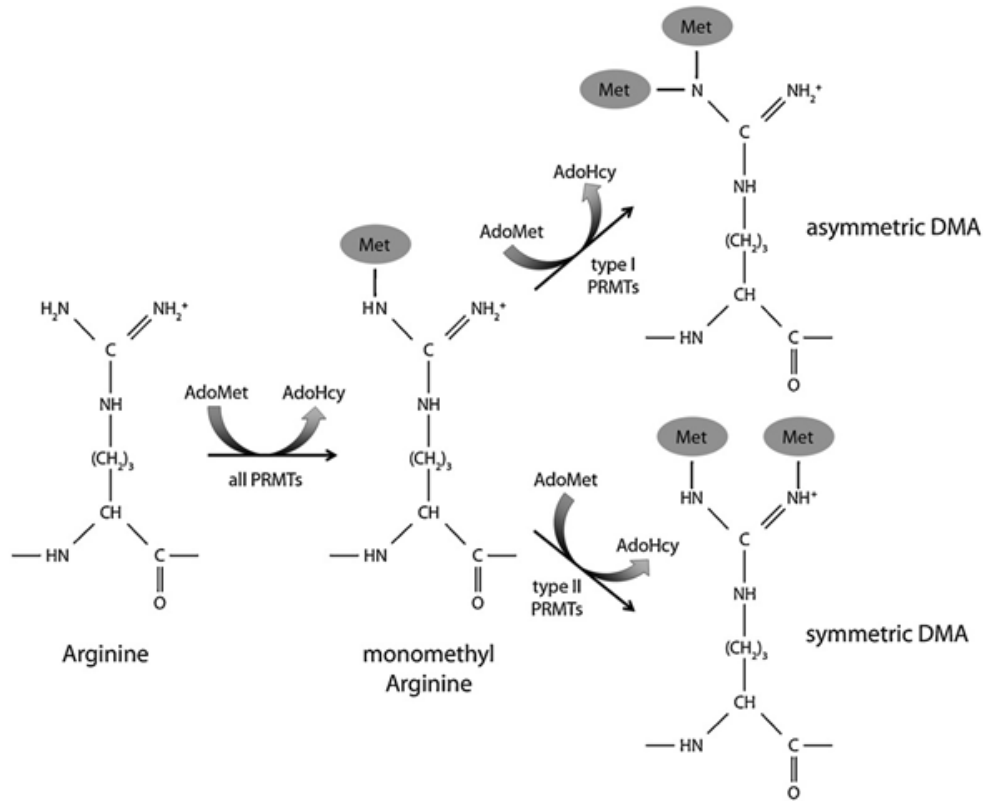
Figure 3.3: The terminal guanidine group of arginine can be methylated in three different ways by PRMTs.[29]

The activity and specificity of PRMTs remains to be largely unexplored, although it appears that over 85% of all PRMT activity is due to PRMT1.[29] PRMT1 is active as a homodimer (MW ~80 kDa) although studies indicate that other oligomers do form in the cell (Figure 3.4).[28] PRMT1 does not recognize specific proteins, rather it recognizes local amino acid sequences (including glycine and arginine rich motifs).[29] One of these common methylation sites for PRMT1 is histones, specifically histone 4 at arginine 3 (H4R3).[29] Methylation of H4R3 by PRMT1 has been linked to transcriptional activation,

while methylation of the same site by PRMT5 (that methylate arginine symmetrically rather than asymmetrically) results in transcriptional repression.[31] According to the crystal structure of PRMT1, there are a number of different possible binding sites suggesting the versatility of its methylation capacity.[30] Other PRMT1 substrates include transcription factors such as STAT1, RUNX1 and FOXO1.[31]
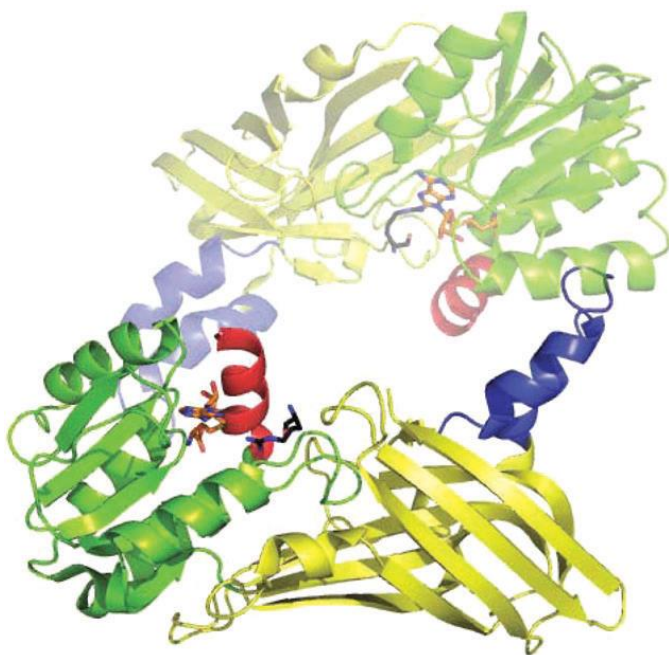
Figure 3.4: A representation of the structure of the PRMT1 dimer with bound SAH (in orange) and substrate arginine (in black). The SAH binding domain is colored green. (PDB 1OR8).[30]

PRMT1 has a high affinity for its substrate (K_m in the low/sub micromolar range), but a relatively slow rate (a k_{cat} of 1/min for a decent substrate).[30] The current proposed catalytic mechanism involves two glutamate residues of PRMT1 on the AdoMet binding domain that induce an electron redistribution through contact with the guanidino group of the arginine to be methylated which permits the attack of a nitrogen on the methyl group of AdoMet.[30] A current debate on the order of AdoMet and peptide binding has supported

the idea of a random distribution in which each methyl transfer in a demethylation is a separate event.[30] Kinetics studies have demonstrated that the methylation of H4R3 is a multiple-step process that includes 1) an ultra-fast binding of substrate (H4), 2) a moderately fast formation of the PRMT1-AdoMet-H4 complex, 3) and the rate-limiting methylation.[31]

A.3 PRMT1 in the Human Body

Arginine methylation by PRMT is not a static post-translation modification, but is a rapid modification of protein function.[29] This contributes to the ever changing activity inside the cell, and as a result, PRMT is critical to proper cell and organism growth. Mice that contain a complete loss-of-function of PRMT1 are embryonic lethal, suggesting the crucial role PRMT1 places in proper signal transduction and organismal survival.[29] Moreover, PRMT has been suggested to play a role in insulin signaling and glucose metabolism.[28] Currently, there is still a tremendous amount that remains unknown about the specificity, mechanism, and regulation of these enzymes.[28] However, improper regulation of PRMT1 has been linked to cancer, cardiovascular disease, oculopharyngeal muscular dystrophy amyotrophic lateral sclerosis (ALS), and several other diseases.[29] While many of the mechanisms remain to be elucidated, the relation of PRMT1 to cardiovascular disease has been suggested to be through the *in vivo* inhibition of nitric oxide synthase (NOS) as nitric oxide functions as a potent vasodilator.[28] Proteolysis of dimethylated protein can produce the metabolite asymmetrically dimethlated arginine (ADMA) which is a competitive inhibitor of NOS, thus inhibition of this enzyme can have major consequences on the cardiovascular system and could also result in diabetes mellitus, kidney failure, and chronic pulmonary diseases.[29]

Due to this importance in PRMT1 function and regulation, there is a large effort to develop inhibitors to be able to study the function and targets of PRMT1 and to combat these diseases due to PRMT1 disregulation. The first inhibitor of PRMT1 was identified in 1978 that was an analog of AdoMet called sinefungin (Figure 3.5).[29] Currently there are two classes of selective PRMT1 inhibitors: peptide derivatives (mostly used to study PRMT1 rather than drug candidates) and organic small molecules.[32] The small organic molecules are difficult to predict, so virtual screens of up to 300,000 compounds have been used to yield a few compounds that have promising inhibition of PRMT1.[33] However, due to the importance of PRMT1 function, most inhibitors seek to reduce PRMT1 activity back to normal levels rather than complete inhibition.[29] The most promising current inhibitors seek to specifically target the PRMT1 substrate rather than the enzyme itself to provide targeted inhibition of the specific system.[33]
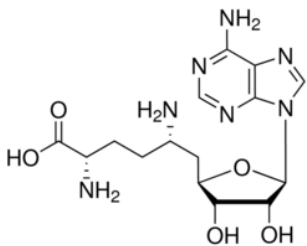

Figure 3.5: The structure of the first identified inhibitor of PRMT1, sinefungin, an analog of AdoMet.

B. UAA: ONBY

Protein function can be controlled utilizing light-removable protecting groups called "caging-groups." These protecting groups typically contain a nitrobenzyl moiety and are installed on amino acids side chains or organic small molecules in such a way that the

native function of the protein is altered (Figure 3.6).[34] Irradiation with light (usually UV) results in the photolytic cleavage of the caging-group so that the original biomolecule is restored. Caging groups have been used in a variety of studies to determine protein function and mechanisms through the directed spatiotemporal control of activity.[34] Caging groups have been used to study gene expression regulation by small molecules effectors such as a caged toyocamycin, a natural antibiotic that inhibits ribozyme function.[35] Protein activity can also be controlled through incorporation of caged amino acids into the protein. Traditional methods utilize non-specific reaction of surface lysines with a photolabile reagent. However, this method is tedious and inefficient as it involves purification of the protein of interest, non-specific caging of surface lysines, reintroduction of the caged protein into the biological system, and the caging of important residues for activity is not even guaranteed.[34] The site-specific incorporation of caged UAA's proffers a solution to target specific amino acids critical to protein function.
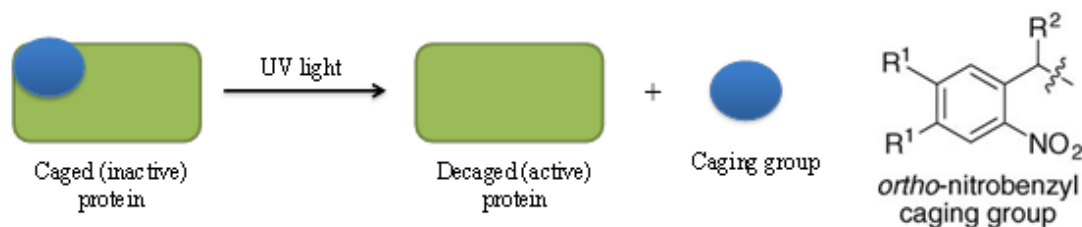
Figure 3.6: The decaging reaction in which UV light catalyzes the photocleavage of the caging group (seen here as the *ortho*-nitrobenzyl group moiety).

This study offers a novel mechanism to spatially and temporally control PRMT1 activity, without the use of external inhibitors. UAAs allow the control of PRMT1 through the targeted incorporation of caged amino acids that can permit the control of PRMT1 function. A caged-tyrosine group can be prepared through the reaction of o-

nitrobenzyl bromide with the amino acid tyrosine. This o-nitrobenzyl-tyrosine (ONBY) confers photo-reactivity, providing ultimate spatial and temporal control over protein function. Brief irradiation with UV light results in the cleavage of the nitrobenzyl group to afford the original tyrosine structure. While the nitroso-aldehyde produced is cytotoxic, it would be at such low levels to have limited effect. Light provides a unique activation of cleavage as it can be easily localized to the target, and the timing and amplitude can be easily regulated.[34] In addition, light does not involve the use of additional reagents like acids or bases that are often used in cleavage procedures.

For many proteins, the hydroxyl moiety of tyrosine is key to function, thus the caging blocks the hydroxyl from participating in its normal function. This control of protein function has been demonstrated in GFP. If the ONBY group is incorporated at Tyr66 of GFP, then the fluorescence is quenched. Only after light irradiation does the cage-group remove and fluorescence is restored (Figure 3.7).[16] This method has also been used in the regulation of Cre recombinase in which Y324 hydroxyl group is essential for catalysis.[35] Human embryonic kidney cells were transfected with the caged Cre recombinase to demonstrate more than 70% catalytic activity recovery after UV irradiation to cleave the cage group. [35,36]
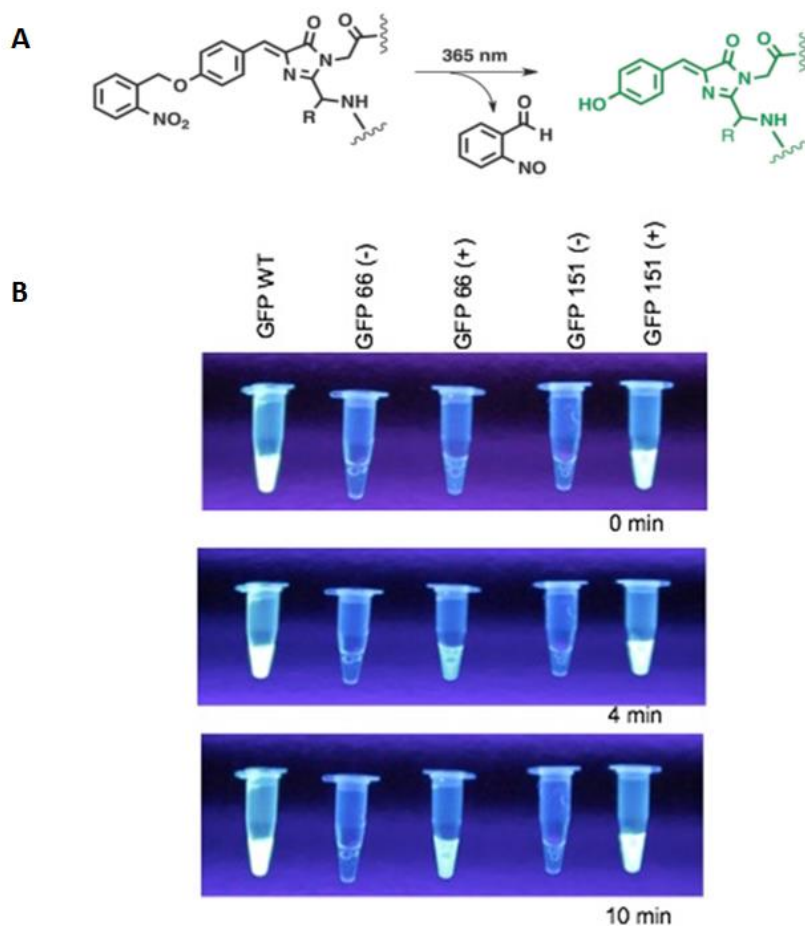
Figure 3.7: Incorporation of GFP with ortho-nitrobenzyl-tyrosine (ONBY) permits recovery of fluorescence after the photo-induced removal of the o-nitrobenzyl moiety. (A) Scheme of the decaging of ONBY in GFP fluorophore. (B) Demonstration of GFP fluorescence with ONBY incorporation at position Y66 and Y151. GFP with ONBY incorporation at Y66 recover fluorescence after only 4 min of 395 nm irradiation while Y151 fluorescence is unaffected as the fluorophore is unaffected. GFP without ONBY incorporation (-) are nonfunctional and exhibit no fluorescence.[16]

A similar mechanism is proposed for PRMT1 control. The tyrosine at position 291 has been shown to be important for protein activity as phosphorylation at this site removes enzymatic activity.[37] It was hypothesized that the site-specific incorporation of ONBY at Y291 would also result in regulated control over PRMT1 activity.

C. Methods

General

Solvents and reagents were obtained from either Sigma-Aldrich or Fisher Scientific and used without further purification. All PRMT1 proteins were purified according to manufacturer's protocols using a Qiagen Ni-NTA Quik Spin Kit. Enzyme activity was measured with a 265 nm SAM methyltransferase assay by GBiosciences. Absorbance was measured at 265 nm on a BioTek Syntergy HT microplate reader on Greiner Bio-One UV 96-well plates. Human recombinant Histone H4 was obtained from New England Biolabs.

Expression of PRMT1-ONBY

Escherichia coli BL21(DE3) cells were co-transformed with a pET-PRMT1-TAG-291 plasmid (0.5 µL) and pEVOL-ONBY plasmid (0.5 µL) using an Eppendorf electroporator. Cells were then plated on LB-agar plates supplemented with kanamycin (10 mg/mL) and chloramphenicol (34 mg/mL) and grown at 37°C. After 16 h, a single colony was selected and used to inoculate LB media (4 mL) supplemented with kanamycin and chloramphenicol. The culture was grown at 37°C for 12 h. The culture was used to begin an expression culture of LB media (100 mL) at OD_{600} 0.1, then incubated at 37°C, to an OD_{600} of ~0.6, at which point cells were induced with 1 M IPTG

(100 µL), 20% arabinose (100 µL) and 100 mM (1000 µL) of o-nitrobenzyl-tyrosine (ONBY). Cultures were grown for an additional 48 h at 25°C, then harvested by centrifugation (10 min at 10,000 rpm). The media was decanted and the cell pellet placed in the -80°C freezer for at 20 min. Purification was accomplished using commercially available Ni-NTA spin columns and according to manufacturer's protocol. Protein yield and purity was assessed by SDS-PAGE, and spectrophotometrically using a Nanodrop spectrophotometer.

Synthesis of ONBY

ONBY was synthesized according to previously described methods by Jackie McKenna. O-nitrobenzyl-bromide (61 mg, 1.5 eq, 0.513 mmol) was added to a suspension of cesium carbonate (220 mg, 2 eq, 0.677 mmol), Boc-Tyrosine-OMe (100 mg, 1 eq, 0.339 mmol), and dimethylformamide (10 mL). The reaction vial was covered with aluminum foil and stirred for 20 hours at room temperature. The reaction was extracted with dichloromethane and brine. The organic layers were dried over $MgSO_4$ and concentrated. Column chromatography (silica gel, 3:1 hexanes/ethyl acetate) afforded the desired product (57 mg, 0.132 mmol, 39%).

To deprotect the unnatural amino acid, 500 µL 1M dioxane and 500 µL 1M lithium hydroxide were added to the reaction vial on ice. The vial was covered in aluminum foil and stirred at room temperature for 2 hours. The dioxane was evaporated and 6M hydrochloric acid was added dropwise until a pH of 4 was achieved. The product was extracted with ethyl acetate in a vial, and the organic extracted was dried with $MgSO_4$ and concentrated. 1 mL of 50% trifluoroacetic acid in DCM was added to the vial on ice

and stirred for 1 hour at room temperature. The DCM was evaporated to afford the desired product (31 mg, 0.098 mmol, 74%).

PRMT1 Assay

A commercially available methyltransferase assay by GBiosciences was used to determine the relative activity of the PRMT1 mutants and wild type. A BioTek Synergy HT microplate reader was used to measure the absorbance at 265 nm. Readings were taken every minute for an hour. Either before or after initial reading, the plate was irradiated at 365 nm for 10 min to decage the ONBY group. Absorbance rates were converted to enzyme activity following the manufacture's protocol. This was done by finding the slope of the change in absorbance of adenine (a product of the enzymatic reaction provided in the kit), and subtracting the slope of the negative control. Beer-Lambert law and adenine's molar absorptivity was then used to convert changes in absorption to changes in molarity.

PRMT1 protein samples were prepared by concentrating in PBS using Corning Spin-X UF 500 concentrator columns to a concentration of ~0.400 mg/mL. Samples were made in triplicate. Substrate for PRMT1 methylation was human recombinant histone 4 (1 mg/ML) from New England Biolabs. Negative controls were made by excluding histidine or PRMT1 protein. Positive control was provided by manufactures and diluted (1:10) in provided buffer (5 μL of positive control in 45 μL of assay buffer). 14 μL of PRMT1 WT or ONBY protein was added to each well (or 14 μL of buffer for the negative control). The histone4 substrate (1 μL) was then added to give a total volume of 15 μL/ well. 5 μL of the diluted positive control was diluted again with 10 μL of buffer in the well. Following sample preparation, the enzyme master mix (supplied in the assay kit) was

46

prepared by adding 3300 µL of buffer, 100 µL SAM (lyophilized SAM dissolved in 20mM HCl) and 200 µL of the kit's enzyme mix. Immediately before the start of the assay, 100 µL of the master mix was added to each well and the plate was placed in the plate reader to immediately begin measuring absorbance.

Absorbance measurements were taken every minute for an hour at 37ºC. After an hour, the plate was irradiated with UV light (365 nm) for 10 min to decage ONBY. Absorbance measurements were then taken for an additional 20 min for every minute. A second trial was then run in which samples were irradiated at 365 nm for 10 min before the addition of the enzyme master mix. The absorbance was then measured for an hour as before. PRMT1 activity was then calculated from the changes in absorbance as described above. This assay was completed five times to verify results.

D. Results

As demonstrated by previous studies of PRMT1, interference of the hydroxyl group of Y291 should limit PRMT1 activity.[37] Consequently caging the hydroxyl group of Y291 as ONBY should eliminate all PRMT1 activity when incorporated at Y291. A commercial SAM methyltransferase assay by GBiosciences was utilized to determine the effect of the UAA insertion on PRMT1 activity. The PRMT1 protein (pET-PRMT1-TAG291) was expressed for two days at room temperature in *E. coli* with an orthogonal aaRS (pEVOL-ONBY) to permit incorporation of ONBY at Y291. Protein was purified to a concentration of about 0.400 mg/mL for analysis of activity. Each trial of the SAM methyltransferase assay was run in triplicate, with ONBY decaging occurring both before

and after an hour assay run-time. For each trial ONBY was decaged through irradiation with UV light (365 nm) for 10 min.

The assay measures the change in absorption at 265 nm, the absorption wavelength of a product, adenine, to correlate with SAM methyltransferase activity. As described in the assay scheme, PRMT1 utilizes S-Adenosylmethionine (AdoMet/SAM) as a methyl donor to transfer a methyl group to Histone 4 (Figure 3.8). Due to PRMT1's slow enzymatic turnover, the assay measures the reaction product formation rather than reactant depletions. In this case, enzymes provided in the kit catalyzed the conversion from the demethylated S-Adenosylhomocysteine (SAH) into S-Ribosylhomocysteine and adenine which removes the feedback inhibition of SAH on PRMT1. Another enzyme catalyzed the deaminiation of adenine into hypoxanthine, which is observed by a decrease in absorption. Because these last two enzymatic reactions are so rapid, the measurement of the change in absorption of adenine provides an indirect measure of PRMT1 activity.
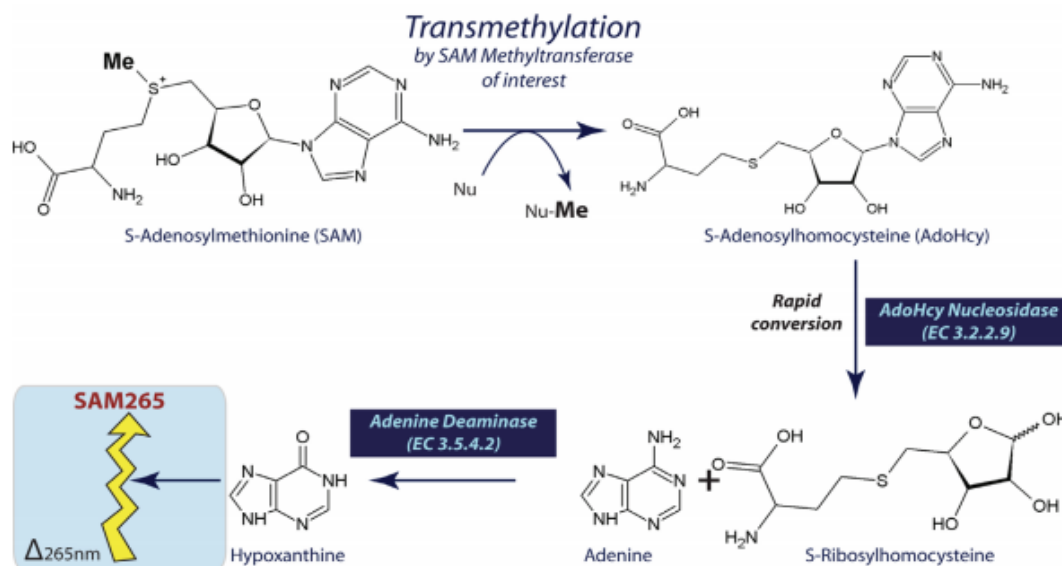
Figure 3.8: The general process of the G-Biosciences SAM 265 nm Methyltransferase Assay. Nu represents the methyltransferase substrate (in this case H4R3). Additional enzymes such as AdoHcy Nucleosidase and Adenine Deaminase were provided in the enzyme mix of the assay kit. The rapid conversion done by these enzymes aids in alleviating the feedback inhibition of SAH on PRMT1. PRMT1 activity is measured by the change in absorbance at 265 nm resulting from the conversion of adenine to hypoxanthine.

The calculation to convert the absorbance measurement of adenine over the time of the assay to the activity of PRMT1 is relatively simple. The absorbance that the plate reader gives is in milliOD at each time point. For each of the three sample trials, the absorbance at 265 nm was graphed over time and the slope was found using the best fit line. The average slope the three trials was taken, and normalized by subtracting the slope of the negative control (no substrate added). The rate, currently in milliOD/min, can be converted to A/min by simply dividing by 1000. Using the Beer-Lambert law where

A=εbc, the rate can be converted to M/min. The extinction coefficient (ε) for adenine is 13.4 mM^{-1}cm^{-1} and the pathlength (b) of the solution in the well is 0.577cm. As the protein sample was diluted by the addition of reagents, the dilution factor of (0.115mL/0.014mL) must be multiplied to get the final activity of PRMT1 in mM/min. To obtain the specific activity of PRMT1, the activity can be divided by the concentration of the protein sample. This data analysis was performed for all trials.

Five full trials of the assay were run, to ensure that the results seen were evident of the PRMT1 activity observed. Initial trials were invalid due to the failure to use a 96-well plate that could be used in the UV range. Preliminary results appeared promising as there was a significant difference in PRMT1-ONBY activity before and after UV irradiation (Figure 3.9).

However, attempts to repeat this experiment with fresh protein did not proceed as planned. A lag time between protein expression and the assay resulted in ONBY decaging before the experiment (Figure 3.10). While disappointing, this trial demonstrated the importance of properly protecting the PRMT1-ONBY protein in aluminum foil during storage as light over the course of a month may be sufficient to decage the ONBY.

Additional trials of the assay produced similar results to the first experiment. However, it was noted that the negative controls (those lacking either his4 or PRMT1-WT) had absorbance slopes similar to the positive samples. While it is still not yet understood what this means, we are attempting alternative data analysis methods and will repeat the experiment again to verify results.
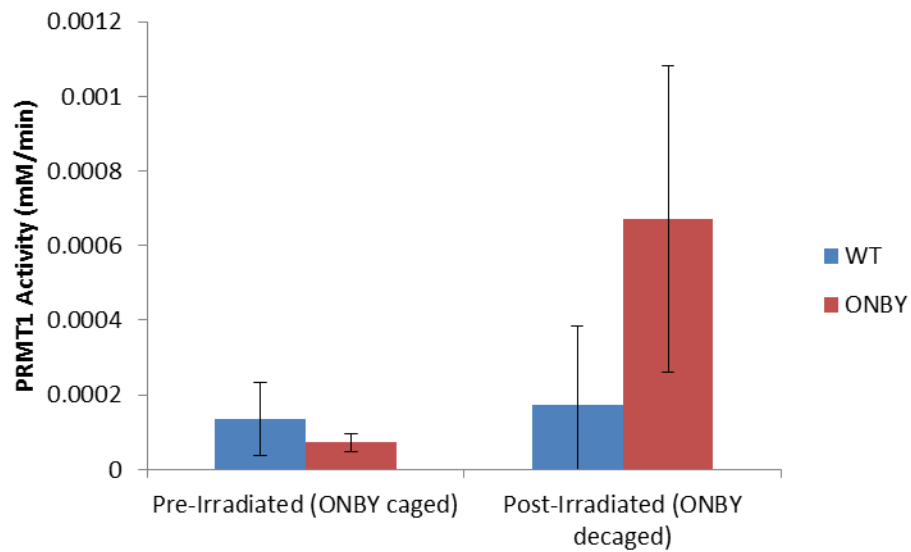
Figure 3.9: PRMT1 activity from SAM methyltransferase assay. UV irradiation results in the photocleavage of the ONBY to permit PRMT1 methyltransferase activity. PRMT1-TAG291-ONBY activity was measured pre-irradiation (PRMT1 caged) and post-irradiation (PRMT1 decaged). Error bars indicate standard deviation between the triplicate samples.
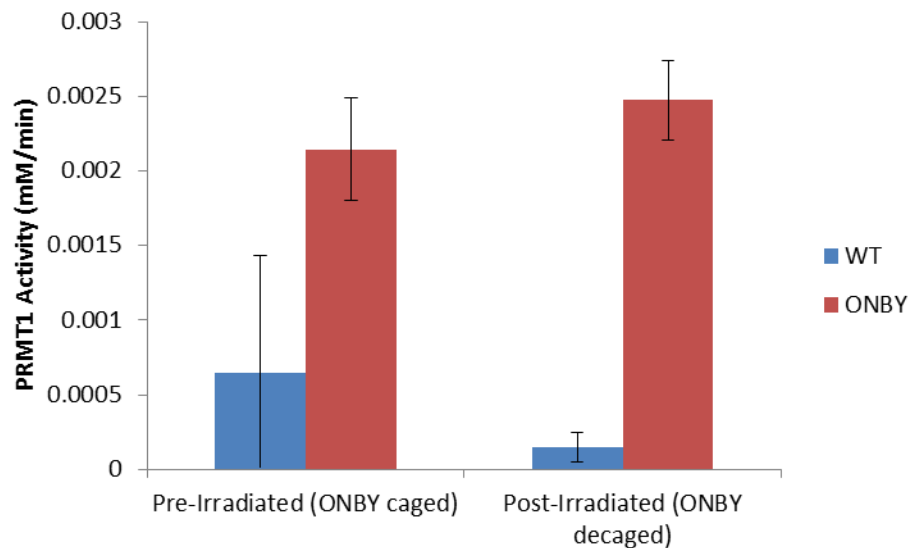
Figure 3.10: PRMT1 Activity from SAM Methyltransferase Assay. After a month in the refrigerator, the PRMT1-ONBY was decaged as seen by the similar activity to PRMT1-WT both before and after decaging. Error bars indicate standard deviation between the triplicate samples.

E. Conclusions

Caging of proteins provides precise spatial and temporal control over protein activity. The use of UAAs presents a more site-specific method than that of the previously utilized global incorporation, in which the targeting of precise sites like the active site is random and uncontrollable. These methods are indispensable in the determination of a protein's characterization. Despite the importance of PRMT1 in the human body, it has yet to be fully characterized. This study presents a method to allow a greater determination of PRMT1's function through the spatial and temporal control of its enzymatic activity.

This study demonstrated that PRMT1's activity can be controlled through the site-specific incorporation of ONBY to Y291. The fully caged PRMT1 demonstrated limited

activity in the methyltransferase assay, but its activity was recovered after the decaging of ONBY. Other caging groups could also be explored to regulate PRMT1 activity, to be able to better elucidate the mechanisms of this misunderstood enzyme. ONBY caging, while providing complete inactivation followed by complete activation after photocleavage, is irreversible. A reversible mechanism could be utilized to both turn on and turn off PRMT1 activity. The fusion of LOV domain, commonly found in plant phototropin proteins, to PRMT1 would allow reversible photochemical switching of protein activity [34]. In this system, when the LOV domain is irradiated with blue light (450-470 nm), it triggers a confirmation; change in the attached protein. Once the irradiation ceases, the protein slowly assumes its original confirmation. However, this method may be difficult to predict the optimum LOV fusion site of PRMT1 to allow photoswitching of activity.

Once this technique has been fully optimized, it would be interesting to incorporate it into a model organism to better determine the activity of PRMT1. Just as PRMT1's activity was controlled *in vitro*, an *in vivo* assay would prove useful to spatially or temporally control the enzyme's activity to determine its function at specific areas of the body or time of life.

IV. Glaser-Hay Bioconjugation

A. Introduction to Bioconjugates

A.1 Bioconjugates and Their Uses

Bioconjugation, in the simplest definition, is the attachment of two molecules (one typically biological in nature) through a covalent bond (Figure 4.1).[38] As a synthetic technique, both the molecules and the coupling mechanism can be directly chosen for the specific application. While dozens of chemical reactions are available, less than ten different reactions are typically utilized to create bioconjugates due to several key requirements for their preparation including the mild reaction conditions to produce of a physiologically stable linkage.[38] However, these same coupling reactions have been successfully utilized to create a diverse range of bioconjugates for further use, demonstrating the versatility of the bioconjugation methods. Primary uses of bioconjugates include fluorescent/biological probes to study protein activity in normal and diseased systems and development of therapeutic techniques such as anti-bodies as drug delivery systems.[38–40]

Conjugation Reaction

Conjugate A   Conjugate B
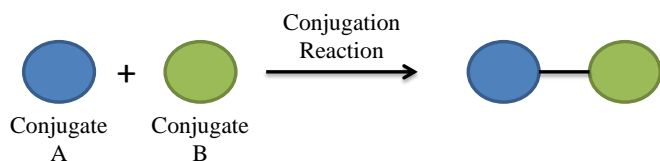
Figure 4.1: The basic process of conjugation. For bioconjugations, one of the conjugates would be biological in nature.

For study of biological systems, the coupling reactions that are utilized typically must be biorthogonal (although the coupling could also occur outside of the system). These biorthogonal reactions are chemical reactions that are able to proceed in a

biological system without perturbing or reacting with any of the endogenous components of the cell or system, and rely on chemical functionalities that are not normally present in the biological system.[39] These reactions must meet several requirements to be useful in the generation of bioconjugates including biologically compatible and mild reaction conditions, production of a physiologically stable linkage, and a degree of chemoselectivity.[38] So far, a number of biorthogonal reactions have been utilized in developing diagnostics and therapeutics including Staudinger ligation, Huisgen 1,3-dipolar Cycloaddition (click) reaction, tetrazine ligation, oxime formations, and photocrosslinking reaction.[39] One of the most commonly used reactions is the click reaction involving an alkyne and azide to yield a highly stable triazole linker (Scheme 4.1). A wide range of alternative reactions have developed from this including those in the absence of catalysts to increase its biocompatibility.[38] However, additional development of novel methodologies will permit extended versatility and maximized applications for bioconjugates.
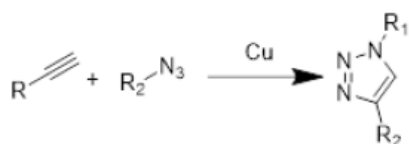
Scheme 4.1: 1,3 dipolar cycloadditions ("Click" Reaction)

A.2 Glaser-Hay Biorthogonal Reaction

Despite the number of biorthogonal reactions available, additional reactions should be assessed for further use and study. This study proposes the utilization of the Glaser-Hay coupling of terminal alkynes as a novel biochemical conjugation strategy.

The Glaser-Hay reaction provides the coupling of two terminal alkynes under a copper-iodine/TMEDA catalyst system (Scheme 4.2). This reaction affords an ideal

conjugation strategy through its formation of a highly stable and rigid carbon-carbon bond that can be formed under mild conditions and without a potentially photosensitive azide. Additionally, the reagents/catalysts are cost efficient and numerous alkyne linkers and conjugation partners are commercially available. The product formed is a highly oxidized linear diyne that is capable of numerous additional reactions. Moreover, this reaction tolerates a wide range of functional groups, which makes it applicable in the development of therapeutics.

$$2 \text{ R}\text{---}\!\!\equiv\!\!\text{---H} \xrightarrow[\text{O}_2\text{, Solvent}]{\text{CuI/TMEDA}} \text{R}\text{---}\!\!\equiv\!\!\text{---}\!\!\equiv\!\!\text{---R}$$
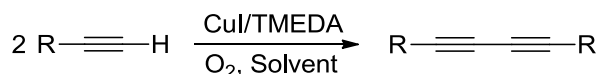
Scheme 4.2: The Glaser-Hay Reaction in which a copper/base catalyst system is utilized to produce a linear diyne product.

B. UAA: Alkyne Handle of Propargyloxyphenylalanine

Despite the usefulness of bioconjugates, they are severely limited through the lack of control in the number and location of conjugation sites. Typical methods utilize the available lysine and cysteine residues present in the protein conjugates. However, this method does not offer site-specific control over the location of the bioconjugates linkage due to the high amount of lysine and cysteine residues present in proteins. The utilization of biorthogonal handles through UAA incorporation proffers an efficient method to incorporate site-specific control over the conjugation reaction. UAAs can be developed that contain terminal alkyne functionalities that can be site-specifically incorporated into the protein of interest. The length of the handle can also be modified to extend the reach of the bioconjugation reaction to allow for optimum Glaser-Hay reaction conditions. For GFP, residue Y151 was an ideal site to incorporate a biorthogonal alkyne handle as

position 151 is on a surface exposed portion of the rigid β-barrel and does not impact GFP fluorescence.

The alkyne handle utilized in this study was propargyloxyphenylalanine due to its previously described synthesis and availability of a previously evolved aaRS (pEVOL-*p*PrF) to incorporate it into protein (Figure 4.2).[41]
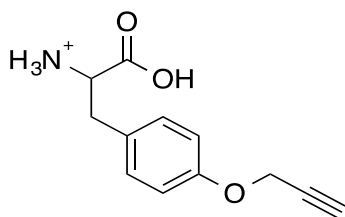
Figure 4.2: The alkyne handle UAA, propargymoxyphenylalanine, utilized in Glaser-Hay bioconjugations

C. Methods

General

Solvents and reagents, including the AlexaFluor 488 Alkyne, were obtained from either Sigma-Aldrich or Fisher Scientific and used without further purification. Epoxy-activated Sepharose 6B was obtained from GE Healthcare. Reactions were conducted under ambient atmosphere with solvents directly from the manufacturer. All GFP proteins were purified according to manufacturer's protocols using a Qiagen Ni-NTA Quik Spin Kit. Samples were analyzed on an Agilent 6520 Accurate-Mass Quadrupole-Time-of-Flight (Q-TOF) mass spectrometer equipped with an electrospray (ESI) ionization source and liquid chromatography (LC) (Agilent). Ionization settings were: positive mode; capillary voltage 3500 kV; fragmentor voltage 200 V; drying gas temperature 350 °C. Instrument was set to standard 2 GHz, extended dynamic range and deconvolution was performed by

Agilent MassHunter Qualitative Analysis software using the maximum entropy setting. To separate analyte a 2.1x150 mm, C8 reverse phase, wide pore (5 μm, 300 Å, Phenomenex) column was used with a water (A)/acetonitrile (B) (0.1% formic acid) gradient (2% B for 3 min, followed by a 2-95% B gradient over 15 min, and 95% B for 7 min).

Glaser-Hay Aqueous Conditions

Aqueous conditions for the Glaser-Hay reaction were optimized by preparing a phenylacetlyene homodimer. To a vial containing H_2O (3 mL) was added TMEDA (10 μL, 0.06 mmol) and CuI (10 mg, 0.05 mmol), forming the catalyst complex. Phenylacetylene (37 mg, 0.364 mmol) was then added and the reaction was allowed to stir at room temperature for 16 h. The reaction was extracted using EtOAc and H_2O washes (4 x 5 mL ea.), concentrated and dried *in vacuo*. The product was obtained as a white solid: 88 mg, 0.349mmol, 96% yield; ^{1}H NMR (400 MHz; CDCl$_3$): ∂ 7.38-7.56 (m, J = 7.2 Hz, 10H).

Expression of GFP-*p*PrF-151

Escherichia coli BL21(DE3) cells were co-transformed with a pET-GFP-TAG-151 plasmid (0.5 μL) and pEVOL-*p*PrF plasmid (0.5 μL) using an Eppendorf electroporator. Cells were then plated on LB-agar plates supplemented with ampicillin (50 mg/mL) and chloramphenicol (34 mg/mL) and grown at 37°C. After 16 h, a single colony was selected and used to inoculate LB media (4 mL) supplemented with ampicillin and chloramphenicol. The culture was grown at 37°C for 12 h. The culture was used to begin an expression culture of LB media (10 mL) at OD_{600} 0.1, then incubated at 37°C, to an OD_{600} of ~0.6, at which point cells were induced with 1 M IPTG (10 μL), 20% arabinose

(10 μL) and 100 mM *p*PrF (100 μL). Cultures were grown for an additional 16 h at 37°C, then harvested by centrifugation (10 min at 10,000 rpm). The media was decanted and the cell pellet placed in the -80°C freezer for at 20 min. Purification was accomplished using commercially available Ni-NTA spin columns and according to manufacturer's protocol. Protein yield and purity was assessed by SDS-PAGE, LC/MS, and spectrophotometrically using a Nanodrop spectrophotometer.

Protein-Fluorophore Glaser Hay Bioconjugation

The expressed GFP-*p*PrF-151 was coupled to AlexaFluor 488 alkyne using Glaser-Hay reaction conditions. In an eppendorf tube, 500 mM CuI (5 μL) and 5uL 500 mM TMEDA (5 μL) were mixed and equilibrated at 37°C. After 10 minutes, 1 mM AlexaFluor 488 alkyne (10 μL) was added and equilibrated at 37°C for 10 min. Finally, GFP-pPrF-151 (20 μL, 0.5 mg/mL) was added. A control reaction was also prepared with the same concentrations of fluorophore and protein, but with the catalyst system replaced with PBS buffer (10 μL). The reactions were incubated for various times at 37 °C. Reactions were then purified through centrifugal concentration on Spin-X UF colums (Corning), with wash cycles of PBS buffer (5 x 100 μL) until flow-through was free of fluorophore. The protein was then analyzed by SDS-PAGE gel to verify coupling of the fluorophore to the protein. Timecouse experiments were analyzed by comparing densiometry of fluorescent bands to their coomasie stained bands using a Biorad Molecular Imager Gel Doc XR+ system.

Immobilization of Alkynes onto Sepharose Resin

Epoxy-activated 6B Sepharose (GE Healthcare, 200 mg) was added to a filter syringe and washed with dH$_2$O (5 x 3 mL). Alkyn-ol (700 μmol) and coupling buffer (3.5 mL, pH

13.0) were added to a 15 mL falcon tube followed by the resin. The mixture was incubated at room temperature for 16 h. The resin was then transferred to a filter syringe and washed with coupling buffer (5 x 4 mL), and transferred to a 15 mL falcon tube with ethanolamine (3.5 mL). The resin was incubated at 30 $^{\circ}$C for 4 h then washed in a filter syringe with 10 mM acetate buffer (pH 4) and Ttris-HCl buffer (pH 8) for 3 cycles (4 mL ea.)

Attaching Immobilized resin to Protein

The expressed GFP-alkyne was then coupled to the immobilized resin using the Glaser-Hay reaction conditions. To eppendorf tube, 5 μL CuI (500 mM) and 5 μL TMEDA (500 mM) were combined and equilibrated at 37°C for 10 minutes. To the Eppendorf was added 30 mg of immobilized resin and again equilibrated at 37°C. After an additional 10 minutes, 20 μL of concentrated GFP-alkyne was added. A control reaction was also set up using 30 mg immobilized resin, 20 μL GFP-alkyne, and 10 μL PBS buffer. The reactions were allowed to shake at 37°C for six hours, 170rpm. After six hours, reactions were washed on filter columns using PBS buffer (200 μL aliquots x 10). The washed resin-protein mixture was transferred to an Eppendorf tube and imaged for fluorescence.

Resin-Fluorophore Glaser Hay Conjugation

The alkyne derivatized resin was reacted under previously described conditions with the AlexaFluor 488. To an eppendorf tube, 500 mM CuI (5 μL) and 500 mM TMEDA (5 μL) were mixed and equilibrated at 37 °C for 10 min. Alkyne derivatized Sepharose (30 mg) was added and equilibrated at 37 °C. After an additional 10 minutes, 1 mM AlexaFluor 488 alkyne (5 μL) was added. A control reaction was also performed using 30 mg immobilized resin, AlexaFluor 488 alkyne (5 μL), and 10 μL 100 mM PBS buffer. The

reactions were allowed to incubate at 37 °C for 6 h. The reactions were then washed on filter columns using 100 mM PBS buffer (10 x 200 µL). The fluorophore derivatized resin was transferred to an eppendorf tube and imaged for fluorescence using a BioRad Molecular Imager Gel Doc system.

Protein-Resin Glaser Hay Immobilziation

The alkyne derivatized resin was reacted under previously described conditions with the GFP-*p*PrF-151 protein. To an eppendorf tube, 500 mM CuI (5 µL) and 500 mM TMEDA (5 µL) were mixed and equilibrated at 37 °C for 10 min. Alkyne derivatized Sepharose (30 mg) was added and equilibrated at 37 °C. After an additional 10 minutes, GFP-*p*PrF-151 (5 µL, 0.5 mg/mL) was added. . A control reaction was also performed using 30 mg immobilized resin, GFP (5 µL), and 10 µL 100 mM PBS buffer. The reactions were allowed to incubate at 37 °C for 6 h. The reactions were then washed on filter columns using 100 mM PBS buffer (10 x 200 µL). The protein immobilized resin was transferred to an eppendorf tube and imaged for fluorescence using a BioRad Molecular Imager Gel Doc system.

D. Results

To assess the feasibility of employing the Glaser-Hay reaction for bioconjugation, the reaction needed to be optimized for compatibility in aqueous conditions. Gratifyingly, the proof-of-concept reactions utilizing the homodimerization of either phenylacetylene or propargyl alcohol proceeded to completion in aqueous solvent after 16 hours at room temperature with a CuI/TMEDA catalyst. These conditions were based on previously

optimized Glaser-Hay conditions in organic solvents and were able to produce greater than 95% yields.[42]

With the aqueous conditions for the Glaser-Hay reaction determined, the next step was to utilize it within the context of a protein. Propargyloxyphenylalanine was chosen as the alkynyl UAA to serve as a biorthogonal handle, as both its synthesis and an aaRS that recognizes it have been previously developed.[41] GFP was chosen as the protein for UAA incorporation due to its fluorescent properties and well-defined role as a reporter system. If the UAA were not correctly site-specifically incorporated into GFP, the nonsense codon would result in termination of the protein and GFP fluorescence would be lost. Also, the natural fluorescence of GFP allows for easy tracking of protein production and conjugation.

The first experiment to determine the feasibility of bioconjugation utilizing the Glaser-Hay reaction was done between the mutant GFP and an AlexaFluor-488 modified alkyne. Utilizing the Cu/TMEDA catalyst, the reaction was incubated at either 37°C or 4°C to determine the optimum temperature of incubation. Following reaction completion, the catalyst and excess fluorophore were removed by centrifugation and the solution exchanged into PBS. The reactions were analyzed using SDS-PAGE to compare the gel under fluorescent imaging and Coomassie blue staining. Because the GFP was denatured in the gel, the only fluorescent bands that should be seen are those of the successful bioconjugation with the fluorophore.

Initial trials employed an overnight reaction time; however, it was found that this resulted in decreased protein concentrations in the SDS-PAGE gel, suggesting protein

degradation. A time course employing slower reaction times was then attempted (from 0.5 to 12 hrs at 4ºC or 37ºC) to determine the optimum conditions (Figure 4.3-4.4)
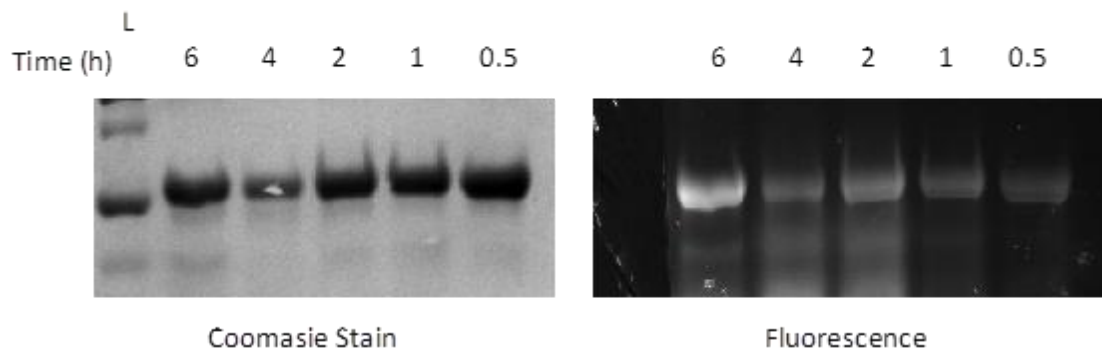
Figure 4.3: SDS-PAGE analysis of Glaser-Hay bioconjugations of GFP-pPrF-151 and the fluorophore. Reaction time varied from 0.5-6 hrs. Gels were first imaged for fluorescence, then stained with Coomassie. The densiometry of each band was measured to generate a ratio of total protein to fluorescently labeled protein.
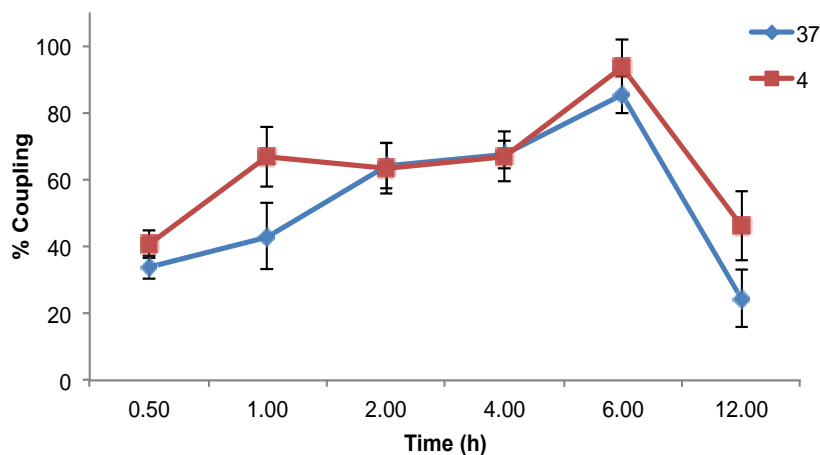
Figure 4.4: Glaser-Hay timecourse optimization. The ratio of the intact GFP to fluorescence of SDS-PAGE gel bands was taken to give a % coupling. Error bars indicate standard deviation between the three trials. Results indicate that the optimum conditions are 6 hr at 4°C. Extended reaction times lead to protein degradation.

Based on these results, 6 hours at 4°C appear to be the optimum conditions for the Glaser-Hay bioconjugation. Extended couplings led to decreased fluorescence and overall lower intact GFP upon staining, which could be a result of Cu(I) oxidative degradation of the protein over time. Indeed in a trial with a reduction of the concentration of the Cu(I) catalyst, there was more intact GFP upon staining. However, this reduced concentration in the reaction conditions yielded decreased fluorescent labeling (Figure 4.5).
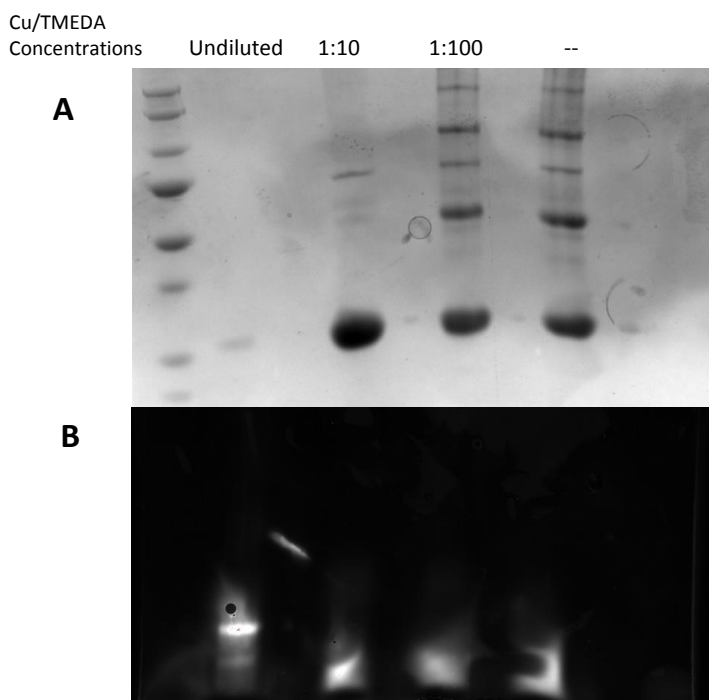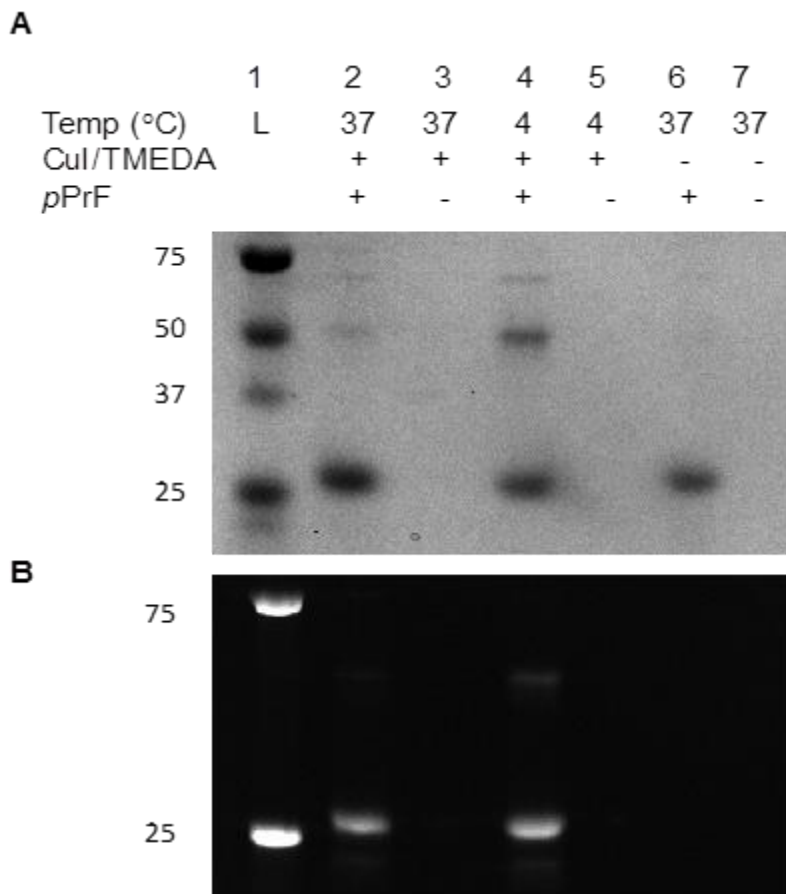

Figure 4.5: SDS-PAGE analysis of Glaser-Hay bioconjugations of GFP-pPrF-151 and fluorophore with dilutions of the catalyst system. Dilutions (1:10 or 1:100) were made to both the copper and TMEDA. (A) SDS-PAGE gel analysis after Coomassie Blue staining displays the limited protein band in the normal concentrations of catalysts. (B) SDS-PAGE gel analysis demonstrates high amount of coupling in lane 2 with no dilutions, and very limited coupling in any lanes with dilutions.

The optimal conditions were utilized to demonstrate the utility of the Glaser-Hay bioconjugation. Fluorescent labeling of the protein was observed at both temperatures, with 4 °C affording slightly higher fluorescence (Figure 4.6, Lanes 2 and 4). Control reactions lacking the Cu/TMEDA catalyst system demonstrated protein presence, but no fluorescent labeling (Figure 4.6, Lane 6). This control demonstrates that the fluorescent labeling seen with the catalyst system present was due to the Glaser-Hay coupling rather than non-covalent interactions of the protein with the fluorophore.

Figure 4.6: Glaser-Hay bioconjugation reactions for 6 hr. at 37°C and 4°C. (A) SDS-PAGE gel analysis after Coomassie Blue staining. Successful incorporation of *p*PrF demonstrated by the presence of the band in lane 6 and not in lane 7. The bands in lanes 2 and 4 demonstrate the presence of the conjugated GFP-*p*PrF-151 after the Glaser-Hay reaction completion. (B) SDS-PAGE gel analysis indicating successful conjugation of protein to fluorophore. Fluorescence of bands in lanes 2 and 4 demonstrates the successful coupling. Lanes 3 and 5 indicate no attached fluorophore.

To probe the utility of this novel bioconjugation in regards to solid-supported reactions, the Glaser-Hay reaction was used to couple the fluorophore and GFP to a solid-support. A propargyl alcohol derivatized Sepharose 6B resin was reacted under similar

conditions with the AlexaFluor-488 alkyne. A control was run in the absence of the fluorophore and the catalyst system. After several PBS washes, the fluorescence of the resin was observed, and only the reactions with both the catalyst and fluorophore resulted in visible fluorescence (Figure 4.7). From these results and the previous bioconjugation conditions, attempts were made to translate this reaction to the immobilization of GFP to a solid-support (Figure 4.8). Utilizing the GFP-pPrF-151 mutant and the optimized conditions, GFP was immobilized on both a propargyl alcohol and 1-hexynol resin (Figure 4.9). Additional controls lacking with Cu(I)/TMEDA catalyst system presented little to no detectable fluorescence, demonstrating that non-covalent interactions between protein and resin are not the cause of the increased fluorescence of the reaction with the catalyst system. From this it can be concluded that the Glaser-Hay coupling was able to successfully immobilize GFP containing an alkyne handle to a solid-support.
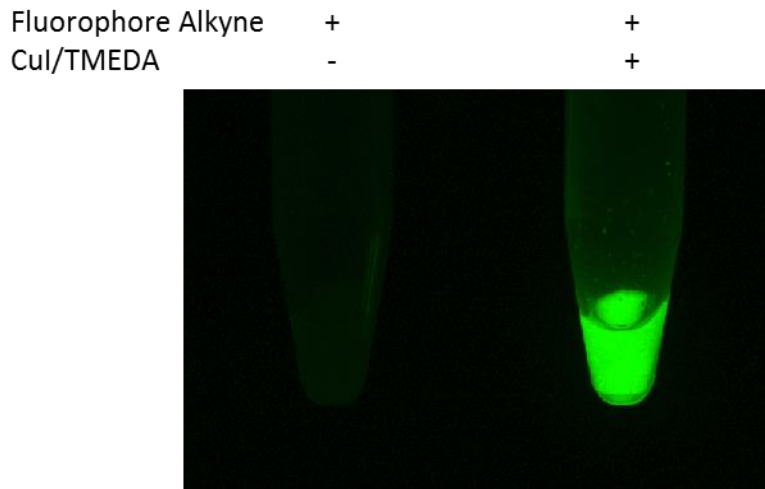
Figure 4.7: Propargyl alcohol immobilized Sepharose resin reacted with an alkynl fluorophore. Successful Glaser-Hay coupling of the fluorophore to the resin is demonstrated by the highly fluorescent resin with the catalyst (right) as compared to the resin with absence of catalyst (left) and no Glaser-Hay coupling.
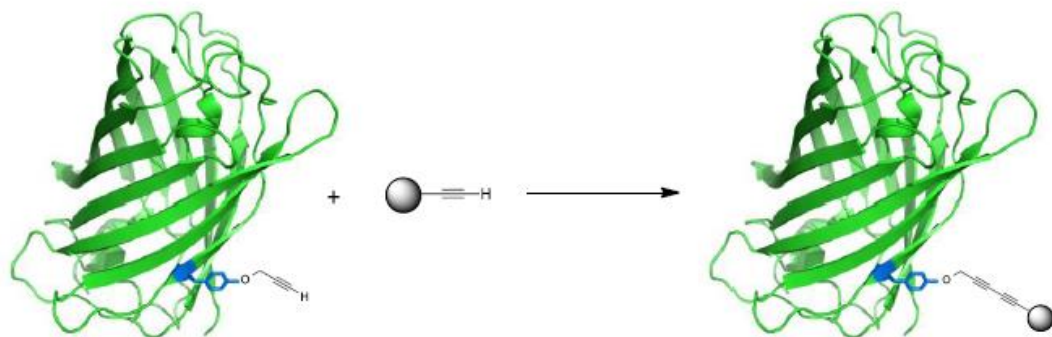
Figure 4.8: GFP with alkynyl UAA insertion at the surface-exposed position 151 undergoing the envisioned Glaser-Hay reaction to successfully immobilize GFP to a solid support.
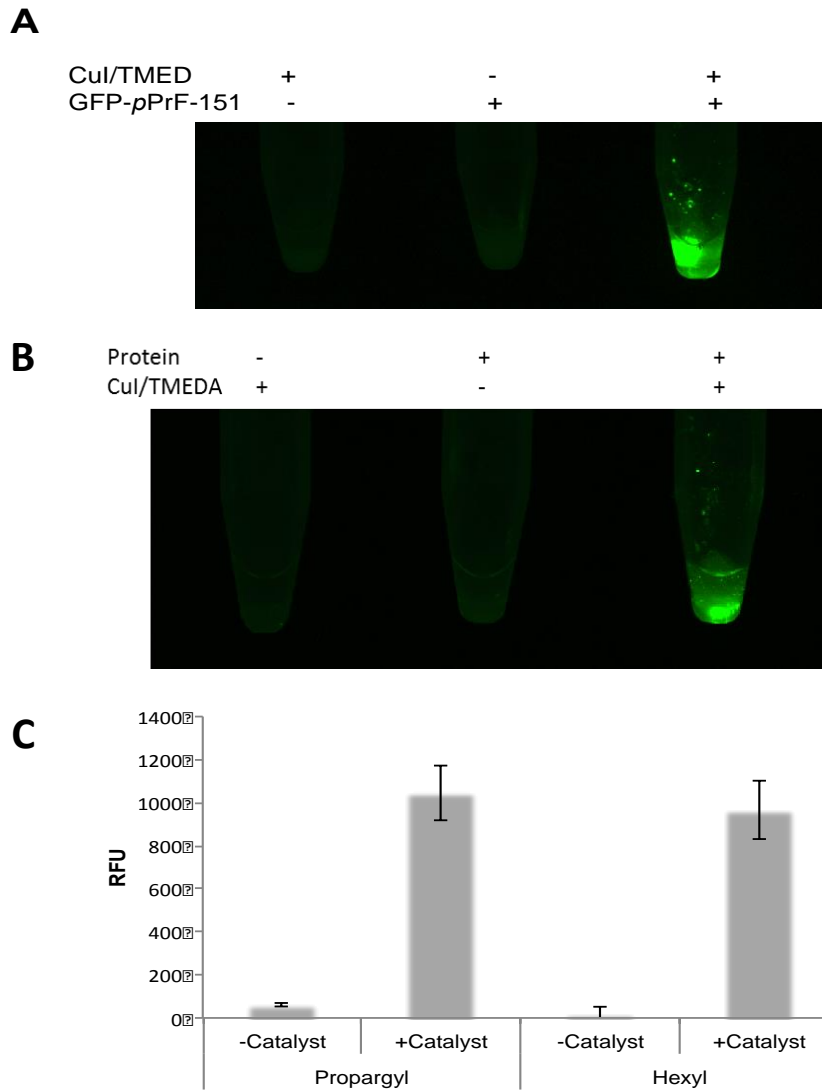
Figure 4.9: Immobilization of GFP to resin using Glaser-Hay coupling. Fluorescence is only observed in presence of both GFP and the catalyst system, indicating successful Glaser-Hay coupling to both (A) immobilized propargyl alcohol Sepharose resin and (B) the hexynol derivatized Sephorose 6B resin. (C) Fluorescence data of completed reactions with both propargyl alcohol and hexynol loaded Sepharose resins. Controls lacking catalyst system demonstrate low background fluorescence only attributed to the resin, while GFP-*p*PrF-151 reacted with the catalyst and resin indicate strong fluorescence. Error bars indicate standard deviations.

E. Conclusions

As discussed previously, bioconjugations have become widely functional in biological applications. This research demonstrates the adaptation of the Glaser-Hay reaction to an aqueous environment and a novel bioconjugation method to immobilize proteins to solid-support utilizing UAA technologies. The formation of the linear carbon-carbon bond affords a highly stable linkage utilizing mild reaction conditions and photochemically inert starting materials. Additionally, this research demonstrates the breadth of the applicability of this method through applications in small molecule dimerizations, protein-fluorophore conjugations, fluorophore-resin conjugations, and protein-resin immobilizations. As seen, this method presents a viable alternative to click conjugations and will certainly be widely utilized.[43]

Future work aims to expand the scope of the conjugations through applications as therapeutic agents. Other proteins should be tried to determine the full effect of immobilization on enzyme activity and its application. Further experiments manipulating the length of the alkyne handle are currently underway to determine its effect on reaction conditions. Overall, this novel method of bioconjugation presents a widespread field of future developments and uses.

V. **Conclusions**

The expanded amino acid library gives a host of new protein functionalities that would not be possible with just the canonical twenty amino acids. This development allows a greater specificity and efficiency in protein production to serve both the health and engineering industries. Unnatural amino acids provide an expanded application of proteins to use as therapeutics, biosensors, structural determinism, and mechanisms of unique chemistry that would otherwise be impossible through previous methods.

This study has demonstrated the utility of three amino acids in altering fluorescence to develop biological probes, spatial and temporal control over protein activity, and as linkers for protein immobilization. The site-specific alterations that UAA's can enact upon proteins permits the development of novel biochemical tools for bioimaging, enzymatic study, and bioconjugation that would otherwise be difficult to create. Future works will certainly embrace the wide range of functionality that UAA incorporation has to offer.

References

1. Allison, L. A. *Fundamental Molecular Biology*. (John Wiley & Sons, Inc., 2012).

2. Zhang, W. H., Otting, G. & Jackson, C. J. Protein engineering with unnatural amino acids. *Curr. Opin. Struct. Biol.* **23,** 581–7 (2013).

3. Kim, C. H., Axup, J. Y. & Schultz, P. G. Protein conjugation with genetically encoded unnatural amino acids. *Curr. Opin. Chem. Biol.* **17,** 412–419 (2013).

4. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).

5. Voloshchuk, N. & Montclare, J. K. Incorporation of Unnatural Amino Acids for Synthetic Biology. *Mol. Biosyst.* **6,** 65–80 (2010).

6. Maurya, I. K. *et al.* Mechanism of action of novel synthetic dodecapeptides against Candida albicans. *Biochim. Biophys. Acta* **1830,** 5193–203 (2013).

7. Amblard, M., Fehrentz, J. a, Martinez, J. & Subra, G. Methods and Protocols of modern solid phase peptide synthesis. *Mol. Biotechnol.* **33,** 239–254 (2006).

8. Coin, I., Beyermann, M. & Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat. Protoc.* **2,** 3247–3256 (2007).

9. Hendrickson, T., de Crecy-Lagard, V. & Schimmel, P. Incorporation of non-natural amino acids into proteins [Review]. *Annu. Rev. Biochem.* **73,** 147–76 (2004).

10. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35,** 225–249 (2006).

11. Opitz, C. a *et al.* Damped elastic recoil of the titin spring in myofibrils of human myocardium. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 12688–93 (2003).

12. Wilkins, B. J. *et al.* Site-specific incorporation of fluorotyrosines into proteins in Escherichia coli by photochemical disguise. *Biochemistry* **49,** 1557–9 (2010).

13. Clancy, S. & Brown, W. Translation: DNA to mRNA to protein. *Nat. Educ.* **1,** 101 (2008).

14. Wang, K., Schmied, W. H. & Chin, J. W. Reprogramming the genetic code: from triplet to quadruplet codes. *Angew. Chem. Int. Ed. Engl.* **51,** 2288–97 (2012).

15. Young, D. D. *et al.* An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50,** 1894–1900 (2011).

16. Niu, W. & Guo, J. Expanding the chemistry of fluorescent protein biosensors through genetic incorporation of unnatural amino acids. *Mol. Biosyst.* **9,** 2961–70 (2013).

17. Ward, W. W. in *Green Fluorescent Protein: Properties, Applications, and Protocols* (eds. Chalfie, M. & Kain, S. R.) 39–65 (John Wiley & Sons, Inc., 2006).

18. Cotton, G. J. & Muir, T. W. Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol.* **7,** 253–61 (2000).

19. Craggs, T. D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev.* **38,** 2865–75 (2009).

20. Tsien, R. Y. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **67,** 509–44 (1998).

21. Pond, M. P., Wenke, B. B., Preimesberger, M. R., Rice, S. L. & Lecomte, J. T. J. 3-Fluorotyrosine as a Complementary Probe of Hemoglobin Structure and Dynamics : A 19 F-NMR Study of Synechococcus sp . PCC 7002 GlbN. *Chem. Biodivers.* **9,** 1703–1717 (2012).

22. Votchitseva, Y. A., Efremenko, E. N. & Varfolomeyev, S. D. Insertion of an unnatural amino acid into the protein structure : preparation and properties of 3 f luorotyrosine containing organophosphate hydrolase. *Russ. Chem. Bull.* **55,** 369–374 (2006).

23. Ren, X. *et al.* Kinetic and structural characterization of human manganese superoxide dismutase containing 3-fluorotyrosines. *J. Mol. Struct.* **790,** 168–173 (2006).

24. Rappaport, F. *et al.* Probing the coupling between proton and electron transfer in photosystem II core complexes containing a 3-fluorotyrosine. *J. Am. Chem. Soc.* **131,** 4425–33 (2009).

25. Minnihan, E. C., Young, D. D., Schultz, P. G. & Stubbe, J. Incorporation of Fluorotyrosines into Ribonucleotide Reductase Using an Evolved, Polyspecific Aminoacyl-tRNA Synthetase. *J. Am. Chem. Soc.* **133,** 15942–15945 (2011).

26. Seyedsayamdost, M. R., Reece, S. Y., Nocera, D. G. & Stubbe, J. Mono-, di-, tri-, and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis. *J. Am. Chem. Soc.* **128,** 1569–79 (2006).

27. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).

28. Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33,** 1–13 (2009).

29. Nicholson, T. B., Chen, T. & Richard, S. The physiological and pathophysiological role of PRMT1-mediated protein arginine methylation. *Pharmacol. Res.* **60,** 466–74 (2009).

30. Wahle, E. & Moritz, B. Methylation of the nuclear poly(A)-binding protein by type I protein arginine methyltransferases - how and why. *Biol. Chem.* **394,** 1029–43 (2013).

31. Luo, M. Current chemical biology approaches to interrogate protein methyltransferases. *ACS Chem. Biol.* **7,** 443–463 (2012).

32. Hu, H. *et al.* Exploration of cyanine compounds as selective inhibitors of protein arginine methyltransferases: synthesis and biological evaluation. *J. Med. Chem.* **58,** 1228–43 (2015).

33. Wang, J. *et al.* Pharmacophore-based virtual screening and biological evaluation of small molecule inhibitors for protein arginine methylation. *J. Med. Chem.* **55,** 7978–87 (2012).

34. Riggsbee, C. W. & Deiters, A. Recent advances in the photochemical control of protein function. *Trends Biotechnol.* **28,** 468–475 (2010).

35. Deiters, A. Light activation as a method of regulating and studying gene expression. *Curr. Opin. Chem. Biol.* **13,** 678–686 (2009).

36. Edwards, W. F., Young, D. D. & Deiters, A. Light-activated Cre recombinase as a tool for the spatial and temporal control of gene function in mammalian cells. *ACS Chem. Biol.* **4,** 441–445 (2009).

37. Rust, H. L. *et al.* Using unnatural amino acid mutagenesis to probe the regulation of PRMT1. *ACS Chem. Biol.* **9,** 649–55 (2014).

38. Hermanson, G. T. *Bioconjugate Techniques*. (Academic Press, 2013). doi:10.1016/B978-0-12-382239-0.00001-7

39. Zheng, M., Zheng, L., Zhang, P., Li, J. & Zhang, Y. Development of bioorthogonal reactions and their applications in bioconjugation. *Molecules* **20,** 3190–205 (2015).

40.   Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chemie - Int. Ed.* **48,** 6974–6998 (2009).

41.   Deiters, A. & Schultz, P. G. In vivo incorporation of an alkyne into proteins in Escherichia coli. *Bioorganic Med. Chem. Lett.* **15,** 1521–1524 (2005).

42.   Tripp, V. T., Lampkowski, J. S., Tyler, R. & Young, D. D. Development of solid-supported glaser-hay couplings. *ACS Comb. Sci.* **16,** 164–7 (2014).

43.   Raliski, B. K., Howard, C. A. & Young, D. D. Site-Speci fi c Protein Immobilization Using Unnatural Amino Acids. *Bioconjug. Chem.* **25,** 1916–1920 (2014).

44.   Figueiredo, L. M., Cross, G. A. . & Janzen, C. J. Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat. Rev. Microbiol.* **7,** 504–513 (2009).