2015

Synthesis of unnatural amino acids for genetic encoding by the pyrrolysyl-tRNA/RNA synthetase system

William A. Knight
knightwa@vcu.edu

Synthesis of unnatural amino acids for genetic encoding by the pyrrolysyl-tRNA/RNA synthetase system

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University

By:
William Arthur Knight
Bachelor of Science, Virginia Polytechnic Institute and State University (Virginia Tech)
2009

Director: Thomas Ashton Cropp,
Associate Professor,
Department of Chemistry

Virginia Commonwealth University
Richmond, Virginia
April, 2015

Acknowledgment

I would like to give acknowledgment to my parents who I love very much, Bill and Denie Knight. You have always been there to support and love me as I have gone through this long process not only in school but, in life – love you both so much. My sister, Erica Knight, for her endless encouragement and kindness; you have provided so much knowledge and guidance as I have grown up over the years. Additionally, my grandparents, Ella Flick and Nancy Wells, for their years of love and inspiration to succeed; especially Ella Flick for everything you have provided me over the years. I truly could not ask for a better family! I would also like to acknowledge my academic mentor, Dr. Ashton Cropp, you have undoubtedly provided me such a fabulous opportunity and have taught me a lifetime of knowledge in the field of chemical biology. Christine, Bill, Courtney, and Megan thank you all for your help and support (especially Christine for your help in the program – could not have done it with out you). I would also like to give acknowledgment to Chris Swift who entered as a classmate and turned into a life long friend – thanks for all of your help, bud! And of course my committee members, thank you all so much for everything you have done for me!

Table of Contents

List of Figures

List of Schemes

List of Tables

List of Abbreviations

1. PTM – Post-translational modification
2. UAA – Unnatural amino acid
3. ATP – Adenosine triphosphate
4. AMP – Adenosine monophosphate
5. Boc – t-butyloxycarbonyl
6. UV – Ultra violet
7. TEV – Tobacco Etch Virus
8. nm – nanometer
9. K_{hib} – 2-hydroxyisobutyryl lysine
10. TLC – Thin-layer chromatography
11. Fmoc – fluorenylmethyloxycarbonyl
12. DCC – N,N'-dicyclohexylcarbodiimide
13. NMR – Nuclear magnetic resonance
14. LC – Liquid chromatography
15. MS – Mass spectrometry
16. Mm – *Methansarcina mazei*
17. Mb - *Methansarcina bakeri*
18. PCR- Polymerase chain reaction

19. MmPylS - *Methansarcina mazei* pyrrolysyl synthetase
20. sfGFP – Superfolder green fluorescent protein
21. His - Histidine
22. ESI – Electrospray ionization
23. CAT – Chloramphenicol acetyltransferase
24. UPP – Uracil phosphoribosyltransferase
25. tRNA – transfer Ribonucleic Acid
26. mRNA – messenger Ribonucleic Acid

Abstract

SYNTHESIS OF UNNATURAL AMINO ACIDS FOR THE GENETIC ENCODING BY THE PYRROLYSYL-TRNA/RNA SYNTHETASE SYSTEM

William Arthur Knight, M.S.

A thesis submitted in partial fulfillment of the requirement for the degree of Master of Science at Virginia Commonwealth University

Virginia Commonwealth University, 2015

Major Director: Dr. Thomas Ashton Cropp, Associate Professor, Department of Chemistry

The complexity of all biomolecules in existence today can be attributed to the variation of the amino acid repertoire. In nature, 20 canonical amino acids are translated to form these biomolecules, however, many of these amino acids have revealed post-translational modifications (i.e. acetylation, methylation) after incorporation. Amino acids that exhibit PTM are known for their involvement in cellular processes such as DNA repair and DNA replication; these PTMs are commonly found on histones within the chromatin complex. Utilization of *in vivo* site-specific incorporation has recently reported functionality of post-translationally modified amino acids.[1]

Here we report the synthesis and *in vivo* site-specific incorporation of the histone PTM, 2-hydroxyisobutyrl lysine (K_{hib}), with the pyrrolysyl tRNA/ RNA synthetase system. This translational machine can better serve to probe K_{hib} for functional benefits. Additionally, this thesis focuses much of its attention on the development of unnatural amino acids (UAA) with optogenetic characteristics. These UAAs, if site-specifically incorporated, can be used to control enzymes and proteins through rapid light perturbation (365nm UV light). Furthermore, discussed is the synthesis of photo-caged threonine and photo-caged serine as potential substrates for the pyrrolysyl translational machinery.

Chapter 1 – Introduction

1.1-Evolution and importance of unnatural amino acids

Complexity of all living organisms can be attributed to the variation of the amino acid repertoire. Each of the 20 canonical amino acids is added sequentially to a growing peptide chain when in response to a unique set of 3 mRNA nucleotides known as a codon. These peptides of amino acids are discovered in all biomolecules in existence today, providing each with functional diversity in nature. Research has identified that some of these amino acids, such as lysine, exhibit post-translational modification (PTM), thereby assisting in the orchestration of cellular processes such as DNA replication, transcription, and DNA repair.[2–6] The proteomic diversity that these PTM amino acids provide, has encouraged researchers to question how site-specific incorporation of post-translationally modified amino acids or new unnatural amino acids (UAA) could benefit in better understanding the functional complexity of biomolecules.

Common UAA groups that have been studied for protein functionality include those that are analogs (2-4) of the known PTM UAA, pyrrolysine (1),[1,7] crosslinker PTMs (i.e. O-linked B-N-acetylglucosamine),[8] or protecting groups removed by light, known as photo-caged (5).[9,10] These UAAs provide information into dynamic processes in cells and organisms. For example, photo-caged groups are useful in their ability to control protein activation through rapid light perturbation, while PTM-mimicking UAAs can be used to study the role that natural PTMs play in protein function.[10–12] The use of site-specific technology has made studying of these and many newly recognized PTMs possible.

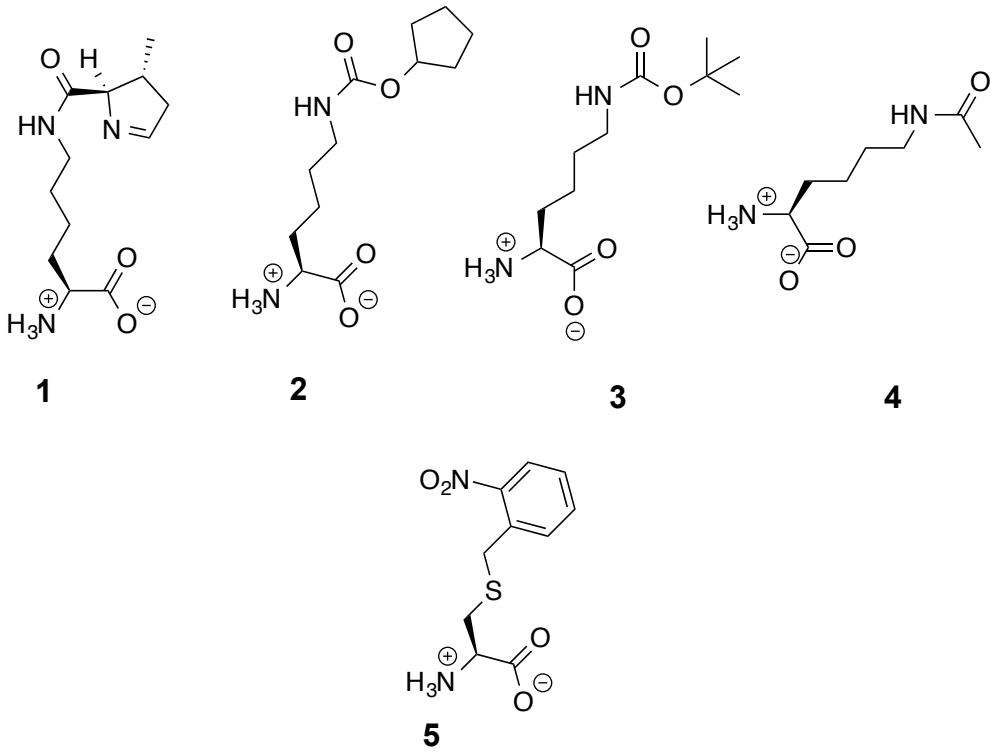


Figure 1 - Represents common UAAs studied for protein functionality

1.2-Site-specific incorporation of unnatural amino acids

Techniques that site-specifically incorporate unnatural amino acids were developed as a tool for analyzing protein function in both *in vitro* and *in vivo*.[2] *In vitro* methods used cell-free expression systems that consist of extracts from rabbit reticulocytes or *Escherichia coli*. These systems site-specifically incorporate into proteins by the addition of chemically aminoacylated suppressor tRNAs to protein synthesis reactions, which are programmed to encode a gene that includes a preferred amber stop codon (figure 2).[2,13]

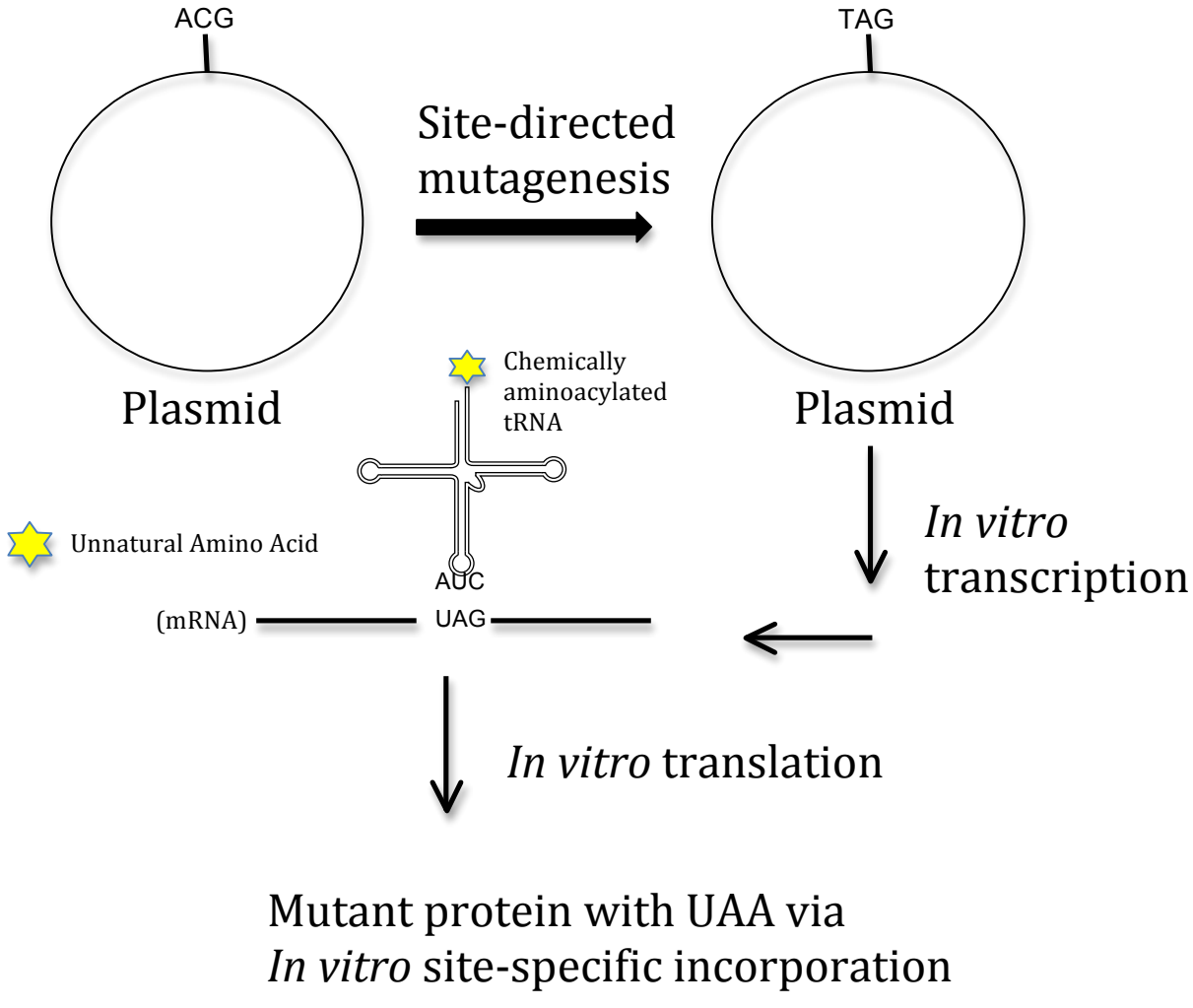


Figure 2 - *In vitro* site-specific incorporation of unnatural amino acids

Newer methods (*in vivo*) require additional biosynthetic machinery that is not recognized by the endogenous components found in the organism performing translation, however, they are still capable of encoding an UAA at a nonsense codon (figure 3). Schultz and co-workers developed this method by importing a tyrosine translational system from the archaebacterium, *Methanococcus jannaschii*.[2] They found the tRNA/RNAsynthetase system to be orthogonal in *E. coli*, meaning 1) the tRNA is not recognized by any endogenous aminoacyl-tRNA synthetases 2) the tRNA synthetase is required to aminoacylate the orthogonal tRNA and none of the endogenous tRNAs 3) the tRNA synthetase must only aminoacylate the tRNA with the desired UAA 4) the system must recognize a nonsense codon to incorporate at – tRNA amber suppressor .[2] This orthogonal system was successful in fulfilling all of these criteria, making it a perfect fit for cross-species aminoacylation. Additionally, after careful selective mutation, Schultz and co-workers developed a mutant synthetase, TyrRS, to encode O-methyl-L-tyrosine at a nonsense amber codon.[2] This ground breaking technology has led researchers to seek out other orthogonal systems that are capable of incorporating new UAAs with novel biological, chemical and physical properties (i.e. PTMs, photo-caged groups).[2,3,5,10,11,14]

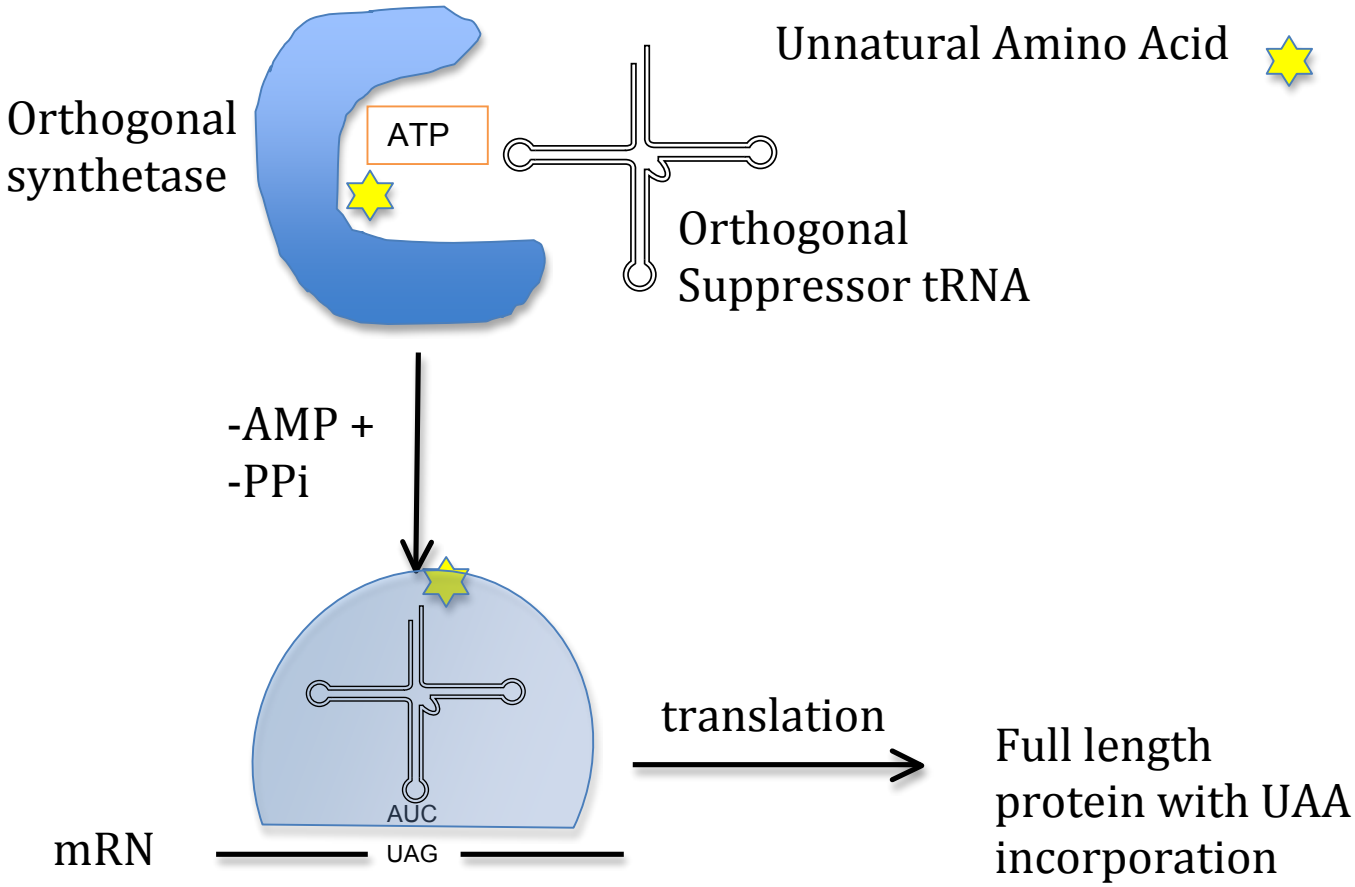


Figure 3 - *in vivo* site-specific incorporation

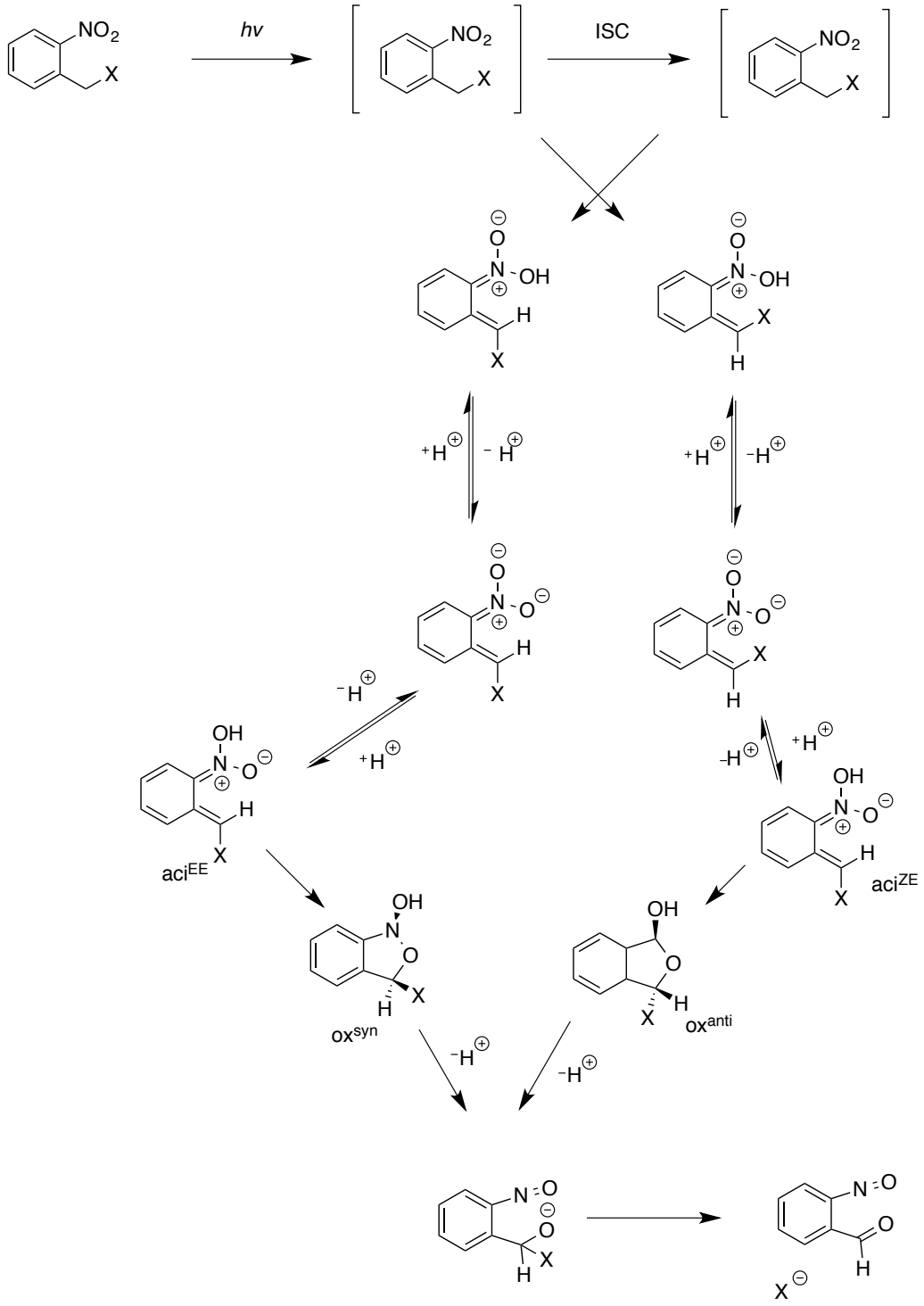
1.3 - Post-translationally modified amino acids

It is well known that PTMs play a crucial role in eukaryotic protein-protein interactions.[3–5,11,15] Inclusive of these proteins are those that decorate the chromatin complex – consistently seen are those found on histones. [11] The “regulatory potential” of histone marks is corroborated by the vast number lysine PTMs. For example, lysine trimethylation of histone H3K4 marks active promoters, whereas enhancers are marked by monomethylation, in the human genome.[11,16] Studying of lysine and other amino acids that exhibit PTM is difficult as many require cellular reactions that involve enzymes with high levels of specificity.[5] Without site-specific incorporation, much of this research was detailed by proteomic profiling techniques that have limitations. However, in the case of lysine, Chin and co-workers described an approach that site-specifically incorporates Nε-(t-butyloxycarbonly-L-Lysine), a PTM of lysine, into a target protein; a process that subsequently allows for site-selective isopeptide formation.[3,5] This approach opened the window for site-specific incorporation of other PTM lysine UAAs that are found on the chromatin complex.

1.4 - Photo-caged unnatural amino acids

Photo control of biomolecules has provided heighten advantages to understanding the functionality of proteins, kinases, enzymes, and post-translational sites; ultimately allowing one to follow physiological events in real time.[10,12] The most popular photo-caged groups are those that contain *o*-nitrobenzyl.[12,17] These groups have been attached to multiple amino acids of various protein structures and are easily cleaved when irradiated with 365nm of UV light.[9,10,12,17] The reaction allows for an amino acid to be returned to its natural state, therefore activating the enzyme or protein of interest. For instance, Chin and co-workers were successful in site-specific incorporation of photo-caged cysteine into a Tobacco Etch Virus (TEV) in mammalian cells, there after allowing for rapid activation of single cells by 365nm illumination.[10]

Scheme 1 Photon absorption and decaging reaction for 2-nitrobenzyl derivatives; X represents the heteroatom



The most common decaging mechanism favors the idea of proton abstraction from the benzylic position to the nitro-group after the system has been promoted to the singlet excited stage[12] - the system assumes a singlet excited stage when absorption of a photon from 365nm of UV light takes place. This proton shift leads to the formation of a short-lived *aci*-nitro intermediate giving the E and Z conformations respectively.[18] The resulting intermediate cyclizes to a benzisoazoline followed by ring opening to give the final 2-nitroso-benzaldehyde product.[12] It is through this mechanism that allows for control of biomolecule activity. Over time researchers have derived new photo-caged structures that provide faster cleavage when irradiated. These new structures introduce electron donating oxygen atoms that highly influence the excitation of π-orbitals in the aromatic ring to the π^*-orbitals of the nitro group.[18]

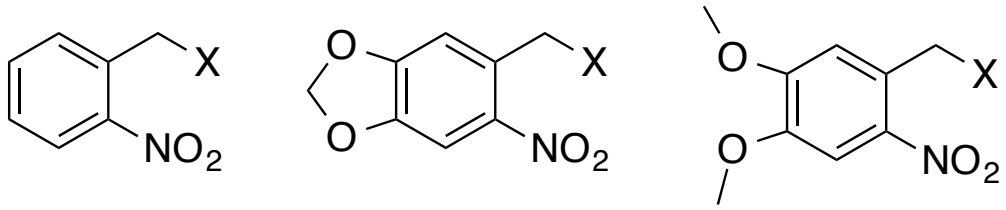


Figure 4 - Different caged compounds, with and without the oxygen electron donators; X represents the heteroatom

With the use of these photo-caged groups and UV light, the opportunity to explore the control of enzyme activity is at the frontline of bio-molecular research.

Chapter 2 – Synthesis and genetic encoding of 2-hydroxyisobutyryl lysine

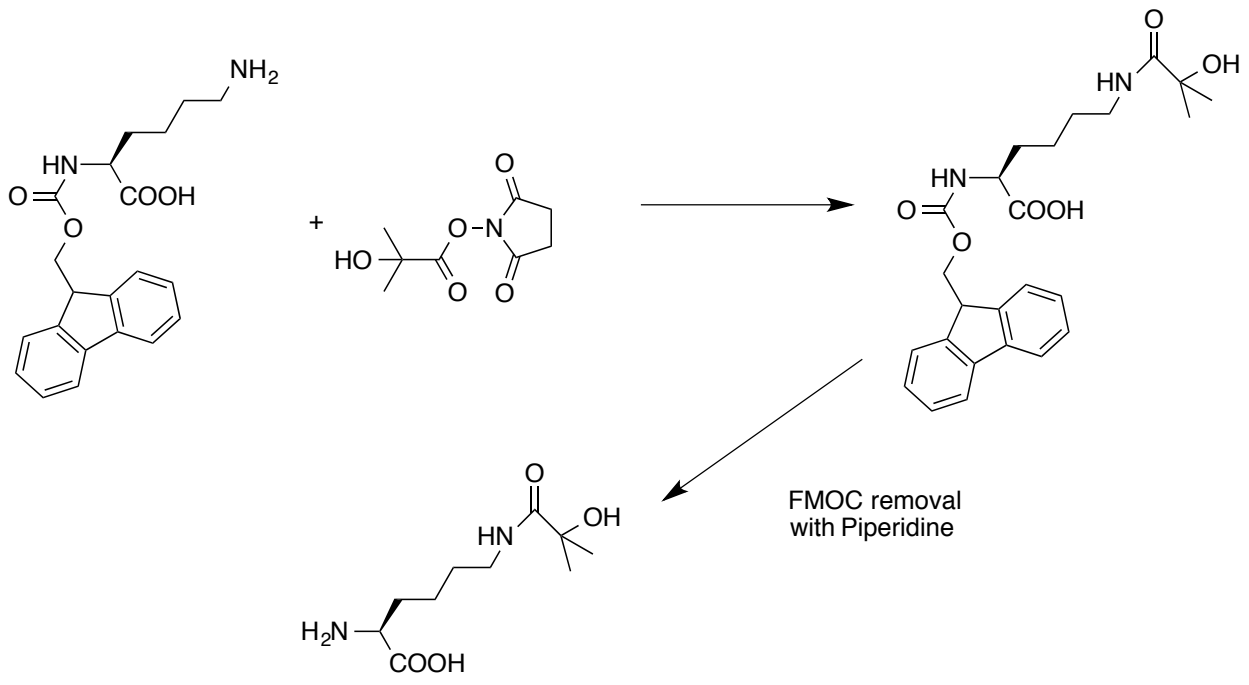
2.1 – Importance of 2-hydroxyisobutyryl lysine

Cellular DNA exists in the form of chromatin, a tightly wound structure that contains repeat units known as nucleosomes. Nucleosomes exist as an octamer of four core histone proteins (H2A, H2B, H3, H4) at which 147 bases of DNA wrap around.[19] These core histones are well known for their array of covalent modifications, such as acetylation, methylation, phosphorylation, and ubiquitylation.[19] Recently, a new histone mark, 2-hydroxyisobutyryl lysine (K_{hib}), was discovered as an indicator of gene transcriptional activity that is directly associated with meiotic and post-meiotic male germ cells.[11] However, this research is based on proteomics profiling in different cell types, and has yet to be studied *in vivo* with the site-specific technology for regulatory enzymes that remove the 2-hydroxyisobutyryl group from 2-hydroxybutyrylated lysine. This study can be conducted with the orthogonal tRNA/RNA synthetase site-specific technology originally developed by Schultz and co-workers. This thesis reports the synthesis of 2-hydroxyisobutyryl lysine and its site-specific incorporation in *E. coli*.

2.2 – Synthesis of 2-hydroxyisobutyryl lysine (K_{hib})

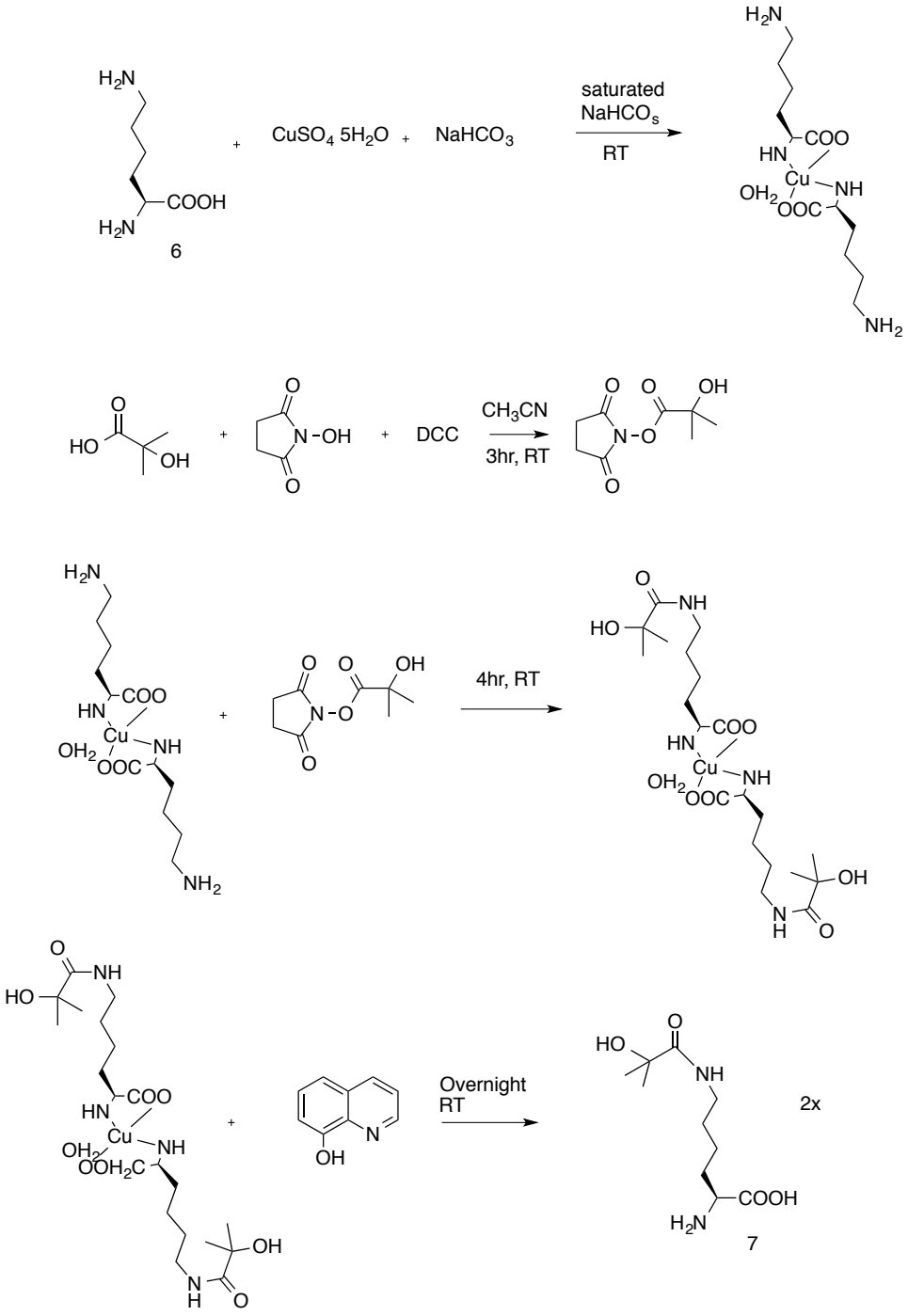
The first attempt of K_{hib} synthesis proceeded by treatment of commercially protected fluorenylmethyloxycarbonyl (Fmoc) lysine with 2-hydroxyisobutyryl-O-succiminide ester. However, after attempting to purify the compound by simple silica column chromatography, it was apparent by TLC that the starting material and final product closely overlapped and a desired elution system was not found. Additionally, the TLC exhibited low product yield due to the reaction's inefficiency in covalent addition of the 2-hydroxyisobutyryl group.

Scheme 2 - First attempt at synthesizing 2-hydroxyisobutyryl lysine



As an alternative, the next approach generated a copper protected form of lysine; this allowed for selective acylation of the ε-nitrogen with 2-hydroxyisobutyryl-O-succinimide ester. The 2-hydroxyisobutyryl-O-succinimide ester was previous prepared using N,N'-dicyclohexylcarbodiimide (DCC), a coupling reagent. Cleavage of the copper complex to give the final product is accomplished by chelation with 8-quinolinol. **Scheme 3** displays the entire synthetic formulation of K_{hib}. Nuclear magnetic resonance (NMR) and mass spectrometry (MS) confirmed synthetic development of K_{hib}. (see appendix for full synthesis and detailed structure analysis)

Scheme 3 - Synthesis of 2-hydroxyisobutyryl lysine



2.3 – Site-specific incorporation of 2-hydroxyisobutyryl lysine in *E. Coli*

As noted earlier, Chin previously described a method that site-specifically incorporated Nε-(t-butyloxycarbonyl)-L-lysine (Boc-lysine) into *E. coli*.[5] The orthogonal system used, pyrrolysyl-tRNA synthetase/$tRNA_{CUA}$, was imported from the archaebacteria, *Methanosarcina barkeri* (Mb), which naturally incorporates pyrrolysine, . However, other variants of this pyrrolysyl-tRNA/RNA synthetase system exist and function as efficient substrates for BOC-lysine, *Methansarcina mazei* (Mm). Here the wild-type synthetase from Mm is screened for relaxed substrate specificity for the genetic encoding of K_{hib} in *E. coli*.

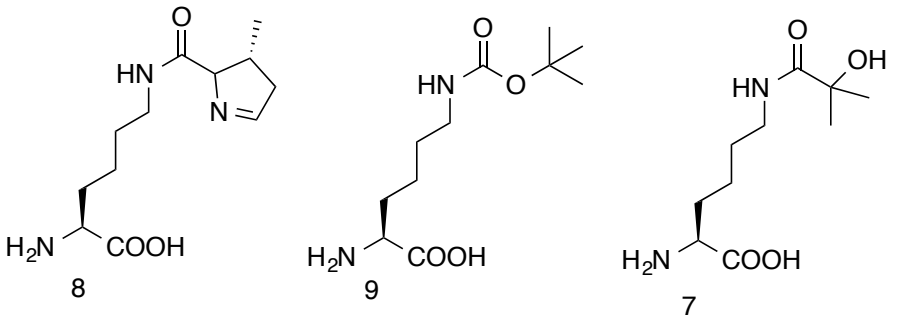


Figure 5 – Pyrrolysine analogs, used for *in vivo* site-specific incorporation

This translational system performs by inserting at a nonsense amber codon that was previously mutated by polymerase chain reaction (PCR) at a native codon-site.[3,5] Given the leniency of the synthetase's specificity and the structural similarities in PTM groups, K_{hib} was screened against the same synthetase (MmPylS) that successfully encodes BOC-lysine, **9,** and pyrrolysine**, 8**. The screening utilized an expression plasmid for superfolder green fluorescent protein (sfGFP) containing an amber stop codon, TAG, in

place of the codon for Y151. This same plasmid also included the pyrrolysyl tRNA (pylT) that is orthogonal to MmPylS.

Successful incorporation and generation of full-length sfGFP yields fluorescent *E. coli* cells when irradiated with UV light. The synthetases tested are as follows with their results as (see appendix for plates and synthetase mutations)

Synthetase	Fluorescent cells
EV_16-5	No
EV_20	No
MmPylS	Yes
EV_17	No

Table 1- pyrrolysyl synthetases used to screen K_{hib} incorporation

Positive results were obtained from the wild type MmPylS synthetase, which subsequently resulted in future testing. The expression was scaled (medium) to increase His-tagged sfGFP production (see appendix); the cell lysate was purified by Ni^{2+} affinity chromatography to give pure sfGFP, Figure 6. The gel represents the process by which two different cultures of *E. coli* are being purified for detection of sfGFP. The columns labeled +AA incorporates K_{hib} during protein translation, and conversely -AA no unnatural amino acid is incorporated. Elution 1 shows evidence of full-length sfGFP with the incorporation of K_{hib} Further analysis by ESI-MS provided

inconclusive results, so a tryptic digestion was performed for fragmentation analysis with LC/MS/MS. Fragmentation of the protein not only displayed evidence of successful full-length sfGFP expression, but the exact location of $\mathbf{K_{hib}}$, **Figure 7.** The difference between y(5) and y(6) is 214, which details the incorporation of K_{hib} – molecular weight after peptide coupling.

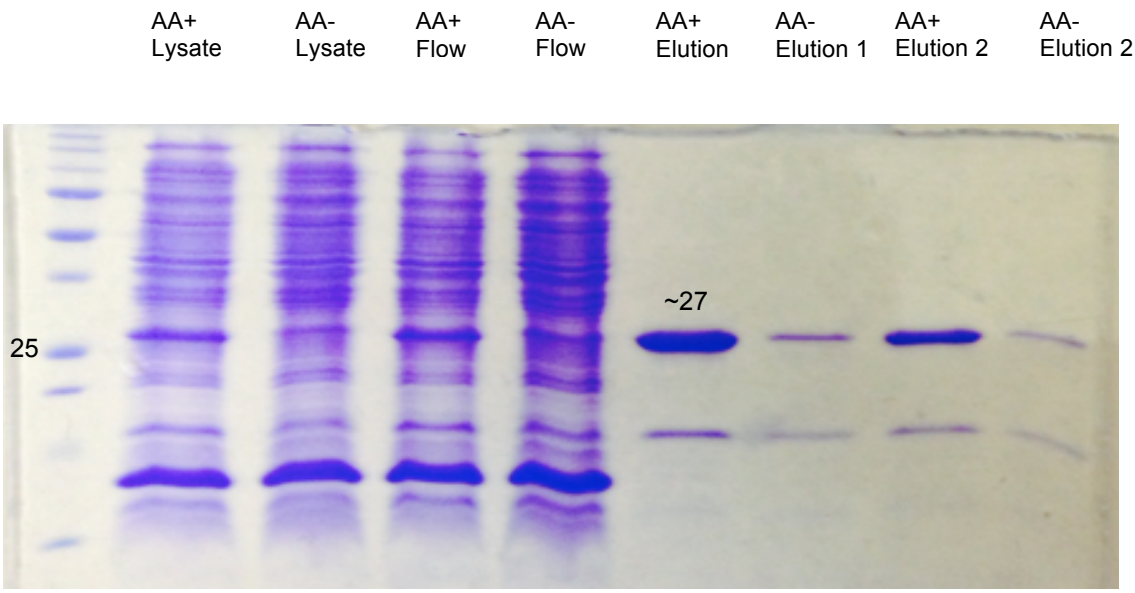


Figure 6 - Protein gel of nickel purification – sfGFP; Flow, lysate after Nickel incorporation; Elution 1 large band full-length sfGFP, smaller band truncated protein

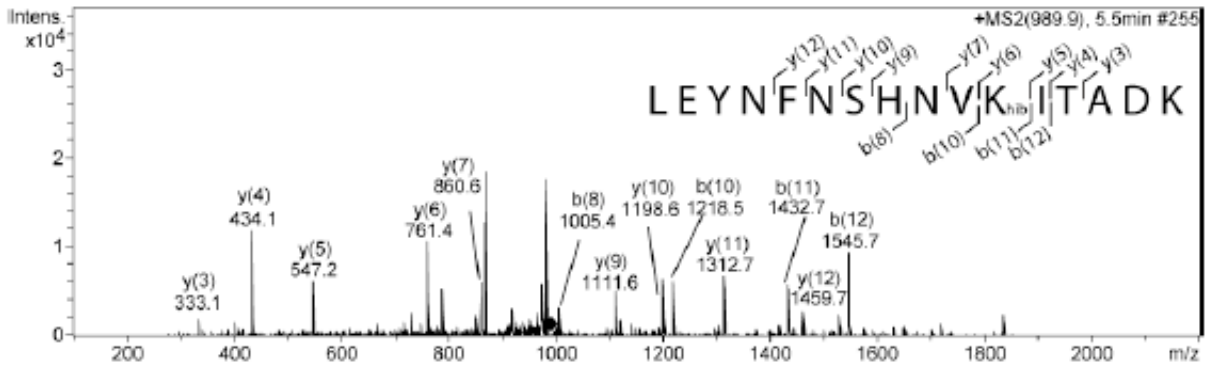


Figure 7 - LC/MS/MS of K_{hib} tryptic digestion

2.4- Conclusion to 2-hydroxyisobutyryl lysine work

In conclusion, the post-translational modification, 2-hydroxyisobutyryl lysine, was synthesized and genetically encoded by the pyrrolysyl tRNA/RNA synthetase system in full. MS, NMR and synthesis data are provided as a foundation for future. As well, this site-specific incorporation can be of use for future histone modification work, benefitting all those that would like to research the true impact this histone post-translational modification provides.

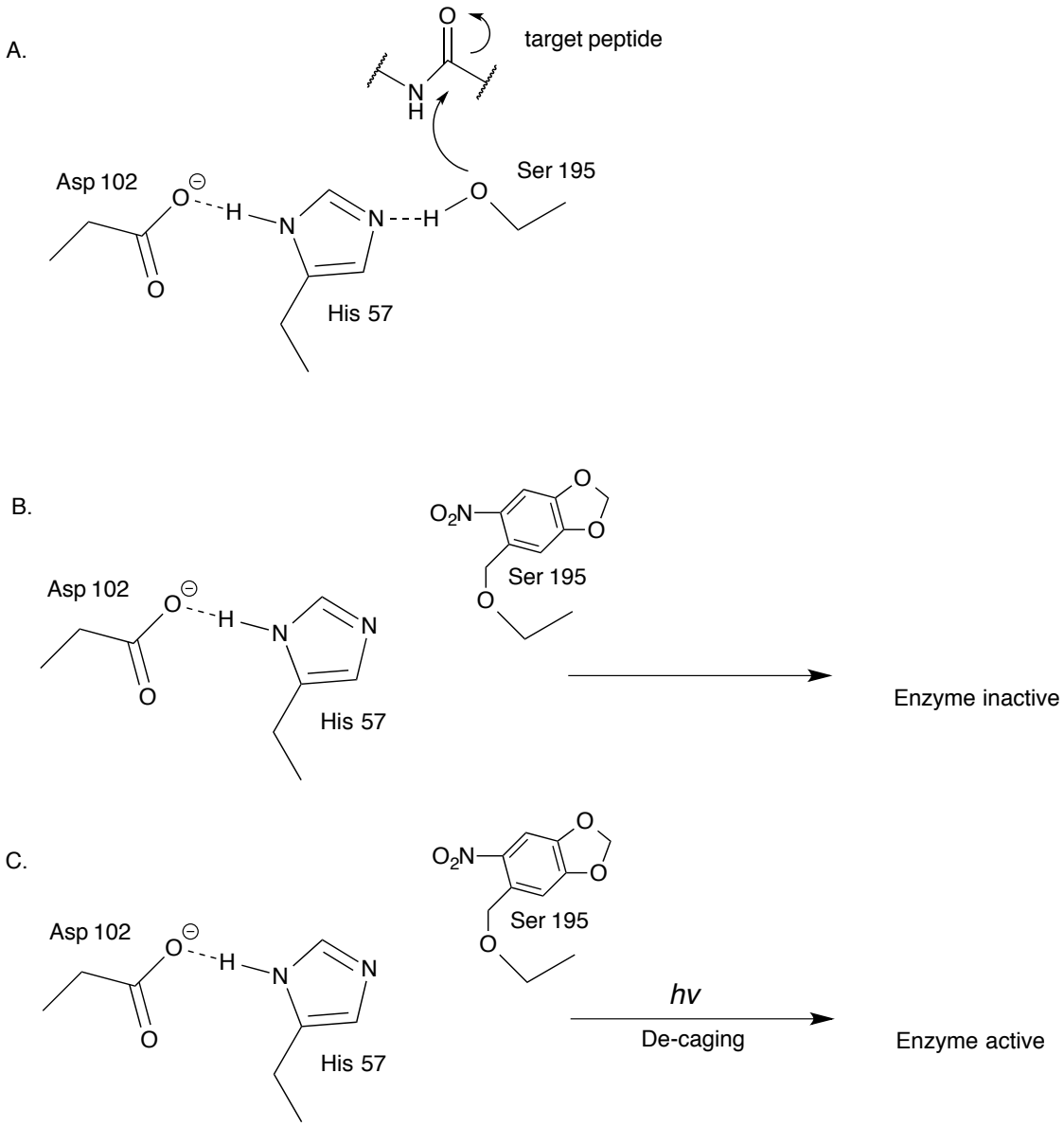
Chapter 3 – Photo-caged amino acid synthesis

3.1 – Importance of photo-caged amino acids

Optogenetics, a term commonly used to describe the genetic encoding of proteins with light-responsive characteristics. With this approach researchers have learned to control biological function such as transcription, protein localization, post-translational modification, and ion channel activity.[20] Most recent technology for the genetic encoding of light-responsive proteins utilizes site-specific incorporation of photo-caged UAAs. Orthogonal biosynthetic machinery is evolved to encode such UAAs at a non-sense amber stop codon. The attached caging group is thought of as a covalently linked chromophore that is easily cleaved with UV light exposure, there by restoring the native properties of the protein.[9,17,20] This approach also allows for one to target a specific amino acid within a protein of interest. Other usages focus much of the attention on serine proteases, for example a known protease of this class, chymotrypsin, relies on a specific set of three amino acids for substrate specificity. In this enzyme this set of amino acids is referred to as "the catalytic triad" – Asp102, His57, Ser 195.[21,22] Each of these amino acids is arranged in a favorable way to conduct proton and charge movement efficiency in the active site. This allows for the hydroxyl group on serine to become more reactive – carbon-oxygen bond formation between a targeted peptide and serine (Scheme 4).[22] By using the site-specific incorporation technology one can target serine at position 195 and in theory control the activity of the enzyme with light de-caging. Along with many other proteases that exude the catalytic triad structure, serine and in more rare cases threonine, this technology can function to help determine purpose and mechanism of action in the cellular environment. Additionally, with this new

technology research is able to circumvent the use of gene knockout technology and further unveil new residues that play a crucial role in enzyme functionality.[10]

Scheme 4 - Catalytic triad control in chymotrypsin; A) active-site of the catalytic triad found in chymotrypsin, B) caged serine with out light perturbation, C) de-caging of serine with light perturbation



3.2 – Synthesis of photo-caged serine

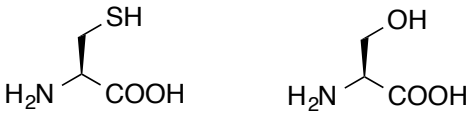
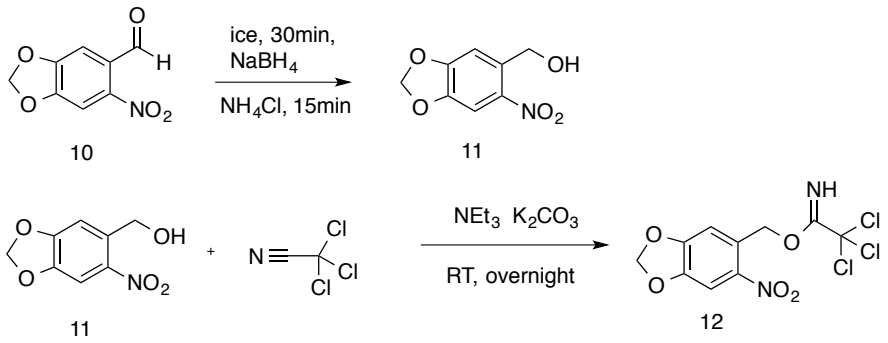


Figure 8 – Cysteine and serine amino acids

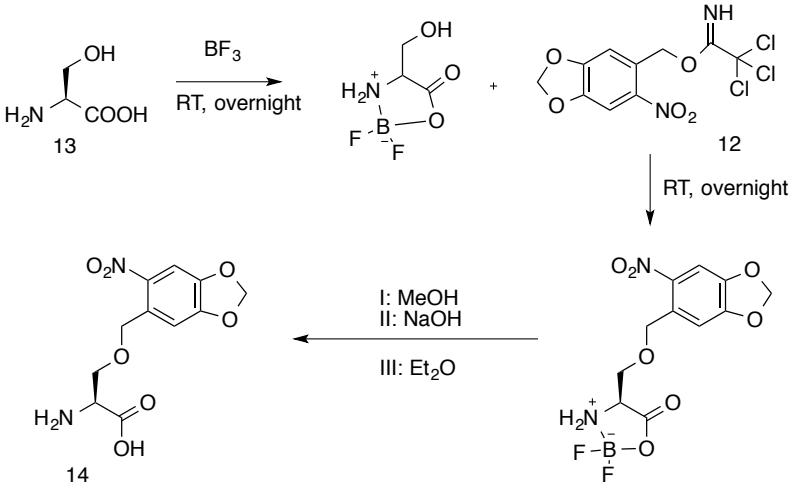
Synthesis of photo-caged serine in the past has used protection groups (i.e. FMOC, BOC) that are chemically cleaved upon purification[23], however, the removal of these groups are typically tedious or fail to provide product in good yield. The greater nucleophilic character of the nitrogen can confer the favoritism for the amino group over the hydroxyl group in nucleophilic substitutions, thus the reason for a sufficient protecting group. This is in stark contrast to that of cysteine, given that the thiol group is more nucleophilic than the amino group. Furthermore, to overcome the burden of typical protection groups, a previously reported synthesis to covalently bind the hydroxyl group to a benzyl was discovered. The synthesis uses boron trifluoride to protect the amino group of serine and threonine.[24] The protected boron trifluoride amino acid can then be derivatized with a synthesized trichloroacetamidate photo-caged group.[24] The synthesis of this trichloroacetamidate is as follows:

Scheme 5 – Synthesis of photo-caged trichloroacetamidate group



10, is commercially available and is reduced by sodium borohydride, producing an alcohol, **11**. **11** then undergoes reaction with trichloroacetonitrile, producing **12**, the trichloroacetamidate of interest. As stated before, this structure will only undergo substitution on serine if the amino group is protected. Serine is protected by boron trifluoride and is then attacked by the trichloroacetamidate product generated above. The synthesis is displayed below:

Scheme 6 – Synthesis of photo-caged serine



Boron trifluoride is added to a stirring solution of serine, **13,** in THF. Afterwards the mixture is rotary evaporated and re-suspended in dioxane. **12,** is then added directly to the stirring dioxane solution to liberate photo-caged protected serine . The reaction is quenched with MeOH and NaOH; and washed with diethyl ether. The compound is purified by XAD resin to give, **14**. **14,** was structurally analyzed by TLC, NMR, and DART MS in high resolution (see appendix MS and NMR). The TLC displays after irradiation and ninhydrin staining (that respective order). Photo-caged serine blocks fluorescence from the plate, displaying the presence of the photo-caged group. Serine

and photo-caged serine both stain with ninhydrin, which is representative of the amino group.

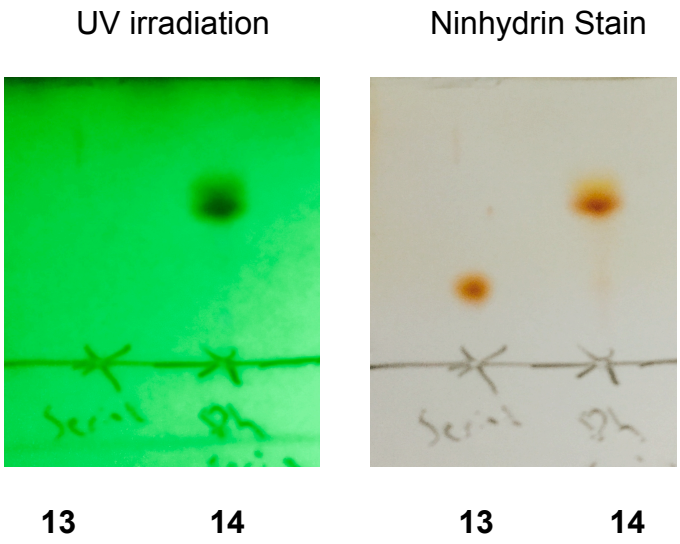
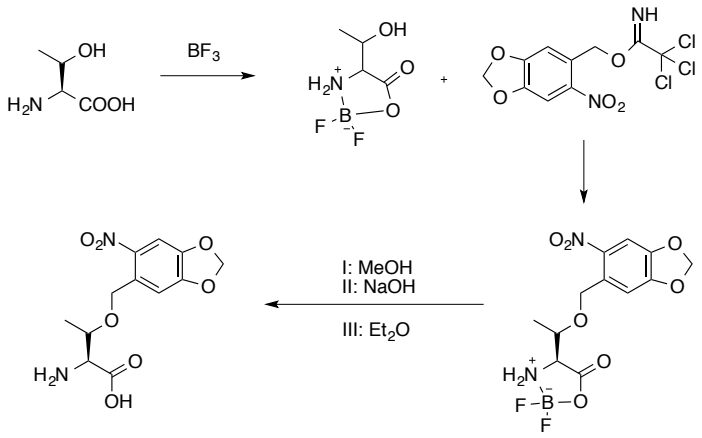


Figure 9 - UV irradiation and ninhydrin staining of serine and photo-caged serine

3.3 – Synthesis of photo-caged threonine

The same approach in **Scheme 6,** can be used to synthesize threonine and allo-threonine. The synthesis can be found below (see appendix for mass spec data on photo-caged threonine):

Scheme 7 – Synthesis of photo-caged threonine



3.4 – Conclusion to photo-caged amino acid synthesis

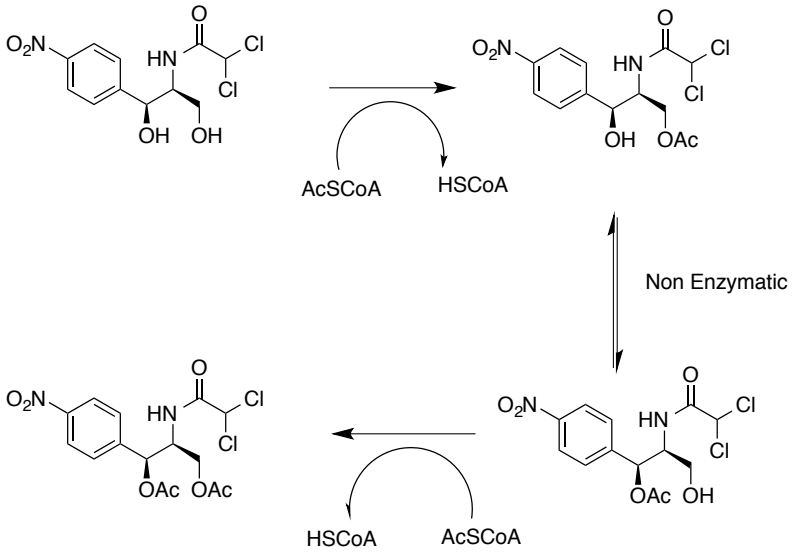
This thesis provides knowledge in the use of photo-caged amino acids, as well, details the structural synthesis and purification of photo-caged serine and photo-caged threonine. The selection process is in the works, as it has potential for controlling protease activity with 365nm of UV light.

Chapter 4 – Future work/conclusion

4.1 – Site-specific incorporation of photo-caged serine, threonine and allo-threonine

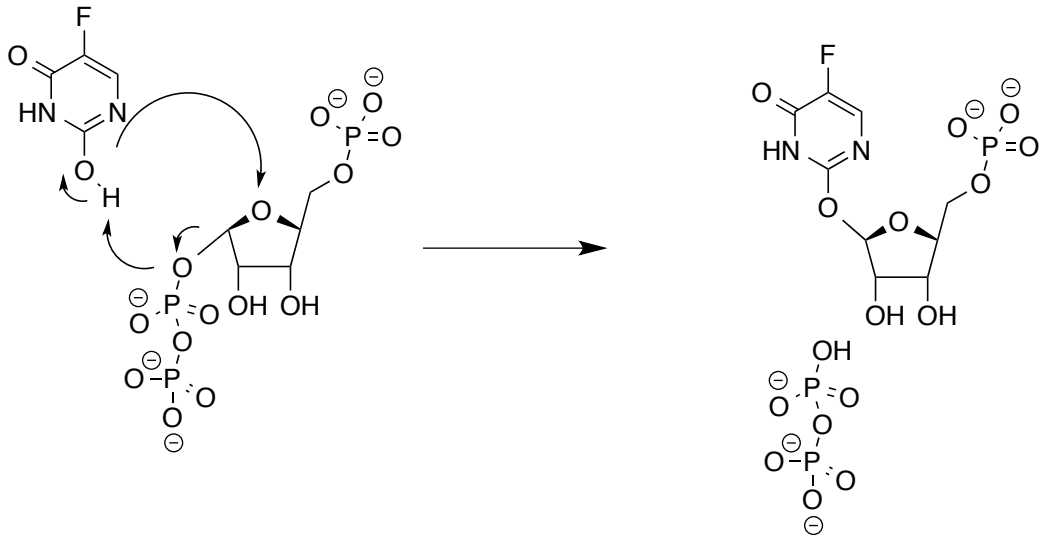
To date a version of photo-caged serine, 4,5-dimethoxy-2-nitrobenzyl serine, has only been site-specifically incorporated into *Saccharomyces cervisiae*[23], making the incorporation into *E. coli* a novel approach. For the most part the tRNA has already been modified properly from prior research; however, evolution of the synthetase would be necessary and can be conducted by alternating positive and negative selection from a developed library in *E. coli* (See appendix for active site mutations). The rounds of selection are accomplished through the use of a fusion gene known as CATUPP, (CAT-chloramphenicol acetyltransferase) and

Scheme 8 – Positive selection in synthetase selection for photo-caged serine, threonine and allo-threonine



(UPP-uracil phosphoribosyltransferase). The positive aspect of the test determines those modified synthetases that successfully survive the antibiotic chloramphenicol, thus by producing a full-length chloramphenicol resistance gene with an in-frame TAG codon, **Scheme 8**. While the negative selection tests for those aaRS that non-specifically suppress amber stop codons in a gene that is toxic in bacteria. In the absence of an unnatural amino acid the toxic gene would be expressed ultimately causing cell death, **Scheme 9**. The clones that survive the positive test and sequentially die for the negative test are selected for.

Scheme 9 – Negative selection mechanism in synthetase selection for photo-caged serine, threonine and allo-threonine



After discovery of a synthetase that optimally charges serine, testing can be tested directly into the sfGFP protein structure. Being that the orthogonal system used recognizes a stop codon, a genetically modified variant of gene that encodes for sfGFP has previously been developed – used for 2-hydroxyisobutyryl lysine above. As for 2-hydroxyisbutyryl lysine, sfGFP can be expressed in chemically competent genehogs

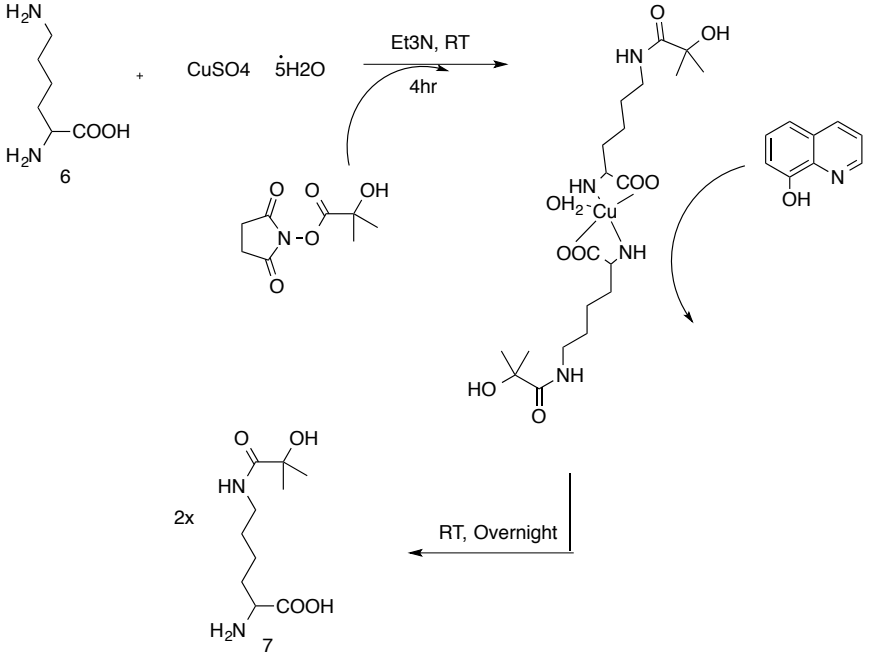
and purified by nickel resin. Afterwards SDS can confirm the protein of interest and further tryptic digestion in collaboration with LC/MS/MS. Decaging of the unnatural amino acid will be confirmed by a m/z shift after irradiating with 365nm of light. The result obtained would be feasible for enzymatic testing in the future.

4.2 – Conclusion

This thesis provides background knowledge in the field of unnatural amino acid incorporation of post-translational modifications and photo-caged amino acids. One can utilize synthetic schemes provide to covalently alter other similar unnatural amino acids of interest. Additionally, as stated this thesis provides a successful synthetic approach to developing and site-specifically incorporating 2-hydroxyisobutyryl lysine. This research can be used for future work in the field of histone post-translational modifications. This thesis also provides a new method for creating photo-caged serine, threonine, and allo-threonine. If successfully incorporated into the *E. coli* genome, functional understanding of proteases and enzymes can be researched through light perturbation.

Appendix I – Synthesis of UAAs

Scheme 10 - Synthesis of 2-hydroxyisobutyryl lysine



2-hydroxyisobutrylysine (**7**): 2-hydroxyisobutryic acid (3.35g, 32mmol) and N-hydroxysuccinimide (3.75g, 32.6mmol) were combined in 30 mL of CH_3CN. After stirring for 10 minutes, N,N'-dicyclohexylcarbodiimide (DCC) (6.72g, 32.6mmol) was added and allowed to stir. After three hours the solution was filtered and the solvent removed to give 2-hydroxyisobutyryl-O-succinamide ester as a yellowish oil (3.24g, 50%) that was used directly without further purification. Separately, L-lysine monohydrochloride (980 mg, 5.37mmol) was added to 5mL of saturated aqueous $NaHCO_3$. The mixture was stirred at room temperature for ten minutes and $CuSO_4$ (669mg, 2.68mmol) added. 2-hydroxyisobutyryl-O-succinamide ester from the first step (3.24g) was re-suspended in 10 mL of acetone, and added drop-wise to the lysine complex over four hours to produce a blue slurry. The reaction was quenched with

methanol and evaporated to dryness. The solid was re-suspended in water (50mL) and stirred at room temperature. To the stirring solution was added 8-quinolinol (1.95g, 13.43mmol) and stirred vigorously overnight to produce a green slurry. The green slurry was filtered, and the filtrate washed with ethyl acetate (3 x 30mL). The aqueous layer was then evaporated to dryness to give **1** (967mg, 77%). R_f=0.36 (3:1:1, 1-butanol:acetic acid:water, pink with ninhydrin stain) ^{1}H NMR (300 MHz, D_2O): σ = 1.19, (8H, m),1.35, (2H, m), 1.68, (2H, m), 3.01 (2H, t, J = 6Hz), 3.60 (1H, t, J = 6Hz), ^{13}C NMR (75 MHz, D_2O): σ = 21.6, 25.2, 26.3, 28.1, 29.9, 38.6, 54.3, 73.0, 176.3, 179.1. MS-ESI (m/z) $[M+H]^+$, calc $C_{10}H_{21}N_2O_4$ 233.1501, found 233.1492, MS-ESI (m/z)

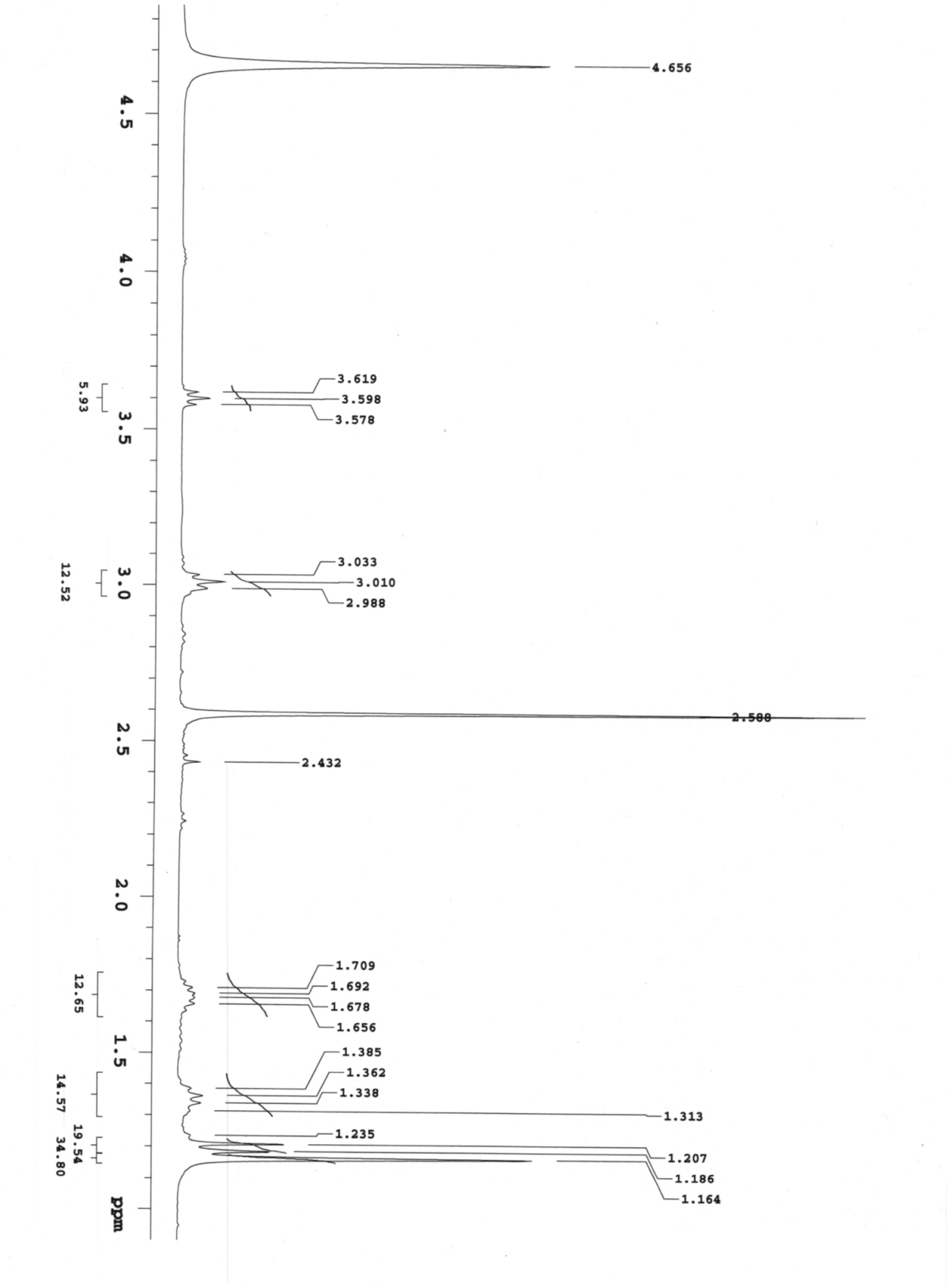


Figure 10 - ^{1}H NMR of 2-hydroxyisobutyryl lysine

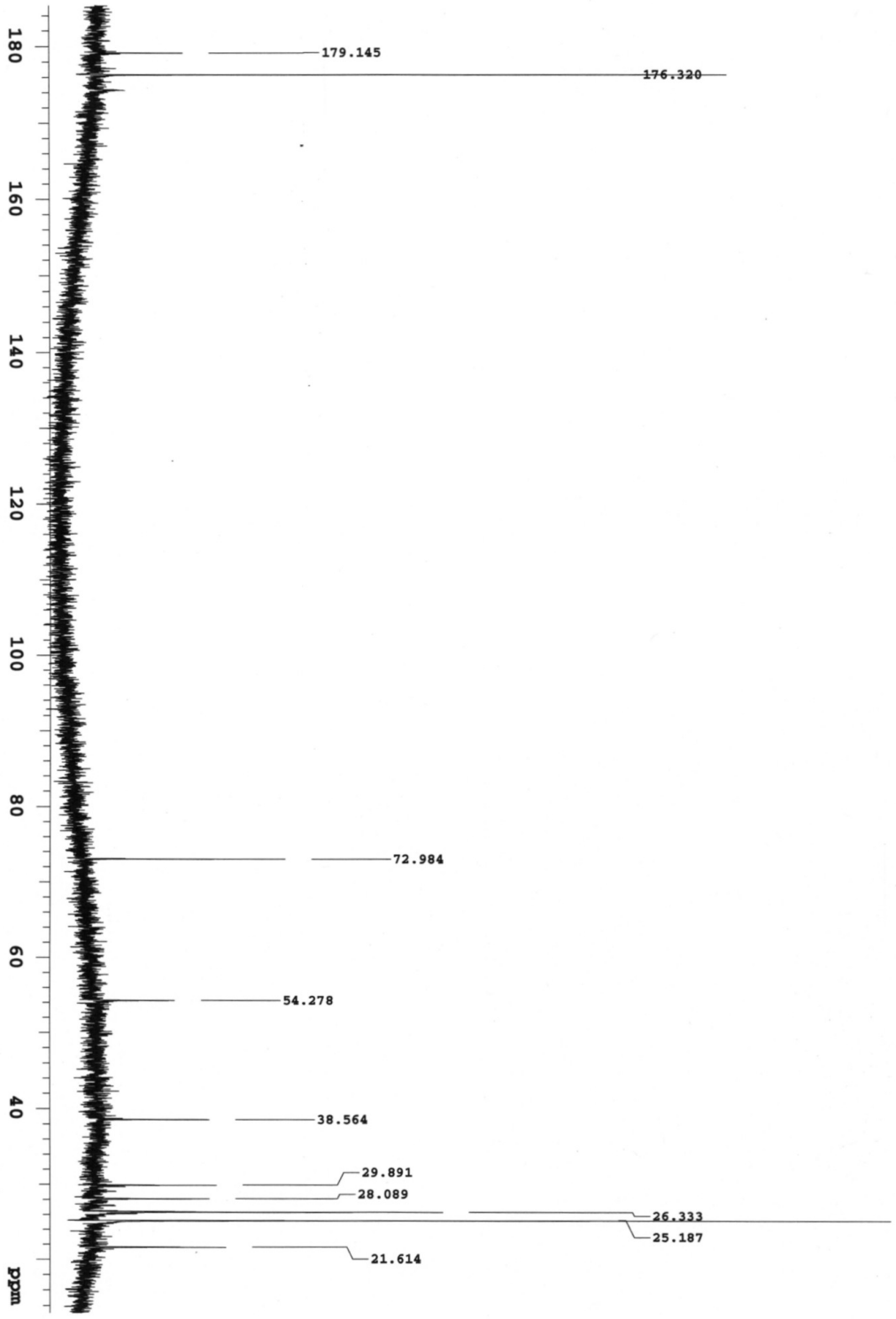


Figure 11 – ^{13}C NMR of 2-hydroxyisobutyryl lysine

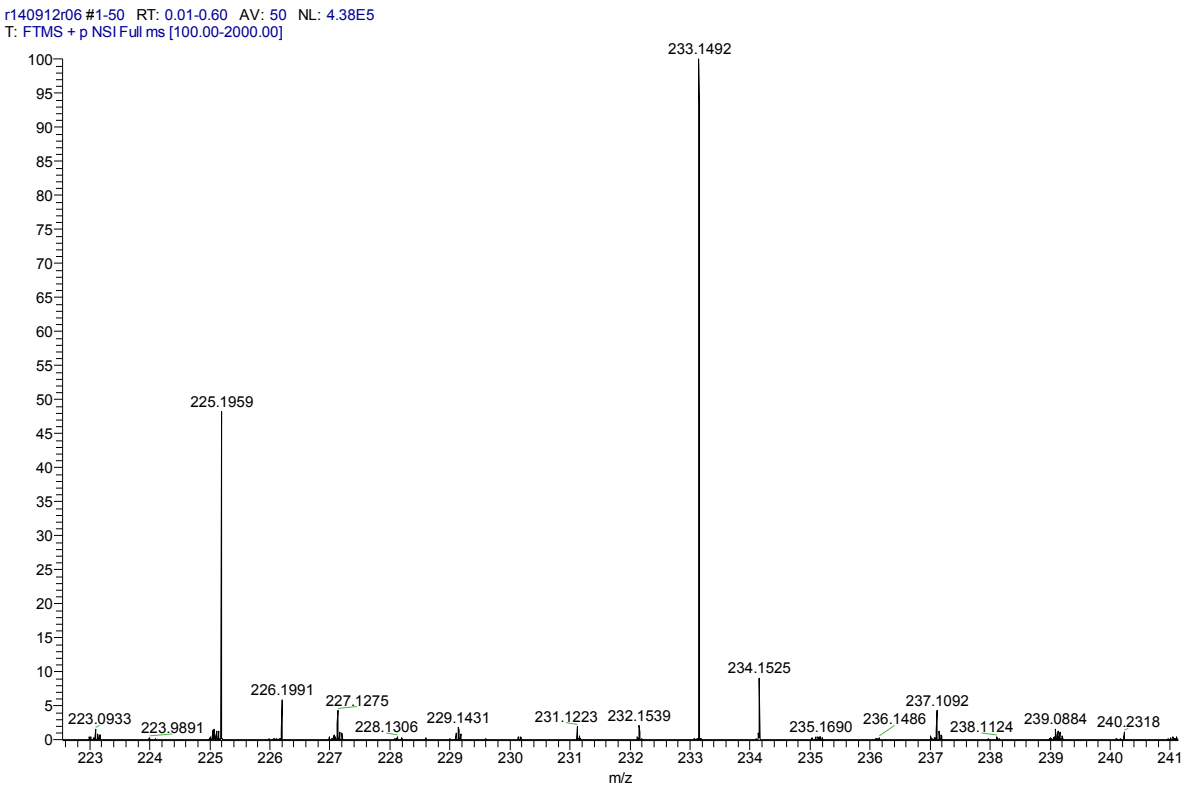


Figure 12 - Mass Spec Data on 2-hydroxyisobutyryl lysine

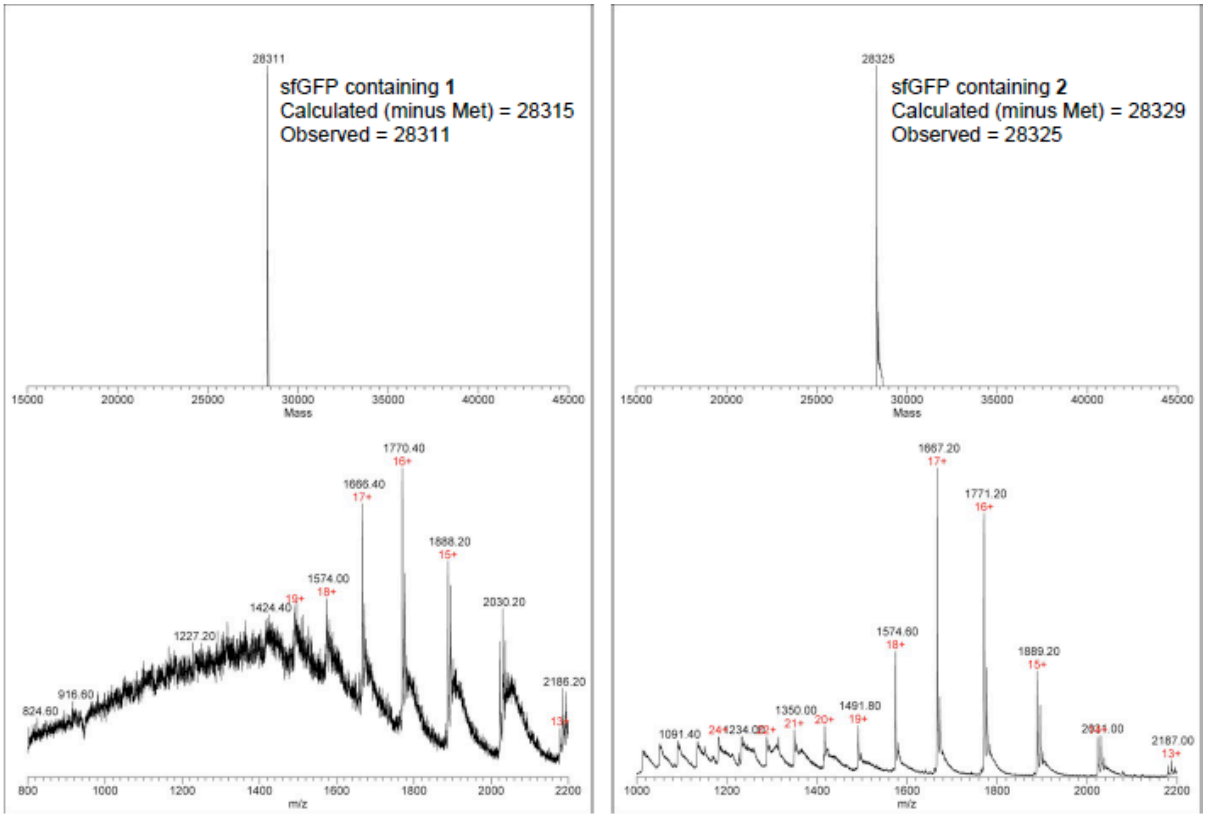
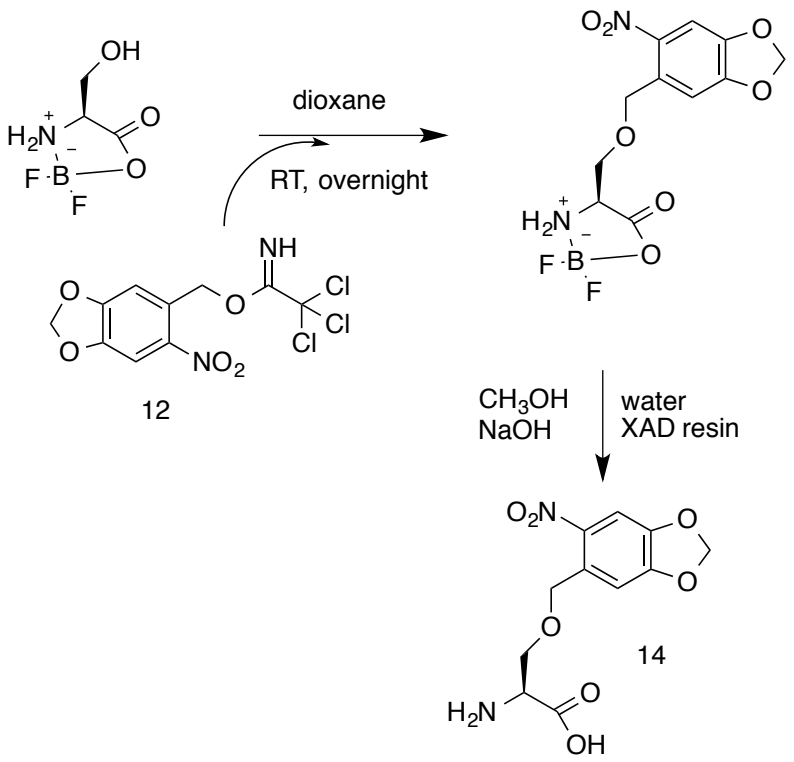


Figure 13 – sfGFP ESI DATA for incorporation of 1) K_{hib}, 2) Boc-lysine

Note: Intact protein ESI-MS of sfGFP. Calculated masses correspond to the protein sequence minus the N-terminal methionine (131) and GFP chromophore maturation (20). Calculated mass difference is 14, observed mass difference is 14.

Scheme 11 - Synthesis of photo-caged serine



10 (20g, 102.49mmol, 1eq) is added to 225mL of methanol while on ice. Sodium borohydride (1.71 g, 45.09 mmol, 0.4eq) is added over 30 minutes to the reaction. The reaction was monitored by TLC (DCM, product spot R_f = 0.31). 115mL of saturated aqueous ammonium chloride was added to the stirring reaction, and allowed to stir for 15 minutes. The reaction was extracted three times each with 140 mL of methylene chloride, washed with brine, and dried over magnesium sulfate. The filtrate was then evaporated to dryness giving, **11. 11** (15 g, 76.08mmol, 1 eq) and potassium carbonate (26.29g, 190.2mmol, 2.5eq) were suspended and stirred at room temperature in 500mL of methylene chloride. To the stirring solution, trichloroacetonitrile (13 mL, 129.34 mmol, 1.7 eq) and triethylamine (12.6 mL, 91.30mmol, 1.2 eq) were added and allowed

to stir overnight. The next day 100 mL of DCM and allowed to continue stirring for 30 minutes. Afterwards the solution was washed twice with 250mL of .5M HCl and once with 250mL of saturated NaCl. The solution was dried with magnesium sulfate, filtered and evaporated to dryness **12** (22.98g, 88%). **13** (5g, 47mmol, 1 eq) was resuspended in 75mL of THF and stirred at room temperature. To the reaction was added 30mL of boron trifluoride diethyl etherate, and allowed to react overnight. The reaction was rotary evaporated the next day to give protected serine. The oil, protected serine**,** was then re-suspended in 200mL of dioxane and stirred for 10 minutes until everything is completely in solution. **12** (18.44g, 54.05mmol, 1.15eq) was added to the reaction and continued to stir overnight at room temperature. The reaction was quenched with methanol (15mL), NaOH (90mL), and evaporated to dryness under vacuum. The solid was re-suspended in a minimal amount of water and purified by XAD resin; this was done by exhausting the resin with water and checking for serine removal by TLC. The resin is then eluted with 1:1, ethanol:water, and then further eluted with 3:1:1, acetone:ethanol:water, **14**. The purification process is closely monitored by TLC (3:1:1, 1-butanol:acetic acid:water, R_f = .53) ^{1}H NMR (300 MHz, $CDCl_3$) = (approximate values) 7.6, 6.9, 6.0, 1.58, 1.40, 1.23; DART-MS (m/z) $[M+H]^+$, calc $C_{11}H_{13}N_2O_7$ 285.0723, found 285.0790

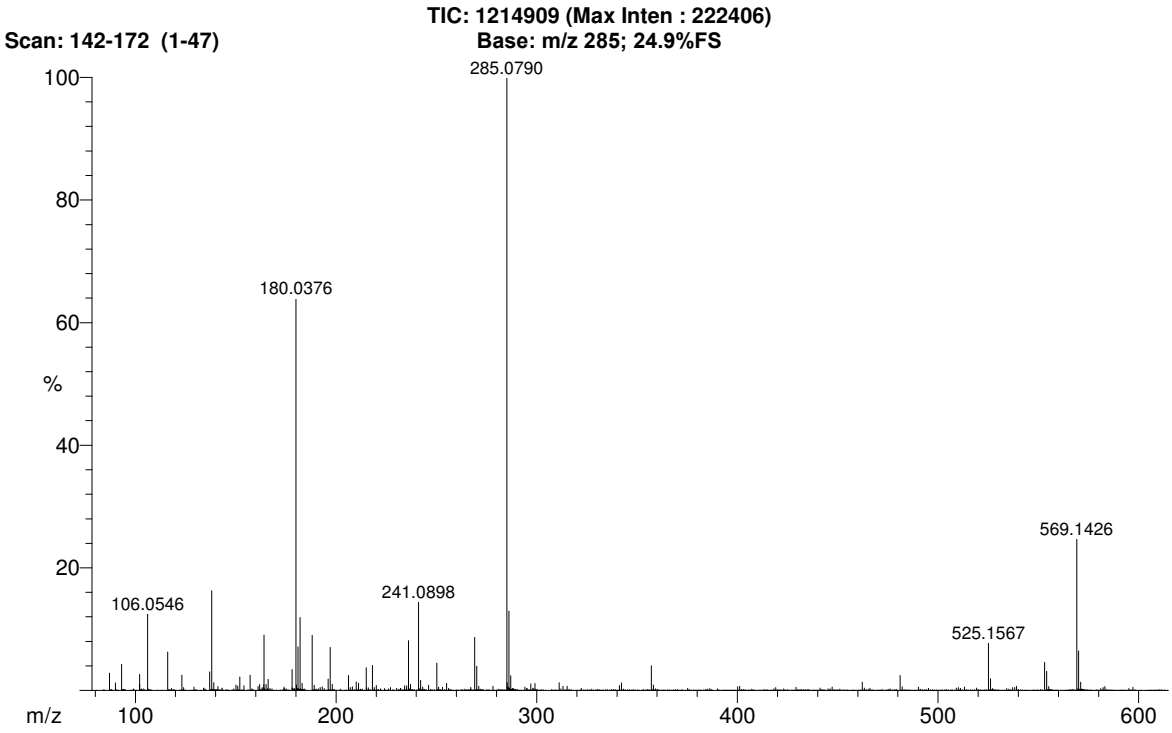


Figure 14 - DART data for photo-caged serine

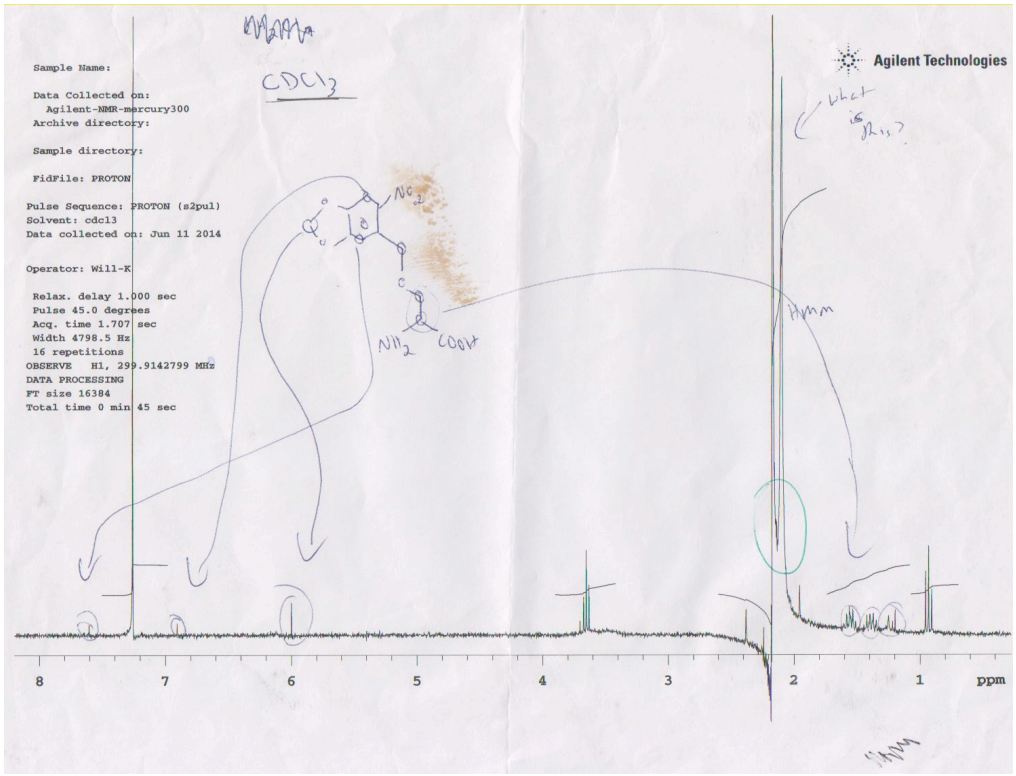


Figure 15 – Preliminary NMR data for the synthesis of photo-caged serine

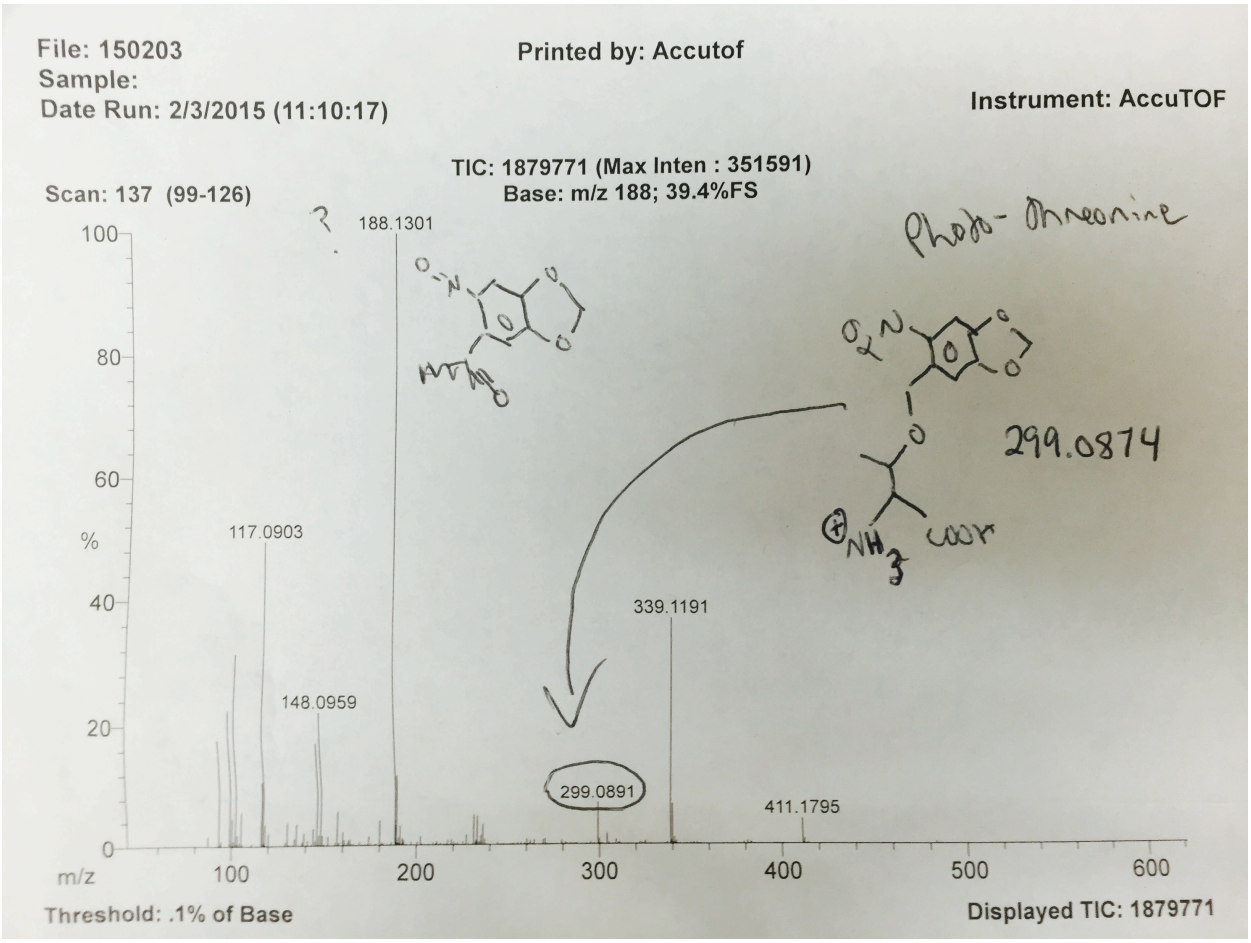


Figure 16 - DART data for photo-caged threonine

Appendix II - Synthetase sequence for the incorporation of 2-hydroxyisobutyryl lysine with screening

M. mazei pyrrolysyl-tRNA synthetase

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAG
GACCGGAACAATTCATAAAATAAAACACCACGAAGTCTCTCGAAGCAAAATCTATAT
TGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCCAGGAGCAGCAGGACT
GCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTGCAAACGCTGCAGGGTTT
CGGATGAGGATCTCAATAAGTTCCTCACAAAGGCAAACGAAGACCAGACAAGCGT
AAAAGTCAAGGTCGTTTCTGCCCCTACCAGAACGAAAAAGGCAATGCCAAAATCCG
TTGCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTC
TGGATCTAAATTTTCACCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCC
CGGCATCTGTTTCAACATCAATATCAAGCATTTCTACAGGAGCAACTGCATCCGCA
CTGGTAAAAGGGAATACGAACCCCATTACATCCATGTCTGCCCCTGTTCAGGCAAG
TGCCCCCGCACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTTAAACCCA
AAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATT
GCTCTCTCGCAGAAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAAT
TATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCT
GGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTG
ATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGA
GACCCATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGGCCCT
GCCTGATCCAATAAAAATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCGACG
GCAAAGAACACCTCGAAGAGTTTACCATGCTGAACTTCTGCCAGATGGGATCGGG

ATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAA

TTGATTTCAAGATCGTAGGCGATTCCTGCATGGTCTATGGGGATACCCTTGATGTA

ATGCACGGAGACCTGGAACTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACC

GGGAATGGGGTATTGATAAACCCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCT

TCTAAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTT

ACTATAACGGGATTTCTACCAACCTG

Synthesis Screening

Chemically competent *E.* coli DH10-B were doubly-transformed with pBK_sfGFP151TAG and each of four different synthetase plasmids (pBK_MmPylS, pBK_EV16-5, pBK_EV17, pBK_EV20). A single colony from each plate was picked to inoculate 4mL liquid culture. Once the liquid cultures became saturated, they were streaked on three different plates: plate 1 (LB, Kan 50μg/mL, Tet 15μg/mL, 0.2% arabinose), plate 2 (LB, Kan 50μg/mL, Tet 15μg/mL, 0.2% arabinose, 2mM Nε-Boc-lysine) and plate 3 (LB, Kan 50μg/mL, Tet 15μg/mL, 0.2% arabinose, 1mM 2-hydroxyisobutyryl lysine). Irradiated and screened for fluorescent *E. coli* cells.

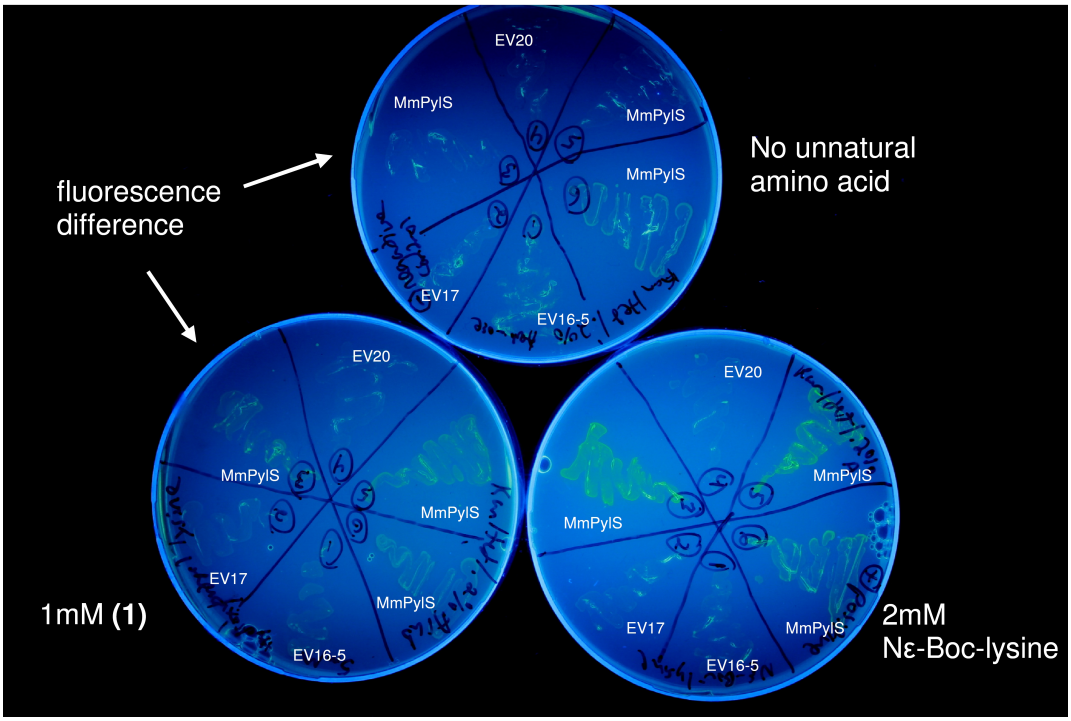


Figure 17 - Screening of 2-hydroxyisobutyryl lysine; plates

Numbers	PylRS synthetases[a]	Mutations								
		276 (241)*	302 (267)	306 (271)	309 (274)	346 (311)	348 (313)	384 (349)	401 (366)	417 (382)
1	WT	M	A	Y	L	N	C	Y	V	W
2	BocLys	M	A	Y	L	N	C	F	V	W
3	CbzLys	M	A	A	L	N	C	F	V	W
4	OnbLys	M	A	M	A	N	A	F	V	W
5	Interm[b]	M	A	M	A	N	C	F	V	W
6	WT+	M	A	Y	L	A	A	Y	V	W
7	BocLys+	M	A	Y	L	A	A	F	V	W
8	CbzLys+	M	A	A	L	A	A	F	V	W
9	OnbLys+[c]	M	A	M	A	A	A	F	V	W
10	Interm+[c]	M	A	M	A	A	A	F	V	W
EV17(Mb)		M	A	M	A	N	C	F	V	W
EV16-5 (Mb)		M	A	M	A	A	C	F	V	W
EV2		F	S	C	M	N	C	Y	V	W
Chin PCC-1		M	A	Y	L	M	Q	Y	G	N
Chin PCC-2		M	A	Y	L	Q	A	Y	M	W

Table 2 – Detailed active site-mutations of MmpylS[25]

Note: [a] N346A/C348A mutations were incorporated in the five synthetases to create the "+" series of constructs.

[b] This was an intermediate construct when OnbLys synthetase was generated from BocLys synthetase.

[c] These two variants are containing the same mutations due to incorporation of N346A/C348A double mutations.

* numbers in parentheses symbolize the *M. bakeri* mutations

Protein expression of sfGFP in E. coli – 2-hydroxyisobutyryl

From the 4mL culture containing the pBK_MmPylS synthetase plasmid, 1mL was used to inoculate an expression cultures in 200mL of LB media containing tetracycline and kanamycin. After the culture reached an OD600=0.6, arabinose (final percentage of 0.2%) and 5mM 2-hydroxyisobutyryl lysine (**1**). A negative control culture was treated the same, with no unnatural amino acid added. The cultures were incubated overnight and then the cells pelleted at 6,000 x g, and re-suspended in 10mL of lies buffer (50mM NaH_2PO4, 300mM Nalco, 10mM imidazole, pH=8) with 5mg of lysozyme. After 30 minutes each culture was sonicated 3 x 1 minute at 50% power with a 30 second rest. The resulting pellets were centrifuged at 15,000 x g and the lysate transferred to a clean 15mL Falcon tube. 500μl of Promega His Bind was added to each tube and shaken for 30 minutes. The resin was pelleted by a centrifugation and sequentially washed 3 times with 12mL of lysis/binding buffer. The resin was eluted with 500μl elution buffer (50mM NaH_2PO_4, 300mM NaCl, 250mM imidazole, pH=8). Samples for analysis by SDS-PAGE were 20μL

Appendix III – MmpylS mutations for photo-caged amino acid selection

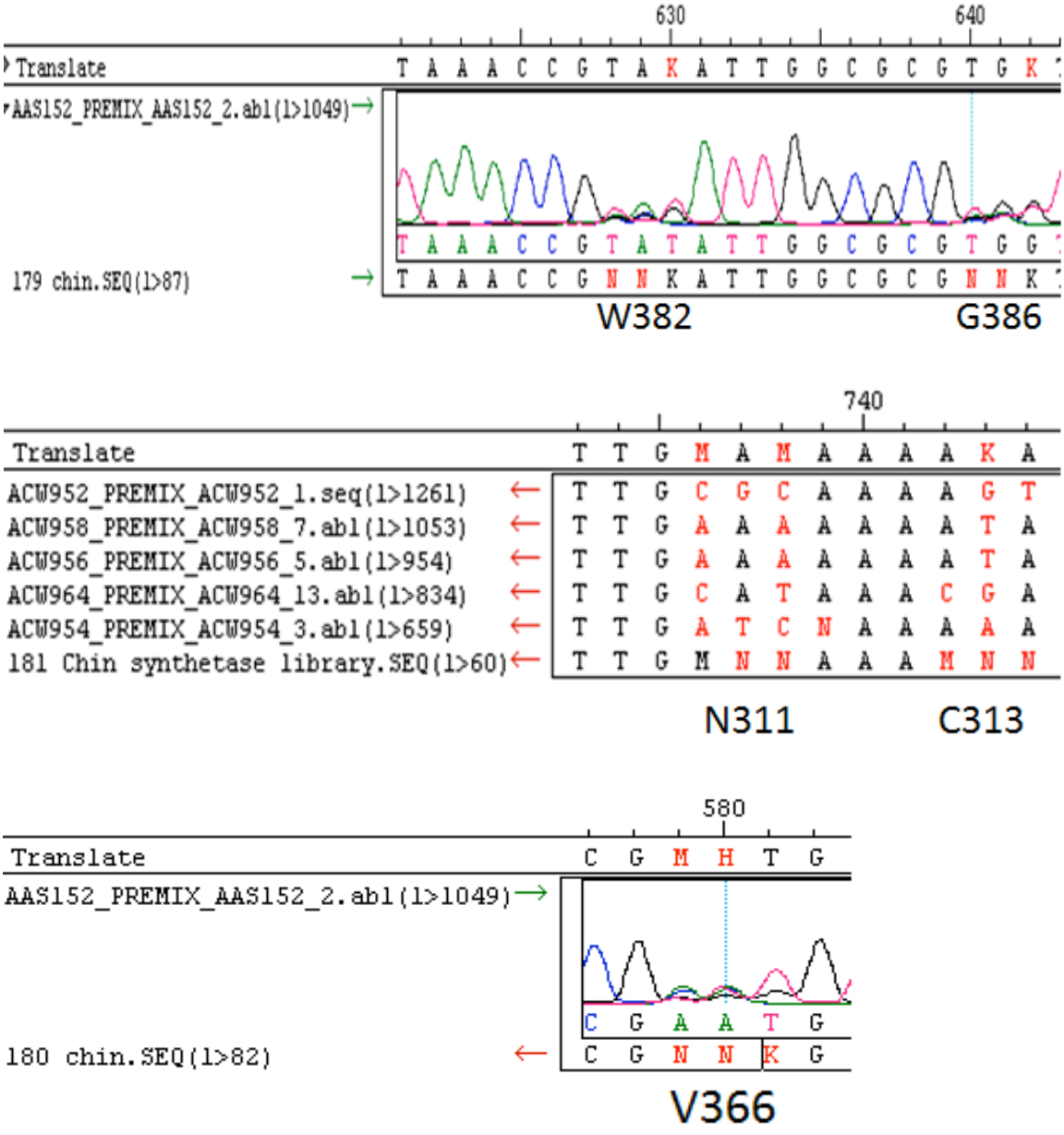


Figure 18 - MmpylS mutations for photo-caged amino acid catalog selection

Bibliography

(1) Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822.

(2) Wang, L.; Brock, a; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498–500.

(3) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232–234.

(4) Li, X.; Foley, E. a; Molloy, K. R.; Li, Y.; Chait, B. T.; Kapoor, T. M. *J. Am. Chem. Soc.* **2012**, *134*, 1982–1985.

(5) Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W. *J. Am. Chem. Soc.* **2011**, *133*, 10708–10711.

(6) Chi, P.; Allis, C. D.; Wang, G. G. *Nat. Rev. Cancer* **2010**, *10*, 457–469.

(7) Polycarpo, C. R.; Herring, S.; Bérubé, A.; Wood, J. L.; Söll, D.; Ambrogelly, A. *FEBS Lett.* **2006**, *580*, 6695–6700.

(8) Yu, S.-H.; Boyce, M.; Wands, a. M.; Bond, M. R.; Bertozzi, C. R.; Kohler, J. J. *Proc. Natl. Acad. Sci.* **2012**, *109*, 4834–4839.

(9) Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 14306–14307.

(10) Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W. *J. Am. Chem. Soc.* **2014**, *136*, 2240–2243.

(11) Dai, L.; Peng, C.; Montellier, E.; Lu, Z.; Chen, Y.; Ishii, H.; Debernardi, A.; Buchou, T.; Rousseaux, S.; Jin, F.; Sabari, B. R.; Deng, Z.; Allis, C. D.; Ren, B.; Khochbin, S.; Zhao, Y. *Nat. Chem. Biol.* **2014**, *10*, 365–370.

(12) Pelliccioli, A. P.; Wirz, J. *Photochem. Photobiol. Sci.* **2002**, *1*, 441–458.

(13) Cornish, V. W.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, 435–462.

(14) Neumann, H.; Peak-chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 2007–2009.

(15) Castañeda, C.; Liu, J.; Chaturvedi, A.; Nowicka, U.; Cropp, T. A.; Fushman, D. *J. Am. Chem. Soc.* **2011**, *133*, 17855–17868.

(16) Heintzman, N. D.; Stuart, R. K.; Hon, G.; Fu, Y.; Ching, C. W.; Hawkins, R. D.; Barrera, L. O.; Van Calcar, S.; Qu, C.; Ching, K. a; Wang, W.; Weng, Z.; Green, R. D.; Crawford, G. E.; Ren, B. *Nat. Genet.* **2007**, *39*, 311–318.

(17) Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G. *Angew. Chemie - Int. Ed.* **2006**, *45*, 2728–2731.

(18) Schaper, K.; Etinski, M.; Fleig, T. *Photochem. Photobiol.* **2009**, *85*, 1075–1081.

(19) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838–849.

(20) Baker, A. S.; Deiters, A. *ACS Chem. Biol.* **2014**, *9*, 1398–1407.

(21) Cera, E. Di. *IUBMB Life* **2009**, *61*, 510–515.

(22) Blow, D. M. *Nature* **1969**, *224*, 177–178.

(23) Lemke, E. a; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. *Nat. Chem. Biol.* **2007**, *3*, 769–772.

(24) Wang, J.; Okada, Y.; Li, W.; Yokoi, T.; Zhu, J. *J. Chem. Soc. Perkin Trans. 1* **1997**, 621–624.

(25) Liu, J. The development and application of transposon mutagenesis and unnatural amino acid mutagenesis technologies, University of Maryland, 2012.

Vita

William Arthur Knight was born on October 7,1986 in Richmond, Virginia. Graduated from James River High School in 2005. He holds a Bachelor of Science degree in Biological Sciences and a minor in Chemistry from Virginia Polytechnic Institute and State University (Virginia Tech) from May, 2009.