PROTEIN ENGINEERING: AMENDING PROTEINS WITH CHEMICAL FUNCTIONALITIES

A Dissertation

by

KETURAH AMARKIE ODOI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe Liu
Committee Members,	James Batteas
	Steve Lockless
	Tadhg Begley
Head of Department,	Simon North

December 2016

Major Subject: Chemistry

ABSTRACT

Directly incorporating noncanonical amino acids (NCAAs) into proteins in living cells is an indispensible technique for investigating structures and functions of proteins. This co-translational insertion of NCAAs, which is similar to post-translational modifications (PTM), amends proteins with new functionalities. With the idea of synthesizing proteins with defined modifications to generate proteins with enhanced properties, the first goal was to assess the capacity of codon recognition of the original or derivative *Methanosarcina mazei* pyrrolysyl (Pyl)-tRNA synthetase (PylRS)-$tRNA^{Pyl}$ pair, and also expand the current NCAA-coding codons to sense or quadruplet codons. These systematic studies expounded on the cross recognition between nonsense or sense codons with various $tRNA^{Pyl}$ anticodons in *E. coli*.

One interesting result that was obtained showed that *E. coli* itself displays a relatively high opal suppression level; tryptophan (Trp) is incorporated at an opal mutation site. Although the PylRS-$tRNA^{Pyl}_{UCA}$ pair can be used to encode N^{ε}-tert-butoxycarbonyl-l-lysine (BocK) at an opal codon, the pair fails to suppress the incorporation of Trp at the same site. For the sense and quadruplet codons, their tRNA did not fully deliver BocK at the AGG and AGGA codon sites when co-expressed with PylRS in *E. coli*, hence contamination by natural amino acids were spotted.

Based on the above studies carried out, the next goal was to use the highly orthogonal pair, mutant PylRS-$tRNA^{Pyl}_{CUA}$, to genetically encode five tyrosine derivatives. These derivatives contain either a norbornene, trans-cyclooctene, or cyclooctyne entity

encoded into proteins for site specific protein labeling. These NCAAs were successfully labeled with a fluorescein tetrazine dye and a diaryl nitrilimine both *in vitro* and in living cells. Detailed kinetic analyses of inverse electron-demand Diels–Alder cycloaddition and nitrilimine-alkene/alkyne 1,3-diploar cycloaddition reactions were conducted and the reactions were applied for rapid protein bioconjugation. When reacted with a tetrazine or a diaryl nitrilimine, strained alkene/alkyne entities including norbornene, trans-cyclooctene, and cyclooctyne displayed rapid kinetics.

Lastly, the study of histone PTM is a crucial step in chromatin states and epigenetic regulations of gene expressions in all eukaryotes because of its complex network of possible PTM combinations which, to this day, is being deciphered. Here, the goal of the research is to profile different lysine sites on histones by mapping out the effects of epigenetic erasers such as HKDMs on the sites. Results showed catalytic activity of LSD1 on various lysine sites on H2A, H2B and H3 histone substrates using a peroxidase-coupled assay of the catalytic activity of LSD1.

DEDICATION

To my dear big brother

Abednego Amartey Ocloo-Lee

Thank you for your love and support from the very beginning. This is also your dream come true! May you continue to rest in perfect peace, I love you!

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Wenshe Liu, and my committee members, Dr. Batteas, Dr. Begley, Dr. Hilty, and Dr. Lockless, for their guidance and support throughout my graduate studies. To the faculty who have taught me a great deal of chemistry, to my teaching supervisors and to the staff members that have helped me in any way possible I say thank you!

I also thank all my lab colleagues from the Liu research lab group for all their intellectual contribution and brainstorming on the various research projects I have worked on. Not forgetting all the wonderful friends I made while at Texas A&M University. Thank you for making Texas A&M home away from home; these are friendships I will cherish for a lifetime.

To the various campus organizations that I have been a part of: Believers' LoveWorld (BLW) Campus Ministry–TAMU, National Organization for the Professional Advancement of Black Chemists and Chemical Engineers (NOBCChE), Ghanafuo Association-TAMU, TAMU Center for Teaching Excellence. Each organization gave me the opportunity to serve in different leadership roles, thank you for that privilege.

To my church family, Christ Embassy-Houston and Texas Region BLW Campus Ministry, thank you for all your encouragement and prayers, I know without them I would not have accomplished all that I have been able to accomplish.

A big thank you to my loving and supportive parents, siblings and close friends and family who have stood there with me throughout this whole journey cheering me on to the finish line, I am eternally grateful.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

Page

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Protein Biosynthesis

Protein biosynthesis that involves 20 canonical amino acids is a well-established system in living organisms. This occurs when the amino acid sequence in a protein is translated from its corresponding nucleotide sequence in mRNA. For every amino acid, a nucleotide triplet in mRNA known as the codon, codes for the amino acid. Several amino acids use more than one codon for their coding which results in 61 triplet codons being use to code for 20 canonical amino acids (Figure I-1). The other 3 triplet codons code for the protein translation termination, making it a total of 64 codons for the protein biosynthesis.[1]

For protein biosynthesis, translation occurs when an aminoacyl-tRNA synthetase (aaRS) acylates its cognate tRNA with an amino acid by esterification. Each aaRS is characterized under two classes based on their binding patterns when acylating their respective tRNAs. Class I aaRSs interact with their tRNA acceptors from the tRNA minor grove while class II enzymes associates with their tRNA acceptors from the major groove.[2] Of the 20 amino acids, methionine, valine, isoleucine, cysteine, glutamate, leucine, lysine, arginine, tryptophan and tyrosine all belong to class I and the rest of the 20 amino acids belong to class II. To acylate the tRNA, the amino acid must first be

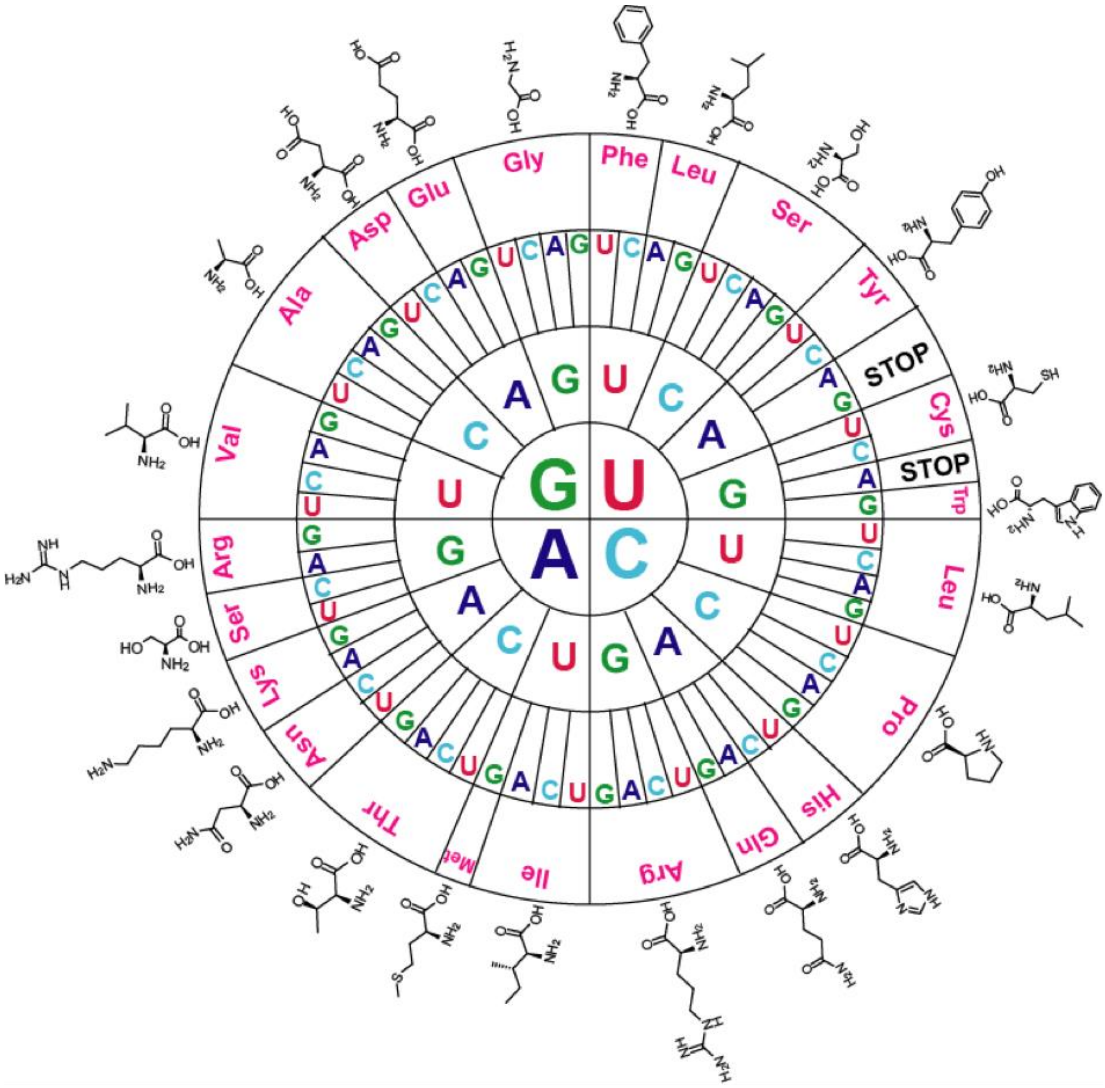


Figure I-1. The standard genetic code.

activated by ATP to produce the aminoacyl-AMP. In the case of class I aaRSs, the amino acid acylation occurs at the 2' OH of the ribose of the last nucleotide (A76) of tRNAs while for class II aaRSs the acylation occurs at the 3' OH.[3] This occurs because the aaRS possesses an ATP binding site, amino acid activation site, and tRNA binding domain.[2]

After acylation, the tRNA is transferred to the ribosome where protein synthesis actually takes place. The ribosome which carries out the translation has two main units which differs in prokaryotes and eukaryotes. Over at the ribosome site, the anticodons matches with their complementary codons through base-pairing. This shows that the tRNA has two distinct ends; one binds to its corresponding amino acid and the other end binds to its matched mRNA codon. Through the base-pairing of a particular anticodon with its successive codon, the amino acid is linked together to extend the growing peptide chain.

Protein Post-translational Modification

As previously mentioned, there are 20 canonical amino acids that form the building blocks of proteins. It is in the specific sequence of the amino acids that allows the protein to fold in a particular three-dimensional shape to carry out several cellular processes such as DNA replication, transcription, translation, metabolism etc. Many of these proteins require additional modifications on their amino acids via a process called post-translational modification (PTM) to carry out their native functions in the cells. Many of these modifications occur covalently on the side chains of the amino acids. The distinct functional groups on the amino acids gives it the name noncanonical amino acids

(NCAAs). These modifications include acetylation, methylation, phosphorylation, sulfation, glycosylation, hydroxylation, lipidation, ubiquitination etc on amino acids such as lysine, arginine, tyrosine etc (Figure I-2). These modifications expand the functions of the protein such as regulating signal transduction, controlling gene expression etc.

Discovery of Two New Amino Acids

Aside the known 20 amino acids, two other amino acids have been found: selenocysteine and pyrrolysine (Figure I-3). Selenocysteine (Sec) is the 21st amino acid found in 1976 in *Clostridium stricklandii* glycine reductase where Stadtman and her group found out that the selenium moiety on the clostridial glycine reductase selenoprotein is in fact from a Sec residue.[4] Sec is structurally similar to cysteine (Cys) with the exception of selenium instead of the sulfur group in cysteine. Its pKa (5.47) is lower than Cys. On the other hand, the 22nd amino acid, Pyrrolysine (Pyl), was discovered in 2002 in methylamine methyltransferase in *Methanosarcina barkeri*.[5, 6]

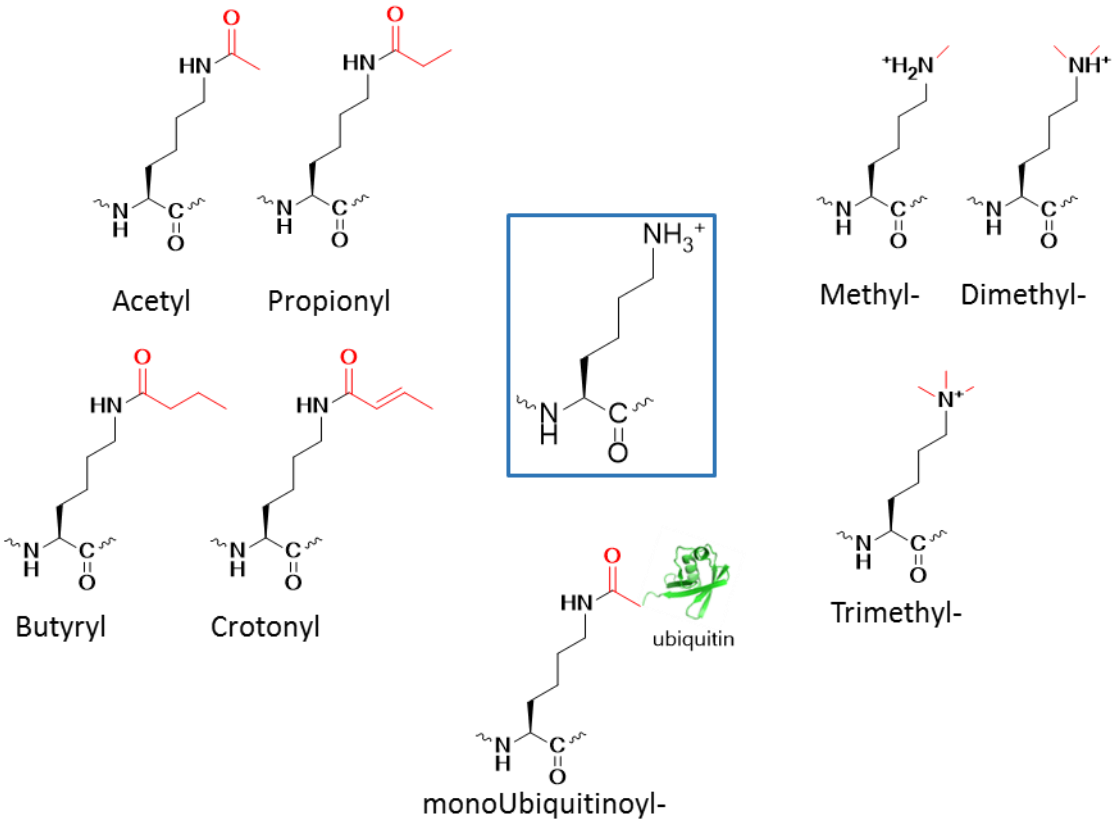


Figure I-2. Lysine post-translational modifications on proteins.

Selenocysteine Incorporation System

Selenium is a well-known trace element essential for cell growth and development but can be toxic to cells when in high dosage. It is a part of the glutathione cycle which is shown to detoxify free radicals by activating the key enzyme, glutathione peroxidase to protect against oxidative damage.[7, 8] The first sequence of selenoprotein sparked a great interest in Sec which is found in the catalytic site of glutathione peroxidase.[9] Earlier research demonstrated that Sec was incorporated post-translationally but a decade after, results showed that it was actually found to be co-translationally incorporated into proteins by using an in-frame UGA opal stop codon.[10, 11] Selenoprotein biosynthesis is now known to occur in all three domains of life: eukaryotes, prokaryotes, and archaea.[12, 13]

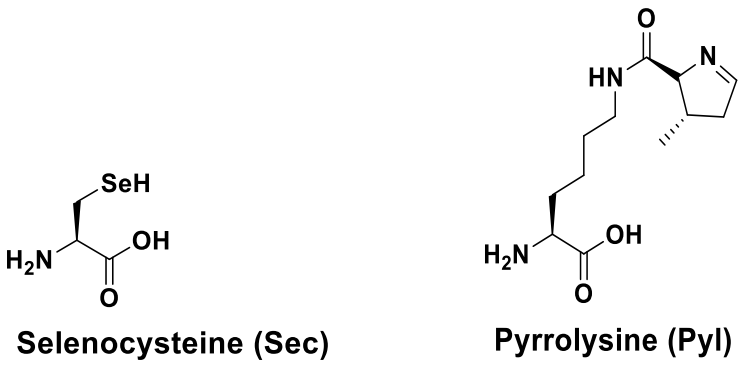


Figure I-3. Structure of Selenocysteine and Pyrrolysine.

Sec has its own tRNA, Selenocysteinyl tRNA ($tRNA^{Sec}$), but lacks its own aaRS. The $tRNA^{Sec}$ is used to synthesis selenocysteine-containing proteins in bacteria. This occurs indirectly when SelC gene encodes $tRNA^{Sec}$ followed by aminoacylation of a non-cognate amino acid, L-serine, by a class II aaRS, seryl-tRNA synthetase (SerRS), on the $tRNA^{Sec}$.[14] The Seryl- $tRNA^{Sec}$ is then converted to Sec-$tRNA^{Sec}$ using selenocysteine synthase (SelA), a PLP-dependent enzyme.[15]

Selenium activated by one ATP molecule is obtained from selenophosphate synthase (SelD) to produce selenophosphate.[16, 17] To avoid translational termination at the UGA opal stop codon, a specific elongation factor, selB, is required.[18] Selenocysteine insertion sequence (SECIS), cis -acting mRNA element which possesses a stem-loop structure lies just downstream from the UGA opal stop codon.[19] This element leads the cell to translate the opal stop codon as selenocysteines (Figure I-4).[13, 20]

Unlike the biosynthesis of Sec-$tRNA^{Sec}$ in bacteria, the Ser-$tRNA^{Sec}$ in archaea and eukaryotes is phosphorylated by O-phosphorylseryl-$tRNA^{Sec}$ kinase (PstK) to form Sep-$tRNA^{Sec}$ after the $tRNA^{Sec}$ has been first aminoacylated with Ser.[21-23] Sep-tRNA:Sec-tRNA synthase (SepSecS), a PLP dependent enzyme, modifies the *O*-phosphoserine (Sep) to Sec on the tRNA (Figure I-5).[23, 24] Compared to the bacteria, eukaryotes also require the SECIS element for the UGA opal stop codon to incorporate Sec instead of terminate the protein. The difference between them is the location of the SECIS. In eukaryotes, it is located at the 3` untranslated region of the mRNA.[25] The distance between the UGA opal stop codon and this element is naturally variable

between 500 and 5300 nucleotides. Due to the distance, a special elongation factor (EFSec) and a SECIS-binding protein (SBP2) form a complex to facilitate the binding of SECIS with Sec-tRNA^{Sec} for the correct recognition of an UGA opal codon in mRNA for eukaryotes.[26, 27] In archaea, for the Sec-tRNA^{Sec} to be recognized by the UGA opal stop codon, SelB must be present to assist SECIS in that process.[28-30]

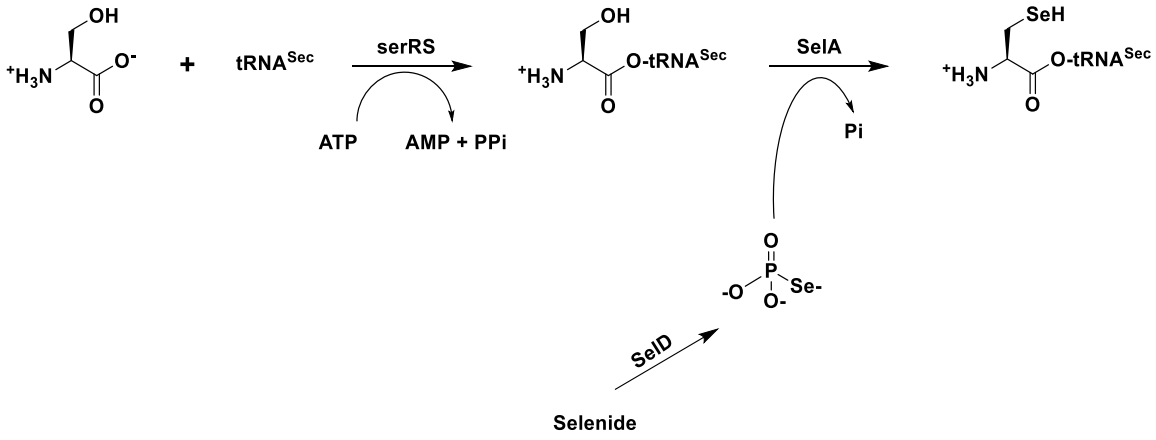


Figure I-4. Sec-tRNA^{Sec} biosynthesis in bacteria.

Pyrrolysine Incorporation System

Two possible routes were proposed for the biosynthesis of the Pyl. One route is by post-translational modification occurring on a Lys and the other route is co-translationally incorporate Pyl. Studies showed that Pyl is actually co-translationally inserted by its own pyrrolysyl-tRNA, $tRNA^{Pyl}$ (pylT) and its own pyrrolysyl-tRNA synthetase (PylRS) which is a class II aminoacyl-tRNA synthetase. Pyl is coded by a UAG amber stop codon in the open reading frame of monomethylamine methyltransferase (MtmB) in *Methanosarcina barkeri,* methane producing species that does not prevent translation of the full-length protein (Figure I-6). [5, 31, 32] Others have been discovered in six archaeal genomes which belong to the *Methanosarcinaceae* family: *Methanosarcina acetivorans*, *Methanosarcina mazei*, the cold-adapted *Methanococcoides burtonii*, halophilic methanogens *Methanohalobium evestigatum*, *Methanohalophilus mahii*, and *Methanosalsum zhilinae*.[33]

In bacteria, the PylRS and pylT is found in the class of Clostridia: *Desulfitobacterium hafniense, Desulfotomaculum acetoxidans, Acetohalobium arabaticum, Thermincola sp. JR*; and the δ- proteobacteria *Bilophila wadsworthia* and metagenomeidentified δ-proteobacterial symbiont of gutless worm *Olavius algarvensis*.[32-34]

For *M. barkeri*, the pylT gene is among a gene cluster (pylT, PylRS, pylB, pylC, pylD) to synthesize Pyl.[32] The pylT varies but the similarities include the three-base variable loop, small D-loop, long anticodon stem etc.[32] The genes coding for pylRS was

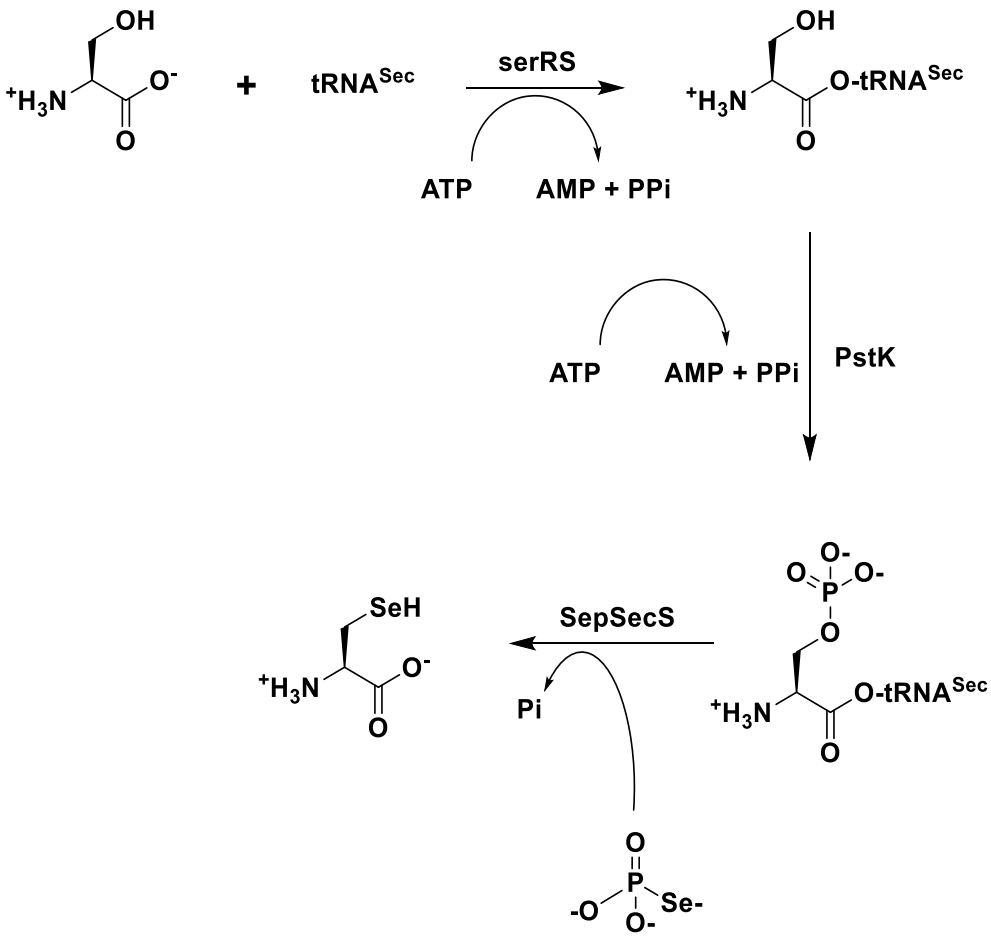


Figure I-5. Sec-tRNA^{Sec} biosynthesis in eukaryotes.

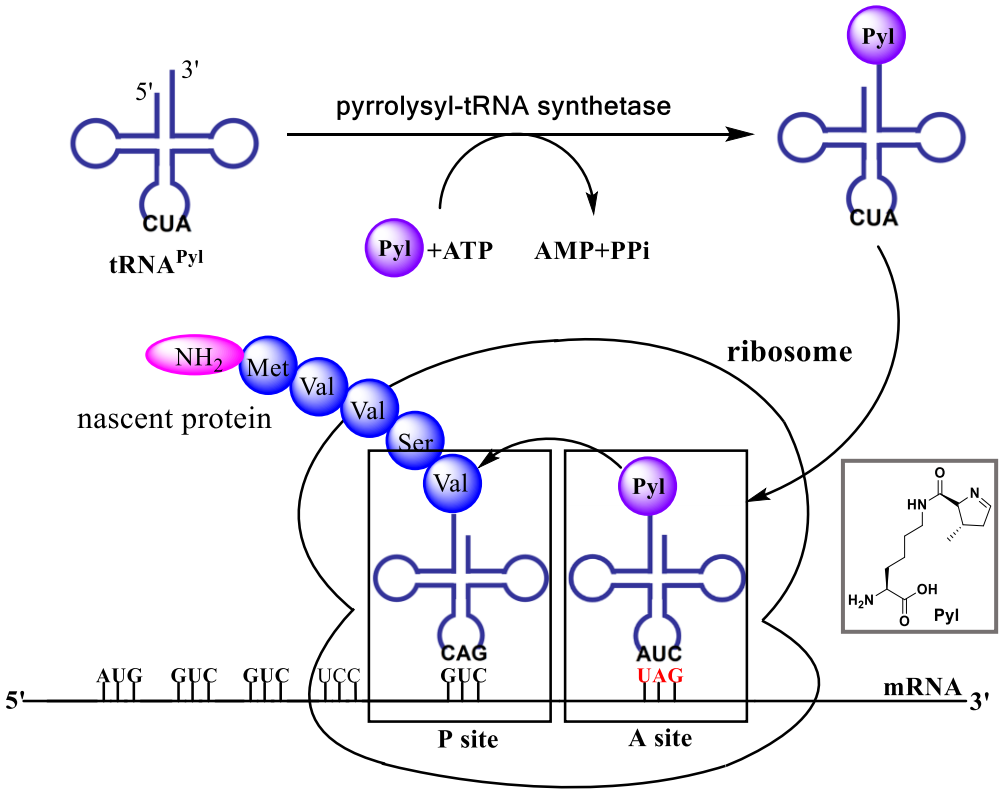


Figure I-6. The pyrrolysine incorporation machinery.

found to be in two domains: C-terminal (PylRSc) which is the catalytic domain of the PylRS and N-terminal (PylRSn). PylRSc is highly conserved in different *Methanosarcina* strains unlike the N-terminal domains which vary in length.[35, 36] The catalytic core has a unique fold structure with typical β sheets mixed with α helices which forms a binding for ATP. [3] The synthesis of the Pyl amino acid include the conversion of lysine to 3 –methylornithine by the SAM-dependent enzyme, pylB. Another lysine is added on by pylC, then pylD promotes the oxidation to yield Pyl[37] (Figure I-7).

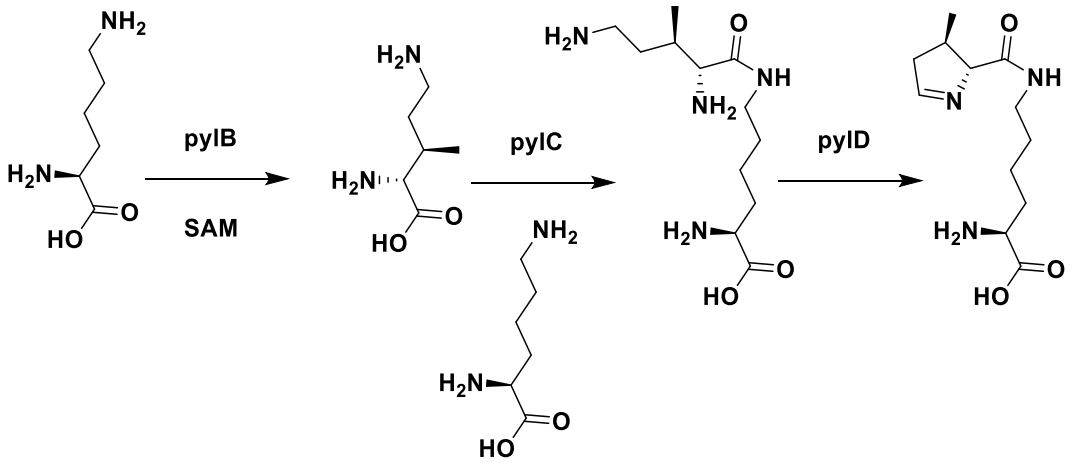


Figure I-7. Biosynthetic pathway of Pyrrolysine**.**

Studies showed that for the crystal structure of *M. mazei* PylRS, Pyl was present in the active site, where its side chain is bound to the deep hydrophobic binding pocket (A302, L305, Y306, L309, N346, C348, and W417). The pocket provides a bulky cavity to house the long side chain. The side chain was determined to be a (4R,5R)-4-methyl-pyrroline-5- caboxylate, with the carboxylate in the amide linkage to the Lys side chain. The hydroxyl group of Tyr384 forms a hydrogen bond with nitrogen of the pyrrole ring.[6, 36, 38] Another interaction involves of the hydrogen bonding between the nitrogen of Asn346 and the Pyl side chain oxygen.[37] From the above interaction between the PylRS and Pyl, it shows the hydrophobic interactions between the binding pocket and the pyrroline region. NCAA substrates similar to Pyl containing large hydrophobic side chains may fit the binding pocket due to the bulky cavity of PylRS (Figure I-8).

Aside the non-specific hydrophobic interactions with other PylRS substrates, PylRS does not possess an editing domain as compared to other aaRS. Editing domains provides high translation fidelity by preventing misacylation of its tRNAs. The lack of an editing domain in PylRS makes it promiscuous and allows misacylated $tRNA^{Pyl}$ possible to incorporate the linked NCAA at the UAG amber stop codon.

Polycarpo *et al*. demonstrated that N^{ε}-cyclopentyloxycarbonyl-L-lysine (Cyc), N^{ε}-D-prolyl-L-lysine (D-prolyl-lysine) are charged with PylRS but cannot recognize lysine. This gave insight on the structure-activity relationship between Pyl and its synthetase, PylRS. The pyrrolysine ring on Pyl plays a crucial role during its activation by PylRS. For instance, the Pyl analog D-prolyl-lysine was recognized by PylRS but not in the case of analogLL-prolyl-lysine because of the C-5 stereocenter on the pyrrolysine

ring.[39, 40] Additionally, Li *et al*. also used different Pyl analogs to determine the interactions necessary for recognition of PylRS substrates. These analogs include 2-amino-6-((R)-tetrahydro-furan-2-carboxamido) hexanoic acid (2-Thf-lys), and 2-amino-6-(cyclopentanecarboxamino)hexanoic acid (Cpn-lys)[41] (Figure I-9).

Not only does PylRS show low selectivity towards the side chain of the Pyl, it also has low selectivity towards the tRNA anticodon meaning the anticodon region has no direct interactions with PylRS. The PylRS has a unique tRNA-binding domain, a special C-terminal, tail, a loop region etc. Mutating the anticodon on the pylT does not affect the aminoacylation by the PylRS.

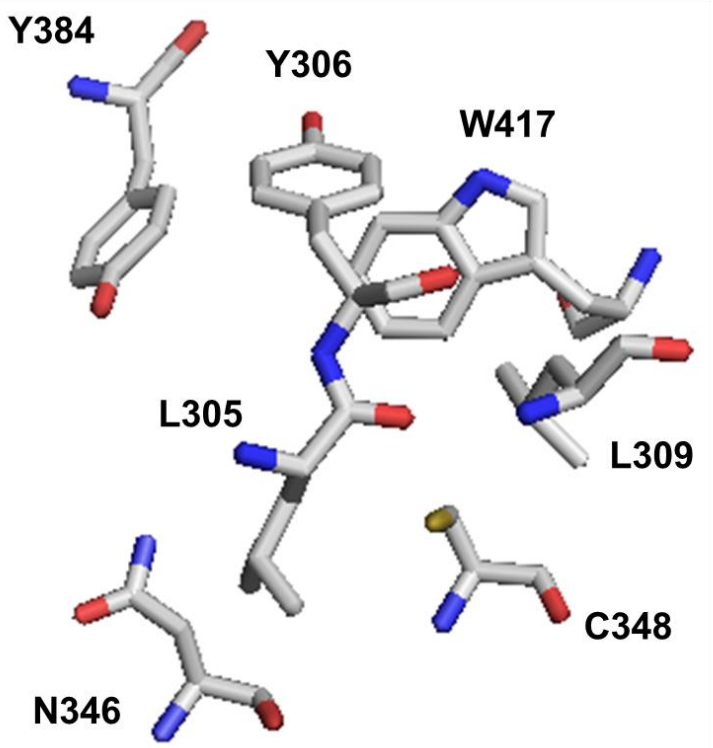


Figure I-8. The active site of PylRS. Structure obtained from PDB entry: 2Q7E

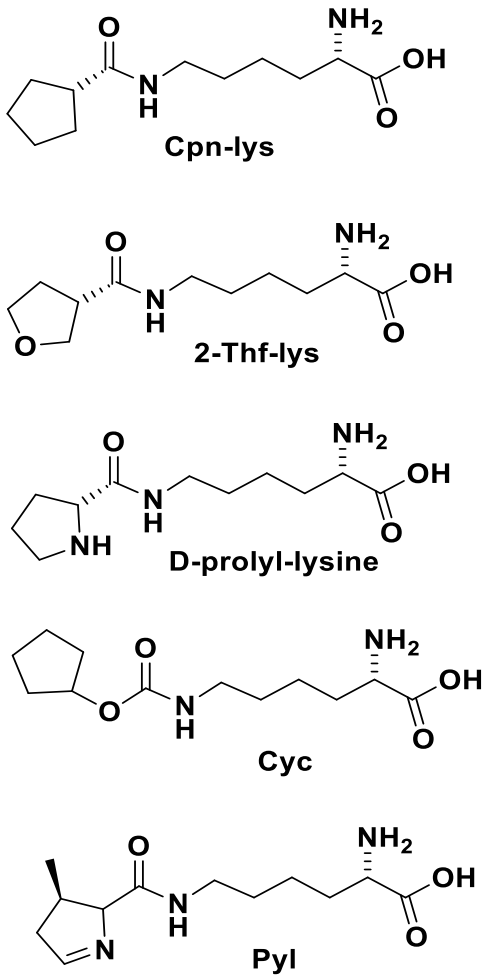


Figure I-9. Pyrrolysine and its analogs activated PylRS.

Synthesis of NCAAs into Proteins

As previously mentioned, many proteins require additional modifications to provide their native functions that goes beyond the 20 amino acids.[42] These covalent modifications such as acetylation, methylation, and phosphorylation occurs on the side chain of the amino acid. Due to the mixture of these modifications, it is a challenge to study their function. A homogeneous protein with specific modification(s) is needed to understand the function of a particular PTM in cellular activities. Several approaches such as cysteine-mediated native chemical ligation and expression protein ligation have been used to install several NCAAs into proteins.[43, 44] Though these methods have been successfully used to incorporate NCAAs, there are a few drawbacks associated with the methods. For instance, the semi synthetic techniques is limited to *in vitro* studies and cannot be applied to *in vivo*. Considering the fact that many proteins have over 100 residues, it is not feasible to synthesize these using the solid phase peptide synthesis because of the tedious process and low yield.[45]

Proteins bearing special chemical groups such as alkenes, alkynes, hydrazine groups are also a huge focus in research because of its protein labeling properties under physiological conditions via click-chemistry, copper-catalyzed azide-alkyne cycloaddition to study protein functions such as protein folding, protein-protein interactions etc.[46-50] Our group has attempted to build proteins with unique functionalities for biophysical and biochemical studies by genetically incorporating the NCAAs into proteins *in vivo* using the PylRS-$tRNA^{Pyl}$ pair.

Genetic Code Expansion Tool

To genetically incorporate NCAAs into proteins, one of the factors that needs to be addressed is the orthogonality of the aaRS-tRNA pair. To consider a pair as orthogonal, the pair must not cross-interact with the endogeneous aaRS-tRNA pair. The tRNA should not be recognized by any endogenous aaRS or decode any codon except one corresponding to its anticodon and assigned to its amino acid. Also, the aaRS should not acylate any endogenous tRNA except its orthogonal tRNA with the amino acid associated.

The archaea PylRS-$tRNA^{Pyl}$ pair is orthogonal in both bacterial and eukaryotic cells hence evolving the pair is made possible.[38] Over a decade, the wild-type and evolved PylRS-$tRNA^{Pyl}$ pair have used the UAG amber stop codon to site specifically incorporate a variety of NCAAs into proteins in different cells *Escherichia coli (E. coli)*, *Saccharomyces cerevisiae*, mammalian cells etc. As of now over 100 different NCAAs with the natural PTMs and unique biofunctionalities such as ketone, strained alkene and alkyne, azide, hydroxylamine, aldehydes, phenylhalides, etc have been incorporated into the proteins.[32, 51, 52]

Genetic Incorporation of NCAAs

Several aaRS-tRNA system have been successfully used to genetically incorporate NCAAs into proteins. A suppressor $tRNA^{Tyr}$ and mutant tyrosyl-tRNA synthetase (TyrRS) pair from *E. coli* was evolved to incorporate 3-iodo-l-tyrosine into proteins in mammalian cells.[53] Another TyrRS-$tRNA^{Tyr}$ from *Methanococcus jannaschii* TyrRS-$tRNA^{Tyr}$ incorporated *O*-methyl-l-tyrosine.[54] Yeast suppressor

tRNA^{Phe} and phenylalanyl-tRNA synthetase incorporates p-fluoro-phenylalanine in *E. coli*.[55] *B. subtilis* TrpRS-tRNA^{Trp} pair has been used to selectively introduce 5-hydroxytryptophan (5-HTTP) into proteins in 293T cells.[56] For all the mentioned aaRS-tRNA pair, there was no recognition between the aaRS and the endogeneous tRNAs, while the endogeneous aaRSs did not recognize the suppressor tRNA.[57]

Now the use of the PylRS-tRNA^{Pyl} has gained great attention because of its orthogonality in diverse species. The main requirement is that the pair found in bacteria should be orthogonal to the pairs in archaea and eukaryotes. In the case of the PylRS-tRNA^{Pyl} pair, it is orthogonal in both *E. coli* and eukaryotes making it a great tool for genetic code expansion. For example, in 2008, Yokoyama's group developed a site-specific incorporation of L-lysine derivatives such as N^{ε}-acetyllysine (AcK), N^{ε}-tert-butyloxycarbonyl-L-lysine (BocK), N^{ε}-benzyloxycarbonyl-L-lysine (ZK) into proteins in mammalian cells.[58] Pyrroline-carboxy-lysine (Pcl), a demethylated form of pyrrolysine was incorporated into the proteins in both *E. coli* and in mammalian cells using the native PylRS-tRNA^{Pyl} pair.[59] A cyclic pyrrolysine N^{ε} -(((1R,2R)-2-azidocyclopentyloxy)carbonyl)-L-lysine (ACPK) was also incorporated into protein using an evolved mutant PylRS-tRNA^{Pyl} pair.[60] Another evolved mutant pair was specific for the N^{ε}-photocaged lysine analogue, o-nitrobenzyloxycarbonyl-N^{ε}-l-lysine (ONBK).[61] With the ability to evolve the PylRS to specifically recognize specific NCAA by using selection occurs in the active site where the active site residues are randomly mutated to determine which evolved pair can take a particular NCAA. There has been more diversity in NCAAs because of the use wild-type or mutant PylRS. As of now,

over a number of 100 different kinds of amino acids has been found in the proteins due to the PTM processes that occur covalently (Figure I-10).

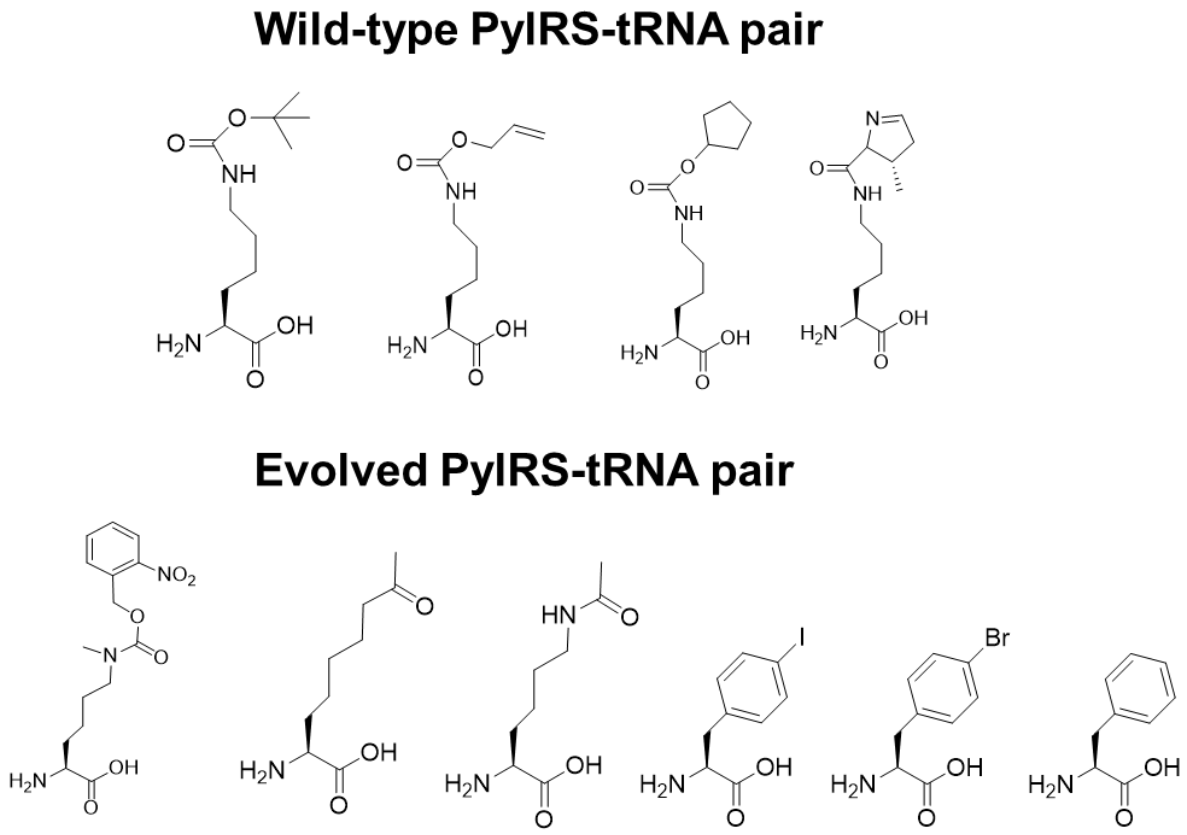


Figure I-10. Amino acid derivatives incorporated by PylRS-$tRNA^{Pyl}_{CUA}$.

For the research projects discussed in this writing, the PylRS-tRNA^{Pyl} pair for incorporation of diverse NCAAs into proteins in *E.coli* was utilized. Alternative codons were explored (ochre, opal, rare and even four base codons) and reassigned because of the non-interaction between PylRS and its tRNA anticodon.[62, 63] With the explored alternative codons, dual labeling can be set up to use two different aaRS-tRNA pair to incorporate two distinctive NCAAs for protein labeling to study protein folding or dynamics via Förster resonance energy transfer (FRET) studies.

Another work discussed here is a detailed kinetic analyses of inverse electron-demand Diels–Alder cycloaddition and nitrilimine-alkene/alkyne 1,3-diploar cycloaddition reactions conducted when a tetrazine or a diaryl nitrilimine is reacted with strained alkene/alkyne entities including norbornene, *trans*-cyclooctene, and cyclooctyne to display rapid kinetics. These unique strained alkyne and alkyne functionalities were incorporated using a mutant PylRS.[48] Lastly, the synthesis of histones with post-translational lysine methylation for probing substrate specificities of histone demethylases are shown.

CHAPTER II

ALTERNATIVE CODON STUDY FOR GENETIC CODE EXPANSION IN *ESCHERICHIA COLI**

Introduction

Pyrrolysine (Pyl, Figure II-1), the 22nd proteinogenic amino acid originally discovered in methanogenic methylamine methyltransferase, was genetically encoded by the RNA nucleotide triplet UAG, a stop codon that halts translation of mRNA during a regular protein translation process[6]. The delivery of Pyl to ribosome was mediated by a unique tRNA, $tRNA_{CUA}^{Pyl}$, that is specifically aminoacylated by a unique aminoacyl-tRNA synthetase (aaRS), pyrrolysyl-tRNA synthetase (PylRS)[32] (Figure I-7).

$tRNA_{CUA}^{Pyl}$ contains a special nucleotide triplet CUA, an anticodon that recognizes a UAG stop codon in mRNA. Unlike $tRNA^{Sec}$ that needs a special elongation factor (SelB in *E. coli* and EFsec in mammalian cells) and an mRNA secondary structure for its binding to the ribosome A site and recognition of a UGA stop codon for the delivery of selenocysteine, $tRNA_{CUA}^{Pyl}$ hijacks the regular translation elongation process to suppress a UAG codon for the incorporation of Pyl[64-66].

---

*Reprinted from "Nonsense and sense suppression abilities of original and derivative *methanosarcina mazei* pyrrolysyl-tRNA synthetase-$tRNA^{pyl}$ pairs in the *Escherichia coli* BL21(De3) cell strain", **Odoi, K. A**., Huang, Y., Rezenom, Y. H., Liu, W.R. *PLoS ONE*. 2013. 8, e57035. Copyright 2013, with permission from PLOS.
I performed all experiments except protein labeling of GFPuv with ProK.

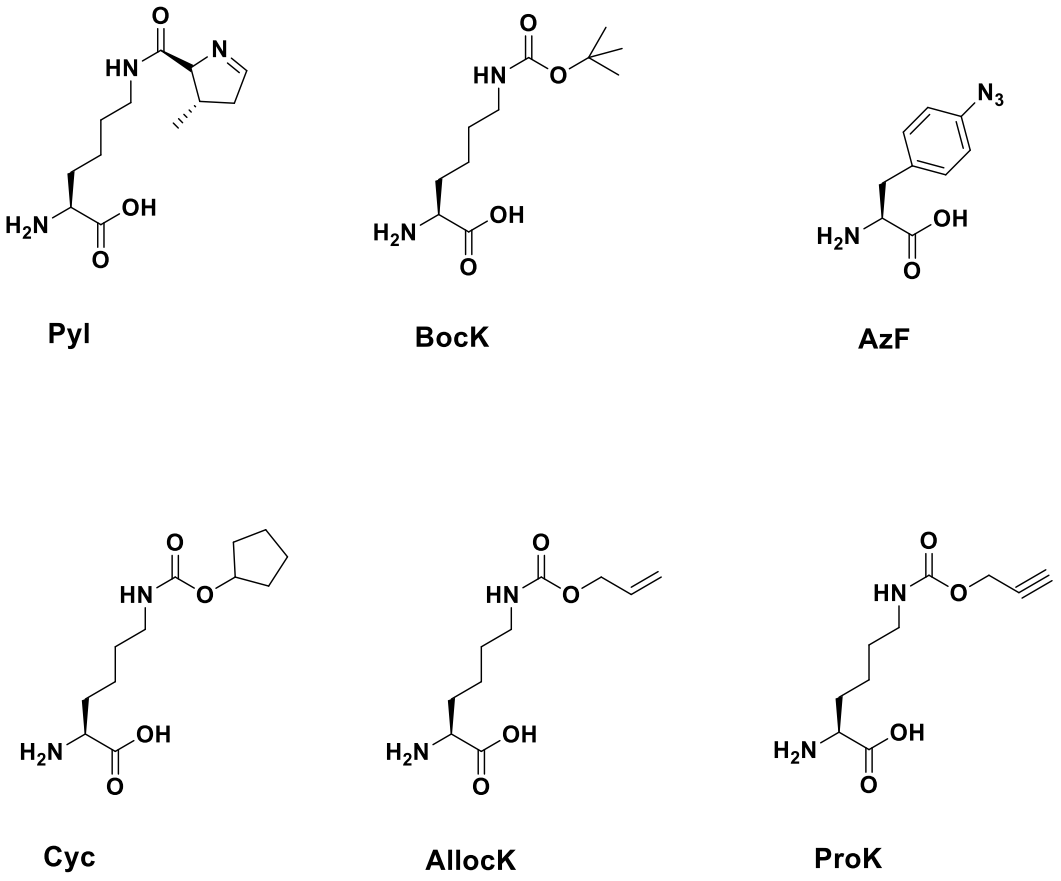


Figure II-1. Structures of NCAAs incorporated.

.

Previous studies also demonstrated that PylRS shows remarkably high substrate promiscuity charges $tRNA_{CUA}^{Pyl}$ with a variety of noncanonical amino acids (NCAAs). For these reasons and due to the naturally high orthogonality of the PylRS-$tRNA_{CUA}^{Pyl}$ pair in bacteria, yeast, and mammalian cells, this pair has been directly transferred to *E. coli*, *Saccharomyces cerevisiae*, and human cells for the genetic incorporation of more than ten lysine derivatives including N^{ε}-tert-butoxycarbonyl-L-lysine (BocK) into proteins at amber mutation sites[67-73].

Engineered PylRS-$tRNA_{CUA}^{Pyl}$ pairs have also been used to genetically encode other lysine derivatives and even phenylalanine derivatives that are structurally distinctive from Pyl[74-82]. The genetic incorporation of these NCAAs into proteins and their following modifications allow a variety of biochemistry studies such as the functional investigation of protein post-translational modifications, protein folding dynamic analysis, biosensor development, tracking signal transduction processes, and probing enzyme mechanisms[83]. Pioneered by Schultz and coworkers on the incorporation of two different NCAAs[84], two methods were recently developed independently in the Chin Group and our group that couple the wild-type or a derivative PylRS-$tRNA^{Pyl}$ pair with another suppressing aaRS-tRNA pair for the genetic incorporation of two different NCAAs into one protein in *E. coli*[85, 86]

The method developed in the Chin Group used one amber codon and one quadruple AGGA codon that were suppressed by the PylRS-$tRNA_{CUA}^{Pyl}$ pair and an evolved *M. jannaschii* tyrosyl-tRNA synthetase *Mj*TyrRS-$tRNA_{UCCU}^{Tyr}$ pair, respectively to

code two different NCAAs. A specially engineered ribosome, Ribo-Q1, was used to improve the AGGA suppression level. Our method relied on the suppression of two stop codons, namely one amber UAG codon and one ochre UAA codon, which was achieved by genetically encoding an evolved *Mj*TyrRS-$tRNA_{CUA}^{Tyr}$ pair for amber suppression and a wild type or evolved PylRS-$tRNA_{UUA}^{Pyl}$ pair for ochre suppression in *E. coli*. The genetic incorporation of two different NCAAs into one protein can be potentially applied to install a FRET pair in a protein for conformation and dynamic studies as demonstrated in a separate publication in our lab[87]. The synthesized proteins with two different post-translational modifications can also be used for functional investigation, building phage-displayed peptide libraries with the expanded chemical diversities.

Although the PylRS-$tRNA_{CUA}^{Pyl}$ pair has been used extensively for the genetic incorporation of different NCAAs in the past few years, two questions related to the pair have not been fully addressed. While our lab and other groups have showed that mutating the anticodon of PylRS-$tRNA_{CUA}^{Pyl}$ does not significantly affect its interaction with the catalytic domain of PylRS[88], further investigation need to be done to determine whether the mutant $tRNA^{Pyl}$ forms can be directly used for the genetic incorporation of NCAAs at an opal or ochre codon or even a quadruplet codon or a sense codon in *E. coli*. Another study is necessary to clarify whether an aminoacylated $tRNA_{UUA}^{Pyl}$ can lead to amber suppression since a UUA anticodon can recognize a UAG codon based on the wobble hypothesis[89]. In this study, two questions were addressed and all the experiments

were carried out in *E. coli* BL21(DE3) cell strain which has been broadly used for the genetic incorporation of NCAAs.

Experimental Section

Materials

Phusion high-fidelity DNA polymerase, T4 DNA ligase, T4 polynucleotide kinase, and restriction enzymes were purchased from New England Biolabs. Oligonucleotide primers were ordered from Integrated DNA Technologies. Ni-NTA superflow resins were purchased from Qiagen. All polymerase chain reactions (PCRs) were performed using Phusion high-fidelity DNA polymerase. BocK was purchased from Chem Impex. *p*-Azido-L-phenylalanine (AzF) (Figure II-1) was synthesized according to a revised literature procedure[90].

DNA and Protein Sequences

DNA Sequence: pylT

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

DNA Sequence: sfGFP wild type (different mutation sites are indicated in red)

Atggttagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtgatgtgaatggccataaattta
gcgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcaccaccggtaaactgccggttc
cgtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcatatgaaacgccatgatttcttt
aaaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaacccgtgcggaagt
taaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaagatggcaacattctgggtcataaac
tggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaactttaaaatccgtcac
aacgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggtgctgctgccgga

taatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttgttaccg

ccgcgggcattacccacggtatggatgaactgtataaaggcagccaccatcatcatcaccattga

Protein Sequence: sfGFP wild type (different mutation sites are indicated in red)

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKL

PVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY

KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGI

KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKR

DHMVLLEFVTAAGITHGMDELYKGSHHHHHH

DNA Sequence: GFPuv149AGG

atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtca

gtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggcca

acacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatg

cccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggt

gatacccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaactat

aactcacacagggtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagat

ggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtc

gacacaatctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacac

atggcatggatgaactctacaaagagctccatcaccatcaccatcactaa

Protein Sequence: GFPuv149AGG

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP

VPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYK

TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGI

KANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKR

DHMVLLEFVTAAGITHGMDELYKELHHHHHH

DNA Sequence: Methanosarcina mazei PylRS

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacac

cacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggact

gcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcac

aaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatcc

gttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgataccg

gtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactg

gtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactga

caggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctct

ctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccag

gttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgat

accgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagacccatgcttgctccaaacctttacaactacctgcg

caagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgctacagaaaagagtccgacggcaaagaac

acctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaaatcttgaaagcataattacggactt

cctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctatggggataccttgatgtaatgcacggagac

ctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttc

gggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggatttctaccaacctgtaa

Protein Sequence: Methanosarcina mazei PylRS

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

DNA Sequence: Methanosarcina barkeri PylRS

atggataagaagccactggatgttctgatttccgctactggtctgtggatgtctcgcactggtactctgcacaagatcaaacaccacgaagtatctcgctccaagatctacattgaaatggcttgtggtgatcacctggttgttaacaactcccgctcttgtcgcaccgctcgtgcgtttcgccaccacaaatatcgtaaaacctgcaaacgctgccgcgtgagcgatgaggatattaacaacttcctgactcgctccaccgagagcaagaactctgtgaaagttcgcgtagtttctgctccgaaagttaagaaagctatgccgaagtctgttagccgtgctccgaaaccgctggagaactctgtgtccgcgaaagcgagcaccaacaccagccgttctgttccgtctccagcgaaatctactccgaactctagcgtgccagcttccgctccagcgccgtctctgacccgctcccagctggaccgtgtagaggctctgctgtctccagaagataagatcagcctgaacatggcgaaaccatttcgtgagctggagccggagctggtgacccgccgcaagaacgacttccagcgtctgtacaccaacgaccgcgaagactacctgggcaaactggaacgcgacatcaccaaattcttcgtagatcgtggtttcctgg

agatcaaaagcccgatcctgatcccggccgaatatgtggaacgcatgggcatcaacaacgacactgagctgtctaaacaaatc

ttccgcgtagataagaacctgtgtctgcgtccgatgctggcgccgaccctgtataactacctgcgcaaactggaccgcatcctg

ccgggtccaatcaaaatcttcgaagttggtccgtgctatcgtaaagaaagcgatggtaaagagcatctggaagagttcactatgg

taaacttctgccagatgggttctggttgcacccgtgaaaacctggaggcgctgatcaaggaatttctggattatctggaaatcgac

tttgagattgtgggcgactcttgcatggtgtacggtgacactctggacattatgcatggtgatctggagctgtcttctgctgtagtg

ggcccagtttccctggatcgtgagtggggtatcgataaaccgtggattggtgcgggcttcggcctggagcgtctgctgaaagtg

atgcatggctttaagaacatcaaacgtgctagccgtagcgagtcttattacaatggcatctccactaacctgtaa

Protein Sequence: Methanosarcina barkeri PylRS

MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSC

RTARAFRHHKYRKTCKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAM

PKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNSSVPASAPAPSLTRSQLDRV

EALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDREDYLGKLERDITKF

FVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYNYL

RKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFCQMGSGCTRENLEALIK

EFLDYLEIDFEIVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPWI

GAGFGLERLLKVMHGFKNIKRASRSESYYNGISTNL

Construction of Plasmids

Construction of pETtrio-pylT-sfGFP with amber mutation

Plasmid pETtrio-pylT(UUA)-PylRS-MCS was derived from pPylRS-pylT-GFP1TAG149TAA[91] and carries a gene (a C34U mutant form of $tRNA^{Pyl}_{CUA}$) under control of the lpp promoter and the rrnC terminator, the wild type *Methanosarcina mazei*

PylRS gene under control of the glnS promoter and terminator, and multiple cloning sites including NcoI, NotI, SalI and KpnI targeted sites under control of the T7 promoter and terminator. To construct pETtrio-pylT(UUA)-PylRS-MCS, two oligonucleotide primers forward primer MCS-F 5'-AGCGCGGCCGCGTCGACGGTACCCTCGAGTCTGGTAAAG-3' and reverse primer MCS-R 5'- ATTGCGGCCGCCCATGGTATATCTCCTTCTTATACTTAAC-3' were used to undergo PCR to amplify pPylRS-pylT-GFP1TAG149TAA. The blunt-end PCR product was phosphorylated at its two 5' ends and directly ligated using T4 DNA ligase to form pETtrio-pylT(UUA)-PylRS-MCS

Plasmid pETtrio-pylT(UUA)-PylRS-sfGFP134TAG that carries an additional superfolder green fluorescent protein (sfGFP) gene with an amber mutation at N134 (sfGFP134TAG) was constructed by cloning the sfGFP134TAG gene to the NcoI and KpnI sites of pETtrio-pylT(UUA)-PylRS-MCS. The sfGFP134TAG gene was PCR amplified from pBAD-sfGFP134TAG using two oligonucleotide primers sfGFP134TAG-F 5'- GATATACCATGGTTAGCAAAGGTGAAGAACTG-3' and sfGFP134TAG-R 5'- CTCGAGGGTACCTCAATGGTGATGATGATGGTG-3'. The PCR for the sfGFP134TAG gene amplification was carried out in the following conditions: 10 µL of 5X Phusion buffer, 1 µl of 10 mM dNTP mix, 0.5 µL of each primer (sfGFP134TAG-F and sfGFP134TAG-R), 0.5 µL of template (pBAD-sfGFPN134TAG), 0.4 µL Phusion polymerase, and 37 µL of water to get 50 µL total volume. The cycling conditions are as follows: (Step 1) 94 ºC – 5 minutes, (Step 2) 95 ºC – 30 seconds, (Step 3) 50 ºC – 30 seconds, (Step 4) 72 ºC – 60 seconds, (Step 5)

Repeat Step 2, 29 times, Step 6) 72 ºC – 5 minutes. 1% agarose gel electrophoresis was used to confirm the sfGFP gene. QIAquick gel extraction kit was used to extract and clean the PCR product. The amplified gene was digested by NcoI and KpnI restriction enzymes and ligated to pETtrio-pylT(UUA)-PylRS-MCS that was predigested by the same enzymes.

Construction of pETtrio-pylT-sfGFP with Variant Codon and Anticodon Mutation

Three plasmids pETtrio-pylT(CUA)-PylRS-sfGFP134TAG, pETtrio-pylT(UUA)-PylRS-sfGFP134TAA, and pETtrio-pylT(UCA)-PylRS-sfGFP134TGA that vary at the anticodon of $tRNA^{Pyl}$ and have different nonsense mutations at N134 of the sfGFP gene with a C-terminal 6×His tag were derived from pETtrio-pylT(UUA)-PylRS-sfGFP134TAG. Constructions of these plasmids were carried out using a site-directed mutagenesis protocol that was based on Phusion DNA polymerase. In brief, two oligonucleotide primers, one of which covers the mutation site were used to amplify the whole plasmid of pETtrio-pylT(UUA)-PylRS-sfGFP134TAG to give a blunt-end PCR product. This PCR product was phosphorylated by T4 polynucleotide kinase and then self-ligated using T4 DNA ligase.

Plasmid pETtrio-pylT(CUA)-PylRS-sfGFP134TAG carries genes coding $tRNA^{Pyl}_{CUA}$, PylRS, and sfGFP134TAG; plasmid pETtrio-pylT(UUA)-PylRS-sfGFP134TAA carries genes coding $tRNA^{Pyl}_{UUA}$, PylRS, and sfGFP with an ochre mutation at N134 (sfGFP134TAA); and plasmid pETtrio-pylT(UCA)-PylRS-sfGFP134TGA carries genes coding $tRNA^{Pyl}_{UCA}$, PylRS, and sfGFP with an opal mutation at N134

(sfGFP134TGA). Plasmid pETtrio-pylT(CUA)-PylRS-sfGFP134TAA that carries genes coding $tRNA_{CUA}^{Pyl}$, PylRS, and sfGFP134TAA was derived from plasmid pETtrio-pylT(UUA)-PylRS-sfGFP134TAA. Primers pylT-F 5'-GTCCATTCGATCTACATGATCAGGTT-3' and pylT-TAG-R 5'-TCTAAATCCGTTCAGCCGGGTTAG-3' were used to construct pETtrio-pylT(CUA)-PylRS-sfGFP134TAG. Primers sfGFP134-F 5'- GGCAACATTCTGCATAAACTGGA-3' and sfGFP134TAA-R 5'- TTATTCTTTAAAATCAATACCTTTCAGTTCAATGC-3' were used to construct pETtrio-pylT(UUA)-PylRS-sfGFP134TAA Two set of primers: (1) pylT-F 5'- GTCCATTCGATCTACATGATCAGGTT-3' and pylT-TGA-R 5'- TTCAAATCCGTTCAGCCGGGTTAG-3'and (2) sfGFP134-F 5'-GGCAACATTCTGCATAAACTGGA-3' and sfGFP134TGA-R 5'-TCATTCTTTAAAATCAATACCTTTCAGTTCAATGC-3' were used to run two consecutive site-directed mutagenesis reactions to obtain plasmid pETtrio-pylT(UCA)-PylRS-sfGFP134TGA The pylT(UCA) gene sequence was confirmed using the primer PylT-SphI-F 5'-GGAATGGTGCATGCTCGAACTTTT-3'. Plasmid pETtrio-pylT(CUA)-PylRS-sfGFP134TAA was made from plasmid pETtrio-pylT(UUA)-PylRS-sfGFP134TAA by the same site-directed mutagenesis using two primers pylT-F 5'-GTCCATTCGATCTACATGATCAGGTT-3' and pylT-TAG-R 5'-TCTAAATCCGTTCAGCCGGGTTAG-3. *Duet-F* 5'-TTGTACACGGCCGCATAATC-3') was used to confirm each mutation at the 134 position on sfGFP gene.

Construction of pETtrio-sfGFP with Opal Codon Mutation

Plasmid pETtrio-sfGFP134TGA was derived from pETtrio-pylT(UCA)-PylRS-sfGFP134TGA by digesting it with SphI to remove the $tRNA_{UCA}^{Pyl}$ gene and parts of the PylRS gene and self-ligating the purified digested plasmid backbone. Plasmid pETtrio-pylT(UCA)-PylRS-sfGFP2TGA carries genes coding $tRNA_{UCA}^{Pyl}$, PylRS, and sfGFP with an opal mutation at S2 and a 6×His tag at its C-terminus (sfGFP2TGA). To construct this plasmid, the sfGFP2TGA gene was used to replace the sfGFP134TGA gene in pETtrio-pylT(UCA)-PylRS-sfGFP134TGA. the sfGFP gene in pBAD-sfGFP was PCR amplified using primers sfGFP2TGA-F 5'- AAAGGTGAAGAACTGTTTACCGGCGT-3' and sfGFP2TGA-R 5'- TCAAACCATGGTATATCTCCTTCTTAT-3'.

The PCR product was digested by NcoI and KpnI restriction enzymes and cloned into the same two sites of pETtrio-pylT(UCA)-PylRS-sfGFP134TGA which was precut by NcoI and KpnI restriction enzymes to remove the sfGFP134TGA insert. The sfGFP2TGA gene sequence was confirmed using the sfGFP134TAG-R 5'-CTCGAGGGTACCTCAATGGTGATGATGATGGTG-3'

Construction of pETtrio-pylT-sfGFP with Quadruplet Codon and Anticodon Mutation

Plasmid pETtrio-pylT(UCCU)-PylRS-sfGFP134AGGA was constructed from pETtrio-pylT(CUA)-PylRS-sfGFP134TAG.

Construction of pETtrio-pylT-sfGFP with Sense Codon and Anticodon Mutation

Plasmid pETtrio-pylT(CCU)-PylRS-sfGFP2AGG was constructed from pETtrio-pylT(CUA)-PylRS-sfGFP2TGA using two consecutive Phusion DNA polymerase-based

site directed mutagenesis steps. Primers pylT-F 5'-GTCCATTCGATCTACATGATCAGGTT-3' and pylT-AGG-R 5'-TCTTAATCCGTTCAGCCGGGTTAG -3' were used to construct the CCU anticodon on the pylT sfGFP2AGG-F 5'- AAAGGTGAAGAACTGTTTACCGGCGT-3' and sfGFP2AGG-R 5'-CCTAACCATGGTATATCTCCTTCTTATACTTAACTAATATACTAAG -3' were used to construct the AGG codon on the sfGFP at the S2 position.

Construction of pETDuet-mbPylRS-pylT(CCU)-GFPuv149AGG

Further studies were performed by Dr. Yu Zeng from my research group to make PylRS-$tRNA^{Pyl}_{CUA}$ competitively recognize AGG during translation to incorporate NCAA such as BocK, AllocK and ProK (Figure II-1). pETDuet-mbPylRS-pylTCCU-GFPuv149AGG was derived from pETDuet–AcKRS–pylTCUA-GFP149TAG which contains a pylT(CCU) gene under control of the lpp promoter and the rrnC terminator, the codon optimized *Methanosarcina barkeri (M.barkeri)* PylRS gene and GFPuv gene with an AGG mutation at position 149, both are under control of the T7 promoter and terminator. The standard QuikChange site-directed mutagenesis was used to generate PstI and AgeI restriction sites of pETDuet-AcKRS-pylTCUA-GFP149TAG (pDUET-pstI-QC-F, 5'-CTGTAACTGCAGGTCGACAAGCTTGCGGCC-3' and pDUETAgeI-QC-R, 5'-TGACATACCGGTGGTATATCTCCTTCTTAAAGTTAAAC-3'), then ligated with optimized *M.barkeri* PylRS (mbpylRS-AgeI-F, 5'-CTGTAAACCGGTATGGATAAGAAGCCACTGGATGTTC-3' and mbpylRS-PstI-R,

5’-GTCGACCTGCAGTTACAGGTTAGTGGAGATGCCATTG-3’ with corresponding restrictive sites to generate pETDuet -mbPylRS- pylT(CUA)-GFP$_{UV}$149TAG first.

The resulting plasmid then underwent the second QuikChange site-directed mutagenesis to mutate the TAG codon at position 149 of GFPuv to the AGG codon (GFPuv-AGG-QC-F, 5’-AGGGTATACATCACGGCAGACAAACAA-3’ and GFPuv-AGG-QC-R, 5’-GTGTGA GTTATAGTTGTACTCGAGTT-3’). And PylT(CUA) was replaced by PylT(CCU) from SphI sites of pETtrio-pylT(CCU)-PylRS-sfGFP2AGG.[92]

Construction of pETtrio-sfGFP with Variant $tRNA^{Pyl}_{UCA}$ Mutation on G73

Three plasmids pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA, pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA, and pETtrio-pylT(UCA)G73A-PylRS-sfGFP134TGA carry mutations that change G73 of $tRNA^{Pyl}_{UCA}$ to C, U, and A, respectively. These plasmids were derived from pETtrio-pylT(UCA)-PylRS-sfGFP134TGA using Phusion DNA polymerase-based site-directed mutagenesis. To make pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA, two primers pylT-G73-F 5’-GGAAACCCCGGGAATCTAACCCGG-3’ and pylT-G73C-R 5’-CCCACTGCCCATCCTTAGCGAAAGC-3’ were employed. Primers pylT-G73-F 5’-GGAAACCCCGGGAATCTAACCCGG-3’and pylT-G73U-R 5’-TCCACTGCCCATCCTTAGCGAAAGC-3’ were used to construct pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA. Primers pylT-G73-F 5’-GGAAACCCCGGGAATCTAACCCGG-3’and pylT-G73A-R 5’-

ACCACTGCCCATCCTTAGCGAAAGC -3’ were used to construct pETtrio-pylT(UCA)G73A-PylRS-sfGFP134TGA.

sfGFP Protein Expression and Purification

Amber, Opal, and Ochre Suppression

Plasmids pETtrio-pylT(CUA)-PylRS-sfGFP134TAG, pETtrio-pylT(UCA)-PylRS-sfGFP134TGA, and pETtrio-pylT(UUA)-PylRS-sfGFP134TAA were individually used to transform *E. coli* BL21(DE3) cells. For each plasmid, a single colony was selected and allowed to grow in 5 mL of LB medium with 100 µg/mL ampicillin at 37 °C overnight. The overnight culture was inoculated into 200 mL of 2YT medium with 100 µg/mL ampicillin and allowed to grow at 37 °C to OD_{600}~1.2. 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 5 mM BocK were then added to the medium to induce expression of sfGFP. Control experiments in which only 1 mM IPTG was added to the medium were also carried out. The induced cells were allowed to grow at 37 °C overnight and then collected by centrifugation (4,200 rpm for 20 minutes). The collected cells were resuspended in 35 mL of lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by sonication in an ice water bath. The lysed cells were clarified by centrifugation (10,000 rpm for 1 hour). The supernatant was decanted and let bind to 5 mL of Ni-NTA superflow resins at 4°C for 1 hour.

The mixture of the supernatant and resins was then loaded to an empty Qiagen Ni-NTA superflow cartridge. The resins were washed with 5 volume times of lysis buffer and sfGFP was then eluted with buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). To further purify the expressed sfGFP, the protein was equilibrated

against buffer A (20 mM Bis-Tris, pH 6.1) and then loaded to a monoS column from GE Health Science. The protein was eluted out by running a gradient from buffer A to 100% of buffer B (20 mM Bis-Tris, 1 mM NaCl, pH 6.1). The finally purified protein was then concentrated to a desired volume and analyzed by SDS-PAGE. To analyze the purified protein by electrospray ionization mass spectrometry (ESI-MS) analysis, the buffer of the purified protein was changed to the phosphate buffer saline. Without further indication, protein purification and characterization in the following experiments were the same.

Suppression of an Opal Mutation at S2 Position of sfGFP

Plasmid pETtrio-pylT(UCA)-PylRS-sfGFP2TGA was used to transform *E. coli* BL21(DE3) cells. A single colony was then used to express sfGFP at two induction conditions: (1) 1 mM IPTG and (2) 1 mM IPTG and 5 mM BocK.

Basal Suppression at an Opal UGA Codon

Plasmid pETtrio-sfGFP134TGA was used to transform *E. coli* BL21(DE3) cells. A single colony was then selected to do protein expression that was induced by the addition of 1 mM IPTG.

Mutagenic Analysis of $tRNA_{UCA}^{Pyl}$

Plasmids pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA, pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA, and pETtrio-pylT(UCA)G73A-PylRS-sfGFP134TGA were used individually to transform *E. coli* BL21(DE3) cells. A single

colony for each plasmid was then selected to do protein expression at two induction conditions: (1) 1 mM IPTG and (2) 1 mM IPTG and 5 mM BocK.

Anticodon-Codon Cross Recognition

Plasmids pETtrio-pylT(CUA)-PylRS-sfGFP134TAG, pETtrio-pylT(CUA)-PylRS-sfGFP134TAA, pETtrio-pylT(UUA)-PylRS-sfGFP134TAG, and pETtrio-pylT(UUA)-PylRS-sfGFP134TAA were individually used to transform *E. coli* BL21(DE3) cells. A single colony for each plasmid was then selected and allowed to grow in 5 mL of LB medium with 100 μg/mL ampicillin at 37 °C overnight. This overnight culture was then inoculated into 500 mL of 2YT medium with 100 μg/mL ampicillin and allowed to grow to OD_{600} ~ 1.2. 1 mM IPTG and 5 mM BocK were then provided and cells were allowed to grow at 37 °C overnight. For pETtrio-pylT(UUA)-PylRS-sfGFP134TAG, expression of sfGFP in the absence of BocK was also tested.

Competitive Recognition of the Third Nucleotide of an Amber Codon

Plasmid pEVOL-AzFRS was a gift from Dr. Peter Schultz at Scripps Research Institute[93]. It carries one $tRNA_{CUA}^{Tyr}$ gene under control of a *proK* promoter and a *proK* terminator, an evolved AzF-specific *Mj*TyrRS (AzFRS) gene under control of a *glnS* promoter and a *glnS* terminator, and an AzFRS gene under control of a pBAD promoter. This plasmid together with pETtrio-pylT(UUA)-PylRS-sfGFP134TAG was used to co-transform *E. coli* BL21(DE3) cells. One single colony was selected and allowed to grow in 5 mL of LB medium with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol at 37 °C overnight. This overnight culture was inoculated into 200 mL of 2YT medium with

100 µg/mL ampicillin and 34 µg/mL chloramphenicol and allowed to grow to OD_{600} ~ 1.2. Expression of sfGFP was then induced. Four induction conditions were tested, including (1) 1 mM IPTG only, (2) 1 mM IPTG and 1 mM AzF, (3) 1 mM IPTG and 5 mM BocK, and (4) 1 mM IPTG, 1 mM AzF, and 5 mM BocK.

AGGA Codon Suppression

Plasmid pETtrio-pylT(UCCU)-PylRS-sfGFP134AGGA was used to transform *E. coli* BL21(DE3) cells. A single colony was then selected and allowed to grow in 5 mL of LB medium with 100 µg/mL ampicillin at 37 °C overnight. This overnight culture was then inoculated into 500 mL of 2YT medium with 100 µg/mL ampicillin and allowed to grow to OD_{600} ~ 1.2. Two conditions were used to induce sfGFP expression. One is the addition of 1 mM IPTG and the other is the addition of 1 mM IPTG and 5 mM BocK.

AGG Codon Suppression

Plasmid pETtrio-pylT(CCU)-PylRS-sfGFP2AGG was used to transform *E. coli* BL21(DE3) cells. A single colony was then selected and allowed to grow in 5 mL of LB medium with 100 µg/mL ampicillin at 37 °C overnight. This overnight culture was then inoculated into 500 mL of 2YT medium with 100 µg/mL ampicillin and allowed to grow to OD_{600} ~ 1.2. Two conditions were used to induce sfGFP expression. One is the addition of 1 mM IPTG and the other is the addition of 1 mM IPTG and 5 mM BocK. The cells sued for the protein expression of the plasmid pET-pylT(CCU)-PylRS-GFP_{UV}149AGG were grown in minimal media to limit the amount of Arg present that will promote shifting of the AGG codon by endogenous arginyl-$tRNA^{Arg}$. pET-

pylT(CCU)-PylRS-GFP_{UV}149AGG were grown in GMML to an OD_{600} ~ 0.5. Addition of 0.5 mM IPTG, 0.2 % Arabinose, 10 mM of the respective NCAA (BocK, AllocK and ProK)[92]

Protein Labeling of GFPuv with ProK

GFPuv containing ProK (1mg/ml in 1×PBS buffer, pH 7.4) was incubated with CuSO4 (100 μM), NiCl (1mM), Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (0.5 mM), and FlAz (2.5 mM) sequentially, followed by Sodium ascorbate (5 mM) at room temperature for 3 hours, then TCA (10%) was added to precipitate the proteins on ice for 30 minutes, followed by 2 volume acetone wash, and SDS-PAGE analysis.[92]

Mass Spectrometry

ESI MS experiments were performed using an Applied Biosystems QSTAR Pulsar (Concord, ON, Canada) equipped with a nanoelectrospray ion source. The MS data were acquired in positive ion mode (500-2000 Da) using spray voltage of +2000 V and flow rate of 700 nL/min. Data analysis and protein signal extraction were performed using the BioAnalyst software (Applied Biosystems). A mass range of m/z 500-2000 was used for spectral deconvolution and the output range was 20000 to 30000 Da using a

---

I performed all experiments except protein labeling of GFPuv with ProK which was done by Dr. Zeng. The northern blotting for the aminoacylation assays were carried out in Dieter Soll's research lab during our collaboration. The mass spectra were conducted at the Laboratory for Biological Mass spectrometry at Department of Chemistry, Texas A&M University.

resolution of 0.1 Da and S/N threshold of 20.

Results and Discussion

Amber, Opal, and Ochre Suppression Efficiencies of the PylRS-tRNA^{Pyl} Pairs

To demonstrate amber suppression efficiency of the PylRS-$tRNA^{Pyl}_{CUA}$ pair, *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAG were used to express sfGFP both in the absence and in the presence of 5 mM BocK, a substrate of PylRS. Without BocK in the growth medium, low level of sfGFP expression was observed. On the contrary, the addition of BocK promoted sfGFP overexpression (Figure II-2A). The ESI-MS analysis of the purified sfGFP displayed two major mass peaks at 27,810±1 Da and 27,941±1Da that agree well with the theoretical molecular weights of sfGFP with BocK incorporated at N134 (27,940 Da for the full-length protein; 27809 Da for the full-length protein without the first methionine (M1)) (Figure II-2B1).

E. coli BL21(DE3) cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAA showed an undetectable expression level of sfGFP when BocK was absent in the growth medium. The addition of BocK induced sfGFP expression (Figure II-2A). The ESI-MS analysis of the purified sfGFP showed two mass peaks (27,940±1 Da and 27,809±1 Da) that agree well with the theoretical molecular weights of sfGFP with BocK incorporated at N134 (Figure II-2B2).

In comparison to amber suppression, the expression level of sfGFP using ochre suppression was lower. *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA exhibited a high expression level of sfGFP both in the absence

and in the presence of BocK in the growth medium (Figure II-2A). The ESI-MS analysis of purified sfGFP expressed in the absence of BocK showed a mass peak at 27,899±1 Da that clearly matched a Trp residue at N134 of sfGFP (calculated mass: 27,898 Da) (Figure II-2B3). The ESI-MS analysis of sfGFP expressed in the presence of BocK displayed a very interesting spectrum. Mass peaks for both Trp residue at N134 of sfGFP (27,900±1 Da) and BocK residue at N134 of sfGFP (27,941±1 Da) were observed. The mass peak for the Trp isoform was much more intensive than the BocK isoform (Figure II-2B4).

To test whether the incorporation of Trp at the 134 position is related to the nucleotide contents of mRNA around the UGA codon, plasmid pETtrio-pylT(UCA)-PylRS-sfGFP2TGA was constructed to test the opal suppression. *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP2TGA displayed high sfGFP expression levels both in the absence and in the presence of BocK that were much higher than in cells transformed with plasmid pETtrio-pylT(UCA)-PylRS-sfGFP134TGA and grown in the same conditions (Figure II-3A).

The ESI-MS analyses of purified sfGFP from cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP2TGA displayed similar patterns as sfGFP from cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA. Two mass peaks for sfGFP expressed both in the absence and in the presence of BocK (27,795±1 Da and 27,927±1 Da in Figure II-3B1 and 27,795±1 Da and 27,926±1 Da in Figure II-3B2) match the molecular weights of sfGFP with Trp incorporated at S2 (calculated mass: 27,926 Da for the full-length protein; 27,794 Da for the full-length protein without M1).

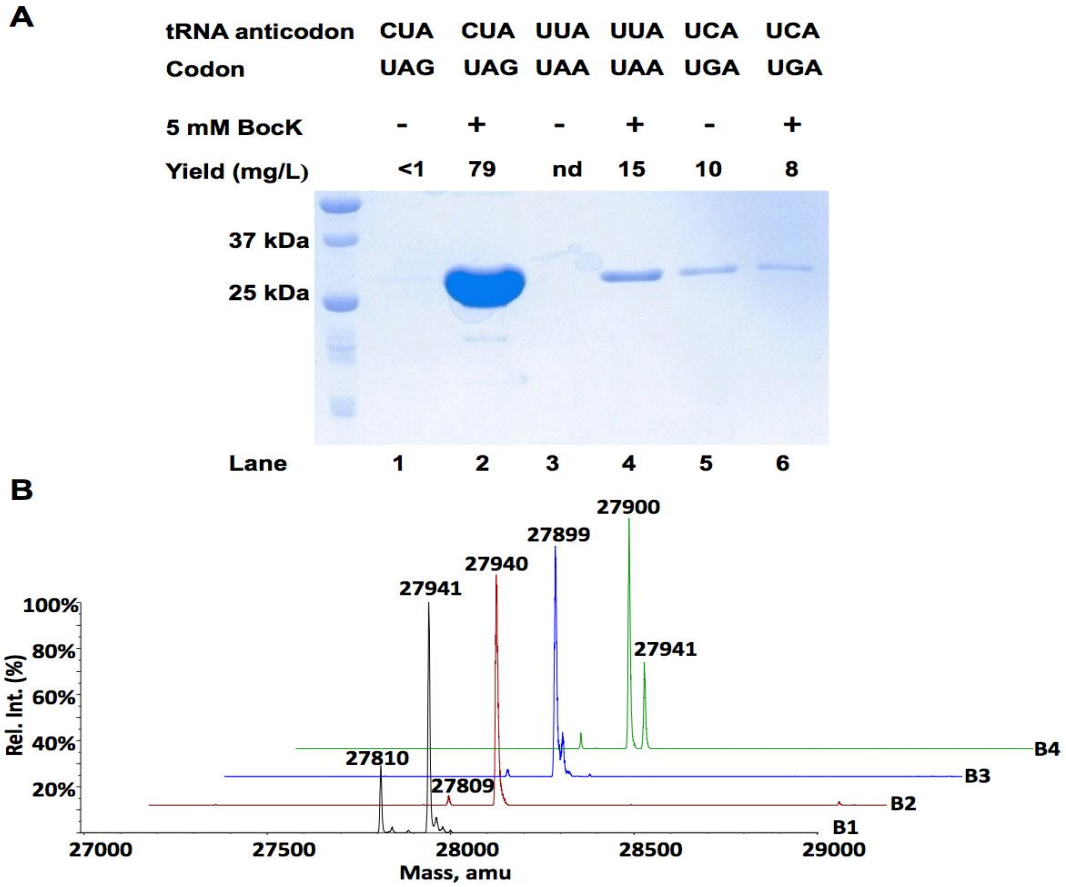


Figure II-2. Suppression of amber, opal, and ochre mutations at N134 of sfGFP by their corresponding PylRS-tRNA^{Pyl} pairs in the absence and presence of BocK. (A) Proteins shown in the gel represent their real relative sfGFP expression levels. Lanes 1 and 2 were transformed with pETtrio-PylT(CUA)-PylRS-sfGFP134TAG; lanes 3 and 4 were transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAA; lanes 5 and 6 were transformed with pETtrio-PylT(UCA)-PylRS-sfGFP134TGA. ESI-MS spectra of sfGFP expressed in cells (B1) transformed with pETtrio-PylT(CUA)-PylRSsfGFP134TAG and grown in the presence of 5 mM BocK, (B2) transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAA and grown in the presence of 5 mM BocK, (B3) transformed with pETtrio-PylT(UCA)-PylRS-sfGFP134TGA and grown in the absence of BocK, and (B4) transformed with pETtrioPylT(UCA)-PylRS-sfGFP134TGA and grown in the presence of 5 mM BocK.

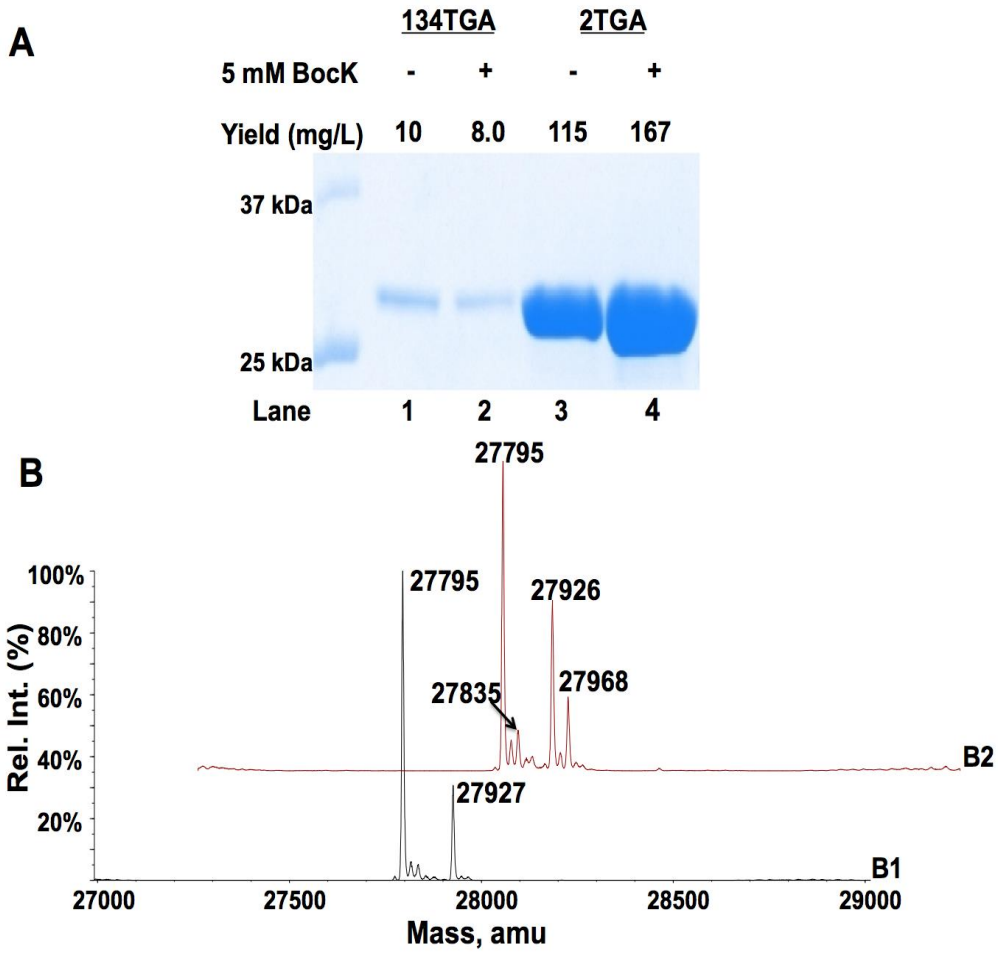


Figure II-3. Suppression of an opal mutation at S2 of sfGFP by the PylRS-tRNA^{Pyl} UCA pair. (A) Expression of sfGFP with an opal mutation. Lanes 1 and 2 were transformed with pETtrio-pylT(UCA)-sfGFP134TGA and grown in the absence or presence of 5 mM BocK; lanes 3 and 4 were transformed with pETtrio-pylT(UCA)-sfGFP2TGA and grown in the absence or presence of 5 mM BocK. Each protein shown in the gel represents their real relative expression levels. ESI-MS of sfGFP expressed in cells transformed with pETtrio-pylT(UCA)-sfGFP2TGA and grown in the (B1) absence or (B2) presence of 5 mM BocK.

For sfGFP expressed in the presence of 5 mM BocK, there were two small peaks at 27,968±1 Da and 27,835±1 Da that agree well with the molecular weights of sfGFP with BocK incorporated at S2 (calculated mass: 27,967 Da for the full-length protein; 27,836 Da for the full-length protein without M1).

To further understand opal suppression in the *E. coli* BL21(DE3) cell strain, plasmid pETtrio-sfGFP134TGA that did not carry opal suppressing $tRNA_{UCA}^{Pyl}$ was used to transform *E. coli* BL21(DE3) cells. The transformed cells displayed a relatively high detectable level of sfGFP expression with a yield of 4 mg/L (Figure II-4A).The ESI-MS analysis of sfGFP resulted from this basal opal suppression showed major mass peaks at 27,899±1 Da and 27,767±1 Da (Figure II-4B). The molecular weight corresponded to sfGFP with Trp incorporated at N134 position.

The suppression efficiencies of three mutated forms of $tRNA_{UCA}^{Pyl}$ with mutations as G73A, G73C, and G73U, respectively were also examined. Cells transformed with either pETtrio-pylT(UCA)G73A-PylRS-sfGFP134TGA or pETtrio-pylT(UCA)G73C-PylRS-sfGFP134TGA displayed similar expression levels of sfGFP both in the presence and in the absence of BocK(Figure II-5A). However, cells transformed with pETtrio-pylT(UCA)G73U-PylRS-sfGFP134TGA showed significantly different expression levels of sfGFP when grown in the absence and in the presence of BocK. The ESI-MS analysis of sfGFP expressed in the absence of BocK showed a major mass peak at 27,895±1 Da that matches sfGFP with Trp incorporated at N134 (Figure II-5B1). The addition of 5 mM BocK promoted the sfGFP expression level to increase. The ESI-MS

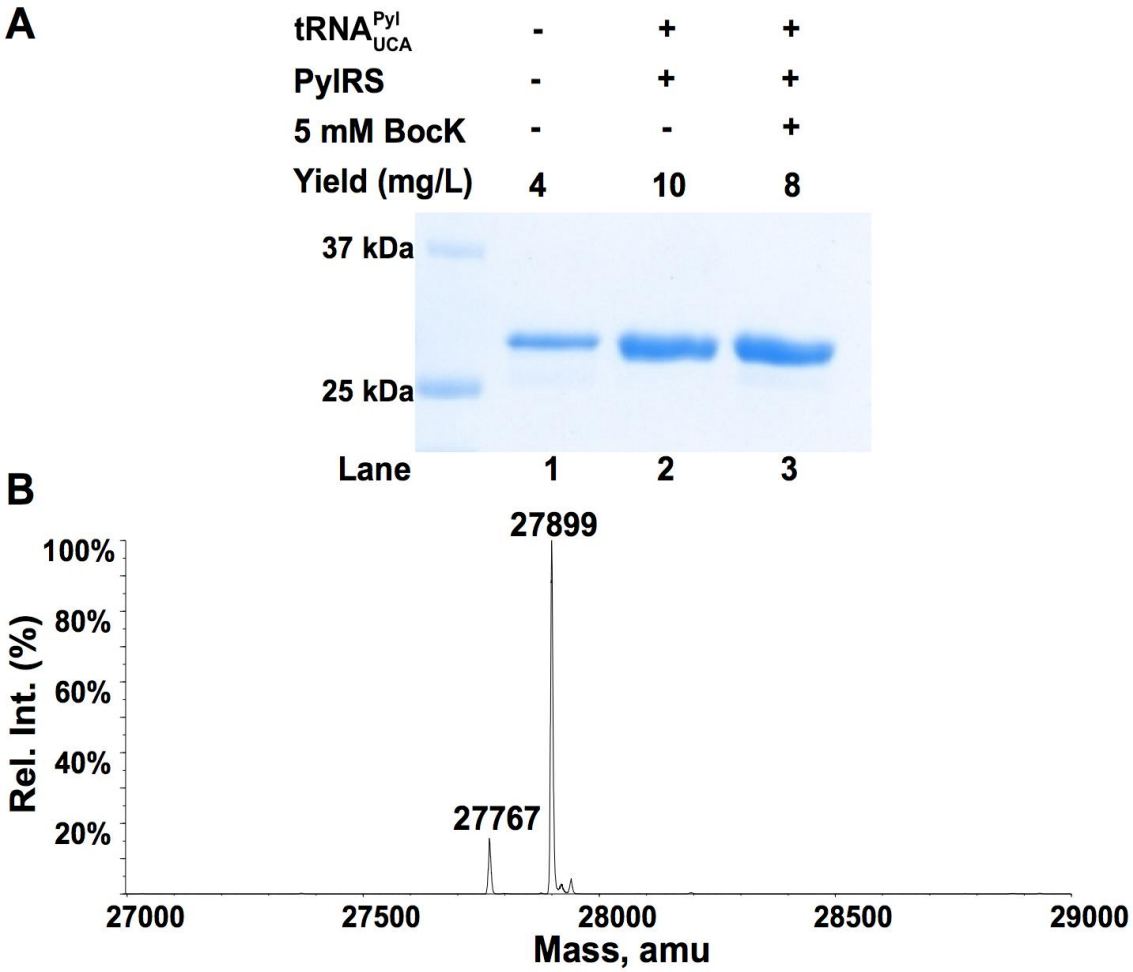


Figure II-4. Suppression of an opal mutation at N134 of sfGFP at different conditions. (A) Proteins shown in the gel represent their real relative expression levels. Lane 1 was transformed with pET-sfGFP134TGA; lanes 2 and 3 were transformed with pET-pylT(UCA)-sfGFP134TGA and grown in the absence or presence of 5 mM BocK. (B) ESI-MS of sfGFP expressed in cells transformed with pETtrio-sfGFP134TGA.

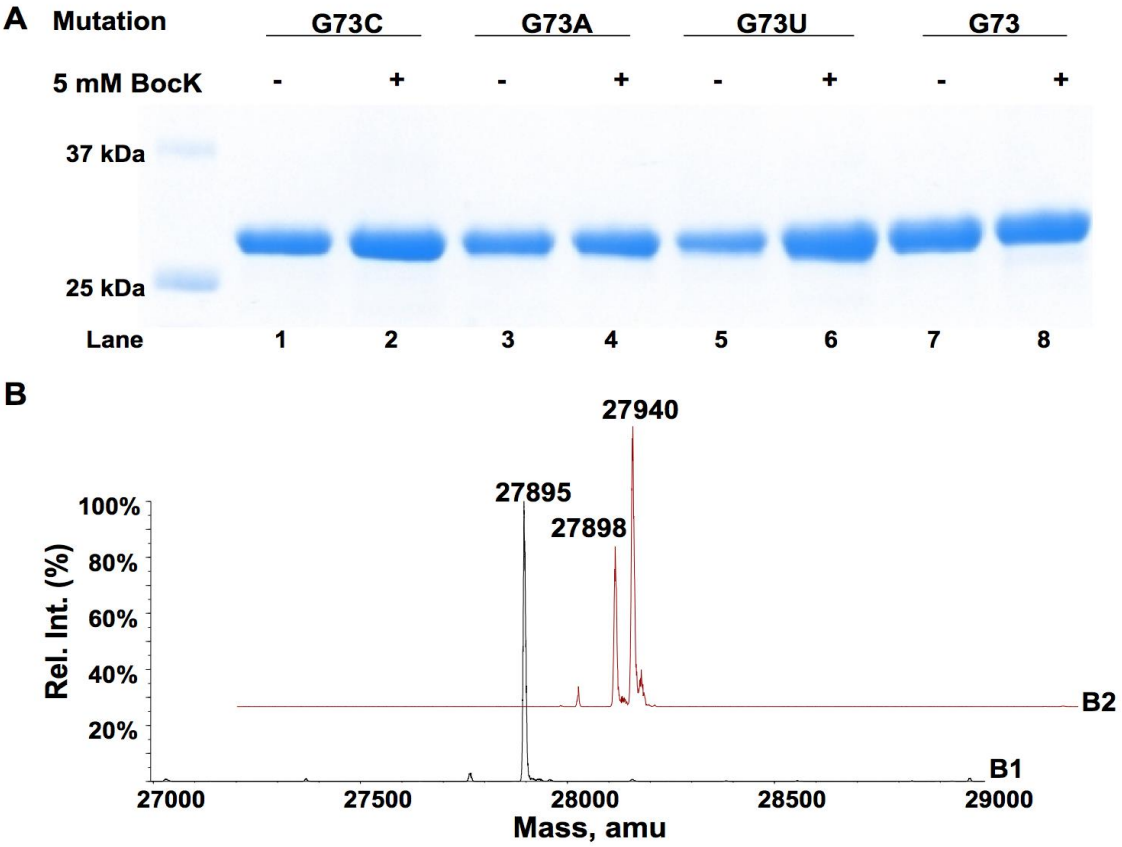


Figure II-5. Suppression of an opal mutation at N134 of sfGFP by different $tRNA^{Pyl}$ UCA variants. (A) Proteins shown in the gel represent their real relative expression levels. Lanes 1 and 2 were transformed with pETtrio-pylT(UCA)G73C-sfGFP134TGA and grown in the absence or presence of 5 mM BocK; lanes 3 and 4 were transformed with pETtrio-pylT(UCA)G73A-sfGFP134TGA and grown in the absence or presence of 5 mM BocK; lanes 5 and 6 were transformed with pETtrio-pylT(UCA)G73U sfGFP134TGA and grown in the absence or presence of 5 mM BocK; lanes 7 and 8 were transformed with pETtrio-pylT(UCA)-sfGFP134TGA and grown in the absence or presence of 5 mM BocK. The ESI-MS analysis of sfGFP expressed in cells transformed with pETtrio-pylT(UCA)G73U-sfGFP134TGA and grown in the (B1) absence or (B2) presence of 5 mM BocK.

analysis of the purified protein confirmed sfGFP with BocK incorporated at N134 became dominant (Figure II-5B2). The intensity of the mass peak at 27,940±1 Da that matched sfGFP with BocK incorporated at N134 was roughly twice of the mass peak at 27,898±1 Da which matched sfGFP with Trp incorporated at N134. Since the mass spectrometry showed that the sfGFP contained both Trp and BocK incorporated at the UGA position further analysis was done in collaboration with Dieter Soll's research group from Yale University. The analysis is to show the source of the Trp read-through and show that the Trp is not ligated to $tRNA_{UCA}^{Pyl}$. *In vitro* aminoacylation studies were carried out on PylRS and TrpRS with α–[^{32}P]-labeled tRNA substrates **(**Figure II-6).[63] Aminoacylation of the $tRNA_{UCA}^{Pyl}$ by the PylRS showed a comparable amount between BocK (52 ± 11%) or Cyc (46 ± 3%) which is similar to the wild type $tRNA_{CUA}^{Pyl}$ containing BocK (45 ± 5%) or Cyc (43 ± 9%) (Table II-1)

The *E. coli* TrpRS did not ligate the Trp with any of the $tRNA^{Pyl}$ variant compared to the active $tRNA^{Trp}$ (Figure II-7). Based on the data shown, TrpRS had no activity towards $tRNA_{UCA}^{Pyl}$. Incorporation of the Trp could result from the near-cognate (mismatch of one codon anticodon) suppresion of UGA by $tRNA_{CCA}^{Trp}$. Though PylRS-$tRNA_{UCA}^{Pyl}$ encoded BocK using UCA, the aminoacylated $tRNA_{UCA}^{Pyl}$ with BocK could not outcompete with the suppression by the $tRNA^{Trp}$ aminoacylated with Trp completely.

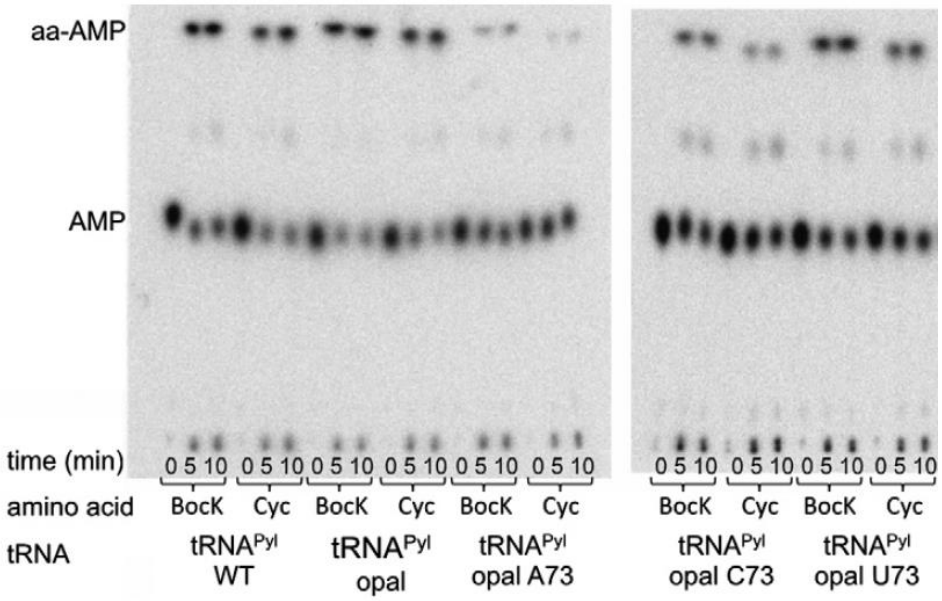


Figure II-6. *In vitro* aminoacylation of $tRNA^{Pyl}$ variants by *M. mazei* PylRS with BocK or N^{ε}-cyclopentyloxycarbonyl-L-lysine (Cyc). An additional spot (representing <5% of total radiolabel) between the aa-AMP and AMP is an unidentified possibly degraded reaction product, often seen in aminoacylation assays.

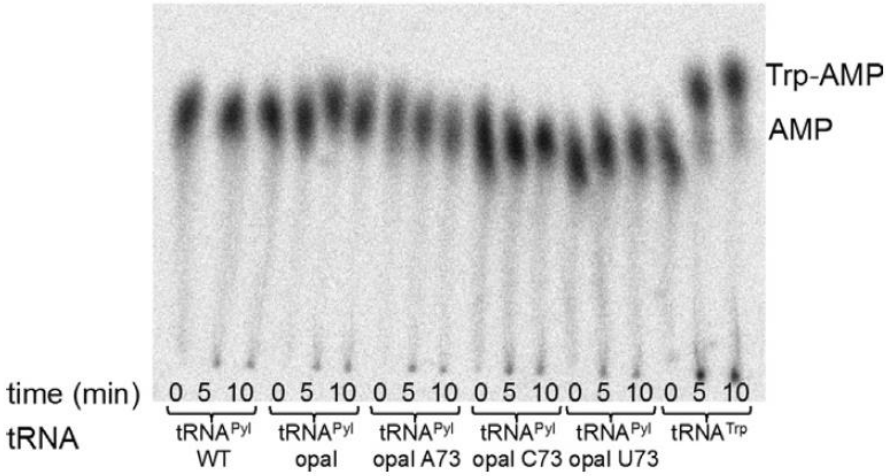


Figure II-7. *In vitro* aminoacylation of $tRNA^{Pyl}$ variants and $tRNA^{Trp}$ by E. coli TrpRS with Trp.

Table II-1. Plateau aminoacylation levels.

Enzyme	Amino acid	$tRNA^{Pyl}$	$tRNA^{Pyl}$ opal	$tRNA^{Pyl}$ opal A73	$tRNA^{Pyl}$ opal C73	$tRNA^{Pyl}$ opal U73	$tRNA^{Trp}$	$tRNA^{Arg}$
PylRS	BocK	45 ± 5%	52 ± 11%	13 ± 2%	25 ± 2%	40 ± 2%		
	Cyc	43 ± 9%	46 ± 3%	6 ± 3	13 ± 4%	29 ± 3%		
TrpRS	Trp	n.d.	n.d.	n.d.	n.d.	n.d.	77 ± 7%	
ArgRS	Arg	n.d.	n.d.	n.d.	n.d.	n.d.		89 ± 1%

Values are in % of total tRNA aminoacylated at plateau level (reached before the 10 minute time point) of the reaction. n.d. – no aminoacylation detected.

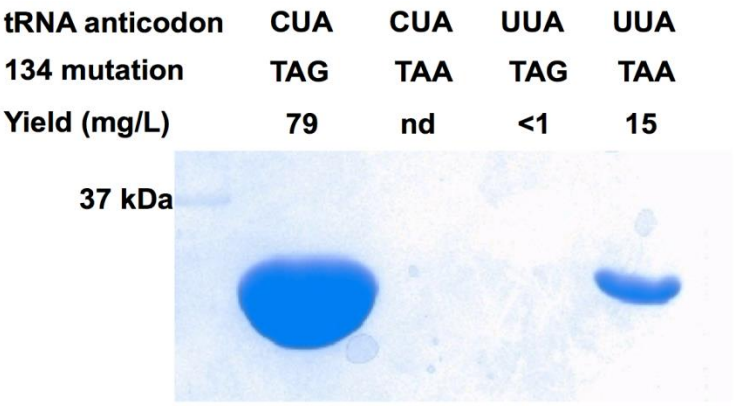


Figure II-8. Cross recognitions between different anticodons of $tRNA^{Pyl}$ and nonsense mutations at N134 of sfGFP. Cells were transformed with pETtrio-PylT(NNN)-PylRS-sfGFP134N'N'N' and grown in the presence of 5 mM BocK (NNN and N'N'N' denote anticodons and codons specified in the figure). Proteins shown in the gel represent their real relative expression levels.

Anticodon-Codon Cross Recognition

Transforming *E. coli* BL21(DE3) cells with pETtrio-pylT(CUA)-PylRS-sfGFP134TAA followed by growing cells in the presence of 5 mM BocK did not lead to detectable expression of sfGFP. *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAG showed a detectable but low level of sfGFP expression (less than 1 mg/L) (Figure II-8).

To see whether BocK promoted suppression at the amber mutation site, *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAG was also grown in the absence of BocK. As shown in Figure II-9, the sfGFP expression levels both in the absence and in the presence of BocK were very similar. Addition of BocK did not lead to significant increase of amber suppression. As shown in Figure II-9B1 & B2, sfGFP expressed in both conditions displayed mass peaks (27,839±1 Da and 27,710±1 Da in Figure II-9B1 and 27,840±1 Da and 27,709±1 Da in Figure II-9B2) that match sfGFP with glutamic acid (Glu), lysine (Lys), or glutamine (Gln) incorporated at N134. Figure II-9B2 did show a mass peak at 27,939±1 Da that matches the molecular weight of sfGFP with BocK incorporated at N134. However, its intensity was much lower than the mass peak of sfGFP at ~27,840±1 Da.

Experiments were also carried out to demonstrate that the PylRS-$tRNA^{Pyl}_{UUA}$ pair does not interfere with suppression of an amber mutation mediated by an evolved *Mj*TyrRS-$tRNA^{Tyr}_{CUA}$ pair.

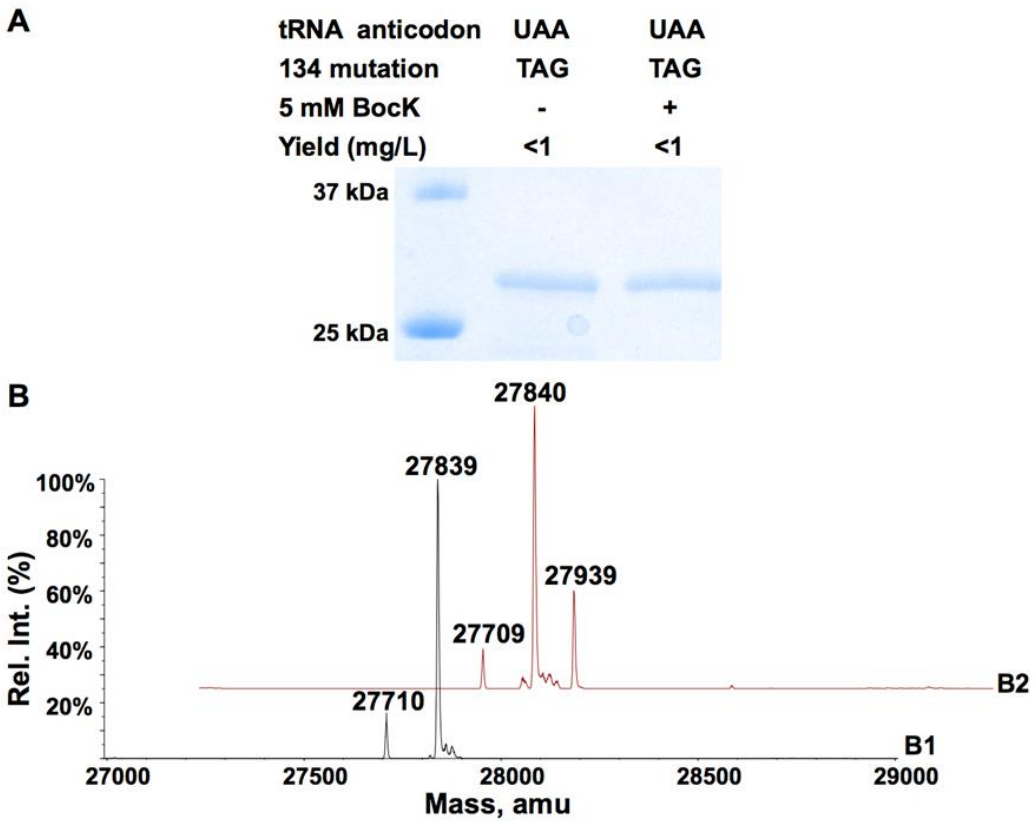


Figure II-9. Expression of sfGFP in cells transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAG. (A) Cells grown in 2YT medium supplemented without or with BocK. ESI-MS of sfGFP expressed in the (B1) absence or (B2) presence of 5 mM BocK.

Two plasmids, pETtrio-pylT(UUA)-pylRS-sfGFP134TAG and pEVOL-AzFRS were used to transform *E. coli* BL21(DE3)cells. As shown in Figure II-10A, growing the transformed cells in four conditions led to different expression levels of sfGFP. When no NCAA or only 5 mM BocK was provided in the medium, only a detectable but very low level of sfGFP expression (less than 1 mg/L) was detected. However, addition of 1 mM AzF to the medium supplemented with or without BocK promoted sfGFP overexpression. The ESI-MS analysis of purified sfGFP in all four conditions displayed very interesting spectra (Figure II-10B). Although a mass peak for sfGFP expressed in two conditions with the supplement of AzF (27,900±1 Da in Figure II-10B2 and 27,901±1 Da in Figure II-10B4) matches the calculated molecular weight (27,901 Da) of sfGFP with AzF incorporated at N134, a mass peak for sfGFP expressed in two conditions without the supplement of AzF (27,859±1 Da in Figure II-10B1 and 27,860±1 Da in Figure II-10B3) does not match the molecular weight of sfGFP with either Lys/Glu/Gln or BocK incorporated at N134. Instead, this mass peak agrees well with the molecular weight of sfGFP with Phe incorporated at N134 (calculated mass: 27,859 Da).

AGGA Codon Suppression

E. coli BL21(DE3) cells transformed with pETtrio-pylT(UCCU)-PylRS-sfGFP2AGGA showed similar sfGFP expression levels both in the absence (138 mg/L) and in the presence (122 mg/L) of BocK (Figure II-11) sfGFP proteins expressed in both conditions displayed one major mass peak at 27,868 Da (Figure II-11) that agreed with the theoretical molecular weight of sfGFP with arginine (Arg) incorporated at N134

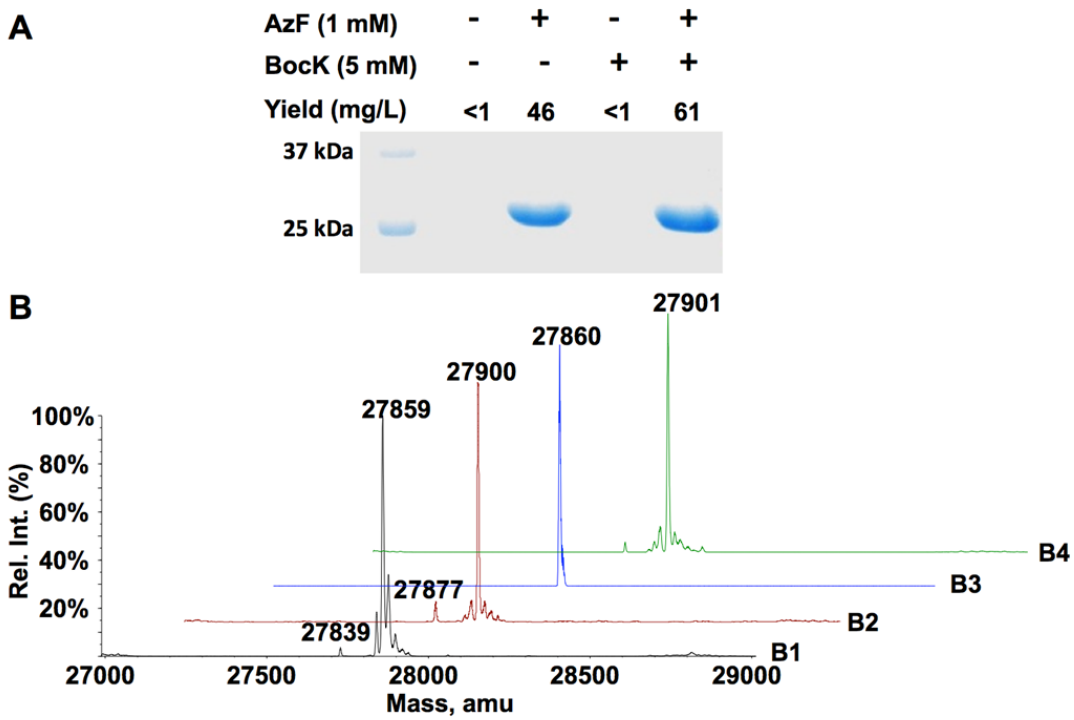


Figure II-10. Expression of sfGFP in cells transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAG and pEVOL-AzFRS. (A)Cells were grown in 2YT medium supplemented with different combinations of NCAAs. ESI-MS of sfGFP expressed in the(B1) absence of both AzF and BocK; (B2) presence of 1 mM AzF; (B3) presence of 5 mM BocK; and (B4) presence of both 1 mM AzF and 5 mM BocK.

(calculated mass: 27,867 Da for the full-length protein). Addition of the BocK to the medium did not chase the ESI-MS patterns. Our collaborators from Dieter Soll research group conducted *in vitro* aminoacylation assays to determine if $tRNA^{Pyl}_{UCCU}$ is a competent substrate for PylRS (Figure II-12) With BocK, aminoacylation for $tRNA^{Pyl}_{UCCU}$ by the PylRS showed a comparable amount between BocK (40 ± 6%) compared to the wild type $tRNA^{Pyl}_{CUA}$ containing BocK (45 ± 5%) (Table II-1). To determine if the Arg signal detected mass from ESI-MS is from the misacylation, $tRNA^{Pyl}$ variants were assayed for Arg aminoacylation activity. Results showed that although *E.coli* ArgRS is highly active with its native substrate, no $tRNA^{Pyl}$ variants could be aminoacylated with Arg **(**Figure II-13).

AGG Codon Suppression

E. coli BL21(DE3) cells transformed with pETtrio-pylT(CCU)-PylRS-sfGFP2AGG showed similar sfGFP expression levels both in the absence and in the presence of BocK (Figure II-14A). sfGFP proteins expressed in both conditions displayed one major mass peak at 27,895 Da that agrees well with the theoretical molecular weight of sfGFP with arginine (Arg) incorporated at S2 (calculated mass: 27,896 Da for the full-length protein) (Figure II-14B). pET-pylT(CCU)-PylRS-GFP_{UV}149AGG showed a higher incorporation of the various NCAA (BocK, AllocK, and ProK) dominating the production of GFP_{UV} with Arg[92] (Figure II-15). The GFP_{UV} was labeled in the absence and presence of ProK using a fluorescein azide (FIAz)[92] (Figure II-16).

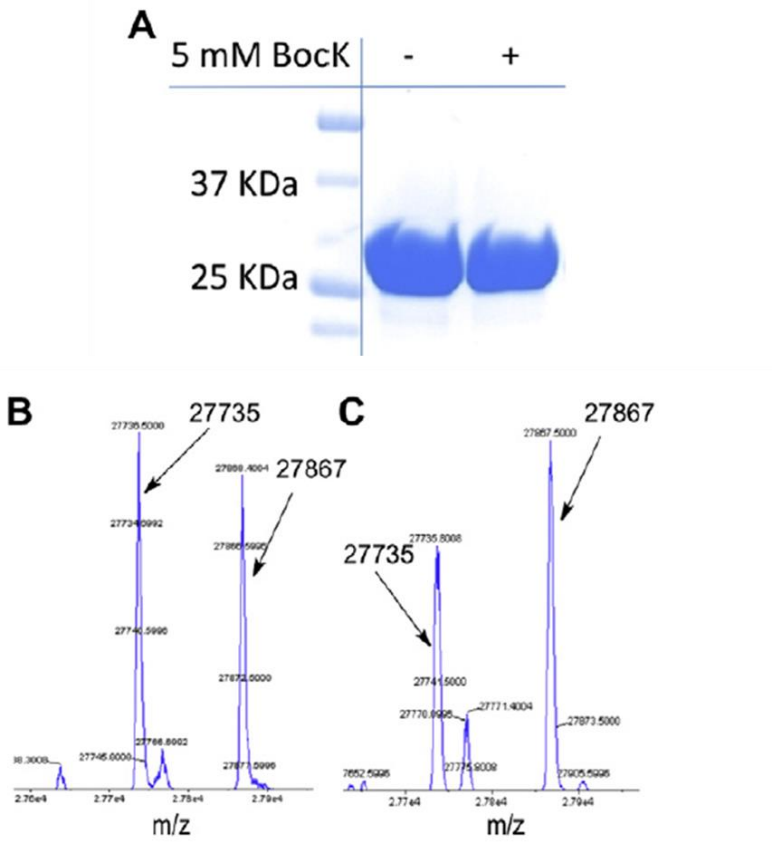


Figure II-11. Quadruplet codon AGGA encodes arginine in vivo in E. coli. (A) Protein produced by BL21(DE3) cells transformed with pETtrio-pylT(UCCU)-sfGFP134AGGA, grown with or without BocK. (B) ESI-MS of sfGFP produced by cells grown with (B) or without (C) BocK.

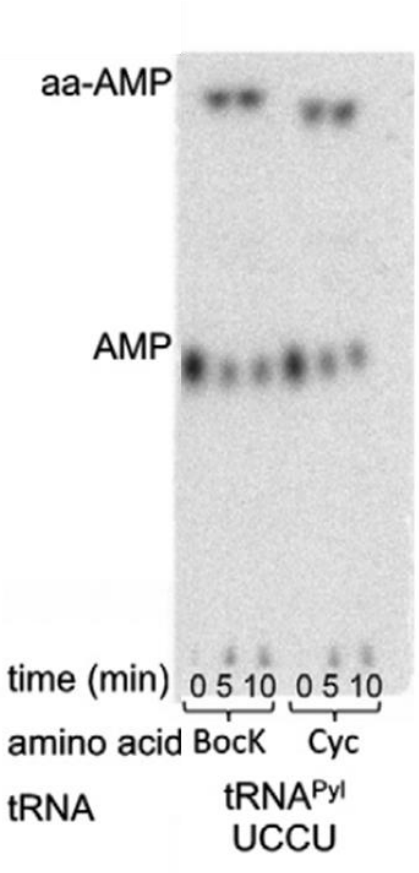


Figure II-12. *In vitro* aminoacylation of tRNA^{Pyl} quadruplet (UCCU) anticodon by *M. mazei* PylRS with BocK or Cyc.

Basal Nonsense Suppression in the *E. coli* BL21(DE3) Cell Strain

E. coli BL21(DE3) cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAG and grown in the absence of BocK yielded a sfGFP expression level close to 1 mg/L. The ESI-MS spectrum of the purified sfGFP clearly indicated a Lys, Glu, or Gln residue at

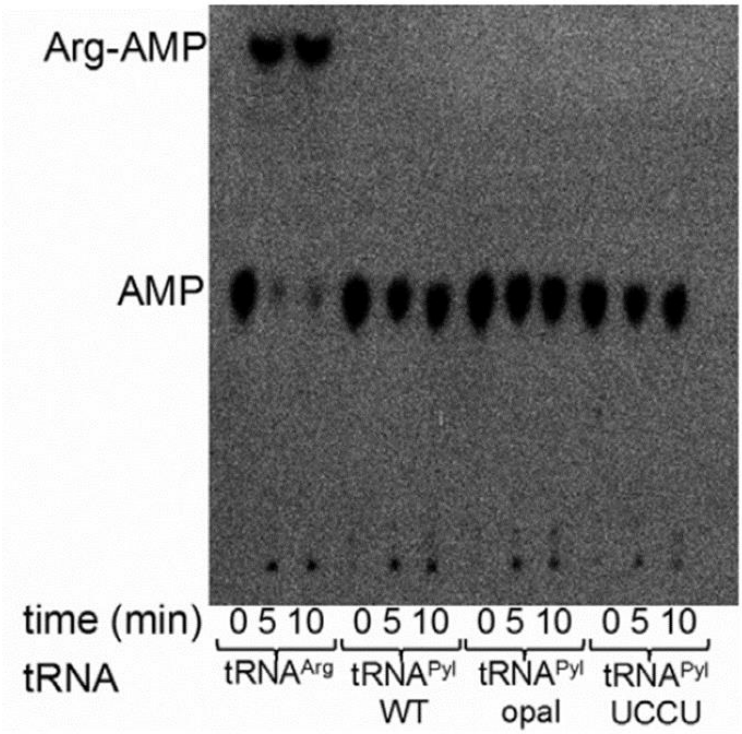


Figure II-13. *In vitro* aminoacylation of $tRNA^{Pyl}$ variants and $tRNA^{Arg}$ by *E. coli* ArgRS.

the amber mutation site. Since PylRS does not recognize Lys, Glu, and Gln and $tRNA_{CUA}^{Pyl}$ itself does not mediate detectable amber suppression in the *E. coli* Top10 cell strain (data not shown), this low but detectable sfGFP expression level was due to the basal amber suppression in the *E. coli* BL21(DE3) cell strain[67, 88]. This basal amber suppression that was also demonstrated in one separate study from us[88] arises possibly from the recognitions of the UAG codon by its near-cognate tRNAs including $tRNA_{UUU}^{Lys}$ / $tRNA_{UUC}^{Glu}$ / $tRNA_{CUG}^{Gln}$ [94, 95].

A similar test with *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA and grown in the absence of BocK yielded a sfGFP expression level of 8 mg/L. The ESI-MS spectrum of the purified protein showed a Trp residue at N134 of sfGFP. To rule out the possibility that $tRNA_{UCA}^{Pyl}$ was charged by *E. coli* tryptophanyl-tRNA synthetase (TrpRS) for the delivery of Trp at the designated opal mutation site. Plasmid pETtrio-sfGFP134TGA that did not contain the $tRNA_{UCA}^{Pyl}$ gene was constructed for further tests. *E. coli* BL21(DE3) cells transformed with this plasmid yielded a sfGFP expression level of 4 mg/L. The ESI-MS spectrum of the purified protein also showed the incorporation of Trp at N134 of sfGFP. Since our research findings recently demonstrated that $tRNA_{UCA}^{Pyl}$ was not recognized by *E. coli* TrpRS,[88] we concluded that the sfGFP expressed in the above two conditions resulted from the basal opal suppression in the *E. coli* BL21(DE3) cell strain. The observed different sfGFP expression levels might be caused by different copy numbers of two plasmids in the

transformed cells. This basal opal suppression is the consequence of the recognition of the opal codon by $tRNA^{Trp}$. *E. coli* $tRNA^{Trp}$ has a CCA anticodon which is a near-cognate tRNA of UGA. It forms two Watson-Crick base pairs with the 5' and middle nucleotides of UGA and a wobble base pair with the 3' nucleotide of UGA, explaining its ability to recognize UGA. We have also excluded the possibility that nucleotide contents around the opal mutation at N134 of sfGFP facilitated the binding of $tRNA^{Trp}$ and induced the Trp incorporation at this site. The *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP2TGA also expressed sfGFP in the absence of BocK and the expressed protein had Trp at the opal mutation site. The sfGFP expression level in *E. coli* BL21(DE3) cells transformed with pylT(UCA)-PylRS-sfGFP2TGA reached to 115 mg/L. This is roughly 20% of the expression level of wild-type sfGFP in *E. coli* BL21(DE3) cells (~500 mg/L, unpublished data). This high efficiency to read across an opal codon with the binding of a near-cognate $tRNA^{Trp}$ may correlate with the short distance from the opal codon to the start codon.

In contrary to UAG and UGA codons, the UAA codon displays high translation termination stringency in the *E. coli* BL21(DE3) cell strain. Cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAA showed an undetectable basal ochre suppression level. This can be explained from several aspects. UAG and UGA are recognized by release factor 1 and release factor 2, respectively, whereas UAA is recognized by both release factor 1 and release factor 2. Its recognition by both release factor proteins, in theory, makes the translation termination at UAA more stringent than

the other two stop codons. Another reason lies at the nucleotide contents of UAA. Unlike UAG and UGA that could involve a GC base pair interaction, UAA could only

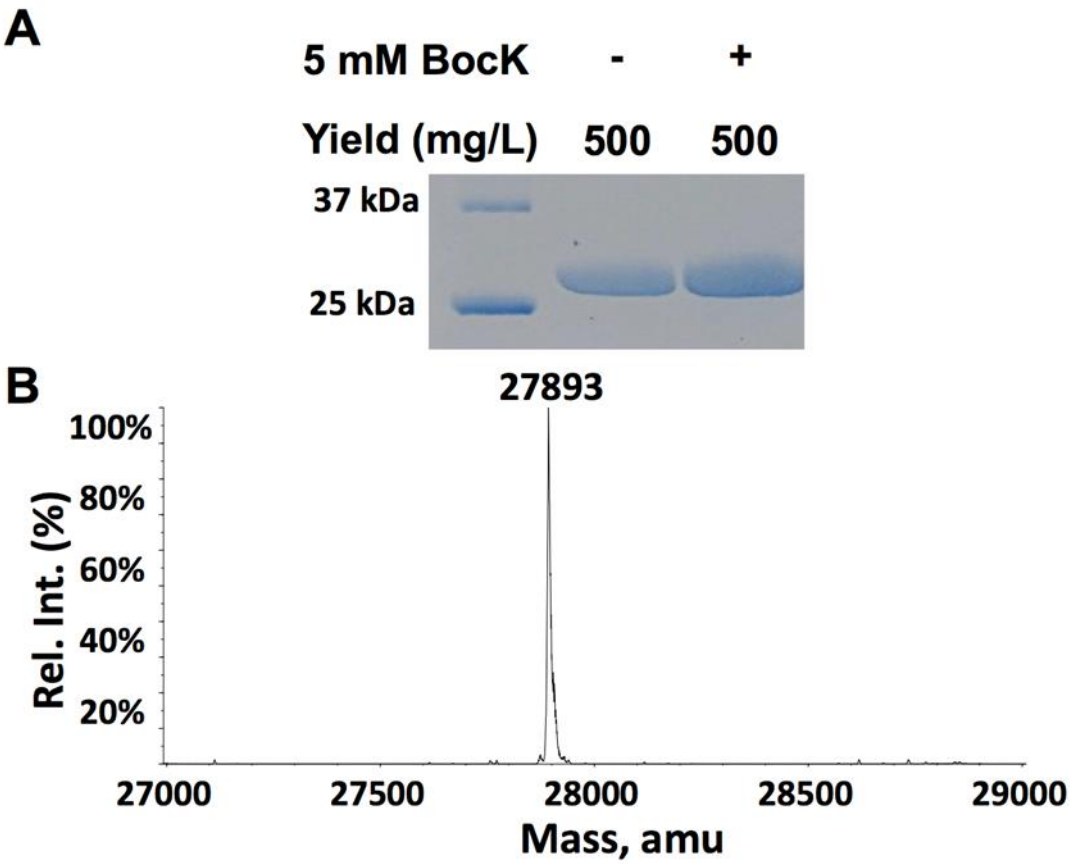


Figure II-14. Suppression of an AGG mutation at S2 of sfGFP by $tRNA^{Pyl}_{CCU.}$ (A) Expression of sfGFP in cells transformed with pETtrio-pylT(CCU)-sfGFP2AGG and grown in the absence or presence of 5 mM BocK. (B) The ESI-MS analysis of sfGFP expressed in the presence of 5 mM BocK.

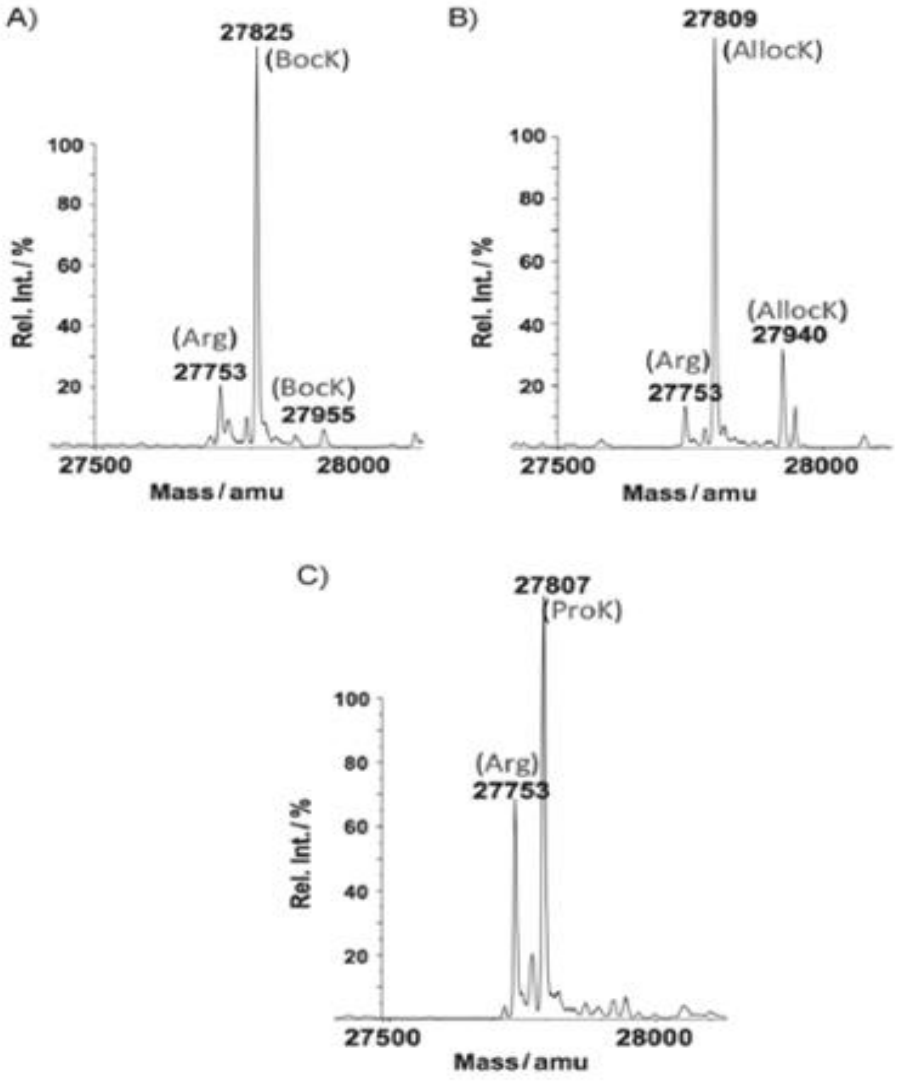


Figure II-15. ESI-MS spectra of GFP_{UV} variants expressed in different NCAA conditions. (10 mm; (A) BocK (B) AllocK and (C) ProK).

form AU pairs or wobble pairs. Its interactions with tRNAs are relatively weak, making its misrecognition less possible than UAG and UGA.

Suppression Efficiencies of the Amber, Opal, and Ochre PylRS-tRNA^{Pyl} Pairs

When co-expressed with PylRS, all three tRNA^{Pyl} isoforms $tRNA^{Pyl}_{CUA}$, $tRNA^{Pyl}_{UCA}$, and $tRNA^{Pyl}_{UUA}$ are capable of delivering BocK at their corresponding codon sites. $tRNA^{Pyl}_{CUA}$ was orthogonal in *E. coli* and displayed the highest efficiency in all three isoforms. Given that *E. coli* BL21(DE3) cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAA did not show a detectable expression level of sfGFP in the absence of BocK, we could conclude that $tRNA^{Pyl}_{UUA}$ is fully orthogonal in *E. coli*. In comparison to sfGFP expressed in cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAG and grown in the presence of BocK, the sfGFP expression level in cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAA and grown in the presence of BocK is five times lower. The low ability of $tRNA^{Pyl}_{UUA}$ to deliver BocK was possibly due to the relative weak base pair interactions between its UUA anticodon, its UAA stop codon and the availability of both release factor 1 and release factor 2 to stop the translation at a UAA stop codon. In any case, the ochre suppression level achieved by the PylRS-$tRNA^{Pyl}_{UUA}$ pair was sufficient to promote overexpression of a protein with an ochre mutation. Although a BocK-aminoacylated $tRNA^{Pyl}_{UCA}$ was able to suppress an opal codon for the incorporation of BocK, it did not inhibit the high incorporation of Trp at the same site. The ribosome could have permitted the codon anticodon mismatch at the

third nucleotide of the codon, so near-cognate suppression of UGA by $tRNA^{Trp}_{CCA}$ aminoacylated with Trp was not unusual.[96, 97]

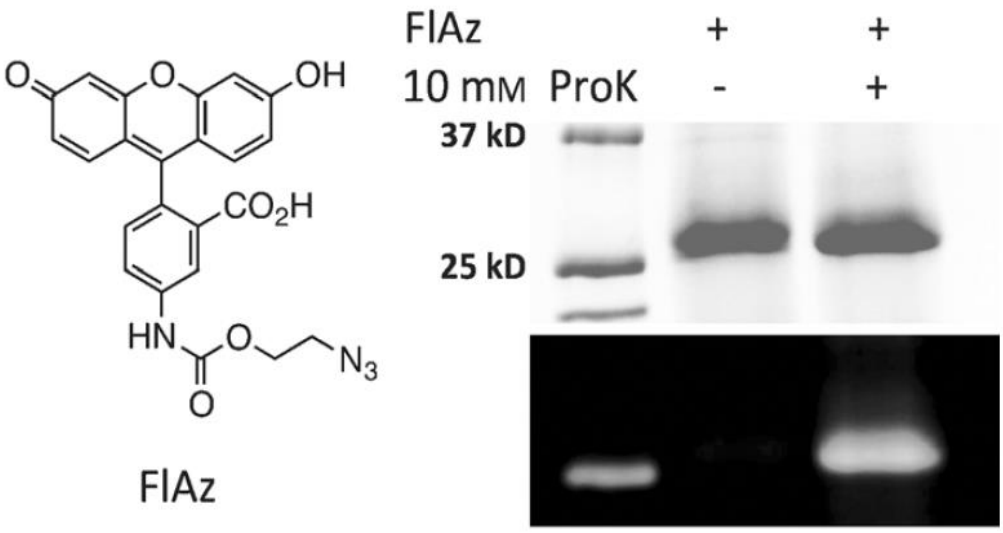


Figure II-16. Labeling of GFP_{UV} expressed in the absence and in the presence of ProK.

As an exogenous tRNA, the sequence and structure of $tRNA_{UCA}^{Pyl}$ may not be optimal for the protein translation process in *E. coli*. When facing a competition from *E. coli* $tRNA^{Trp}$, the recognition of $tRNA_{UCA}^{Pyl}$ by the *E. coli* translation machinery may be inhibited. Since the 73rd nucleotide serves as a strong recognition element for most tRNAs[98], it was mutated in $tRNA_{UCA}^{Pyl}$ to U, A, and C and searched for a mutant that show a higher opal suppression efficiency. When co-expressed with PylRS in the presence of BocK, $tRNA_{UCA}^{Pyl}(G73U)$ led to higher incorporation level of BocK compared to Trp at the same opal mutation site. However, further mutagenesis with $tRNA_{UCA}^{Pyl}(G73U)$ is necessary to fully inhibit the incorporation of Trp at an opal mutation site.

Cross Recognition of Anticodon-Codon Pair

The wobble hypothesis was first introduced by Francis Crick in 1966 to explain the observation that a single tRNA is able to efficiently recognize multiple codons[89]. Based on this hypothesis, an ochre suppressor, pylT (UUA) could recognize an amber UAG codon. This is a concern when both UAG and UAA codons are used to code two different NCAAs. However, cells transformed with pETtrio-pylT(UUA)-PylRS-sfGFP134TAG and grown in the presence of 5 mM BocK showed a sfGFP expression level close to that from the basal amber suppression. This suggests very weak recognition of UAG by $tRNA_{UUA}^{Pyl}$.

Weak base pairing interactions involved with the UUA anticodon may contribute to the weak recognition of UAG. However, this was certainly not the determining factor

since other tRNAs such as $tRNA_{UUU}^{Lys}$ were also involved in weak base pairing interactions to recognize multiple codons. One possible explanation for this weak recognition of UAG by $tRNA_{UUA}^{Pyl}$ is the tRNA modifications.

All cognate tRNAs are known to exhibit similar affinities for the ribosome A site when they bind to corresponding codons[99, 100]. This uniform binding is unexpected as certain codon-anticodon interactions are expected to be more stable than others due to factors such as the GC base pair content. It has been proposed that the specific sequence and post-transcriptional modification status of the tRNA in the region near the anticodon is tuned to ensure nearly indistinguishable binding of tRNAs to the ribosome A site[101-103]. This has been the case for tRNAs such as $tRNA_{UUU}^{Lys}$ in which both nucleotides at 34 and 37 are post-transcriptionally modified to achieve similar recognitions of AAA and AAG codons[103]. Unlike endogenous tRNAs that have corresponding modification enzymes, $tRNA_{UUA}^{Pyl}$ is exogenous and may not be targeted by tRNA modification enzymes in *E. coli*. $tRNA_{UUA}^{Pyl}$ without modifications at its anticodon loop likely has a very weak binding affinity to the ribosome A site to associate UAG.

Results showed that the wobble base pairing at the 3' nucleotide of a codon was not sufficient for recruiting a tRNA to the ribosome A-site. Additional interactions are required. This aspect needs to be further investigated. Since $tRNA_{UUA}^{Pyl}$ has a low ability to recognize UAG, it is feasible to use an amber suppressor aaRS-tRNA pair and a wild

type or evolved PylRS-$tRNA_{UUA}^{Pyl}$ pair to code two different NCAAs at amber and ochre mutation sites, respectively, in *E. coli*.

During the anticodon-codon cross recognition analysis to examine whether $tRNA_{UUA}^{Pyl}$ can compete against $tRNA_{CUA}^{Tyr}$ to bind UAG in the ribosome A site, we noticed that Phe was incorporated at N134 of sfGFP which was expressed in cells transformed with pEVOL-AzFRS and pETtrio-pylT(UUA)-PylRS-sfGFP134TAG and grown either in the absence or in the presence of BocK. It is possible AzFRS could recognize Phe, leading to the misincorporation of Phe. Since AzF is a hydrophobic amino acid and structurally very similar to Phe, one would expect that AzFRS that was originally evolved from *Mj*TyrRS would probably recognize Phe at a relatively low level and charge $tRNA_{CUA}^{Tyr}$ with Phe when AzF is absent in the medium.

Although not clearly addressed in existing literature, most evolved *Mj*TyrRS-$tRNA_{CUA}^{Tyr}$ pairs did show significant background amber suppression even in minimal media[54, 104]. Since most evolved *Mj*TyrRS variants are for Phe derivatives, it is highly possible that background amber suppression caused by these evolved *Mj*TyrRS-$tRNA_{CUA}^{Tyr}$ pairs was due to their recognition of either Phe or Tyr or both. In this study, the background amber suppression induced by the AzFRS-$tRNA_{CUA}^{Tyr}$ pair inhibited both the basal amber suppression level and amber suppression induced by the PylRS-$tRNA_{UUA}^{Pyl}$ pair. This test also provided evidence that amber suppression mediated by the PylRS-

$tRNA_{UUA}^{Pyl}$ pair is too low to be a concern. It also points out that substrate specificities of evolved NCAA-specific aaRSs need to be further characterized.

Codon Suppression for AGGA and AGG

Our lab and other groups have shown that mutating the anticodon of $tRNA^{Pyl}$ does not significantly affect its interactions with *M. mazei* PylRS. Three $tRNA^{Pyl}$ isoforms that are specific for three stop codons are capable to deliver BocK at their corresponding codon sites when co-expressed with PylRS. The ability a $tRNA^{Pyl}$ isoform to suppress a quadruplet or sensing codon was also determined. Though further work needs to be done on the quadruplet codon, the non-standard four base pair allows incorporation of Arg by favoring the three-base coding models. Also, the suppression of the AGG codon was tested since it is a rarely used codon and $tRNA_{CCU}^{Arg}$ is limitedly expressed in the *E. coli* BL21(DE3) cell strain[105]. However, *E. coli* transformed with pETtrio-pylT(CCU)-PylRS-sfGFP2AGG only expressed sfGFP with Arg at its S2 position in the presence of 5 mM BocK. Increasing the BocK concentration to 10 mM did drive the expression of sfGFP with BocK incorporated at S2 of sfGFP, which was confirmed by the ESI-MS analysis of the purified proteins (data not shown).

In comparison to the Arg-containing sfGFP isoform that showed a high intensity in the ESI-MS spectrum of the purified sfGFP, the intensity for the BocK-containing sfGFP isoform was very low. This indicates $tRNA_{CCU}^{Pyl}$ is charged by *M. mazei* PylRS with BocK in *E. coli* and is able to deliver BocK at an AGG codon site. However, the *M. mazei* PylRS- $tRNA_{CCU}^{Pyl}$ pair cannot compete efficiently against the endogenous Arg

incorporation system at the AGG codon, same goes for the PylRS- $tRNA_{UCCU}^{Pyl}$. Similar to $tRNA_{UCA}^{Pyl}$, the sequence and structure of $tRNA_{CCU}^{Pyl}$ may not be optimized for the protein translation machinery in *E. coli*. An approach reassign the rare AGG codon to code for NCAA instead of Arg can be obtained by using a different PylRS- $tRNA_{CCU}^{Pyl}$ from *M. barkeri* and also GFP_{UV} instead of the sequence optimized sfGFP. Results show high efficiency incorporation of BocK, AllocK and ProK at the AGG codon in minimal medium. To support the genetic incorporation of ProK at the AGG codon site, ProK containing a terminal alkyne was labeled with FIAz via a copper catalyzed click reaction.

Conclusion

In summary, the suppression efficiencies of the original and mutant $tRNA^{Pyl}$, the cross recognition between nonsense codons and $tRNA^{Pyl}$ anticodons in the *E. coli* BL21(DE3) cell strain have been investigated. Among all $tRNA^{Pyl}$ isoforms, $tRNA_{CUA}^{Pyl}$ has the highest suppression efficiency for the delivery of BocK at its corresponding codon. Besides its orthogonal nature in *E. coli*, $tRNA_{UUA}^{Pyl}$ does not induce a significant level of suppression at an amber codon. This is contrary to the wobble hypothesis and makes it feasible to use amber suppressing aaRS-tRNA pair and the PylRS-$tRNA_{UUA}^{Pyl}$ pair to code two different NCAAs at amber and ochre codons respectively in *E. coli*.

Data also showed that the PylRS-$tRNA_{UCCU}^{Pyl}$ although competent in BocK it cannot outcompete AGGA suppression by *E. coli*'s native $tRNA^{Arg}$ aminoacylated by

Arg. Our study also demonstrated the PylRS-$tRNA_{CCU}^{Pyl}$ pair cannot efficiently deliver BocK at an AGG codon site. Further work to optimize the sequence and structure of $tRNA_{CCU}^{Pyl}$ for the *E. coli* translation machinery may be necessary to increase the BocK incorporation efficiency and suppress the Arg incorporation at the AGG codon. Further studies showed that rare AGG codon can be reassigned to code for NCAA by using a different PylRS- $tRNA_{CCU}^{Pyl}$ from *M. barkeri* and also GFPuv instead of the sequence optimized sfGFP. This solves the problem of the low incorporation of NCAA and allows different NCAA with unique functionalities to be site-selectively incorporated for applications such as therapeutic protein production.

CHAPTER III

RAPID CATALYST-FREE CLICK REACTIONS FOR IN VIVO PROTEIN LABELING MEDIATED BY GENETICALLY ENCODED STRAIN-PROMOTED ALKENE/ALKYNE FUNCTIONALITIES*

Introduction

First described by K. Barry Sharpless in 1999, the term "click chemistry" defines chemical reactions that are selective, rapid, and highly efficient.[106, 107] Early focus of click chemistry centered on the Cu^{+}-catalyzed azide-alkyne cycloaddition (CuAAC) reaction.[108, 109] Due to the biologically inert nature of the azide and alkyne functional groups, CuAAC is highly bio-orthogonal. Applications of this reaction in chemical biology include activity-based protein profiling and selective biomolecule labeling.[110-114]

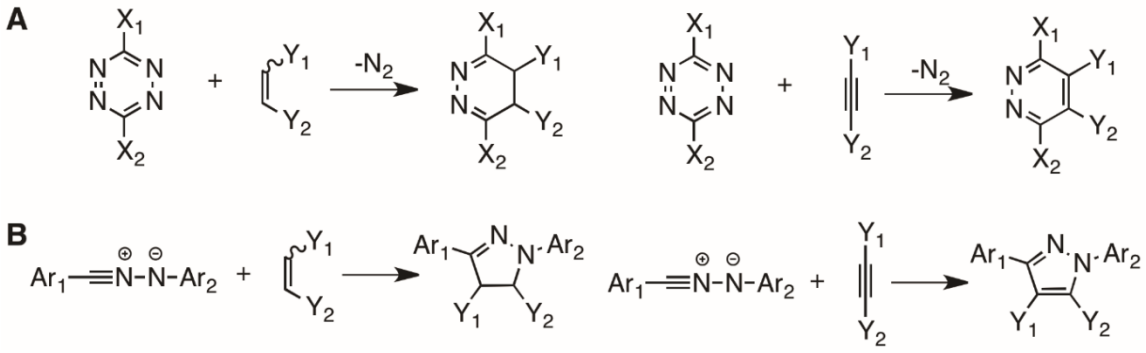
One potential pitfall of using CuAAC in living systems is the cytotoxic effect of Cu^{+} catalysts, which can be avoided using copper-free azide-cyclooctyne cycloaddition.[115, 116] Due to strain promotion, the cyclooctyne triple bond has an elevated energy state and therefore reacts simultaneously with an organic azide in mild aqueous conditions, bypassing the use of a Cu^{+} catalyst. Although originally described as a relative slow reaction, cyclooctyne derivatives that show fast reaction kinetics with organic azides have been developed.[117-121]

---

*Reprinted from "Two rapid catalyst-free click reactions for in vivo protein labeling of genetically encoded strained alkene/alkyne functionalities", Kurra, Y.$^{+}$, Odoi, K.A.$^{+}$, Lee, Y.J., Lu, T., Wheeler, S., Torres-Kolbus, J., Deiters, A., Liu. Bioconjugate Chem. **2014**. 25 (9), 1730-1738 Copyright 2013, with permission from American Chemical Society. $^{+}$ Contributed equally

The pursuit of other fast catalyst-free click reactions that can be applied in living systems for labeling biomolecules at bio-relevant concentrations has recently revitalized two other reactions, namely inverse electron-demand Diels-Alder cycloaddition (EDDAC) and nitrile imine-alkene/alkyne 1,3-dipolar cycloaddition (NIADC) (scheme III-1).[122-126] Both reactions display rapid kinetic features when a strain-promoted alkene/alkyne is involved.

Applications of these two reactions include labeling biomolecules in cells and cancer diagnostics.[127-129] In this work, the report shows the kinetic investigations of EDDAC and NIADC reactions that involve strain-promoted alkene/alkyne functionalities including norbornene, transcyclooctene, and cyclooctyne and their applications in labeling proteins that are genetically incorporated with these functional groups.



Scheme III-1. (A) Tetrazine-alkene/alkyne EDDAC and (B) NIADC reactions.

Experimental Section

Materials

Reagents were purchased from general chemical manufacturers such as Sigma Aldrich, TCI, Alfa Aesar and Chem Impex and used without further purification. Anhydrous DMF and Methanol from Sigma-Aldrich and used as is. All reactions involving moisture sensitive reagents were carried out in oven-dried glassware under argon atmosphere. Thin layer chromatography (TLC) was performed on silica 60F-254 plates and visualized using UV irradiation at 254 nm. Flash chromatography was performed with silica gel (particle size 32-63 μm) from Dynamic adsorbents inc. (Atlanta, GA).

All ^{1}H and ^{13}C spectra were recorded on Varian Mercury-300, INVOA 300 and INVOA-500 MHz spectrometers. NMR spectra chemical shifts were reported in parts per million (ppm) and referenced to solvent peaks: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.8 ppm for ^{1}H). A minimal amount of 1,4-dioxane was added as the reference standard (67.19 ppm for ^{13}C) for carbon NMR spectra in deuterium oxide. ^{1}H NMR spectra are tabulated as follows: chemical shift multiplicity (s= singlet, bs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). Mass spectra were conducted at the Laboratory for Biological Mass spectrometry at Department of Chemistry, Texas A&M University.

Chemical Synthesis

NaH (1.74 g, 72.5 mmol) was washed with dry hexane (15 mL) and then decanted. Dry THF (20 mL) was added and the mixture was allowed to stir at room temperature. 5-Norbornene-2-ol (NOR, 2.0 g, 18.18 mmol) in THF (15 mL) was also added to the flask. Heated to reflux for 1 hour whilst stirring. α-Bromoacetic acid (1.74 g, 18.18 mmol) in THF (12 mL) was then added to the mixture and refluxing continued overnight. The reaction mixture was cooled to room temperature, and concentrated on the rotary evaporator. The residue was cooled in an ice bath and water was added followed by acidification with 3 M HCl. The aqueous layer was extracted with three portions of ether. The extracts were dried with Na_2SO_4 and concentrated in vacuo to form oil. The compound was used directly for the next reaction without further purification.

The above acid compound (1.5 g, 8.92 mmol) was dissolved in DMF (12 mL) followed by adding K_2CO_3 (3.69 g, 26.7 mmol) but methyl iodide (1.14 mL, 17.73 mmol) was then added dropwise. The mixture was stirred at 60 ^oC for 1 hour. The reaction system was diluted with EtOAc, and then washed sequentially with water and HCl (1 M). The organic layer was separated and dried with anhydrous Na_2SO_4. After filtration, the organic layer was concentrated. Column chromatography with hexanes/EtOAc as eluent (9:1 v/v) afforded pure methyl ester product **2** (1.10 g, 68%).

To a stirred solution of LAH (0.230 g, 6.59 mmol) in dry THF (10 mL) at 0 ^oC was added dropwise a solution of **2** (1.0 g, 5.49 mmol) in THF under Argon. The mixture was stirred for 4 hours at room temperature. It was then cooled to 0 ^oC, diluted

with ether, quenched by dropwise addition of a saturated solution of Na_2SO_4 (15 mL), and extracted with EtOAc. The organic layer was dried with Na_2SO_4, filtered and concentrated in vacuo. Purification of crude product by column chromatography afforded the product **3** as a liquid (0.610 g, 72%).

DIAD (1.18 g, 5.84 mmol) was added dropwise slowly to an oven dried flask into a solution of compound **3** (0.6 g, 3.89 mmol), compound **4** (1.72 g, 5.84 mmol) and TTP (1.42 g, 5.45 mmol) in Dry THF (10 mL) and the mixture was cooled to 0 ^oC. The reaction mixture was stirred at room temperature overnight. The reaction mixture was then concentrated and subjected directly to column chromatography (eluent: hexane/EtOAc as 4/1(v/v)). **5** (1.3 g, 78% yield) was afforded as light yellow oil. Compound **5** (1.2 g, 2.78 mmol) was dissolved in THF (10 mL) and aqueous NaOH (1.0 M, 3.61 mL) was added dropwise. The mixture was stirred at room temperature for 2 hours, diluted with water (10 mL), and then extracted with Et_2O (20 mL). The aqueous layer was collected and adjusted to pH = 1 with HCl (3.0 M in water).

The resulting mixture was extracted with EtOAc (10 mL × 2) and the combined organic layers were dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to get acid compound**,** which was used directly in the next step without further purification.

The compound (1**.**0 g, 2.39 mmol) obtained from the above step was dissolved in anhydrous dioxane (15 mL) followed by an injection of HCl/dioxane solution (4.0 M, 1.19 mL). The resulting mixture was stirred at room temperature for 12 hours. The white solid was collected by filtration and washed sequentially with EtOAc and

dichloromethane (DCM), then dried under vacuum using oil pump to afford the final product, **NCAA 1** as a white solid (0.609 g, 62% over two steps) (Scheme III-2).

^{1}H NMR (CD_3OD, 300 MHz) δ 7.32 (d, 2H, *J* = 8.4 Hz), 7.18 (d, 2H, *J* = 8.4 Hz), 6.25 (dd, 1H, *J* = 3.0, 5.7 Hz), 5.93 (dd, 1H, *J* = 3.3, 5.7 Hz), 4.29-4.20 (m, 1H), 4.18-4.16 (m, 1H), 4.11-4.04 (m, 2H), 3.82-3.70 (m, 2H), 3.21 (ABq, 2H, *J* = 7.8, 14.4 Hz), 2.76 (m, 1H), 2.03-1.95 (m, 1H), 1.44-1.29 (m, 3H), 0.87 (m, 1H).

^{13}C NMR (CD_3OD, 75 MHz) δ 169.8, 158.5, 140.3, 137.4, 132.8, 131.0, 130.3, 130.2, 126.0, 121.6, 114.8, 80.8, 67.5, 67.2, 53.8, 48.5, 45.1, 42.0, 35.0, 33.9, 26.5.

DIAD (1.30 g, 6.43 mmol) was added dropwise slowly to an oven dried flask charged with **1** (0.5 g, 4.54 mmol) and **2** (2.0 g, 6.77 mmol), and TTP (1.60 g, 6.10 mmol) in Dry THF (15 mL)) and the mixture was cooled to 0 ^{o}C. The reaction mixture was stirred at room temperature overnight, concentrated, and then subjected directly to column chromatography (eluent: hexane/EtOAc as 4:1(v/v)) to afford **3** (1.08 g , 62% yield) as light yellow oil.

Compound **3** (1.0g, 2.58 mmol) was dissolved in THF (10 mL) and aqueous NaOH (1.0 M, 3.35 mL) was added dropwise. The mixture was stirred at room temperature for 2 hours, diluted with water (10 mL), and then extracted with Et_2O (20 mL). The aqueous layer was collected, and adjusted to pH = 1 with HCl (3.0 M in water). The resulting mixture was then extracted with EtOAc (10 mL × 2) and the combined organic layers were dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to get acid compound**,** which was used directly in the next step without further purification.

The compound (0.7 g, 1.87 mmol) obtained from the above step was dissolved in anhydrous dioxane (10 mL) followed by an injection of HCl/dioxane solution (4.0 M, 1.29 mL). The resulting mixture was stirred at room temperature for 12 hours. The white solid was collected by filtration and washed sequentially with EtOAc and DCM, then dried under vacuum using oil pump to afford the final product **NCAA 2** (Scheme III-3) as white solid (0.33 g, 47% over two steps).

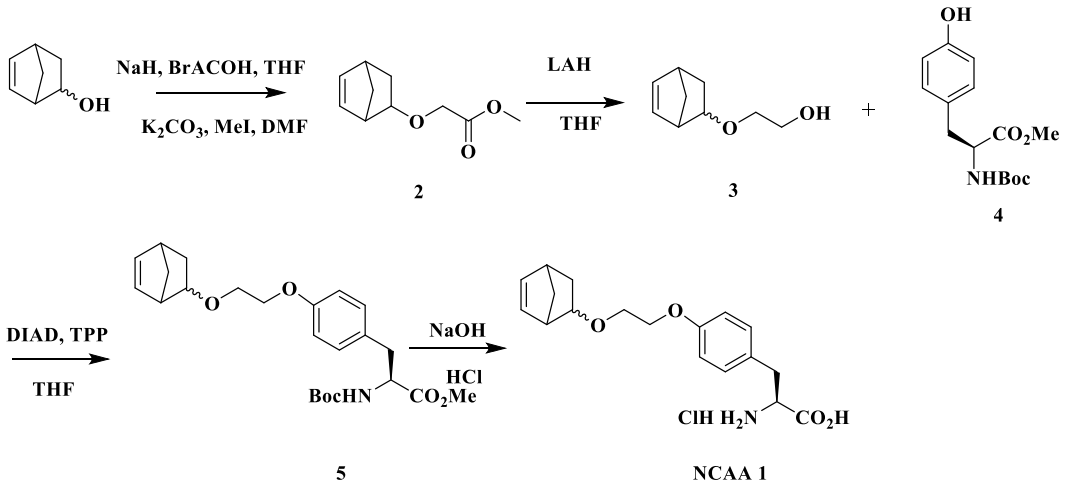
^{1}H NMR (CD_3OD, 300 MHz) δ 7.23 (d, 2H, *J* = 7.2 Hz), 6.93(d, 2H, *J* = 7.5 Hz), 6.35 (m, 1H), 6.07 (m, 1H), 4.61(m, 1H), 4.36 (m, 1H), 3.24 (ABq, 2H, *J* = 6.9, 14.1 Hz), 3.00 (brs, 1H), 2.91(brs, 1H), 1.94-1.83 (m, 1H), 1.63-1.41 (m, 1H), 1.31-1.15 (m, 2H).

^{13}C NMR (CD_3OD, 75 MHz) δ 170.1, 157.4, 140.9, 132.3, 130.7, 125.8, 115.6, 115.5, 69.5, 54.0, 45.0, 40.0, 35.0, 33.0, 29.0.

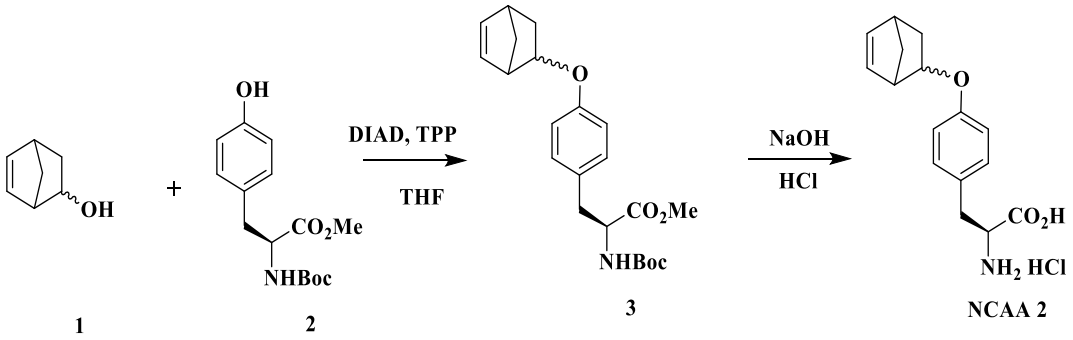
The synthesis of transcyclooctene derivatives followed the procedures developed by Fox *et al.*[130] To a stirred solution of 1,5-cyclooctadiene (10.0 g, 96 mmol, 1.0 equiv) in dry DCM (150 mL) at 0 ºC was added slowly m-CPBA (18.5 g, 113 mmol, 1.2 equiv). The mixture was stirred for 1 hour 30 minutes at room temperature. Then a saturated solution of $Na_2S_2O_3$ (15 mL) was added and the mixture was extracted with DCM. The organic layer was washed with H_2O twice, saturated $NaHCO_3$ twice and then saturated $Na_2S_2O_3$ twice. The organic layer was then dried with Na_2SO_4, filtered and concentrated in vacuo. Purification of crude product by column chromatography afforded the product **3** as liquid (8.6 g, 72%).

To a stirred solution of LAH (2.8 g, 80 mmol) in dry THF (20 mL) at 0 ^{o}C was added dropwise a solution of **3** (7.8 g, 62.9 mmol) in THF under argon. The mixture was stirred for 4 hours at room temperature It was then cooled to at 0 ^{o}C, diluted with ether, and quenched by dropwise addition of a saturated solution of Na_2SO_4 (15 mL). The reaction mixture was extracted with EtOAc. The organic layer was then dried with Na_2SO_4, filtered and concentrated in vacuo. Purification of crude product by column chromatography afforded the product **4** as liquid (5.78 g, 73%).

NaH (2.28 g, 52.8 mmol) was washed with dry hexane (15 mL) and then decanted. Dry THF (20 mL) was added and the mixture was allowed to stir at room temperature cis-Cyclooctene-4-ol **4** (1.668 g, 13.22 mmol) in THF (15 mL) was added to the flask. The mixture was stirred and heated to reflux for 1 hour. α-Bromoacetic acid (1.838 g, 13.22 mmol) in THF (25 mL) was added and the mixture was allowed to reflux overnight. The mixture was cooled to room temperature and then concentrated on the rotary evaporator. The residue was cooled in an ice bath and water was added followed by acidification with 3 M HCl. The aqueous layer was extracted with three portions of Et_2O. The extracts were dried with Na_2SO_4 and concentrated *in vacuo* to provide an oil compound. The title compound was used directly for the next reaction without further purification.



Scheme III-2. Synthesis of NCAA 1.



Scheme III-3. Synthesis of NCAA 2.

The above acid compound (1.8 g, 9.89 mmol) was dissolved in DMF (20 mL). K_2CO_3 (2.71 g, 19.63 mmol) was added followed by dropwise addition of methyl iodide (1.2 mL, 19.78 mmol). The mixture was stirred at 60 ^oC for 1 hour. The reaction system was diluted with EtOAc, and then washed sequentially with water and HCl (1 M). The organic layer was separated and dried with anhydrous Na_2SO_4. After filtration, the organic layer was concentrated. Column chromatography with hexanes/EtOAc as eluent (9:1 v/v) afforded pure methylester product **5** in (1.57 g, 60%).

To a stirred solution of LAH (0.275 g, 7.85 mmol) in dry THF (10 mL) at 0 ^oC was added dropwise a solution of **5** (1.3 g, 6.56 mmol) in THF under argon. The mixture was stirred for 4 hours at room temperature. It was then cooled to 0 ^oC, diluted with ether, and quenched by dropwise addition of a saturated solution of Na_2SO_4 (15 mL). The reaction mixture was extracted with EtOAc. The organic layer was then dried with Na_2SO_4, filtered and concentrated *in vacuo*. Purification of crude product by column chromatography afforded the product **6** as liquid (0.66 g, 60%).

(Z)-2-(cyclooct-4-en-1-yloxy) ethanol **6** (3.8 g, 18.6 mmol) and methyl benzoate (4.98 g, 37.2 mmol) were dissolved in 500 mL of 9:1 ether:hexane in a quartz flask. The photoisomerization was carried out using the flow apparatus described previously.[1,2] The column was packed with 8.5 cm of silica gel, and then silver impregnated silica gel (16 g) on top. The column was flushed with 9:1 ether:hexane (250 mL). The pump was turned on at a flow rate of 100 mL/min and irradiation begun. Photoisomerization of the mixture was carried out for 6 hours.

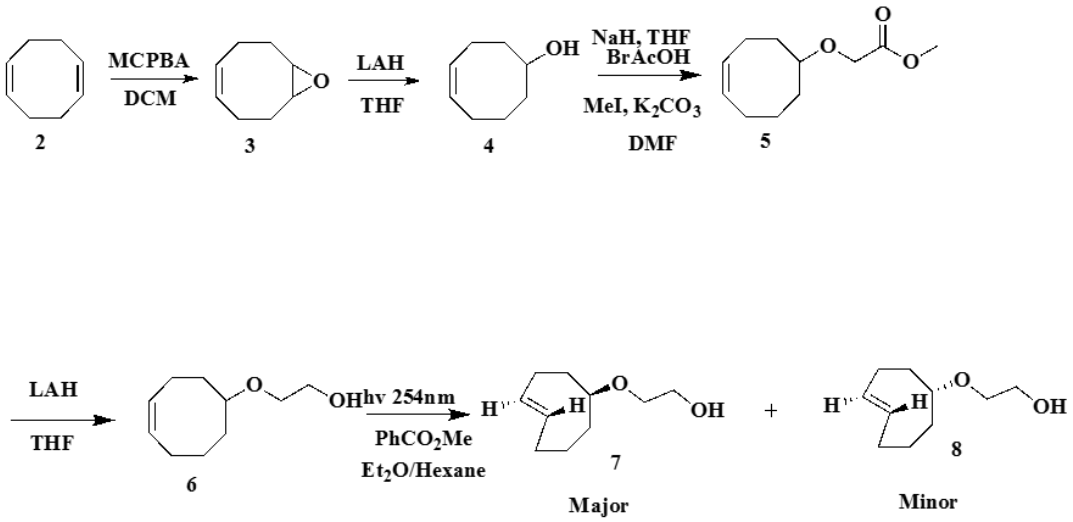
The column was flushed with 9:1 ether:hexane (250 mL) and then dried with compressed air. The silica was placed into an Erlenmeyer flask and stirred with ammonium hydroxide (200 mL) and methylene chloride (200 mL) for 5 minutes. The silica gel was filtered and the filtrate was placed into a separatory funnel. The organic layer was separated, and the ammonium hydroxide layer was extracted three times with methylene chloride. The organic layers were combined and washed with water twice. The organic layers were then dried with $MgSO_4$, filtered, and purified by column chromatography with 5% ether to 30% ether in hexanes. Two diastereomers, 0.989 g (32%) of compound **7** (DS2) and 0.366 g (18%) of compound **8** (DS1) (Scheme III-4) as colorless oils, were isolated.

Spectroscopic Properties of the Major Diastereomer 7

^{1}H NMR ($CDCl_3$, 300 MHz) δ 5.66-5.56 (m, 1H), 5.45-5.36 (m, 1H), 3.70-3.35 (m, 2H), 3.55-3.48 (m, 1H), 3.45-3.40 (m, 1H), 3.08-3.05 (d, 1H, J = 0.9 Hz), 2.41-2.31 (m, 2H), 2.20-1.89 (m, 6H), 1.64-1.48 (m, 2H). ^{13}C NMR (CDCl3, 75 MHz) δ 135.3, 132.3, 86.0, 69.1, 62.0, 40.7, 37.7, 34.5, 32.9, 31.6.

Spectroscopic Properties of the Minor Diastereomer 8

^{1}H NMR (CDCl3, 300 MHz) δ 5.46-5.60 (m, 2H), 3.78-3.67 (m, 2H), 3.65-3.62 (m, 1H), 3.61-3.55(m, 1H), 3.42-3.52 (m, 1H), 2.40-2.02 (m, 4H), 1.88-1.79 (m, 4H), 1.58-1.50 (m, 1H), 1.28-1.16 (m, 1H). ^{13}C NMR (CDCl3, 75 MHz) δ 135.8, 131.3, 74.9, 69.6, 62.1, 40.1, 34.4, 32.9, 29.8, 27.6.



Scheme III- 4. Synthesis of transcyclooctene derivatives.

DIAD (0.533 g, 2.64 mmol) was added dropwise slowly to an oven dried flask charged with compound **8** (0.3 g, 1.76 mmol), compound **9** (0.78 g, 2.64 mmol) and TTP (0.646 g, 2.47 mmol) were dissolved in Dry THF¬ (10 mL)) and the mixture was cooled to 0 oC. The reaction mixture was stirred at room temperature overnight. The reaction mixture was then concentrated and subjected directly to column chromatography with hexane/ EtOAc (4:1) as eluent to afford 10 (0.47 g , 60% yield) as light yellow oil.

Compound **10** (0.4 g, 0.65 mmol) was dissolved in THF (8 mL), and aqueous NaOH (1.0 M, 1.06 mL) was added dropwise. The mixture was stirred at room temperature for 2 hours, diluted with water (5 mL), and then extracted with Et2O (100 mL). The aqueous layer was collected, and adjusted to pH = 1 with HCl (3.0 M in water). The resulting mixture was extracted with EtOAc (10 mL × 2) and the combined organic layers were dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to get acid compound, which was used directly in the next step without further purification.

The compound (0.21 g, 0.50 mmol) obtained from the above step was dissolved in anhydrous dioxane (8 mL) followed by an injection of HCl/dioxane solution (4.0 M, 0.5 mL). The resulting mixture was stirred at room temperature for 12 hours. The white solid was collected by filtration and washed sequentially with EtOAc and DCM, then dried in vacuo using oil pump to afford the final product **NCAA 3** as white solid (0.179 g, 56% over two steps). (Scheme III-5).

^{1}H NMR (CD_3OD, 300 MHz) δ 7.21 (d, 2H, *J* = 5.1 Hz), 6.99 (d, 2H, *J* = 4.8 Hz), 5.65 (m, 2H), 4.30-4.21 (m, 2H), 4.12-4.09 (m, 1H), 3.87 (t, 1H, *J* = 5.1 Hz), 3.82-3.62 (m, 1H), 3.60-3.55 (m, 1H), 3.24 (ABq, 2H, *J* = 6.9, 14.4 Hz), 2.20-2.14 (m, 2H), 2.09-1.98 (m, 2H), 1.95-1.60 (m, 6H).

^{13}C NMR (CD_3OD, 75 MHz) δ 169.8, 158.5, 130.4, 130.3, 114.9, 114.8, 79.2, 67.6, 66.7, 61.8, 53.8, 36.1, 35.0, 32.1, 19.9.

DIAD (1.60 g, 7.94 mmol) was added dropwise slowly to an oven dried flask charged with compound **7** (0.9 g, 5.29 mmol), compound **9** (2.34 g,7. 94 mmol) and TTP (1.94 g, 7.41 mmol) in Dry THF (20 mL) and the mixture was cooled to 0 ^{o}C. The reaction mixture was stirred at room temperature overnight, concentrated, and then subjected directly to column chromatography with hexane/ EtOAc (4:1) as eluent to afford **11** (1.41 g , 60% yield) as light yellow oil.

Compound **11** (1.2g, 1.95 mmol) was dissolved in THF (8 mL) and aqueous NaOH (1.0 M, 3.2 mL) was added dropwise. The mixture was stirred at room temperature for 2 hours, diluted with water (10 mL), and then extracted with Et2O (20 mL). The aqueous layer was collected, and adjusted to pH = 1 with HCl (3.0 M in water). Then the resulting mixture was extracted with EtOAc (10 mL × 2) and the combined organic layers were dried over anhydrous Na2SO4 and concentrated under reduced pressure to get an acid compound, which was used directly in the next step without further purification.

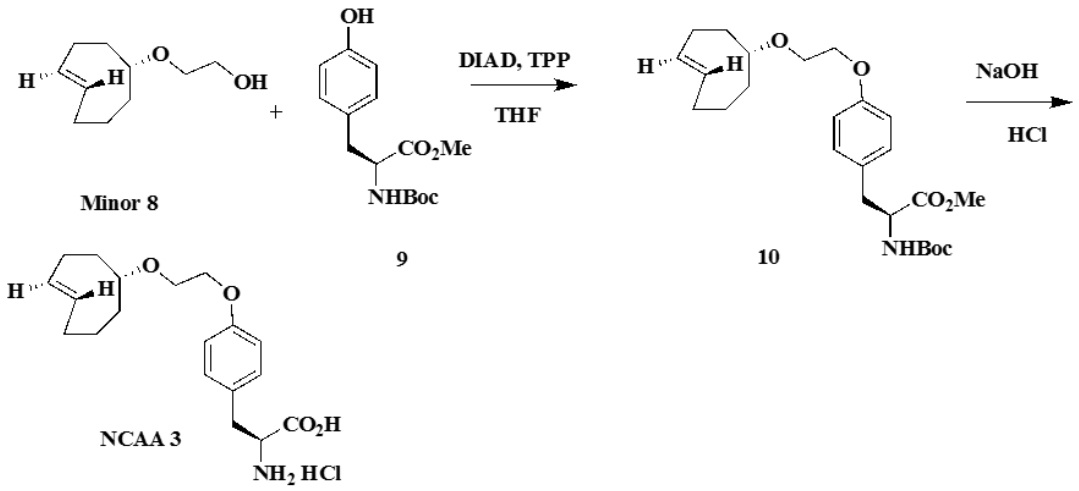
The compound (0.65 g, 1.50mmol) obtained from the above step was all dissolved in anhydrous dioxane (8 mL) followed by an injection of HCl/dioxane solution

(4.0 M, 1.5 mL). The resulting mixture was stirred at room temperature for 12 hours. The white solid was collected by filtration, washed sequentially with EtOAc and DCM, and then dried in vacuo using oil pump to afford the final product **NCAA 4** (Scheme III-6) as white solid (0.538g, 56% over two steps).

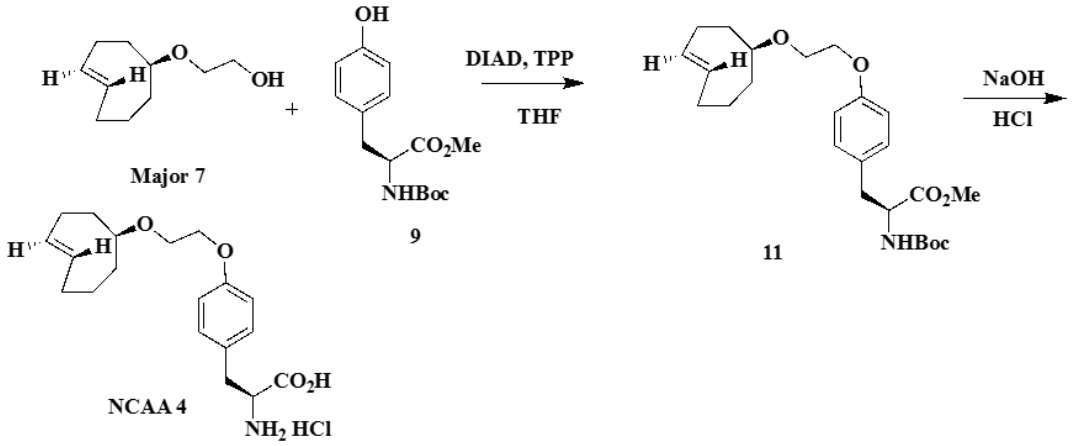
^{1}H NMR (CD_3OD, 300 MHz) δ 7.30 (d, 2H, *J* = 9.0 Hz), 6.94 (d, 2H, *J* = 8.7 Hz), 5.61 (ddd, 1H, *J* = 10.8, 15.9 Hz), 5.41 (ddd, 1H, *J* = 7.5, 16.2 Hz) 4.24-4.20 (m, 1H), 4.11-4.09 (m, 2H), 3.82-3.68 (m, 1H), 3.66-3.62 (m, 2H), 3.20 (ABq, 2H, *J* = 6.9, 14.7 Hz), 2.39-2.23 (m, 2H), 2.19-2.08 (m, 2H), 1.96-1.80 (m, 4H), 1.61-1.47 (m, 2H).

^{13}C NMR (CD_3OD, 75 MHz) δ 169.8, 158.8, 134.9, 131.9, 130.3, 126.1, 114.8, 88.0, 67.5, 66.5, 53.8, 40.4, 37.5, 35.0, 34.0, 32.5, 31.3.

A solution of L-tyrosine (5.45 g, 30 mmol) in methanolic HCl was stirred at reflux overnight under argon and then taken to dryness to give pure methyl tyrosinate hydrochloride (6.84 g, 98%). To a solution of sodium carbonate (3.87 g, 46.5 mmol) in water (70 mL) was added methyl tyrosinate hydrochloride (3.55 g, 15.25 mmol). After formation of a clear solution, Fmoc-O-(N-hydroxysuccinimide) (5.2 g, 15.35 mmol) in dioxane (70 mL) was added dropwise, giving a white suspension. The mixture was stirred under argon overnight, adjusted to pH 2 by addition of 6 N HCl, extracted with EtOAc (3 x 100 mL), washed with brine (2 x 20 mL), and dried (NaS04). Removal of



Scheme III-5. Synthesis of NCAA 3.



Scheme III-6. Synthesis of NCAA 4.

solvent and purification by silica gel chromatography (hexane/EtOAc, 4: 1) afforded methyl -Fmoc-L-tyrosinate **(3)** (6.35 g, 99%).

The synthesis of cyclooctyne derivatives followed the procedures developed by Lemke *et al.*[131] 4.0 g of cycloheptene (41.6 mmol), 9.33 g of Bu ^{t}OK (83.3 mmol), and 9 ml of anhydrous pentane are introduced into a dry round-bottomed flask under argon. The solution, placed in an ice/salt bath (<-5 ^{o}C), is vigorously stirred and then 5.5 ml of bromoform (62.5 mmol) are added dropwise. Once the addition is complete, the mixture is allowed to return to room temperature overnight under argon and with vigorous stirring. Approximately 50 ml of water are subsequently added and the pH is neutralized with 1M HCl. The organic and aqueous phases are separated; the aqueous phase is extracted with 3 × 20 ml of pentane and the pentane phase is washed with 3 × 20 ml of water. After drying over $MgSO_4$, the solvent is evaporated under vacuum to give orange-yellow oil. The product **4** is subsequently purified by filtration through silica with cyclohexane/AcOEt (19:1) as eluent to afford colorless oil (7.80 g, 70% yield).

Compound **4** (6.0 g, 22.3 mmol) and anhydrous ethane-1,2-diol (22.2 mL, 447 mmol, 20 eq) were dissolved in anhydrous acetone (38 mL), Anhydrous $AgClO_4$ (13.9 g, 67.1 mmol) was added in small portions under exclusion of light and stirred at room temperature for 1 hour. After addition of EtOAc (100 mL) and filtration, 1M HCl (100 mL) was added and the aqueous layer was extracted with EtOAc (3x50 mL). The combined organic layer were washed with 1M HCl in saturated NaCl solution (100 mL each) and dried over Na_2SO_4. The solvent was evaporated under reduced pressure and **5** (2.11 g, 38%) was obtained as yellow oil and used without further purification.

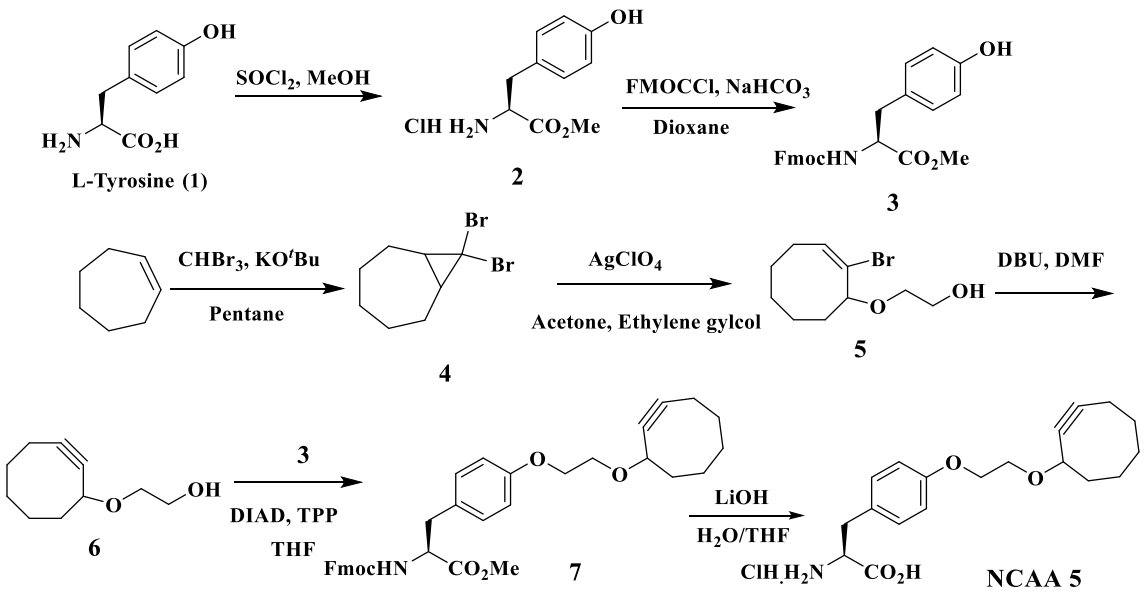
Compound **5** (3 g, 12.0 mmol) was dissolved in DMSO (30 mL) and heated to 60 ^{o}C. DBU (3.66 mL, 24.0 mmol, 2.0 eq) was added, the resulting solution was stirred for 15 minutes and more DBU (14.6 mL, 96.0 mmol, 8.0 eq) was added. The mixture was stirred at 60 ^{o}C overnight and then cooled to room temperature EtOAc (100ml) and water (100 ml) were added. After acidification to pH 1 with concentrated HCl, the aqueous phase was extracted with EtOAc (3 x 50ml). The combined organic layer were washed with 1M HCl in saturated NaCl solution (100 each) and dried over Na_2SO_4. The solvent was evaporated under reduced pressure. The compound **6** (COY, 1.20 g, 60% yield) was afforded after column chromatography as light yellow oil.

DIAD (1.98 g, 9.82 mmol) was added dropwise slowly to an oven dried flask charged with compound **6** (1.10 g, 6.54 mmol), compound **3** (2.89 g, 9.82 mmol) and TTP (2.6 g, 9.82 mmol) dissolved in Dry THF (20 mL) and the mixture was cooled to 0 ^{o}C. The reaction mixture was stirred at room temperature overnight. The reaction mixture was then concentrated and subjected directly to column chromatography (hexane/EtOAc as 4:1) to afford **7** (1.74 g, 60% yield) as light yellow oil.

Lithium hydroxide monohydrate (44 mg, 1.05 mmol) was added to a stirring solution of compound **7 (**0.5 g, 0.881 mmol**)** in THF/H_2O (3:1, 8 mL). The solution was stirred for 4 hours at room temperature and EtOAc (100 mL) and H_2O (100 mL) were added. The aqueous phase was carefully acidified to pH 4 by the addition of HCl and extracted with EtOAc (4 x 50 mL). The aqueous phase was then evaporated under reduced pressure to yield **NCAA 5** (Scheme III-7), (0.2 g, 61.9%**)** as a white solid.

^{1}HNMR (300MHz, CD_3OD) δ 7.27(d 2H, *J* = 7.8 Hz), 6.95 (d 2H, *J* = 7.8 Hz), 4.32-4.28 (m, 1H), 4.10-4.14 (m, 3H), 3.96-3.80 (m, 1H), 3.62-3.75 (m, 1H), 3.20 (ABq, 2H, *J* = 22.0, 6.6 Hz), 2.24-2.04 (m, 2H), 1.98-1.89 (m, 6H), 1.85-1.50 (m, 2H),

^{13}CNMR (75 MHz, CD_3OD): δ 169.9, 156.8, 130.2, 124.5, 115.4, 99.4, 92.2, 72.4, 70.1, 60.7, 53.9, 41.9, 35.0, 34.0, 29.5, 26.0, 19.8.



Scheme III-7. Synthesis of NCAA 5.

NaH (1.5 g, 62.5 mmol) was washed with dry hexane (15 mL) and then decanted. Dry THF (15 mL) was added and the mixture was allowed to stir at room temperature 5-Norbornene-2-methanol (2.0 g, 16.12 mmol) in THF (15 mL) was added to the flask. The mixture was stirred and heated to reflux for 1 hour. α-Bromoacetic acid (2.23 g, 16.12 mmol) in THF (10 mL) was added and the mixture was allowed to reflux overnight. The mixture was cooled to room temperature and then concentrated on the rotary evaporator. The residue was cooled in an ice bath and water was added followed by acidification with 3 M HCl. The aqueous layer was extracted with three portions of Et2O. The extracts were dried over Na2SO4 and concentrated in vacuo to provide an oil-type compound. The title compound was used directly for the next reaction without further purification.

The above acid compound (1.5 g, 8.24 mmol) was dissolved in DMF (20 mL) followed by the addition of K_2CO_3 (2.27 g, 16.48 mmol). Methyl iodide (1.05 mL, 16.48 mmol) was then added dropwise. The mixture was stirred at 60 ^oC for 1 hour. The reaction mixture was diluted with EtOAc, and then washed sequentially with water and HCl (1 M). The organic layer was separated and dried over anhydrous Na_2SO_4. After filtration, the organic layer was concentrated. Column chromatography with Hexanes/EtOAc as eluent (9:1 v/v) afforded pure methylester product **2** (1.11 g, 72%).

To a stirred solution of LAH (0.214 g, 6.12 mmol) in dry THF (10 mL) at 0 ^oC was added dropwise a solution of **2** (1.0 g, 5.10 mmol) in THF under argon. The mixture was stirred for 4 hours at room temperature It was then cooled to 0 ^oC, diluted with Et_2O and quenched by dropwise addition of a saturated solution of Na_2SO_4 (10

mL). The reaction mixture was then extracted with EtOAc. The organic layer was dried with Na_2SO_4, filtered and concentrated *in vacuo*. Purification of crude product by column chromatography afforded the product **3** (0.59 g, 70%) as liquid.

DIAD (0.9 g, 4.41 mmol) was added dropwise slowly to an oven dried flask charged with compound **3** (0.5 g, 2.97 mmol), compound **4** (1.31 g, 4.46 mmol), and TTP (1.09 g, 4.16 mmol) dissolved in Dry THF (10 mL) and the mixture was cooled to 0 ^oC. The reaction mixture was stirred at room temperature overnight. The reaction mixture was then concentrated and subjected directly to column chromatography with Hexane/ EtOAc (4:1) as eluent to afford **5** (0.86 g, 65%) as light yellow oil.

Compound **5** (0.8 g, 1.79 mmol) was dissolved in THF (10 mL), and aqueous NaOH (1.0 M, 2.33 mL) was added dropwise. The mixture was stirred at room temperature for 2 hours, diluted with water (10 mL), and then extracted with Et_2O (20 mL). The aqueous layer was collected, and adjusted to pH = 1 with HCl (3.0 M in water). The resulting mixture was extracted with EtOAc (2 × 10 mL). The combined organic layers were dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to get acid compound**,** which was used directly in the next step without further purification.

The compound (0.6 g, 1.39 mmol) obtained from the above step was dissolved in anhydrous dioxane (10 mL) followed by an injection of HCl/dioxane solution (4.0 M, 0.696 mL). The resulting mixture was stirred at room temperature for 12 hours. The white solid was collected by filtration, washed sequentially with EtOAc and DCM, and

then dried *in vacuo* using oil pump to afford the final product **NCAA 6** (Scheme III-8) as white solid (0.27 g, 53% over two steps).

^{1}H NMR (CD_3OD, 300 MHz) δ 7.20 (d, 2H, *J* = 8.7 Hz), 6.93 (d, 2H, *J* = 7.5 Hz), 6.13-6.07 (m, 1H), 5.92 (dd, 1H, *J* = 3.3, 6.0 Hz), 4.18 (t, 1H, *J* = 5.2 Hz), 4.13-4.07 (m, 2H), 3.79-3.69 (m, 2H), 3.23-3.07(m, 4H), 3.04(m, 1H), 2.88(m, 1H), 2.39-2.34(m, 1H), 1.86-1.76(m, 1H), 1.52-1.24(m, 2H), 0.51(dd, 1H, *J* = 3.3, 7.5 Hz).

^{13}C NMR (CD_3OD, 75 MHz) δ 169.8, 158.5, 136.7, 136.3, 136.1, 132.0, 130.3, 126.1, 114.8, 74.7, 69.1, 67.2, 53.9, 48.9, 44.5, 42.0, 38.4, 35.0, 28.6.

DIAD (1.22 g, 6.03 mmol) was added dropwise slowly to an oven dried flask charged with compound **1** (0.5 g, 4.03 mmol), compound **2** (1.70 g, 5.76 mmol) and TTP (1.60 g, 6.03 mmol) dissolved in Dry THF (15 mL) and the mixture was cooled to 0 ^{o}C. The reaction mixture was stirred at room temperature overnight. The reaction mixture was then concentrated and subjected directly to column chromatography with hexane/EtOAc (4:1) to afford **3** (1.04 g, 65%) as light yellow oil.

Compound **3** (1.0 g, 2.49 mmol) was dissolved in THF (10 mL) and aqueous NaOH (1.0 M, 3.2 mL) was added dropwise. The mixture was stirred at room temperature for 2 hours, diluted with water (10 mL), and then extracted with Et_2O (20 mL). The aqueous layer was collected and adjusted to pH = 1 with HCl (3.0 M in water). The resulting mixture was extracted with EtOAc (2 × 10 mL) and the combined organic layers were dried over anhydrous Na_2SO_4 and concentrated under reduced pressure to get acid compound**,** which was used directly in the next step without further purification.

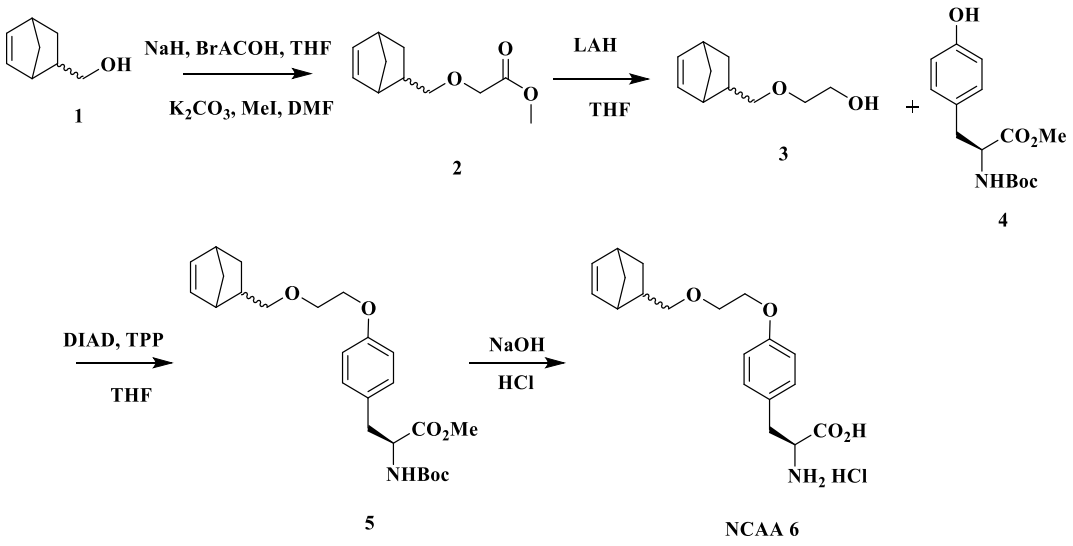
The compound (0.8 g, 2.58 mmol) obtained from the above step was dissolved in anhydrous dioxane (10 mL), and followed by an injection of HCl/dioxane solution (4.0 M, 1.29 mL). The resulting mixture was stirred at room temperature for 12 hours. The white solid was collected by filtration, washed sequentially with EtOAc and DCM, and then dried *in vacuo* using oil pump to afford the final product **NCAA 7** (Scheme III-9) as white solid (0.4 g, 66% over two steps).

^{1}H NMR (CD_3OD, 300 MHz) δ 7.23 (d, 2H, *J* = 5.1Hz), 6.91(d, 2H, *J* = 6.9 Hz), 6.18 (m, 1H), 5.95 (dd, 1H, *J* = 3.0, 5.7 Hz), 4.21 (t, 1H, *J* = 5.7 Hz), 3.73 (t, 1H, *J* = 6.6 Hz), 3.57 (t, 1H, *J* = 9.3 Hz), 3.19 (ABq, 2H, *J* = 7.8, 14.4 Hz), 3.02 (brs, 1H), 2.85 (brs, 1H), 2.85-2.58 (m, 1H), 1.99-1.92 (m, 1H), 1.48-1.31 (m, 2H), 0.65 (ddd, 1H, *J* = 2.4, 4.2, 6.9 Hz).

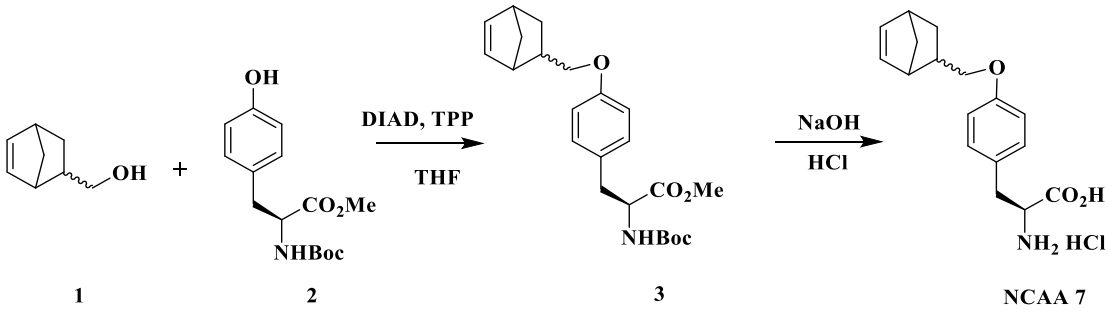
^{13}C NMR (CD_3OD, 75 MHz) δ 169.8, 158.8, 137.1, 131.8, 130.2, 125.0, 114.6, 71.1, 53.8, 48.9, 43.7, 42.0, 38.2, 28.4.

Kinetic Characterizations

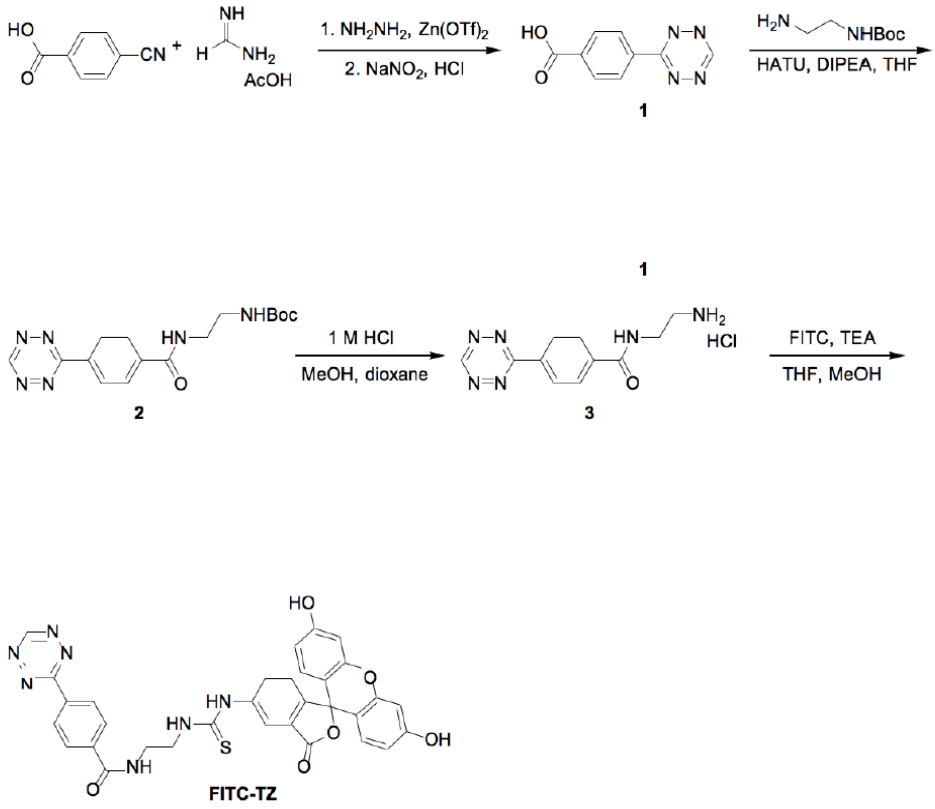
All kinetic measurements were carried out using a PTI QMA40 fluorescent spectrophotometer. For IEDDAC reactions between strained alkenes/alkynes and FITC-TZ (Scheme III-10), fluorescent emission was monitored at 515 nm with an excitation light at 493 nm. For NADC reactions between strained alkenes/alkynes with a hydrazonyl chloride (HZCL) that served as a nitrilimine precursor, fluorescent emission was monitored at 480 nm with an excitation light at 318 nm. Data were collected during a reaction period of 1 hour.



Scheme III-8. Synthesis of NCAA 6.



Scheme III-9. Synthesis of NCAA 7.



Scheme III-10. Synthesis of FITC-TZ.

DNA and Protein Sequences

Gene and protein sequences of pylT, *Methanosarcina mazei* PylRS and sfGFP were listed in the DNA sequence of Experimental Section in Chapter II.

Construction of Plasmid

Plasmid Construction - pEVOL-pylT-N346A/C348A/Y384F/Y306A

The plasmid pEVOL-pylT- Y306A/N346A/C348A/Y384F was derived from pEVOL-pylT-N346A/C348A[132] that carried a $tRNA_{CUA}^{Pyl}$ under the control of the *proK* promoter and terminator and a N346A/C348A form of the *Methanosarcina mazei* PylRS gene under the control of the *glnS* promoter and terminator. Two primers PylRS-Y384F-Fwd: 5'-tggggataccttgatgtaatgcacg-3' and PylRS-Y384F-Rev: 5'-aagaccatgcaggaatcgcctacgat-3' were used to run quick-change mutagenesis to introduce the Y384F mutation. The primers PylRS-Y306A-Fwd: 5'-aactacctgcgcaagcttgacagggc-3' and PylRS-Y384F-Rev: 5'-cgcaaggtttggagcaagcatgggt-3' were used to run quick-change mutagenesis to introduce the Y306A mutation.

sfGFP Protein Expression and Purification

sfGFP Protein Expression

The pET-sfGFPS2TAG plasmid previously reported[132] was co-transformed with pEVOL-pylT-N346A/C348A/Y384F/Y306A into *E. coli* BL21 (DE3) cells. From a single colony, a 5 mL overnight culture was prepared and inoculated into 300 mL LB medium supplemented with 100 µg/mL Ampicillin and 34 µg/mL Chloramphenicol and let grow at 37 °C. When OD_{600} reached 0.6, expression was induced with 1 mM IPTG, 0.2 % arabinose and 1 mM of a NCAA. After induction for 8 hours, cells were harvested

and suspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), sonicated in an ice water bath (ice bath temperature was 10 °C before each 3 minute run of the sonication). The cell lysate was clarified by centrifugation (4 °C for 60 minutes at 11,000 g).

The crude supernatant was decanted and allowed to bind to 5 mL of Ni-NTA superflow resin at 4 °C for 1 hour. The mixture was loaded into a empty cartridge and washed with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0). sfGFP was eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). Eluted fractions were concentrated in 10 k MWCO cut filter unit and dialyzed against 10 mM ammonium bicarbonate. The protein purity was analyzed by 12% SDS-PAGE and characterized using electrospray ionization mass spectrometry (ESI-MS).

Protein Labeling

FITC-TZ Protein Labeling

0.1 mM FITC-TZ was added to each of the solutions of sfGFP incorporated with a NCAA in PBS (pH 7.4) buffer. The reaction was carried out at room temperature for 1 hour. The protein was purified with NI-NTA resin and analyzed with 12% SDS-PAGE. The fluorescence was detected using a regular UV transilluminator (Figure III-1).

HZCL Protein Labeling

0.5 mM HZCL was added to a solution of sfGFP incorporated with a NCAA in PBS (pH 7.4) buffer. The reaction was also carried out at room temperature for 1 hour.

The protein was analyzed with 12% SDS-PAGE. The fluorescence was detected in a Biorad ChemiDoc XRS+ system under UV irradiation (Figure III-1).

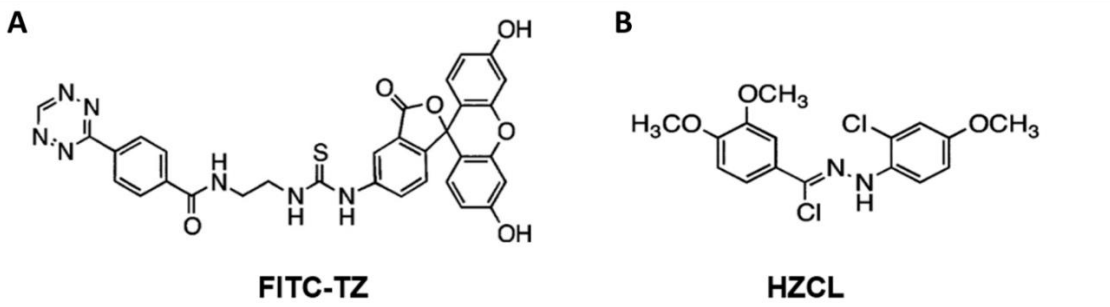


Figure III-1. Structure of a (A) Fluorescein Tetrazine Dye (FITC-TZ) (B) Hydrozonoyl Chloride, HZCL.

Fluorescent Protein Labeling on Living Cell

BL21(DE3) cell was transformed to carry plasmids pEVOL-PylT-Y306A/N346A/C348A/Y384F and pETDuet-OmpXTAG. A medium shift protocol was used to overexpress the OmpX protein. Bl21(DE3) cell was grown to OD_{600} of 1 and pelleted. After 3 washes of the cell pellet with PBS, the cells were resuspended in minimal medium supplied with IPTG (0.5 mM), arabinose (0.2%) and a NCAA **4**, **5** and **6** (5 mM), and was incubated at 25 ^{o}C for 20 hours. The cell was then pelleted and washed with isotonic saline 3 times before incubated with FITC-TZ (0.5 mM) or HZCL (0.5 mM) at room temperature. The incubation time for cell labeling is typically 3 hours. After the reaction, the cell was pelleted by centrifugation, washed with PBS for 4 times and subjected to the fluorescence microscopy for cell imaging. The cell culture without a NCAA supplement in the medium served as the control and treated following the same protocol.

Microscopy

Microscopy was performed using an Olympus IX-81 inverted microscope (Olympus America, Center Valley PA, USA) equipped with UPLSAPO 100x/1.4 oil immersion objective, a Rolera XR CCD camera (Q imaging, Surrey BC, Canada) and a Proscan H117 motorized XY stage (Prior Scientific, Rockland, MA, USA) controlled by the μManager freeware (http://www.micro-manager.org). A field lens of 1.6x magnification was used to achieve Nyquist sampling for imaging. The following fluorescence filter sets (Chroma technology Corp., Bellows Falls, VT, USA) were used, with the central

wavelength and bandwidth of the excitation and emission filters as indicated: GFP (Ex. 470/40; Em. 525/50); DAPI (Ex. 350/50; Em. 460/50).

Results and Discussion

Kinetic characterization of EDDAC reactions of norbornene, transcyclooctene, and cyclooctyne

Kinetic investigations of reactions between strain-promoted alkene/alkyne dienophiles and tetrazines have been reported.[122, 133, 134] However, solvent systems, tetrazines, dienophiles, and methods of kinetic determination varied in previous reports, making it challenging to compare relative reactivities of dienophiles with tetrazines of interest. We chose to characterize EDDAC reactions of strain-promoted alkene/alkyne dienophiles in the PBS buffer at pH 7.4 to best represent their kinetics in living systems.

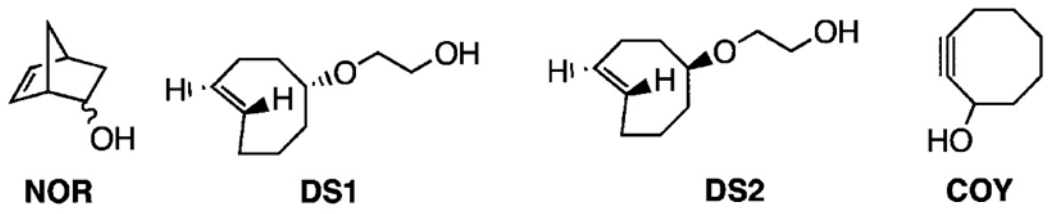


Figure III-2. Structures of four strain-promoted alkene/alkyne molecules.

I performed all experiments except the synthesis of the alkene/alkyne containing compounds (NCAA 1-7) and the FITC-TZ dye which were synthesized by Dr. Yadagiri Kurra. Characterizations of FITC-TZ and HZCl were also carried out by Dr. Yadagiri Kurra. Living cells were fluorescently labeled by Dr. Yan-Juin Lee.

Three strain-promoted alkenes, 5-norbornen-2-ol and two diastereomers of (*E*)-2-(cyclooct-4-en-1-yloxyl) ethanol, and cyclooct-2-ynol (NOR, DS1, DS2, and COY in Figure III-2) as dienophiles and a fluorescein tetrazine (FITC-TZ in Figure III-1) were chosen for kinetic studies. DS1 and DS2 are two diasteromeric products (ratio close to 1:2) of the UV-induced isomerization of (*Z*)-2-(cyclooct-4-en-1-yloxyl) ethanol. As minor products, DS1-like molecules were mostly dismissed in previous reports. Since structure differences between DS1 and DS2 may lead to different reactivities with a tetrazine, they were both included in our studies.

FITC-TZ has a tetrazine entity that efficiently quenches the fluorescence emission of fluorescein. Its reaction with a dienophile will remove this quenching effect, making the reaction to be easily tracked using a fluorometer. Since submicromolar concentrations can be applied in this kind of fluorescent studies, superfast reactions can be effectively analyzed. The reactions of all four dienophiles with FITC-TZ were carried out in pseudo-first-order conditions in which a dienophile was at least 20-fold of FITC-TZ (Figure III-1).

For NOR and COY, FITC-TZ was used at 1 μM for optimal signal-to-noise ratios of collected data. Concentrations of NOR and COY were chosen to give reliable data during a 1 hour reaction for the subsequent analysis. For DS1 and DS2, their reactions with FITC-TZ are so fast that the concentration of FITC-TZ had to be significantly reduced to 250 pM in order to detect the reaction process, leading to relatively low signal-to-noise levels. Data collected at all conditions were fit nicely to a single exponential increase equation $F = F_1 - F_2 \times e^{(-k' \times t)}$, where F is the detected

fluorescent signal at a give time, F_1 the final fluorescence, F_1-F_2 the background fluorescent signal, and *k'* the apparent pseudo first-order rate constant. The determined values of *k'* were plotted against dienophile concentrations and fitted to equation

$$k' = k \times [dienophile] + C$$

where k is a second-order rate constant of a reaction of a dienophile with FITC-TZ. The calculated second-order rate constants for all four dienophiles are shown in Table III-1.

Among all four dienophiles, NOR reacts slowest with FITC-TZ (Figure III-3). The determined rate constant is similar to what have been reported of reactions between

Table III-1. Second-order rate constants of reactions of strain-promoted alkene/alkyne dienophiles with FITC-TZ**.**

Dienophile	NOR	DS1	DS2	COY
$k/M^{-1}s^{-1}$	5.7 ± 0.2	292,000 ± 6000	64000 ± 4000	17 ± 1

norbornenes and tetrazines.[134-136] The determined rate constant for COY is 20-fold lower than what was reported by Lemke and coworkers.[137] This discrepancy may be attributed to different tetrazine dyes used in the two analyses. Both DS1 and DS2 react exceedingly fast with FITC-TZ, several orders of magnitude faster than reactions of NOR and COY. For DS2, its reaction with FITC-TZ has a rate constant in the same range of those of reported rapid transcyclooctene reactions with tetrazines.[133, 134, 138, 139]

The reaction of DS1 with FITC-TZ is almost five-fold of that of DS2. With this fast rate constant, a reaction with both reactants at 1 μM will finish to 97% in 100

seconds. As slightly minor products during synthesis, applications of DS1-like molecules have long been ignored. The current study indicates that DS1 is a much better choice than DS2 for applications that require labeling molecules at very dilute conditions such as in living systems.

To address this reactivity difference between DS1 and DS2, we carried out a detailed computational study at the CPCM-M06-2X/6-311+G(d,p) level of theory[140-142] This level of theory provides accurate reaction barriers for Diels-Alder cycloadditions, and has been employed in similar studies of strain-promoted click reactions by Houk and co-workers.[143] We predict free energy barriers of 14.0 and 15.5 kcal mol^{-1} for the EDDAC of slightly simplified models of DS1 and DS2 with 3-phenyl-1,2,4,5-tetrazine, respectively. This corresponds to a predicted rate for DS2 that is twelve-times that of DS1, in reasonable agreement with the experimentally observed five-fold difference.

The 1.5 kcal mol^{-1} difference in barrier heights can be explained in part through distortion-interaction analyses, in which the energy barrier for bimolecular is decomposed into the energy required to distort the reactants into the geometry of the transition state (E_{dist}) and the stabilization provided by bringing these two distorted reactants together (E_{int}). In particular, we found out that both the interaction energies and distortion energies slightly favor DS2 over DS1. This net 0.8 kcal mol^{-1} difference, combined with a 0.9 kcal mol^{-1} contribution from vibrational and entropy effects, leads to the final free energy difference of 1.5 kcal mol^{-1}.

Kinetic Characterization of NIADC Reactions of Norbornene, Transcyclooctene, and Cyclooctyne

It was demonstrated previously that olefins, serving as dipolarphiles, react with diaryl nitrile imines to form fluorescent pyrazoline products with appreciable reaction rates in mild aqueous conditions.[125] Due to the labile nature of nitrile imines, currently two methods are used to generate them transiently in aqueous conditions. One is the photolysis of tetrazoles; the other is the simultaneous dissociation of hydrozonoyl chlorides in water.[144, 145]

The first method has been significantly expanded by Lin and coworkers for photo-click labeling of biomolecules.[146-149] The second method has been recently applied to label biomolecules that contain norbornene and acrylamide functionalities.[50, 150] We chose the second method due to its relative simplicity during kinetic investigations.

We recently reported that NOR reacts readily with a hydrozonyl chloride shown as HZCL in Figure III-4 with a rate constant as 0.75 $M^{-1}s^{-1}$.[151] Like NOR, both DS1 and DS2 also reacted readily with HZCL, which could be detected by following the fluorescence emission of the final products. To determine second-order rate constants of DS1 and DS2 reactions with HZCL, reactions were set up under pseudo-first-order conditions in PBS/acetonitrile (1:1) at pH 7.4.

Data collected at all conditions were fit nicely to the same single exponential increase equation used in the aforementioned EDDAC analyses. The determined *k'*

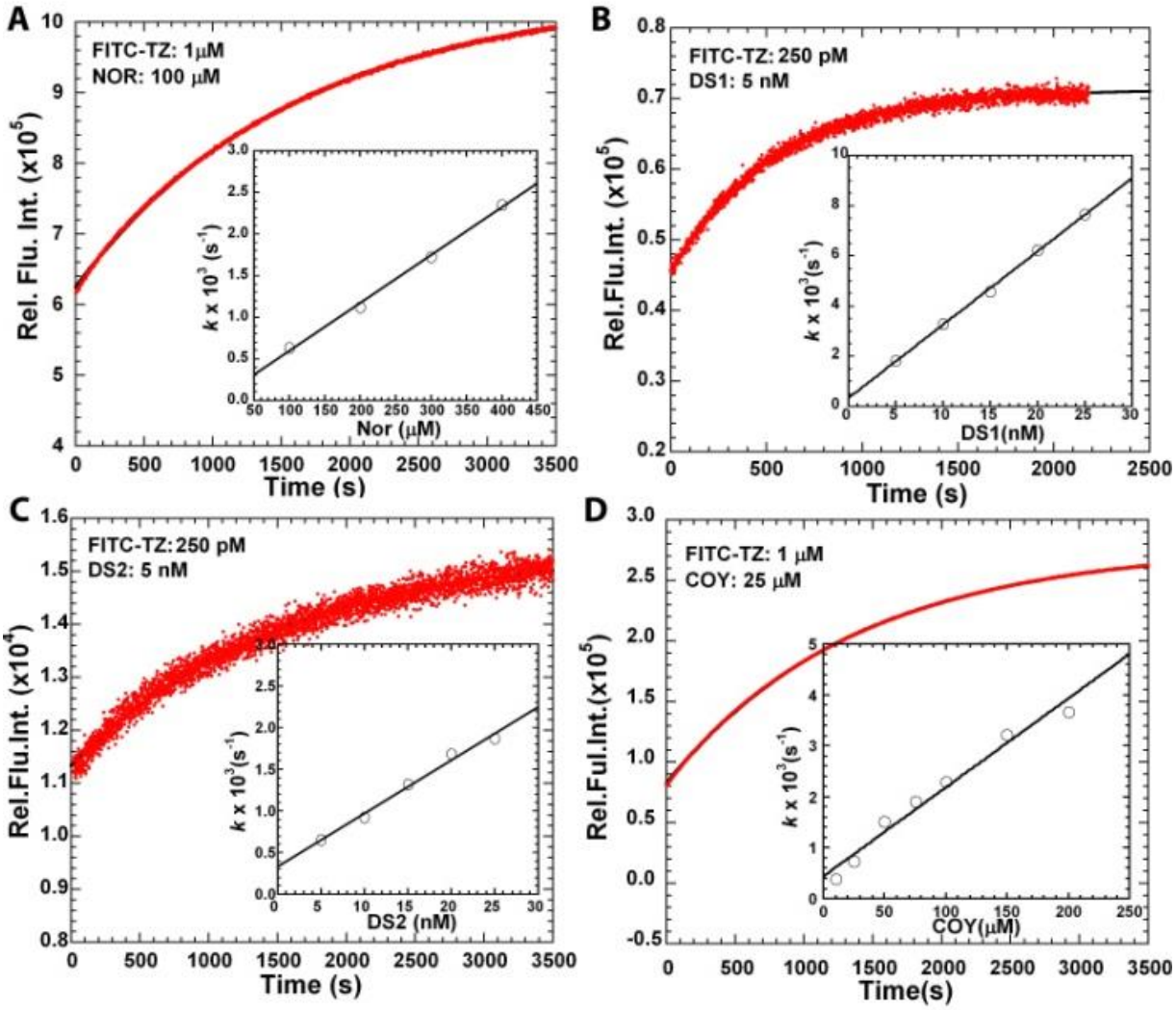


Figure III-3. Characterization of FITC reactions with (A) NOR, (B) DS1, (C) DS2, and (D) COY. All reactions were carried out in the PBS buffer at pH 7.4. The fluorescence emission was detected at 515 nm with the excitation light at 493 nm. For A-D, each presents the fluorescence change as a function of time at a given condition shown in the top left corner. The insets show the linear dependence of pseudo-first-order rate constants of a reaction on dienophile concentrations.

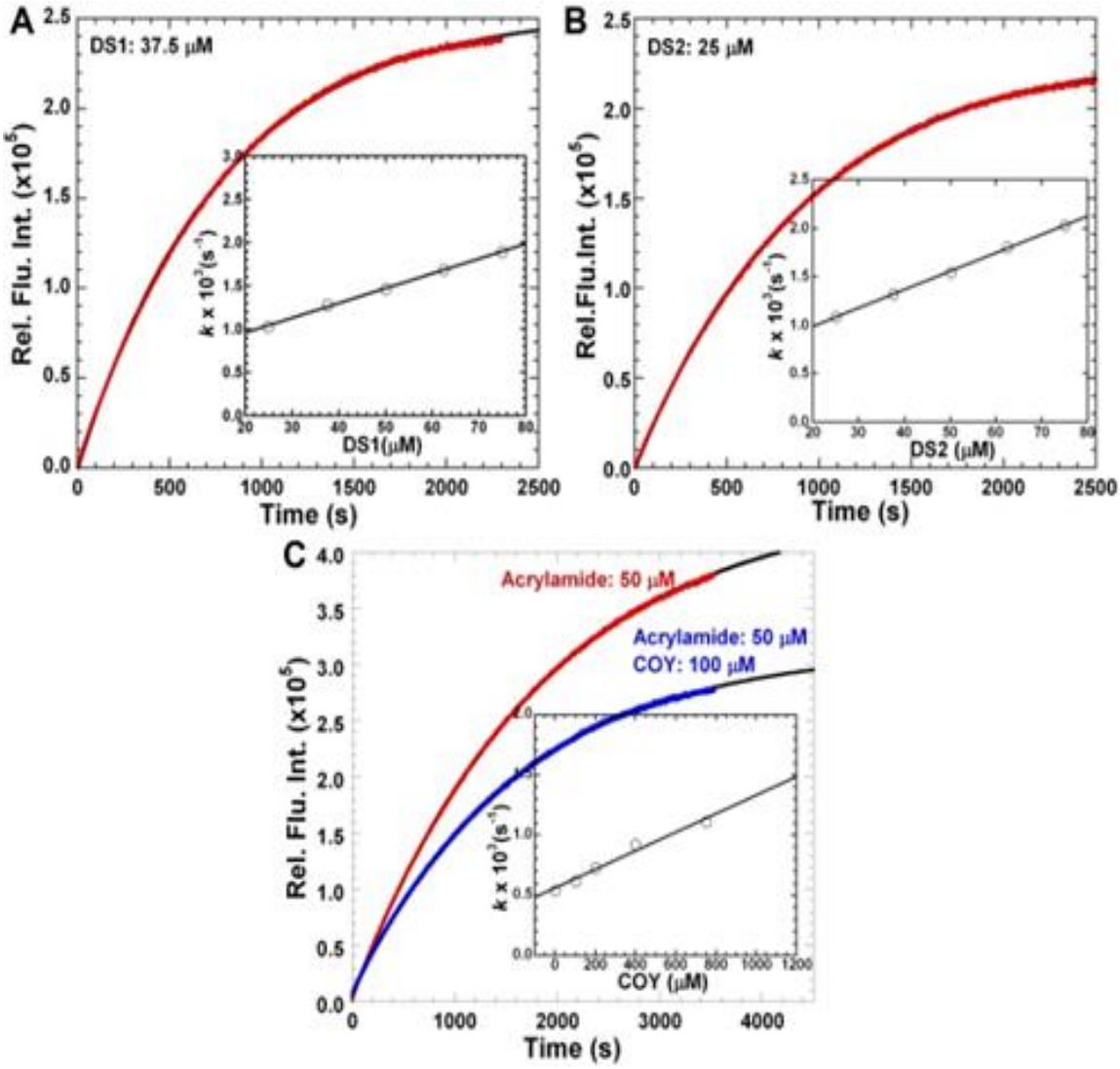


Figure III-4. Characterization of HZCL reactions with (A) DS1, (B) DS2, and (C) COY. All reactions were carried out in PBS/acetonitrile (1:1) at pH 7.4. HZCL was provided at 1 μM. The fluorescence emission was detected at 480 nm with excitation at 318 nm. The detected at 480 nm with excitation at 318 nm. The insets show the linear dependence of the pseudo-first-order rate constants on the dipolarophile concentrations.

values were plotted against dipolarphile concentrations and fitted to equation $k' = k_1 \times [dipolarphile] + k_2$, where k_1 is a second-order rate constant of a dipolarphile reaction with HZCL and k_2 is the rate constant of HZCL hydrolysis. The determined rate constants for DS1 and DS2 are shown in Table III-2. Unlike NOR, DS1, and DS2, COY does not yield a fluorescent product when it reacts with HZCL.

To fluorescently sensitize the COY reaction with HZCL, acrylamide, a dipolarphile that reacts competitively with HZCL to form a fluorescent product, was provided at 50 μM. Reactions at all conditions showed single exponential fluorescent increase over time. The determined pseudo-first-order rate constants were plotted against COY concentrations and fitted to the equation

$$k' = k_1 \times [COY] + k_2'$$

where k_1 is a second-order rate constant of the COY reaction with HZCL and k_2' is the sum of the rate constant of HZCL hydrolysis and the pseudo-first-order rate constant of the HZCL reaction with 50 μM acrylamide. The determined rate constant for COY is also presented in Table III-2.

As shown in Table III-2, COY has a reactivity toward HZCL similarly as NOR. Both DS1 and DS2 react rapidly with HZCL with rate constants more than 20-fold of those for NOR and COY. Although elevated energy states of DS1 and DS2 lead to higher reactivities toward HZCL than NOR and COY, the rate enhancement is not as significant as observed in EDDAC reactions.

Genetic Incorporation of Tyrosine Derivatives with Strain-Promoted Alkene/Alkyne Functionalities into Proteins in *E. coli*

Lysine derivatives that contain strain-promoted alkene/alkyne functionalities were previously incorporated into proteins in living cells using different engineered PylRS-$tRNA^{Pyl}_{CUA}$ pairs.[134, 136, 137, 152, 153] A different route was taken to install strain-promoted alkene/alkyne functionalities into proteins. As previously shown, the engineered N346A/C348A mutant of PylRS (PylRS(N346A/C348A)) demonstrated that this mutant enzyme in coordination with $tRNA^{Pyl}_{CUA}$ was able to incorporate close to 20 tyrosine and phenylalanine derivatives into proteins at amber mutation sites in *E. coli*.[77, 81, 154] Since one tyrosine derivative, *O*-benzyl-tyrosine is spatially close to tyrosine derivatives with *O*-linked norbornene, transcyclooctene, or cyclooctyne, we reasoned PylRS (N346A/C348A) might be applied for the genetic incorporation of some of these NCAAs.

Table III-2. Second-order rate constants of reactions between strain-promoted alkene/alkyne dipolarphiles and HZCL.

Dienophile	NOR[a]	DS1	DS2	COY
$k/M^{-1}s^{-1}$	0.75 ± 0.2	17.2 ± 0.4	19.0 ± 0.3	0.77 ± 0.07

[a] This rate constant was determined by Jiangyun Wang *et al*.[147]

The high substrate promiscuity of PylRS (N346A/C348A) might also allow using just one cellular system for the genetic installation of all strain-promoted alkene/alkyne functionalities, providing an advantage over using lysine derivatives. In addition,

syntheses of tyrosine derivatives are in general easier than those of lysine derivatives with fewer steps involved. Seven *O*-alkylated tyrosine derivatives with strain-promoted alkene/alkyne functionalities, NCAA **1**-**7** in Figure III-5 were synthesized.

Their genetic incorporation was tested in an *E. coli BL21* system that contained two plasmids, pEVOL-pylT-N346A/C348A and pET-pylT-sfGFP2TAG, with genes coding PylRS (N346A/C348A), $tRNA_{CUA}^{Pyl}$, and superfolder green fluorescent protein (sfGFP) with an amber mutation at its S2 position. This E. coli system was not able to express sfGFP in GMML (a minimal medium supplemented with 1% glycerol and 0.3 mM leucine). Providing 2 mM of either of NCAA **1**-**7** did not yield significant sfGFP expression increase. Apparently, the active site pocket of PylRS (N346A/C348A) is not big enough to accommodate the stretched and bulky side chains of NCAA **1**-**7**.

To expand the active site pocket, we introduced two more mutations Y306A and Y384F to PylRS (N346A/C348A) to afford PylRS (Y306A/N346A/C348A/Y384F) and used this gene to replace the PylRS (N346A/C348A) gene in pEVOL-pylT-N346A/C348A to obtain plasmid pEVOL-pylT-Y306A/N346A/C348A/Y384F. *E. coli* BL21 cells transformed with pEVOL-pylT-Y306A/N346A/C348A/Y384F and pET-pylT-sfGFP2TAG were not able to express sfGFP in lysogeny broth (LB) medium.

The Y306A mutation to PylRS (N346A/C348A) apparently decreases the activity of the original mutant enzyme on phenylalanine and reduces the misincorporation of phenylalanine in a rich medium. However, supplementing LB with 1 mM of one of five

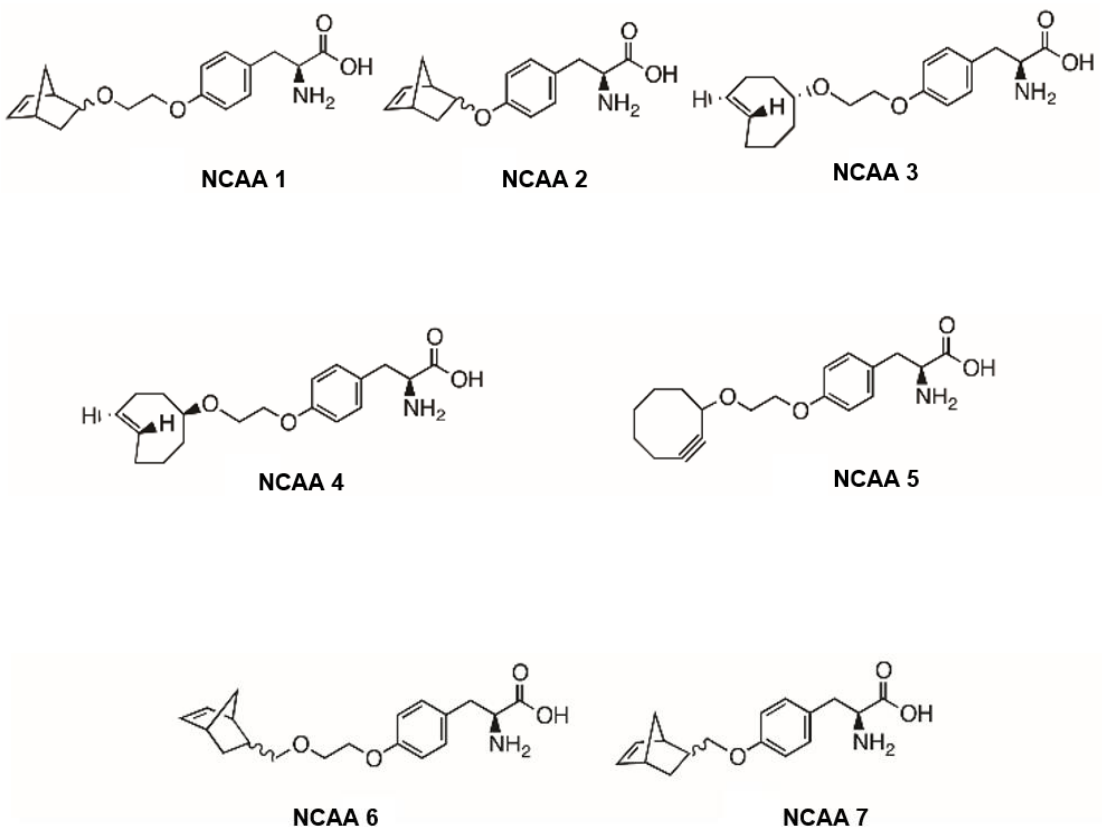


Figure III-5. Structures of O-alkylated tyrosine derivatives that contain strain-promoted alkene/alkyne functionalities.

NCAAs (**1**, **2**, **4**, **5**, and **6**) induced sfGFP overexpression (Figure III-6). The addition of 2 mM **NCAA 3** or **7** into LB medium only led to minimal sfGFP expression increase. All five purified sfGFP variants had molecular weights determined by electrospray ionization mass spectrometry analysis that agreed well with their theoretical molecular weights (Table III-3).

Labeling Proteins Incorporated with Strain-Promoted Alkene/Alkyne Functionalities using EDDAC and NIADC Reactions *In vitro*

With sfGFP proteins incorporated with different NCAAs in hand, we proceeded to test the labeling of the proteins with FITC-TZ and HZCL. To label sfGFP proteins with FITC-TZ, reactions were set up in the PBS buffer with the addition of 100 μM FITC-TZ for 1 hour. As expected, sfGFP proteins incorporated with NCAAs **1**, **2**, **4**, **5**, and **6** were efficiently labeled with FITC-TZ and showed intense fluorescence in a denaturing SDS-PAGE gel under UV irradiation Figure III-7A.

As a control, sfGFP with N^{ε}-Boc-lysine incorporated at its S2 position (sfGFP-BocK) could not be labeled with FITC-TZ under the same conditions. All protein samples were treated with SDS and heated at 100 °C for 20 minutes to totally quench the intrinsic GFP fluorescence before they were analysed by SDS-PAGE. The same treatment was also done with proteins labeled with HZCL. For labeling with HZCL, each protein was incubated with 0.5 mM HZCL in the PBS buffer for 1 hour before they were analyzed on a denaturing SDS-PAGE gel under UV irradiation, Figure III-7B.

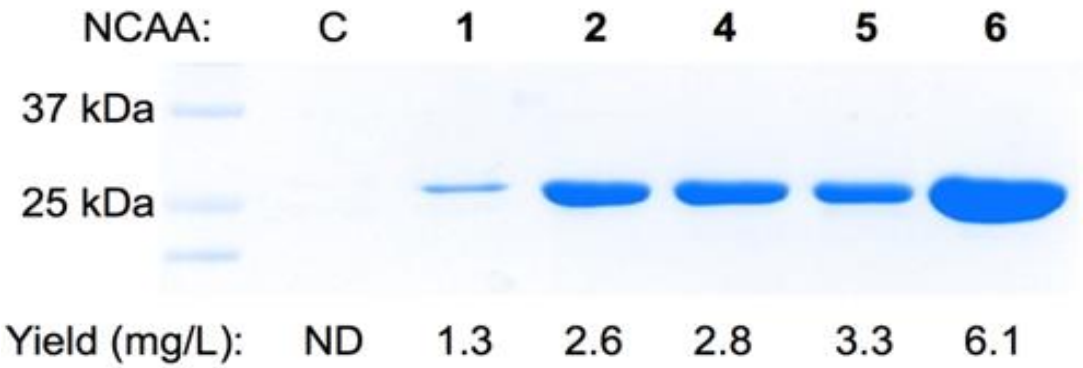


Figure III-6. Site-specific incorporation of five NCAAs that contain strain-promoted alkene/alkyne functionalities into sfGFP at its S2 position. Proteins were expressed in E. coil BL21 cells transformed with pEVOL-pylT-Y306A/N346A/C348A and pET-pylT-sfGFP2TAG. Cells were grown in LB medium supplemented with 2 mM of a NCAA for 8 hours. Supplementing LB with 2 mM of 3 or 7 did not yield sfGFP expression increase. Therefore, data for these two NCAAs are not presented.

Table III-3. Detected and theoretical molecular weights (MWs) of sfGFP proteins with different NCAAs incorporated at the S2 position.

Protein	Detected MW[a]	Theoretical MW
sfGFP-**1**[b]	27,879	27,880
sfGFP-**2**[b]	27,837	27,836
sfGFP-**4**[b]	27,895	27,876
sfGFP-**5**[b]	27,894	27,874
sfGFP-**6**[b]	27,874	27,874

[a] With an error of ±1 Da.

[b] sfGFP with **1**, **2**, **4**, **5**, or **6** incorporated at the S2 position.

.

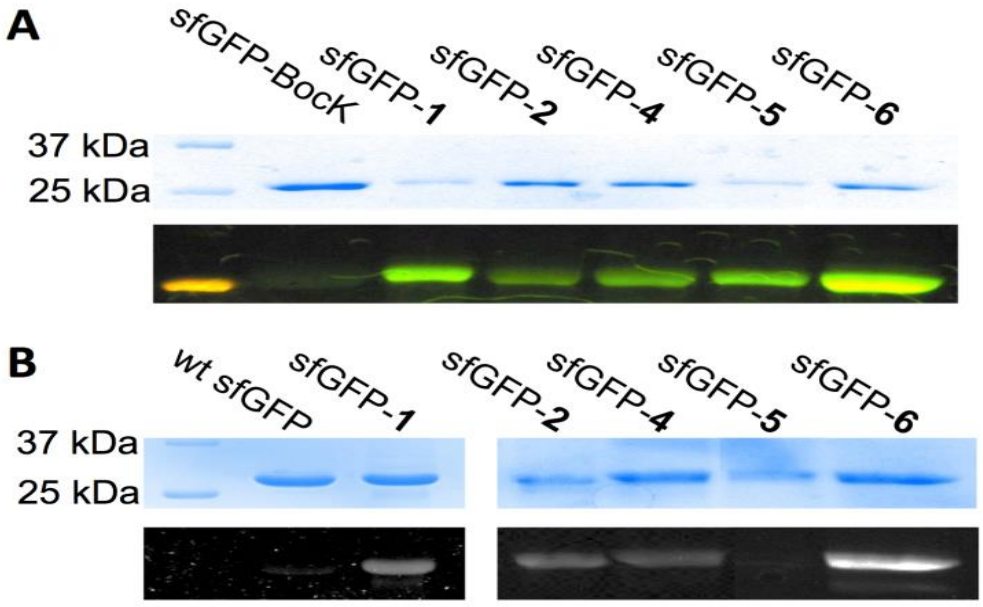


Figure III-7. Selective labeling of sfGFP proteins incorporated with strain-promoted alkene/alkyne functionalities with (A) FITC-TZ and (B) HZCL. In A and B, top panels show denaturing SDS-PAGE analysis of proteins stained with Coomassie blue and the bottom panels are from fluorescent imaging of the same gels before Coomassie blue staining. In A, the fluorescent image was captured by a digital camera and displayed was real color of the emitting light. In B, the fluorescent image was captured by a Bio-Rad ChemiDoc XRS system

sfGFP proteins incorporated with **1**, **2**, **4**, and **6** were fluorescently labeled after their reactions with HZCL. As expected, the labeling product of sfGFP-**5** could not be fluorescently visualized due to the non-fluorescent reaction product of cyclooctyne with HZCL. As a control, wild-type sfGFP could not be labeled with HZCL under the same conditions

Labeling Proteins Incorporated with Strain-Promoted Alkene/Alkyne Functionalities using EDDAC and NIADC Reactions in Living Cells

Additionally, the IEDDAC and NADC reactions were also used to label proteins that bear site-specifically incorporated NCAAs with strained alkene/alkyne functionalities in living cells. pEVOL-pylT-Y306A/N346A/C348A/Y384F and a previously constructed plasmid pETDuet-OmpXTAG that contained an *E. coli* outer membrane protein OmpX gene with an AAAAXAA (A denotes alanine and X denotes the NCAA) insertion between two extracellular residues, 53 and 54, were used to transform E. coli BL21 cells. The transformed cells were grown in LB supplemented with 2 mM of **4**, **5**, or **6** to produce OmpX with these NCAAs site-specifically incorporated.

The subsequent reactions with FITC-TZ fluorescently labeled the expressed proteins at the bacterial cell surface. Similar reactions with HZCL also fluorescently labeled the expressed proteins except for OmpX incorporated with 5, whose labeling with HZCL led to a non-fluorescent product (Figure III-8). Collectively, our assembled in vitro and cell labeling data demonstrate the high efficiency and selectivity of

IEDDAC and NADC reactions in the labeling of newly developed and genetically encoded strained alkene/alkyne NCAAs.

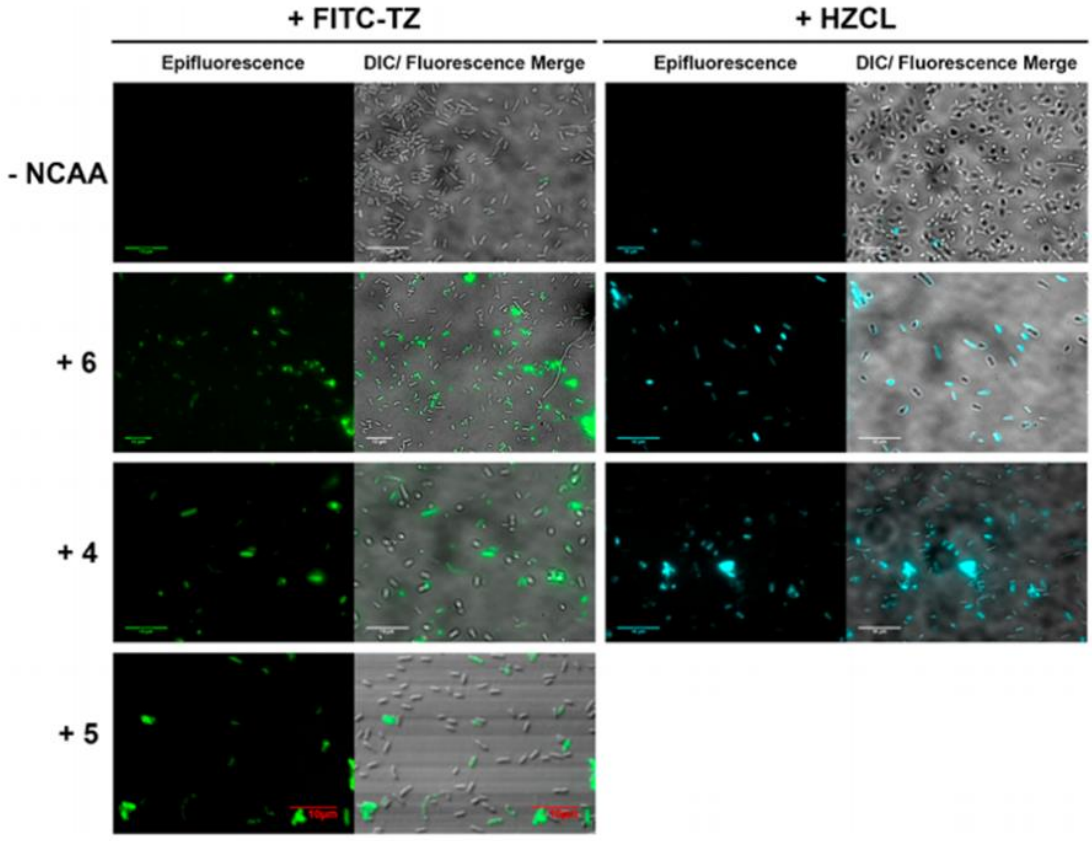


Figure III-8. Selective labeling of sfGFPs that contained site-specifically incorporated NCAAs with strained alkene/alkyne functionalities with (A) FITC-TZ and (B) HZCL in E. coli cells. Cells were labeled with FITC-TZ and HZCL for 3 hours and then washed with PBS buffer 3 times before undertaking epifluorescent and differential interface contrast (DIC) imaging.

Conclusion

In summary, we have kinetically characterized IEDDAC and NADC reactions that involve strained alkene/alkyne functionalities including norbornene, trans-cyclooctene, and cyclooctyne, and applied the two reactions to selectively label proteins bearing these functionalities. Being a rapid catalyst-free click reaction with a high bio-orthogonality, IEDDAC has received substantial attention recently.[128, 130, 138, 155] Our study showed that all three strained alkene/alkyne functionalities undergo rapid IEDDAC reactions with tetrazines and rate constants close to 300 000 M^{-1} s^{-1} could be reached.

Our study also revealed that all three functionalities undergo fast reactions with diaryl nitrilimines at rate constants close to 20 M^{-1} s^{-1}. To the best of our knowledge, the current study is the first report of fast reaction kinetics between trans-cylooctenes and diaryl nitrilimines.

To synthesize proteins with site-specifically encoded strained alkene/alkyne functionalities, a mutant PylRS-$tRNA^{Pyl}_{CUA}$ pair was developed, which has been successfully applied to incorporate five tyrosine derivatives containing norbornene, trans-cyclooctene, and cyclooctyne functionalities into proteins at amber mutation sites in *E. coli*. Proteins bearing these tyrosine derivatives were fluorescently labeled using IEDDAC and NADC reactions both in vitro and in living cells. Our current study has significantly expanded the tool kits for protein labeling. Potential applications of the developed methods include studies such as protein folding/dynamics, protein trafficking, and protein−protein/DNA interactions.

CHAPTER IV

DEMETHYLASE: SUBSTRATE SPECIFICITY ANALYSIS ON HISTONE PROTEINS

Introduction

Nucleosomes, the primary building block of chromatin plays a very important role in epigenetics by regulating cellular processes such as replication, repair, and transcription to maintain the proper function of the cell. In the nucleosome, the histone octamer comprises globular core proteins and a flexible, unstructured N-terminal tail. The globular protein consists of two sets of four histone core proteins which entails a tetramer of H3 and H4 together with dimer of H2A and H2B wrapped around twice with a 147 base pairs of DNA.[156] Numerous PTM such as acetylation, methylation, phosphorylation, ubiquitylation occur on the flexible, unstructured N-terminal tails of the histone core proteins in response to extracellular or intracellular stimuli. This affects the state of the chromatin assembly by either loosening (euchromatin) or tightening (heterochromatin)[157] (Figure IV-1).

Alteration of the chromatin assembly affects the histone-histone interactions or histone-DNA interactions which regulates the gene expression *in vivo.*[158] The acetylation functional group for instance neutralizes the positive charge on the histone and eventually loosens the binding between histone and DNA. This can initiate gene transcription as well as replicate or repair the DNA.[159, 160] In the case of histone H3K56Ac, it blocks formation of higher-order chromatin arrays, but does not destabilize

the nucleosome. Since the addition and removal of the PTM is dynamic, the acetylation is simultaneously removed to regulate the cell cycle progression and avoid DNA damage.[161] Another modification such as methylation can either transcriptionally activate genes at H3K4, H3K36 and H3K79 or silence genes at H3K9, H3K27, and H4K20.[162, 163] Ubiquitination, a larger modification, is known to play a vital role in gene activation or repression.

These modifications on the histone N-terminal tail can also lead to a process called histone crosstalk due to its dynamic process. In the dynamic process, a particular site can either add or remove the modification of another site making the state of the chromatin dependent on the type of PTM or set of PTMs to yield different cellular responses. For example, phosphorylation at H3S10 is known to promote acetylation at the H4K16 site to control the transcription elongation.[164] Lysine N^{ε}-amine is highly diversified with modifications such as acetylation (Ac) and mono-, di-, tri-methylation (M1, M2, M3) present on different sites in the nucleosome. These different modifications not only regulate cellular processes but can cause mis-regulations in gene expression which can lead to several diseases such as cancer, diabetes etc.[158, 165, 166]

Due to histone's high complexity, it is challenging to study the function of a particular modification or its effects on the same core protein or different histone core proteins. Several methods such as native chemical ligation, expressed peptide ligation, and solid state synthesis have been used to obtain the histone PTM mimics but these techniques have drawbacks such as their inability to obtain a homogenous modified histone, low yields and tedious synthetic work.[167, 168]

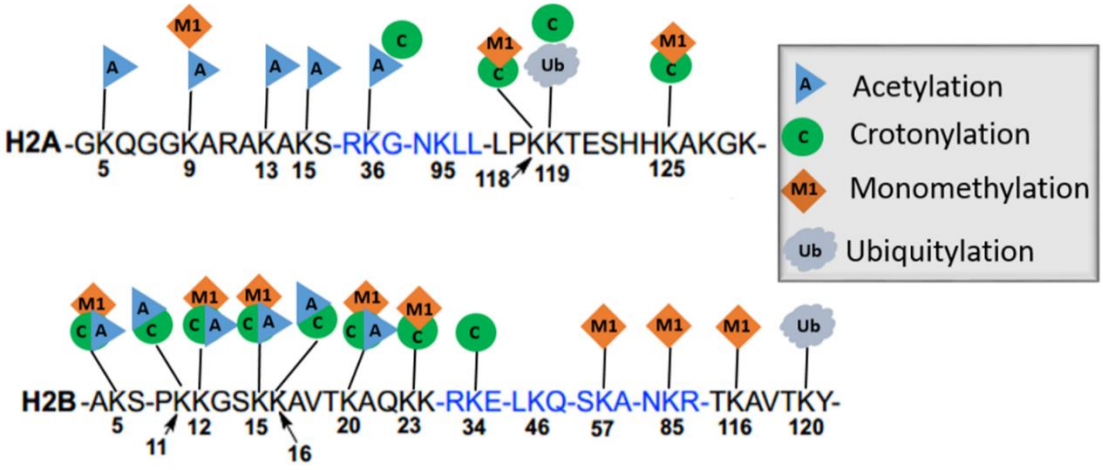


Figure IV-1. Several N^{ε}-amine lysine modification on H2A and H2B histone proteins.

Here we show a simple method by which modifications such as mono- or, dimethyl lysine can be recombinantly synthesized on the histone proteins to study the epigenetic erasers such as histone lysine demethylases (HKDMs) on the M1 and M2 modification at different lysine sites. These PTM variants are very necessary to study the crosstalk pathways present on the same histone (*cis*-tail) or different histones (*trans*-tail) of the four histone proteins for the regulation of gene expression.

Lysine methylation is one of the most studied modifications in histones with several histone methyltransferases (HMTs) and HKDMs enzymes discovered. The level of methylation added (M1, M2, M3) on the lysine residue by HMTs play different roles. For example monomethylation at H3K9, H3K27, and H4K20 are related to transcription activation but trimethylation is related to transcription repression.[169, 170] These methylations on lysine residues (Figure IV-2) are controlled by HKDMs through a reversible process.

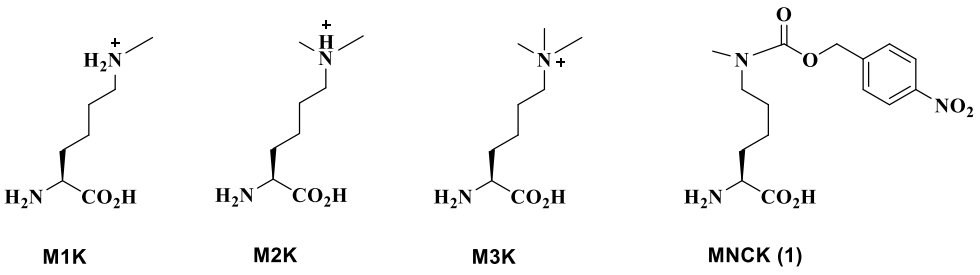


Figure IV-2. Structures of mono-, di-, tri-methyl-L-lysines and compound 1 for M1K synthesis.

Altering these modifications can cause various diseases, therefore investigating how these HKDMs catalyze the removal of methylation can give more insight into discovering several HDAC inhibitors as potential drug targets. Currently, there are two classes of HKDMs regulated by flavin adenine dinucleotide (FAD) -dependent Lysine Specific Demethylase 1 and 2 (LSD1 and LSD2) and α-ketoglutarate (α-KG), Fe (II) and molecular oxygen-dependent Jumonji C (JmjC) family for demethylation. LSD1 and LSD2 can effectively demethylase mono- and di-methylation, while the JmjC family can demethylate all 3 methylation types.[169, 171, 172] Several methods have been used to determine specific sites for methylation, but our approach will yield a full length recombinant protein with desired methylation to study the specificity of the both HKDMs families on different lysine sites. In this study, we will focus on monomethylation on different lysine sites on histone H2A, H2B and H3.

To carry out this modified synthesis on the histones H3, H2A and H2B, the lysine derivatives will be genetically incorporated into protein directly during translation using the Pyl incorporation machinery which is found in *M. mazei* to co-translate Pyl, into protein.[57 33] This naturally occurring system paves a way to incorporate other diverse NCAAs into protein to study and manipulate biological process in histones. This same Pyl incorporation machinery will be used to effectively synthesize histones with defined modifications.

Firstly, one main issue to address is the difficulty to select monomethyl-lysine or M1K from free lysine because they differ by only a methyl group which cannot be selectively taken up by PylRS wild type (Figure IV-2). Several research groups

including our lab have used either wild-type or evolved PylRS-$tRNA_{CUA}^{Pyl}$ pair to produce protein with monomethylation. N^{ε}-Boc-N^{ε}-methyllysine has been successfully incorporated into histone H3 at the K9 position using the wild-type PylRS- pair. 2% TFA was used to deprotect the boc protecting group to obtain the Histone with N^{ε}-methyllysine.[173] Another route to recover N^{ε}-methyllysine is via UV photolysis using a photocaged N^{ε}-methyllysine.[76, 174] N^{ε}-(o-nitrobenzyloxycarbonyl)-N^{ε}-methyllysine was successfully incorporated into protein using he evolved PylRS- pair. The nitrobenzyloxycarbonyl group was deprotected using UV light.

Though these two methods produced the desired monomethylated product, an alternative route is discussed. The monomethyl-lysine precursor, compound **1 (**MNCK) containing nitrobenzyl group was used for selection. According to *Chin et al.* the nitrobenzyl group can be easily removed by using *E. coli*'s nitroreductase which was shown to convert the majority of the nitrobenzylcarbamoyl lysine to lysine when incorporated into proteins.[175, 176] With that idea, we suggest that Compound **1** can also be subjected to enzymatic reduction using *E. coli* nitroreductase and further reduced with $Na_2S_2O_4$ to recover M1K in *E. coli* for further studies (Figure IV-3). In this study, we report the incorporation of M1K at various Lys sites on histones and the display of LSD1 enzymatic activity via peroxidase coupled assay.

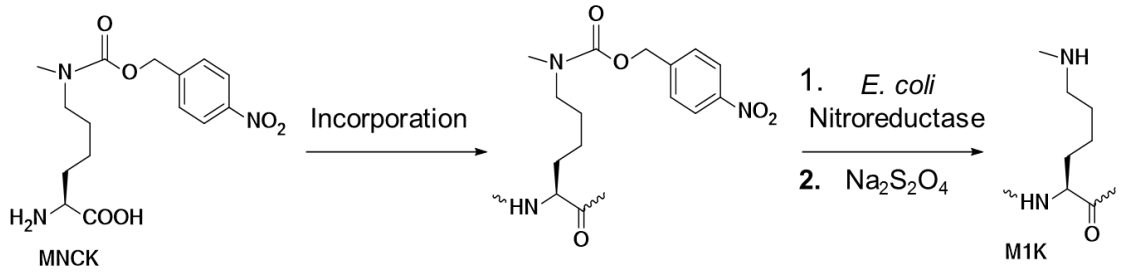


Figure IV-3. Reduction of compound 1 with *E. coli* nitroreductase to recover M1K.

Experimental Section

Materials

Reagents were purchased from general chemical manufacturers such as Sigma Aldrich, TCI, Alfa Aesar and Chem Impex and used without further purification. Methanol from Sigma-Aldrich and used as is. All reactions involving moisture sensitive reagents were carried out in oven-dried glassware under argon atmosphere. Thin layer chromatography (TLC) was performed on silica 60F-254 plates and visualized using UV irradiation at 254 nm. Flash chromatography was performed with silica gel (particle size 32-63 μm) from Dynamic adsorbents inc. (Atlanta, GA).

All ^{1}H and ^{13}C spectra were recorded on Varian Mercury-300, INVOA 300 and INVOA-500 MHz spectrometers. NMR spectra chemical shifts were reported in parts per million (ppm) and referenced to solvent peaks: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.8 ppm for ^{1}H). A minimal amount of 1,4-dioxane was added as the reference standard (67.19 ppm for ^{13}C) for carbon NMR spectra in deuterium oxide. ^{1}H NMR spectra are tabulated as follows: chemical shift multiplicity (s= singlet, bs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). Mass spectra were conducted at the Laboratory for Biological Mass spectrometry at Department of Chemistry, Texas A&M University.

The N-terminal His-tagged human histones (H2A, H2B, H3) plasmids were purchased from Epoch Life Science Inc. The plasmid pGEX-6P-1 vector encoding the LSD1 which was used for the demethylase assay was obtained from Dr. Philip A. Cole at

John Hopkins University School of Medicine. The pan anti mono- or dimethyllysine antibody was purchased from PTM biolabs.

Chemical Synthesis

Copper (II) sulfate pentahydrate (2.5 g, 10 mmol) in water was added into a aqueous solution of lysine hydrochloride (3.7 g, 20 mmol) and sodium bicarbonate (4.7 g, 56 mmol) in small portions to prevent excessive bubble formation. 4- Nitrobenzyl chloroformate (2.53 g, 13.36 mmol) in dioxane (5 mL) was then added dropwise in 5 min, followed by sodium hydroxide (0.49 g, 12.25 mmol) in one portion. The reaction mixture was stirred at room temperature for 16 hours, filtered, washed with water (100 mL), ethanol (50 mL) and diethyl ether (50 mL), and dried in the open air for 1 hour to give the crude copper complex (2.40 g, 97%) as a blue solid. All the above copper complex (2.40 g, ~7.40 mmol) was suspended in sodium hydroxide solution (0.2 *N*, 100 mL, 20 mmol), and a solution of 8-hydroxylquinoline (1.40 g, 9.64 mmol) in 1,4-dioxane (10 mL). The resulting green suspension was stirred at room temperature overnight and filtered. The filtrate was adjusted to pH 3 with hydrochloric acid (3 *N*) and extracted with ethyl acetate (40 mL x 2). The organic extracts were discarded, and the aqueous phase was concentrated to about 20 mL and loaded onto an ion-exchange column made from Dowex 50WX4-400 cation-exchange resin (~14 mL bed volume). The column was washed with excessive water (300 mL) and then eluted with pyridine (1 M, 450 mL) to give a yellow solid upon evaporation, which was suspended in ethanol, filtered, washed ethyl acetate dried to give **6** (1.46 g, 65% for two steps) as a white solid.

[□]D 22 +15.3 (*c* 1.02, 3 *N* HCl). All other characterization data were identical to that of **6** from the longer route (Scheme IV-1).

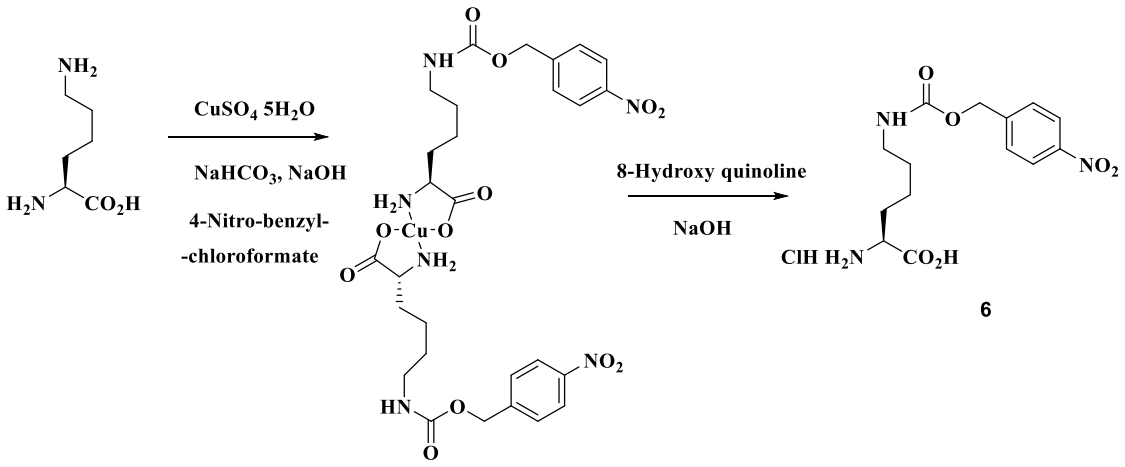
^{1}HNMR (CD_3OD, 300MHz) ; δ 8.23 (d, 2H, *J* = 8.1 Hz), 7.62 (d, 2H, *J* = 8.4 Hz), 5.22 (s, 2H), 3.85-3.78 (m, 1H), 3.19-3.18 (m, 2H), 2.15-1.82(m, 2H), 1.73-1.42 (m, 4H).

^{13}CNMR (CD_3OD, 75 MHz); δ 169.4, 156.8, 147.0, 144.6, 127.8, 123.3, 65.0, 53.4, 40.1, 29.8, 29.0, 22.1.

To a 250 mL of round-bottom-flask was added acid compound (3 g, 7.89 mmol), 4-(dimethyl- amino)pyridine (DMAP, 0.482 g, 3.94 mmol), *p*-toluenesulfonic acid monohydrate (*p*-TSA, 0.757 g, 3.94 mmol) and MeOH (100 mL). *N*,*N*′-dicyclohexylcarbodiimide (DCC, 2.14 g, 9.46 mmol) was dissolved in CH_2Cl_2 (15 mL) and added dropwise to the above solution over 1 hour. Then the resulting mixture was stirred at room temperature for 20 hours. The solvents were removed under reduced pressure and the residue was diluted with EtOAc (50 mL). Then the reaction mixture was filtered through Celite and washed with EtOAc (30 mL × 3). The filtrate was dried over anhydrous Na_2SO_4 and then concentrated under reduced pressure and purified by column chromatography to afford product **2** (2.18 g, 70% yield).

N^{α}-Boc-N^{ε}-Benzyl-N^{ε}-Methyl-L-Lysine Methyl Ester (4)

To a solution of crude **3** (~10.6 mmol) in methanol (100 mL) was added formaldehyde (37% aqueous solution, 3.00 mL, 40.3 mmol), and the reaction mixture was stirred at room temperature for 30 min. The mixture was then cooled in an



Scheme IV-1. Synthesis of intermediate for Compound 1.

ice bath, and sodium borohydride (0.77 g, 20.4 mmol) was added portionwise. The mixture was then stirred further at room temperature for 4 hours, and water (30 mL) was added dropwise to quench the reaction. Most of the methanol was evaporated under a reduced pressure, and the residue was dissolved in ethyl acetate (100 mL), washed with water (30 mL), hydrochloric acid (1 *N*, 30 mL), sodium hydroxide (1 *N*, 30 mL) and brine (30 mL), dried (Na2SO4), evaporated, and flash chromatographed (EtOAc/hexanes, 1:1 then 5% to 10% methanol in dichloromethane) to give **4** (3.32 g, 86% yield for three steps) as a yellow oil.

N^{α}-Boc-N^{ϵ}-4-Nitrobenzyloxycarbonyl -N^{ϵ}-methyl-L-lysine methyl ester (5)

A solution of **4** (2.58 g, 7.07 mmol) in methanol (50 mL) was hydrogenated under a H_2 balloon in the presence of palladium on alumina (10 wt.% Pd, 0.50 g, 0.47 mmol) at room temperature for 5 hours. The mixture was then filtered over a pad of Celite and evaporated to give the crude amine (Boc-Lys(Me)-OMe) as a grey oil.

To a solution of the above amine (~7.07 mmol) and diisopropylethylamine (2.00 mL, 11.5 mmol) in anhydrous dichloromethane (40 mL) cooled in an ice bath was added 4-Nitro-benzyl chloroformate (95%, 1.50 mL, 10.5 mmol) dropwise over 10 minutes, and the mixture was stirred at room temperature for 12 hours. The mixture was then diluted in ethyl acetate (100 mL), washed with sodium hydroxide (0.5 *N*, 40 mL) and brine (40 mL), dried (Na_2SO4), evaporated, and flash chromatographed (EtOAc/hexanes, 1:3) to give **5** (2.59 g, 90% for two steps) as a colorless oil.

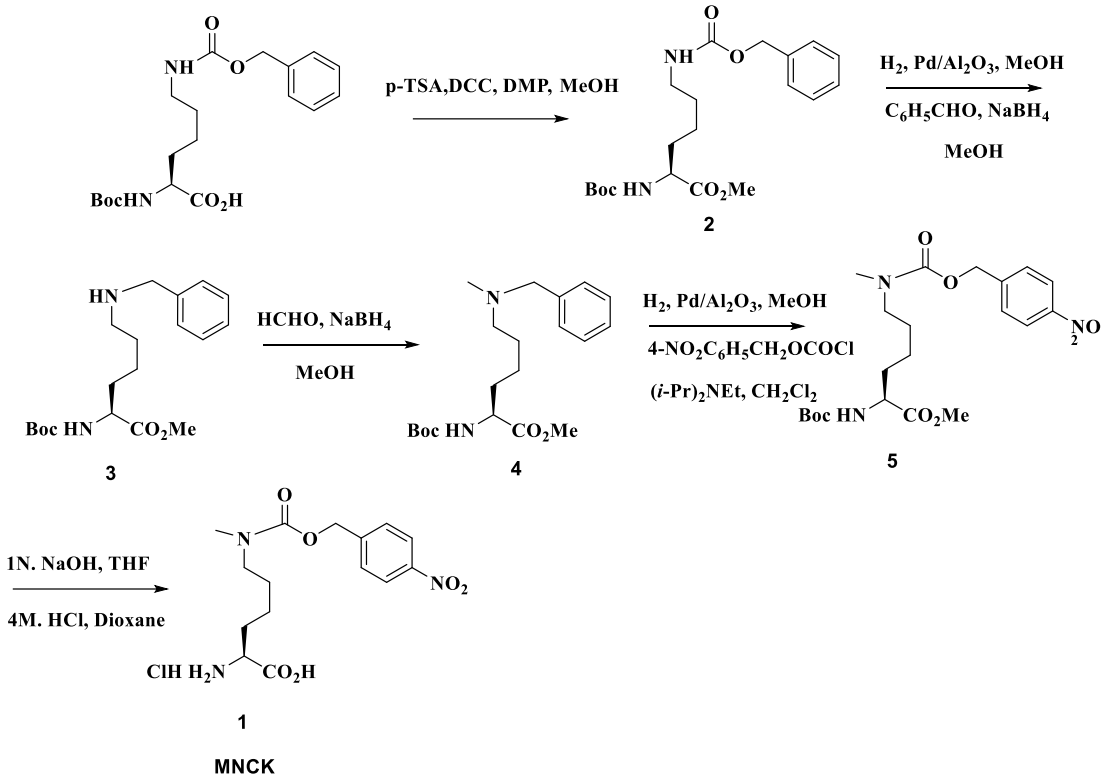
N^{ε}-(2-Nitrobenzyl) Oxycarbonyl-N^{ε}-Methyl-L-lysine (1)

To a solution of **6** (2.59 g, 6.34 mmol) in THF (20 mL) was added lithium hydroxide solution (0.5 M, 25.0 mL, 12.5 mmol), and the mixture was stirred at room temperature for 3 hours. The mixture was diluted in water (20 mL) and extracted with ether (30 mL x 2). The ether extracts were discarded, and the remaining aqueous solution was adjusted to pH 3 with hydrochloric acid (3 *N*), with the concomitant formation of white precipitate. The suspension was extracted with ethyl acetate (50 mL x 2), and the combined organic phases were washed once with brine (30 mL), dried (Na2SO4), and evaporated to give the crude carboxylic acid as a colorless oil, which was directly used without further purification. The above crude acid (~6.34 mmol) was dissolved in 1,4-dioxane (15 mL), and hydrogen chloride in 1,4-dioxane (4.0 M, 5.0 mL, 20.0 mmol) was added. The resulting white suspension was stirred at room temperature for 12 h, evaporated, redissolved in a minimal amount of water, and loaded onto an ion-exchange column made from Dowex 50WX4-400 cation-exchange resin (~14 mL bed volume). The column was washed with excessive water (300 mL) and then eluted with pyridine (1 M, 450 mL) to give **1** (1.51 g, 81% for two steps) as a white powder.

^{1}H NMR analysis showed a 1:1 mixture of rotamers at room temperature. Major rotamer: ^{1}HNMR (CD_3OD, 300MHz) ; δ 8.26 (dd, 2H, *J* = 1.8, 8.7 Hz), 7.44 (d, 2H, *J* = 8.4 Hz), 5.11 (s, 2H), 3.97 (t, 1H, *J* = 3.6 Hz), 3.15-3.36 (m, 2H), 3.02 (s, 3H), 1.99-1.92 (m, 2H), 1.65 (t, 2H, *J* = 3.6 Hz) 1.51-1.44 (m, 2H).

^{13}CNMR (CD_3OD, 75 MHz); δ 169.4, 156.8, 147.0, 144.6, 127.8, 123.3, 65.0, 53.4, 40.1, 29.8, 29.0, 22.1. Characteristic peaks of the minor rotamer: ^{1}H NMR

(CD_3OD, 300MHz) δ 2.95 (s, 3 H); 13C NMR (CD_3OD, 75 MHz δ) 128.2, 67.2, 49.2, 34.2.



Scheme IV-2. Synthesis of Compound **1**also known as MNCK.

DNA and Protein Sequences

Gene and protein sequences of pylT, *Methanosarcina mazei* PylRS and sfGFP were listed in the DNA sequence of Experimental Section in Chapter II.

DNA Sequence of Optimized Histag-TEV-H3

Atgggcagcagccatcaccatcatcaccacagccaggatccggaaaatctgtacttccaggctcgcaccaaacagactgctc
gtaagtccactggcggtaaagcgccgcgtaaacagctggcaaccaaggcagcgcgtaaaagcgctccagctactggcggc
gtgaagaagccgcaccgttatcgcccgggtactgtggctctgcgtgaaatccgccgctaccagaaaagcaccgaactgctgat
tcgcaaactgccatttcaacgtctggttcgcgaaattgctcaggatttcaaaaccgacctgcgcttccagtctagcgctgtgatgg
cactgcaagaggcgtctgaggcatatctggttggcctgttcgaagataccaacctggcagcaatccatgcaaagcgtgtaacca
ttatgccgaaagacatccaactggctcgtcgtatccgtggtgagcgtgcg

Protein Sequence of Optimized Histag-TEV-H3

MGSSHHHHHHSQDPENLYFQARTKQTARKSTGGKAPRKQLATKAARKSAPAT
GGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSS
AVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

DNA Sequence of Optimized Histag-TEV-H2A

Atgggcagcagccatcaccatcatcaccacagccaggatccggaaaatctgtacttccagtctggtcgtggtaaacaaggtgg
taaagcacgtgcaaaggctaagactcgtagcagccgtgccggtctgcagtttccagtgggtcgcgttcaccgtctgctgcgtaa
aggcaactatgctgaacgtgtgggtgctggtgcaccggtttacctggcagctgtactggaatatctgaccgcagagattctgga
gctggcaggtaacgcagctcgtgataataagaagacccgcatcatcccacgtcacctgcagctggccatccgcaacgatgag
gaactgaacaaactgctgggcaaagttactatcgctcaaggtggcgttctgccgaacatccaggcagttctgctgccgaagaa
gaccgaatcccaccacaaagcgaaaggtaagtga

Protein Sequence of Optimized Histag-TEV-H2A

MGSSHHHHHHSQDPENLYFQSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHR
LLRKGNYAERVGAGAPVYLAAVLEYLTAEILELAGNAARDNKKTRIIPRHLQLA
IRNDEELNKLLGKVTIAQGGVLPNIQAVLLPKKTESHHKAKGK*

DNA Sequence of Optimized Histag-TEV-H2B

Atgggcagcagccatcaccatcatcaccacagccaggatccggaaaatctgtacttccagtcagaaccagctaaatctgcacc
ggctccgaagaaaggctctaagaaggctgttaccaaggctcagaagaaagatggtaagaaacgcaaacgttctcgtaaagaa
agctattctgtgtacgtgtataaagttctgaaacaagtacatccagacactggcatttccagcaaagcgatgggcattatgaacag
cttcgttaacgatatcttcgaacgtatcgcaggcgaagcgagccgtctggctcactataacaaacgttctaccatcacctctcgtg
aaattcaaactgcagttcgtctgctgctgccaggtgaactggctaaacacgcggttagcgaaggcactaaagcagttaccaaat
acacttcttccaaatga

Protein Sequence of Optimized Histag-TEV-H2B

MGSSHHHHHHSQDPENLYFQSEPAKSAPAPKKGSKKAVTKAQKKDGKKRKRS
RKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRLAHYNKR
STITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYTSSK

Table IV-1 Primer sequences of Lys site mutations on H2A.

Primer Name	Sequence (5’ 3’)
H2A-K5TAG-F	CTGGTCGTGGTTAGCAAGGTGGTAAAGC
H2A-K5TAG-R	GCTTTACCACCTTGCTAACCAGCACGACCAG
H2A-K9TAG-F	GGTAAACAAGGTGGTTAGGCACGTGCAA
H2A-K9TAG-R	GCCTTTGCACGTGCCTAACCACCTTGTTT
H2A-K13TAG-F	GGTAAAGCACGTGCATAGGCTAAGACTC
H2A-K13TAG-R	GCTACGAGTCTTAGCCTATGCACGTGCTT
H2A-K15TAG-F	AGTCGTAGCAGCCGTGCCGGTC
H2A-K15TAG-R	CTAAGCCTTTGCACGTGCTTTACCACC
H2A-K36TAG-F	GGCAACTATGCTGAACGTGTGGGTG
H2A-K36TAG-R	CTAACGCAGCAGACGGTGAACGCG

Table IV-2. Primer sequences of Lys site mutations on H2B.

Primer Name	Sequence (5’ 3’)
H2B-K5TAG-F	GTCAGAACCAGCTTAGTCTGCACCG
H2B-K5TAG-R	CGGTGCAGACTAAGCTGGTTCTGAC
H2A-K12TAG-F	CGGCTCGCAAGTAGGGCTCTAAGAAG
H2A-K12TAG-R	CTTCTTAGAGCCCTACTTCGGAGCCG
H2A-K16TAG-F	GAAAGGCTCTAAGTAGGCTGTTACCAAG
H2A-K16TAG-R	CCTTGGTAACAGCCTACTTAGAGCCTTTC
H2A-K20TAG-F	GAAGGCTGTTACCTAGGCTCAGAAGAAA
H2A-K20TAG-R	CTTTCTTCTGAGCCTAGGTAACAGCCTTC

Construction of Plasmids

Each of the human histone wild-type gene, H3, H2A and H2B were cloned into petDuet-1 using BamH I and Hind III as the restriction sites. Specific lysine sites were mutated to amber mutation to obtain the desired mutants. His-tag was present on the N-terminal of each wild-type histones followed by a TEV protease digestion sequence. Ten Histag-TEV-H3 mutants were constructed as previously reported (K4, K9, K14, K18, K23, K27, K36, K56, K79, K115).[177] Different lysine positions on the Histag-TEV H2A (K5, K9, K13, K36) and H2B (K5, K12, K17, K20, K23, K34) were mutated to amber mutations using PFU or Phusion Quikchange mutagenesis. (Table IV-1 and Table IV-2). Optimized *M. Mazei* PylRS mutant (NCKRS) possessing the following mutant: L309G, C348V, Y384F was selected to take Compound **1**, MNCK

sfGFP Protein Expression and Purification

Before expressing MNCK in histones, the uptake of this NCAA was first tested in sfGFP. Plasmid pET-sfGFPN13TAG was co-transformed with pEVOL-pylT-NCKRS into *E. coli* BL21(DE3) cells. A single colony was selected and allowed to grow in 5 mL of LB medium with 100 µg/mL ampicillin at 37 °C overnight. The overnight culture was inoculated into 200 mL of Lysogeny broth (LB) medium with 100 µg/mL ampicillin and allowed to grow at 37 °C to OD_{600}~0.8. 1 mM IPTG and 1 mM MNCK were then added to the medium to induce expression of sfGFP. Control experiments in which only 1 mM IPTG was added to the medium were also carried out. The induced cells were allowed to grow at 37 °C for 8 hours and then collected by centrifugation (4,200 rpm for 20 minutes). The collected cells were resuspended in 35 mL of lysis buffer (50 mM

HEPES, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by sonication in an ice water bath. The lysed cells were clarified by centrifugation (10,000 rpm for 1 hour).

The supernatant was decanted and let bind to 5 mL of Ni-NTA superflow resins at 4°C for 1 hour. The mixture of the supernatant and resins was then loaded to an empty Qiagen Ni-NTA superflow cartridge. The resins were washed with 5 volume times of lysis buffer and sfGFP was then eluted with buffer (50 mM HEPES, 300 mM NaCl, 250 mM imidazole, pH 8.0). The purified protein was concentrated to a desired volume and analyzed by SDS-PAGE. For further analysis, western blot was carried out to detect the incorporation of MNCK. The sample was ran on an SDS-PAGE gel and transferred to the nitrocellulose membrane following the western blot protocol. After the transfer, the membrane was coated with 5% fat-free milk for 2 hours at room temperature. The membrane was then washed with PBST buffer (PBS with 0.1% tween-20, 10 mL) on the shaker three times before adding the first antibody, pan antimono-, dimethyl lysine (1:1000) overnight at 4 °C. After first antibody, the membrane was washed again with PBST buffer six times at about 5 minute intervals. The membrane was treated with second antibody (1: 10000) from Jackson ImmunoResearch (West Grove, PA) at room temperature for 1 hour. The membrane then was washed by PBST buffer, three times. Results was visualized with Pierce ECL Western Blotting Substrate. Images were taken by ChemiDoc XRS+ system from Bio-Rad.

Histone Protein Expression and Purification

The above plasmids with histone mutants encoded were individually co-transformed with pEVOL-pylT-NCKRS and transform into *E. coli* BL21(DE3) cells. Same procedure for the protein expression and mutant as compared to the sfGFP protein expression and purification. After expression, the collected cells were resuspended in 35 mL of lysis buffer containing 20 mM Tris (pH 7.5), 500 Mm NaCl, 0.1% NaN_3, 0.1% Triton X-100, then lysed. The cell lysate was centrifuged (6000 rpm for 30 minutes) and pellet was collected and washed with 20 mM Tris (pH 7.5), 500 mM NaCl, 0.1% NaN_3. Next, the pellet was dissolved in urea buffer containing 20 mM Tris (pH 7.5), 500 NaCl, and 6 M Urea. Cells were clarified by centrifugation (10,000 rpm for 1 hour). The supernatant was decanted and purification process was carried out using the Qiagen Ni-NTA superflow cartridge. The protein was eluted with urea buffer containing 300 imidazole.

LSD1 Protein Expression and Purification

pGEX-6P-1-LSD1 plasmid[178] which is GST-tagged human LSD1Δ1 = LSD1 (171-852) was transformed into *E. coli* BL21(DE3) cells. Cells were grown in 1 L 2YT medium to an OD_{600} of 1.0 at 37 °C, induced with 1 mM IPTG and expressed at 16 °C for 20 hours. Cells were harvested by centrifugation (4,200 rpm for 20 minutes). The collected cells were resuspended in 80 mL of ice cold lysis buffer (280 mM NaCl, 5.4 mM KCl, 20 mM Na_2HPO_4, 3.6 mM KH_2PO4, 1 mM EDTA, 10 mM DTT, and 10% glycerol pH 7.4) and lysed by sonication in an ice water bath. The lysed cells were clarified by centrifugation (10,000 rpm for 1 hour). The supernatant was decanted and let

bind to 5 mL of glutathione-Sepharose 4 fast flow column (GE Healthcare). The column was pre-equilibrated with the lysis buffer and the clarified lysate loaded on to column and allowed to bind with resin for 5 hours. Column was then washed with 75 mL lysis buffer and eluted with elution buffer containing 50 mM reduced glutathione added to the lysis buffer (280 mM NaCl, 5.4 mM KCl, 20 mM Na_2HPO_4, 3.6 mM KH_2PO4, 1 mM EDTA, 10 mM DTT, 50 mM GSH and 10% glycerol pH 7.4). The purified LSD1 was dialyzed using buffer containing 20 mM HEPES-NaOH (pH 7.5), 50 mM NaCl in preparation for the demethylase assay.

LSD1 Histone Demethylase Assay

A peroxidase-coupled assay which detects the amount of hydrogen peroxide released was used to determine the amount of H_2O_2 production during LSD1 demethylation process (Figure IV-4). Ampliflu™ Red (Sigma), a colorless non fluorescent produces a fluorescent resorufin upon oxidation.[179] Reaction was carried out in a 96-well plate reader with a total volume of 100 µL. The reaction sample contains 10 µM Ampliflu™ Red, 0.76 µM horseradish peroxidase, 10 µM mono-or dimethylated histone H3 at the Lys4 position (Sigma) in 20 mM HEPES-NaOH (pH 7.5), 50 mM NaCl buffer. Samples were prepared with or without 50 nM LSD1. The changes in fluorescence was monitored with excitation at 530 nm and emission at 590 nm using a Synergy H4 Hybrid Multi-Mode Microplate Reader (BioTek).

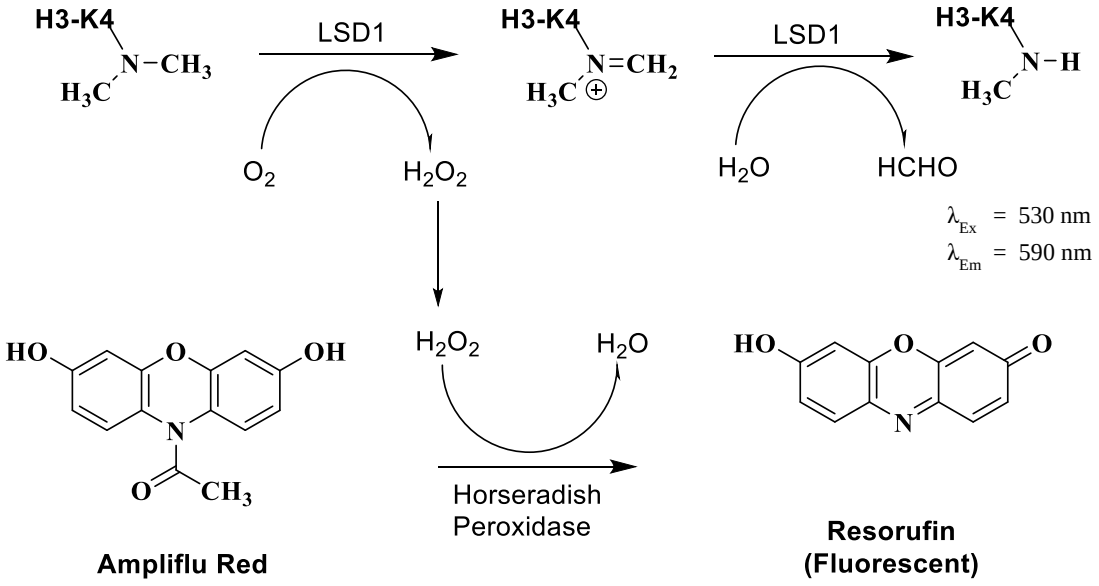


Figure IV-4. Ampliflu Red fluorescence mediated LSD1 demethylase assay used to detect generation of hydrogen peroxide to test the activity of LSD1 on different Lys sites on each histone protein.

I performed all experiments except the synthesis of alkene/alkyne containing compounds (NCAA 1-7), the FITC-TZ dye which was synthesized by Dr. Yadagiri Kurra.

Results and Discussion

Synthesis of Histones with Post-Translational Lysine Methylation (M1K) for Probing Substrate Specificities of Histone Demethylases

In order to profile lysine recognition sites of LSD1 on H2A, H2B and H3 proteins, the selected PylRS variant, NCKRS, was first used to incorporate 1 mM MNCK into sfGFP at the amber mutation site on N134 position. Without MNCK in the growth medium, no level of sfGFP expression was observed. On the contrary, the addition of MNCK promoted sfGFP overexpression. Western blot further confirmed that expressed protein has incorporated MNCK (Figure IV-5A) because there was no detection on sfGFP wild-type. This means MNCK has been successfully incorporated into sfGFP without background.

Based on our previous success in incorporating MNCK in sfGFP, we expect to see similar results in histone H2A, H2B, H3 protein irrespective of the lysine sites in the absence and presence of 1 mM MNCK. So far certain lysine sites on H3 have been successfully incorporated with 1 mM MNCK (Figure IV-6), MNCK have also been incorporated into specific Lys sites on H2A and H2B (Figure IV-7).

To obtain M1K, the p-nitrobenzyl carbamate on the MNCK could be successfully cleaved off enzymatically using *E. coli's* nitroreductase (Figure IV-3). Steady-state kinetics have shown that the *E. coli's* nitroreductase enzyme follows the Ping Pong Bi mechanism. This occurs when either NADH or NADPH donates two electrons to the flavin mononucleotide (FMN) cofactor during its binding to the enzyme and releases NAD^+ or $NADP^+$.[180, 181] Nitro-aromatic group containing substrates such as

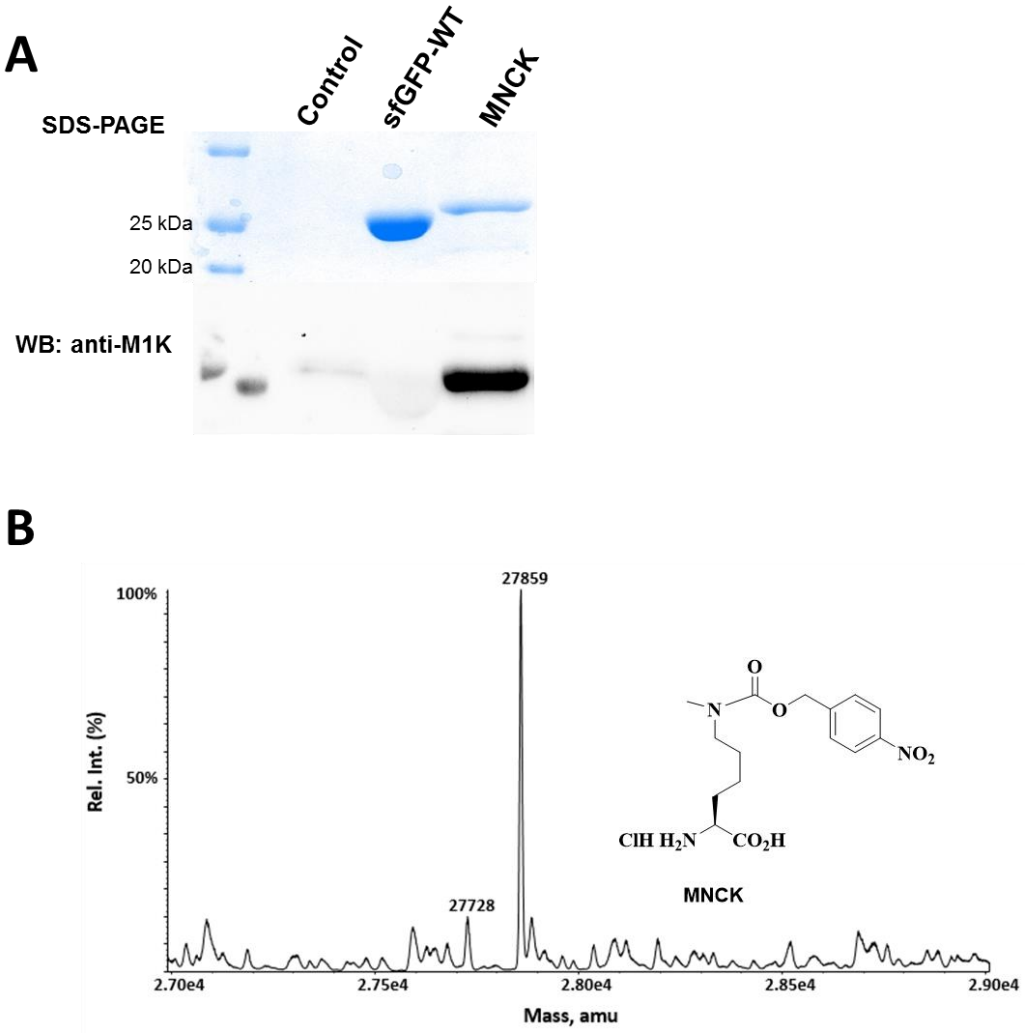


Figure IV-5. Incorporation of MNCK into sfGFP. (A) SDS-PAGE and western blot analysis shows the incorporation of MNCK by NCKRS-tRNA$_{CUA}$. Lane 1 contains no added NCAA, lane 2 is expressed wild-type of sfGFP, and lane 3 has MNCK incorporated. (B) ESI-MS of sfGFP incorporated with MNCK but has been converted to M1K.

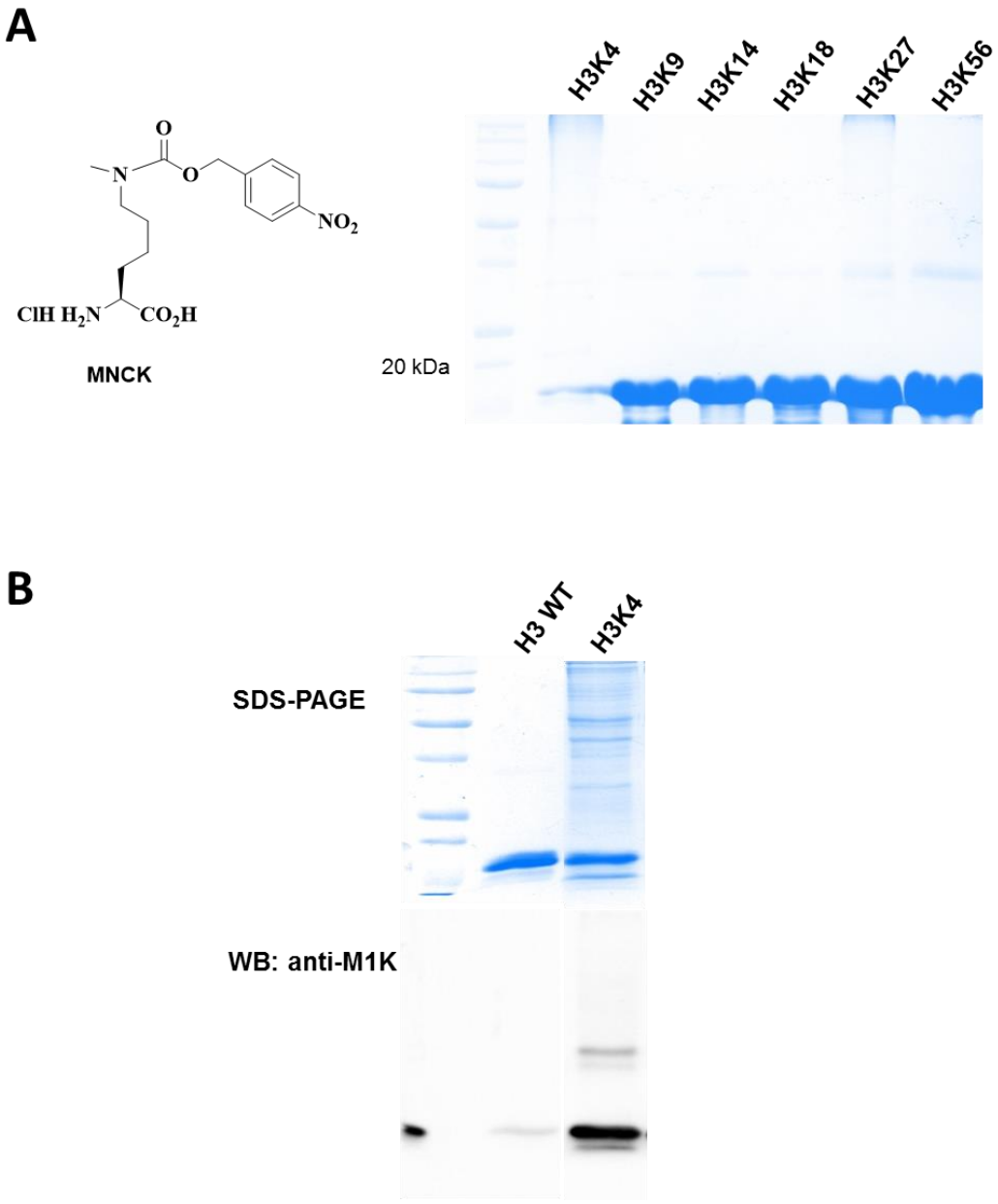


Figure IV-6. Incorporation of MNCK into Histone H3 at various Lys sites. (A) SDS-PAGE analysis shows the incorporation of MNCK by the mutant NCKRS-tRNA$_{CUA}$. Each lane represents the Lys sites which MNCK incorporated. (B) SDS-PAGE ad western blot analysis of MNCK incorporation at H3K4.

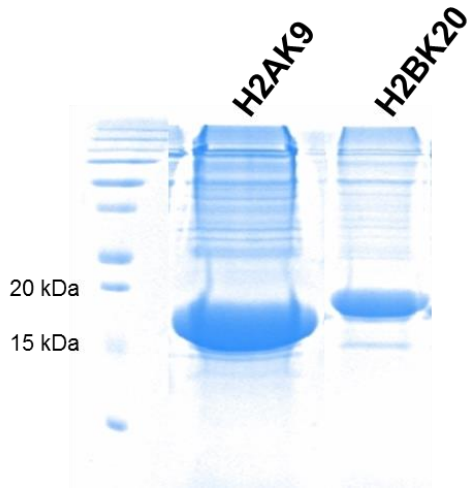


Figure IV-7. Incorporation of MNCK into Histone H2A and H2B Lys sites respectively. SDS-PAGE analysis shows the incorporation of MNCK by the mutant NCKRS-$tRNA_{CUA}$.

MNCK activates the enzyme reduction process to produce M1K, the desired unprotected lysine derivative. Recent studies in our lab showed that *E. coli's* nitroreductase can undergo deprotection of the nitrobenzyl group in sfGFP (Figure IV-5B).

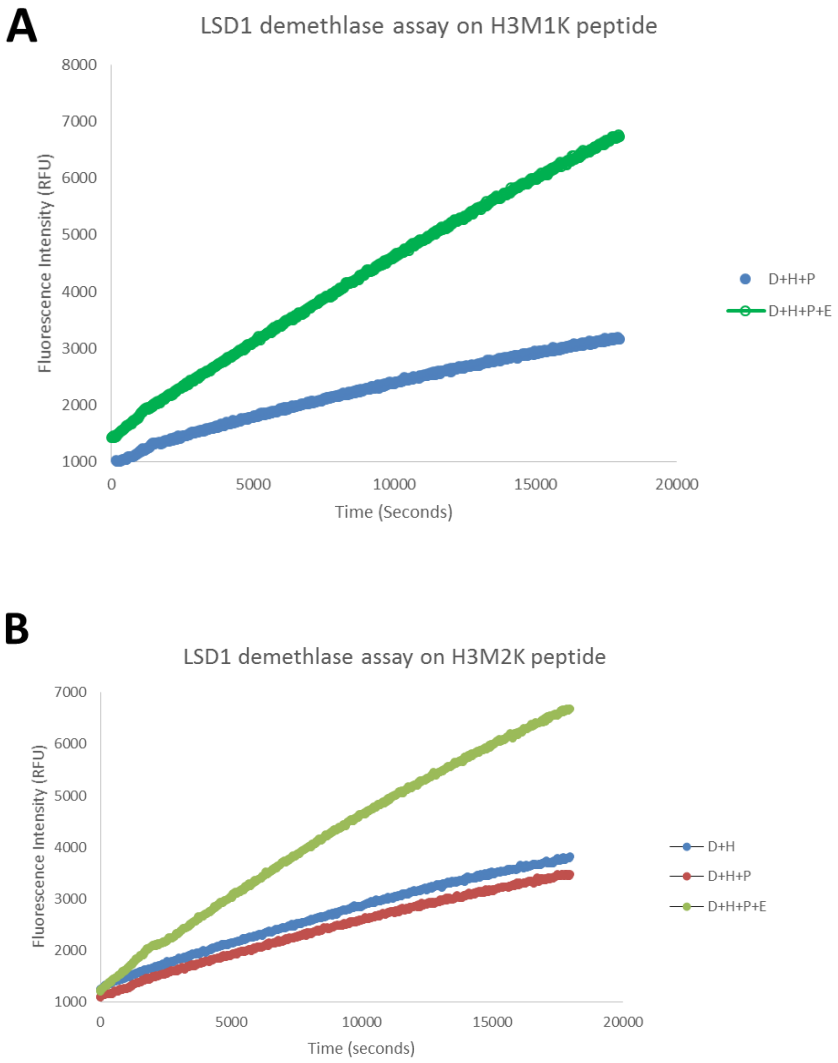
Detailed mechanism of the removal of the nitro-substituted Cbz is currently unknown but it is the likely the reduction of the aromatic nitro moiety before or after cell lysis.[176] Also the nitrobenzyl group can be further reduced using sodium hyposulfite.[182] ESI-MS showed that MNCK converted to M1K based on a clear mass shift. The purified sfGFP displayed two major mass peaks at 27728±1 Da and 27728±1Da that agree well with the theoretical molecular weights of sfGFP with MNCK incorporated at N134 (27858Da for the deprotected full-length protein; 27723 Da for the deprotected full-length protein without the first methionine (M1)) (Figure IV-5B).

HKDMs Activity on each H2A or H2B or H3 Protein Containing M1K

As mentioned previously, it is challenging to incorporate methyl-L- lysine directly onto histone because M1K only differs from lysine by the insertion of methyl group, therefore using M1K derivative is necessary since it undergoes simple methods to recover M1K. All expressed histone H2A, H2B and H3 lysine sites containing M1K can be used to test the activity LSD1 which is one of the HKDMs. LSD1 is known to generate hydrogen peroxide during the demethylation reaction. Using the peroxidase-coupled assay, the LSD1 activity and reaction kinetics can be determined based on the amount of H_2O_2 produced. Ampliflu™ Red is converted to resorufin in the presence of H_2O_2, thus the production of H_2O_2 during a demethylation reaction was detected using this method. Though there was background, there was a clear difference between the reaction with or without LSD1 (Figure IV-8).

Conclusion

In summary, we showed that a methyl-lysine precursor can be genetically incorporated into proteins in E. coli. This precursor molecule undergoes a presumptive enzymatic process to covert to methyl-lysine after its incorporation. This is confirmed with the incorporation of this precursor into sfGFP and its mass spectrometry and Western blot analysis. We have further used this developed approach to synthesize a number of methyl-histone proteins and activities of a histone demethylase, LSD1 on these methyl-histones are under characterization.



D: Ampliflu Red **H**: Horseradish Peroxidase
P: Peptide **E**: LSD1

Figure IV-8. Ampliflu Red based fluorescence studies on LSD1 demethylase assay on (A) mono-methyl-H3K4 peptide (B) di-methyl-H3K4 peptide.

CHAPTER V

CONCLUDING REMARKS AND FUTURE OUTLOOK

Based on the unique Pyl incorporation machinery, extensive research on the PylRS-$tRNA^{Pyl}$ pair has yielded over 100 NCAAs incorporated into proteins. The lack of an editing domain in PylRS makes it promiscuous and allows misacylated $tRNA^{Pyl}$ possible to incorporate NCAA such as BocK. Also, studies showed that the anticodon of the native $tRNA^{Pyl}$ ($tRNA^{Pyl}_{CUA}$) does not significantly affect its interaction with the catalytic domain of PylRS, hence in this study, the abilities of the codon recognition of both the original or evolved *M. mazei* pyrrolysyl (Pyl)-tRNA synthetase (PylRS)-$tRNA^{Pyl}$ pair to encode other BocK was further explored. In addition, the sense or quadruplet codons was also explored to expand the current NCAA-coding codons. This systematic studies elucidated on the suppression efficiencies of the PylRS-$tRNA^{Pyl}$ pair containing either amber, ochre, opal, sense or quadruplet codon-anticodon.

The ochre suppression level achieved by PylRS-$tRNA^{Pyl}_{UUA}$ pair was low compared to the PylRS-$tRNA^{Pyl}_{CUA}$ but it was sufficient to promote overexpression of a protein with an ochre mutation. In the case of the opal suppression with PylRS-$tRNA^{Pyl}_{UCA}$ pair, there was high level of sfGFP expression in the absence of BocK. ESI-MS on this purified protein showed that Trp has been incorporated into the protein. Further analysis confirmed the orthogonality of PylRS-$tRNA^{Pyl}$ pair because the $tRNA^{Pyl}_{UCA}$ was not recognized by *E. coli*'s TrpRS. This concluded that there was no mis-acylation of the

$tRNA^{Pyl}_{UCA}$ with Trp but rather a competition between the two pairs. *E. coli*'s TrpRS-$tRNA^{Trp}$ pair with CCA anticodon is near cognate to the PylRS-$tRNA^{Pyl}_{UCA}$ pair bearing UCA anticodon. Since the discriminator base G73 on the acceptor stem is a major identity element for $tRNA^{Pyl}$, G73 was mutated to either U, A or C to determine its effects on the opal suppression efficiency. $tRNA^{Pyl}_{UCA}$ G73U was found to compete with the $tRNA^{Trp}$. In pursuit of fully inhibiting the incorporation of Trp at the opal mutation site, further mutagenesis on the $tRNA^{Pyl}_{UCA}$ with the G73U mutant is necessary. For reassigned rare sense codon (AGG) to encode NCAA, there was a mixture of BocK and Arg incorporated. In the case of quadruplet (AGGA) codons, there was a misreading of the first three nucleotide of AGGA by the endogenous arginyl—$tRNA^{Arg}$ hence Arg was incorporated along with the NCAA. Further studies will be to replace all of the AGG codons with other Arg codons as well as remove the $tRNA^{Arg}_{CCU}$ to enable the specific incorporation of NCAA at the AGG codon site.

Based on the assessment of the codons previously, PylRS-$tRNA^{Pyl}_{CUA}$ pair displayed the highest efficiency. For the next studies, an evolved pair of PylRS-$tRNA^{Pyl}_{CUA}$ was used to incorporate into proteins unique chemical functionalities such as strained alkene/alkyne entities such as norbornene, trans-cyclooctene, and cyclooctyne. These proteins bearing these unique functionalities were successively labeled with a fluorescein tetrazine dye and a diaryl nitrilimine both *in vitro* and in living cells via rapid catalyst-free click reaction under biologically relevant conditions. This study has

expanded the tool kits of protein labeling for potential applications in protein folding, protein trafficking etc.

Lastly, with the PylRS-$tRNA^{Pyl}$ pair, the histone PTMs which is known to play a very important in chromatin states and epigenetic regulations of gene expressions can be mimicked to help gain understanding on the complex networks of the PTMs. PTMs such as mono-methylation is a challenge to install into proteins because it differs from Lys by just one methyl group making it difficult to directly select or incorporate it. In this particular study, using the PylRS-$tRNA^{Pyl}_{CUA}$ pair, the M1K was successfully incorporated at several Lys sites in histone H2A, H2B, H3 without the concern of Lys misincorporation. This work has allowed the intensive profiling of enzymatic substrate specificity across various mono-methyllysine sites on the histone substrates. This profiling will map out the effects of epigenetic erasers such as HKDMs.

REFERENCES

[1] Ibba, M., and Söll, D. (2000) Aminoacyl-tRNA Synthesis, *Annual Review of Biochemistry 69*, 617-650.

[2] Ibba, M., Curnow, A. W., and Söll, D. (1997) Aminoacyl-tRNA synthesis: divergent routes to a common goal, *Trends in Biochemical Sciences 22*, 39-42.

[3] Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs, *Nature 347*, 203-206.

[4] Cone, J. E., Del Río, R. M., Davis, J. N., and Stadtman, T. C. (1976) Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety, *Proceedings of the National Academy of Sciences of the United States of America 73*, 2659-2663.

[5] Burke, S. A., Lo, S. L., and Krzycki, J. A. (1998) Clustered Genes Encoding the Methyltransferases of Methanogenesis from Monomethylamine, *Journal of Bacteriology 180*, 3432-3440.

[6] Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., and Chan, M. K. (2002) A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase, *Science 296*, 1462-1466.

[7] Zhang, J., Robinson, D., and Salmon, P. (2006) A novel function for selenium in biological system: Selenite as a highly effective iron carrier for Chinese hamster

ovary cell growth and monoclonal antibody production, *Biotechnology and Bioengineering 95*, 1188-1197.

[8] Baker, R. D., Baker, S. S., Larosa, K., Whitney, C., and Newburger, P. E. (1993) Selenium Regulation of Glutathione Peroxidase in Human Hepatoma Cell Line Hep3B, *Archives of Biochemistry and Biophysics 304*, 53-57.

[9] A, G. W., J, S. G., Adelbert, G., A, K. I. M. S.-M., Fritz, Ö., Albrecht, W., and Leopold, F. (1984) The Amino-Acid Sequence of Bovine Glutathione Peroxidase, In *Hoppe-Seyler´s Zeitschrift für physiologische Chemie*, p 195.

[10] Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W., and Harrison, P. R. (1986) The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon, TGA, *The EMBO Journal 5*, 1221-1227.

[11] Zinoni, F., Birkmann, A., Stadtman, T. C., and Böck, A. (1986) Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from Escherichia coli, *Proceedings of the National Academy of Sciences of the United States of America 83*, 4650-4654.

[12] Zinoni, F., Heider, J., and Böck, A. (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine, *Proceedings of the National Academy of Sciences of the United States of America 87*, 4660-4664.

[13] Thanbichler, M., and Bock, A. (2002) Selenoprotein biosynthesis: purification and assay of components involved in selenocysteine biosynthesis and insertion in Escherichia coli, *Methods in enzymology 347*, 3-16.

[14] Hatfield, D. L., and Gladyshev, V. N. (2002) How Selenium Has Altered Our Understanding of the Genetic Code, *Molecular and Cellular Biology 22*, 3565-3576.

[15] Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. (1991) Selenocysteine: the 21st amino acid, *Molecular Microbiology 5*, 515-520.

[16] Forchhammer, K., and Böck, A. (1991) Selenocysteine synthase from Escherichia coli. Analysis of the reaction sequence, *Journal of Biological Chemistry 266*, 6324-6328.

[17] Tormay, P., Wilting, R., Lottspeich, F., Mehta, P. K., Christen, P., and Böck, A. (1998) Bacterial selenocysteine synthase, *European Journal of Biochemistry 254*, 655-661.

[18] Varlamova, E. G., Goltyaev, M. V., Novoselov, S. V., Novoselov, V. I., and Fesenko, E. E. (2013) Selenocysteine biosynthesis and mechanism of incorporation into growing proteins, *Molecular Biology 47*, 488-495.

[19] Hüttenhofer, A., Westhof, E., and Böck, A. (1996) Solution structure of mRNA hairpins promoting selenocysteine incorporation in Escherichia coli and their base-specific interaction with special elongation factor SELB, *RNA 2*, 354-366.

[20] Liu, Z., Reches, M., Groisman, I., and Engelberg-Kulka, H. (1998) The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in Escherichia coli, *Nucleic Acids Research 26*, 896-902.

[21] Kaiser, J. T., Gromadski, K., Rother, M., Engelhardt, H., Rodnina, M. V., and Wahl, M. C. (2005) Structural and Functional Investigation of a Putative Archaeal Selenocysteine Synthase, *Biochemistry 44*, 13315-13327.

[22] Carlson, B. A., Xu, X.-M., Kryukov, G. V., Rao, M., Berry, M. J., Gladyshev, V. N., and Hatfield, D. L. (2004) Identification and characterization of phosphoseryl-tRNA([Ser]Sec) kinase, *Proceedings of the National Academy of Sciences of the United States of America 101*, 12848-12853.

[23] Sheppard, K., Akochy, P.-M., and Söll, D. (2008) Assays for transfer RNA-dependent amino acid biosynthesis, *Methods (San Diego, Calif.) 44*, 139-145.

[24] Yuan, J., Palioura, S., Salazar, J. C., Su, D., O'Donoghue, P., Hohn, M. J., Cardoso, A. M., Whitman, W. B., and Söll, D. (2006) RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea, *Proceedings of the National Academy of Sciences 103*, 18923-18927.

[25] Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L., and Driscoll, D. M. (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs, *The EMBO Journal 19*, 306-314.

[26] Chavatte, L., Brown, B. A., and Driscoll, D. M. (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes, *Nat Struct Mol Biol 12*, 408-416.

[27] Martin Iii, G. W., and Berry, M. J. (2001) Selenocysteine codons decrease polysome association on endogenous selenoprotein mRNAs, *Genes to Cells 6*, 121-129.

[28] Rother, M., Resch, A., Wilting, R., and Böck, A. (2001) Selenoprotein synthesis in archaea, *BioFactors 14*, 75-83.

[29] Berry, M. J., Banu, L., Harney, J. W., and Larsen, P. R. (1993) Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons, *The EMBO Journal 12*, 3315-3322.

[30] Rother, M., Wilting, R., Commans, S., and Böck, A. (2000) Identification and characterisation of the selenocysteine-specific translation factor SelB from the archaeon Methanococcus jannaschii1, *Journal of Molecular Biology 299*, 351-358.

[31] Blight, S. K., Larue, R. C., Mahapatra, A., Longstaff, D. G., Chang, E., Zhao, G., Kang, P. T., Green-Church, K. B., Chan, M. K., and Krzycki, J. A. (2004) Direct charging of tRNACUA with pyrrolysine in vitro and in vivo, *Nature 431*, 333-335.

[32] Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA, *Science 296*, 1459-1462.

[33] Gaston, M. A., Jiang, R., and Krzycki, J. A. (2011) Functional context, biosynthesis, and genetic encoding of pyrrolysine, *Current Opinion in Microbiology 14*, 342-349.

[34] Zhang, Y., and Gladyshev, V. N. (2007) High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm Olavius algarvensis, *Nucleic Acids Research 35*, 4952-4963.

[35] Herring, S., Ambrogelly, A., Polycarpo, C. R., and Söll, D. (2007) Recognition of pyrrolysine tRNA by the Desulfitobacterium hafniense pyrrolysyl-tRNA synthetase, *Nucleic Acids Research 35*, 1270-1278.

[36] Nozawa, K., O/'Donoghue, P., Gundllapalli, S., Araiso, Y., Ishitani, R., Umehara, T., Soll, D., and Nureki, O. (2009) Pyrrolysyl-tRNA synthetase-tRNAPyl structure reveals the molecular basis of orthogonality, *Nature 457*, 1163-1167.

[37] Wan, W., Tharp, J. M., and Liu, W. R. (2014) Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1844*, 1059-1070.

[38] Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T. A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation, *Proceedings of the National Academy of Sciences 104*, 11268-11273.

[39] Polycarpo, C. R., Herring, S., Bérubé, A., Wood, J. L., Söll, D., and Ambrogelly, A. (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase, *FEBS Letters 580*, 6695-6700.

[40] Polycarpo, C., Ambrogelly, A., Bérubé, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., Wood, J. L., and Söll, D. (2004) An aminoacyl-tRNA synthetase

that specifically activates pyrrolysine, *Proceedings of the National Academy of Sciences of the United States of America 101*, 12450-12454.

[41] Li, W.-T., Mahapatra, A., Longstaff, D. G., Bechtel, J., Zhao, G., Kang, P. T., Chan, M. K., and Krzycki, J. A. (2009) Specificity of Pyrrolysyl-tRNA Synthetase for Pyrrolysine and Pyrrolysine Analogs, *Journal of Molecular Biology 385*, 1156-1164.

[42] Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J. (2005) Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications, *Angewandte Chemie International Edition 44*, 7342-7372.

[43] Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J., and Shokat, K. M. (2007) The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones, *Cell 128*, 1003-1012.

[44] Schwarzer, D., and Cole, P. A. (2005) Protein semisynthesis and expressed protein ligation: chasing a protein's tail, *Current Opinion in Chemical Biology 9*, 561-569.

[45] Dawson, P., Muir, T., Clark-Lewis, I., and Kent, S. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776-779.

[46] Speers, A. E., and Cravatt, B. F. (2004) Profiling Enzyme Activities In Vivo Using Click Chemistry Methods, *Chemistry & Biology 11*, 535-546.

[47] Uttamapinant, C., Tangpeerachaikul, A., Grecian, S., Clarke, S., Singh, U., Slade, P., Gee, K. R., and Ting, A. Y. (2012) Fast, Cell-Compatible Click Chemistry

with Copper-Chelating Azides for Biomolecular Labeling, *Angewandte Chemie International Edition 51*, 5852-5856.

[48] Kurra, Y., Odoi, K. A., Lee, Y.-J., Yang, Y., Lu, T., Wheeler, S. E., Torres-Kolbus, J., Deiters, A., and Liu, W. R. (2014) Two Rapid Catalyst-Free Click Reactions for In Vivo Protein Labeling of Genetically Encoded Strained Alkene/Alkyne Functionalities, *Bioconjugate Chemistry 25*, 1730-1738.

[49] Wu, B., Wang, Z., Huang, Y., and Liu, W. R. (2012) Catalyst-Free and Site-Specific One-Pot Dual-Labeling of a Protein Directed by Two Genetically Incorporated Noncanonical Amino Acids, *ChemBioChem 13*, 1405-1408.

[50] Wang, X. S., Lee, Y.-J., and Liu, W. R. (2014) The nitrilimine-alkene cycloaddition is an ultra rapid click reaction, *Chemical Communications 50*, 3176-3179.

[51] Jiang, R., and Krzycki, J. A. (2012) PylSn and the Homologous N-terminal Domain of Pyrrolysyl-tRNA Synthetase Bind the tRNA That Is Essential for the Genetic Encoding of Pyrrolysine, *Journal of Biological Chemistry 287*, 32738-32746.

[52] Bianco, A., Townsley, F. M., Greiss, S., Lang, K., and Chin, J. W. (2012) Expanding the genetic code of Drosophila melanogaster, *Nat Chem Biol 8*, 748-750.

[53] Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I., and Yokoyama, S. (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Research 30*, 4692-4699.

[54] Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of Escherichia coli, *Science 292*, 498-500.

[55] Furter, R. (1998) Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli, *Protein Science : A Publication of the Protein Society 7*, 419-426.

[56] Zhang, Z., Alfonta, L., Tian, F., Bursulaya, B., Uryu, S., King, D. S., and Schultz, P. G. (2004) Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells, *Proceedings of the National Academy of Sciences of the United States of America 101*, 8882-8887.

[57] Wang, L., Xie, J., and Schultz, P. G. (2006) EXPANDING THE GENETIC CODE, *Annual Review of Biophysics and Biomolecular Structure 35*, 225-249.

[58] Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008) Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases, *Biochemical and Biophysical Research Communications 371*, 818-822.

[59] Ou, W., Uno, T., Chiu, H.-P., Grünewald, J., Cellitti, S. E., Crossgrove, T., Hao, X., Fan, Q., Quinn, L. L., Patterson, P., Okach, L., Jones, D. H., Lesley, S. A., Brock, A., and Geierstanger, B. H. (2011) Site-specific protein modifications through pyrroline-carboxy-lysine residues, *Proceedings of the National Academy of Sciences 108*, 10437-10442.

[60] Hao, Z., Song, Y., Lin, S., Yang, M., Liang, Y., Wang, J., and Chen, P. R. (2011) A readily synthesized cyclic pyrrolysine analogue for site-specific protein "click" labeling, *Chemical Communications 47*, 4502-4504.

[61] Chen, P. R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B. H., and Schultz, P. G. (2009) A Facile System for Encoding Unnatural Amino Acids in Mammalian Cells, *Angewandte Chemie International Edition 48*, 4052-4055.

[62] Odoi, K. A., Huang, Y., Rezenom, Y. H., and Liu, W. R. (2013) Nonsense and Sense Suppression Abilities of Original and Derivative Methanosarcina mazei Pyrrolysyl-tRNA Synthetase-tRNA(Pyl) Pairs in the Escherichia coli BL21(DE3) Cell Strain, *PloS one 8*, e57035.

[63] O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R., and Söll, D. (2012) Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion, *FEBS Letters 586*, 3931-3937.

[64] Baron, C., and Böck, A. (1991) The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of Escherichia coli is the determinant for binding to elongation factors SELB or Tu, *Journal of Biological Chemistry 266*, 20375-20379.

[65] Tujebajeva, R. M., Copeland, P. R., Xu, X. M., Carlson, B. A., Harney, J. W., Driscoll, D. M., Hatfield, D. L., and Berry, M. J. (2000) Decoding apparatus for eukaryotic selenocysteine insertion, *EMBO Reports 1*, 158-163.

[66] Yuan, J., O'Donoghue, P., Ambrogelly, A., Gundllapalli, S., Sherrer, R. L., Palioura, S., Simonovic, M., and Soll, D. (2010) Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems, *FEBS Lett 584*, 342-349.

[67] Polycarpo, C. R., Herring, S., Berube, A., Wood, J. L., Soll, D., and Ambrogelly, A. (2006) Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase, *FEBS Lett 580*, 6695-6700.

[68] Mukai, T., Kobayashi, T., Hino, N., Yanagisawa, T., Sakamoto, K., and Yokoyama, S. (2008) Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases, *Biochem Biophys Res Commun 371*, 818-822.

[69] Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification, *Chem Biol 15*, 1187-1197.

[70] Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., and Chin, J. W. (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry, *J Am Chem Soc 131*, 8720-8721.

[71] Ai, H. W., Lee, J. W., and Schultz, P. G. (2010) A method to site-specifically introduce methyllysine into proteins in E. coli, *Chem Commun (Camb) 46*, 5506-5508.

[72] Fekner, T., Li, X., Lee, M. M., and Chan, M. K. (2009) A pyrrolysine analogue for protein click chemistry, *Angew Chem Int Ed Engl 48*, 1633-1635.

[73] Li, X., Fekner, T., Ottesen, J. J., and Chan, M. K. (2009) A pyrrolysine analogue for site-specific protein ubiquitination, *Angew Chem Int Ed Engl 48*, 9184-9187.

[74] Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins, *Nat Chem Biol 4*, 232-234.

[75] Huang, Y., Wan, W., Russell, W. K., Pai, P. J., Wang, Z., Russell, D. H., and Liu, W. (2010) Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications, *Bioorg Med Chem Lett 20*, 878-880.

[76] Wang, Y. S., Wu, B., Wang, Z., Huang, Y., Wan, W., Russell, W. K., Pai, P. J., Moe, Y. N., Russell, D. H., and Liu, W. R. (2010) A genetically encoded photocaged Nepsilon-methyl-L-lysine, *Mol Biosyst 6*, 1557-1560.

[77] Wang, Y. S., Russell, W. K., Wang, Z., Wan, W., Dodd, L. E., Pai, P. J., Russell, D. H., and Liu, W. R. (2011) The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives, *Mol Biosyst 7*, 714-717.

[78] Chen, P. R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B. H., and Schultz, P. G. (2009) A facile system for encoding unnatural amino acids in mammalian cells, *Angew Chem Int Ed Engl 48*, 4052-4055.

[79] Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., Fu, X., Chang, Z., and Chen, P. R. (2011) A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance, *Nat Chem Biol 7*, 671-677.

[80] Takimoto, J. K., Dellas, N., Noel, J. P., and Wang, L. (2011) Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids, *ACS Chem Biol 6*, 733-743.

[81] Wang, Y. S., Fang, X., Wallace, A. L., Wu, B., and Liu, W. R. (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum, *J Am Chem Soc 134*, 2950-2953.

[82] Gautier, A., Nguyen, D. P., Lusic, H., An, W., Deiters, A., and Chin, J. W. (2010) Genetically encoded photocontrol of protein localization in mammalian cells, *J Am Chem Soc 132*, 4086-4088.

[83] Liu, W. R., Wang, Y. S., and Wan, W. (2011) Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach, *Mol Biosyst 7*, 38-47.

[84] Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S., and Schultz, P. G. (2004) An expanded genetic code with a functional quadruplet codon, *Proc Natl Acad Sci U S A 101*, 7566-7571.

[85] Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., and Chin, J. W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome, *Nature 464*, 441-444.

[86] Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P. J., Russell, D. H., and Liu, W. R. (2010) A facile system for genetic incorporation of two different

noncanonical amino acids into one protein in Escherichia coli, *Angew Chem Int Ed Engl 49*, 3211-3214.

[87] Wu, B., Wang, Z., Huang, Y., and Liu, W. R. (2012) Catalyst-free and site-specific one-pot dual-labeling of a protein directed by two genetically incorporated noncanonical amino acids, *Chembiochem 13*, 1405-1408.

[88] O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R., and Soll, D. (2012) Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA(Pyl) for genetic code expansion, *FEBS Lett 586*, 3931-3937.

[89] Crick, F. H. (1966) Codon--anticodon pairing: the wobble hypothesis, *J Mol Biol 19*, 548-555.

[90] Schwyzer, R., and Caviezel, M. (1971) p-Azido-L-phenylalanine: A photo-affinity 'probe' related to tyrosine, *Helvetica Chimica Acta 54*, 1395-1400.

[91] Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P.-J., Russell, D. H., and Liu, W. R. (2010) A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli, *Angewandte Chemie International Edition 49*, 3211-3214.

[92] Zeng, Y., Wang, W., and Liu, W. R. (2014) Towards Reassigning the Rare AGG Codon in Escherichia coli, *ChemBioChem 15*, 1750-1754.

[93] Young, T. S., Ahmad, I., Yin, J. A., and Schultz, P. G. (2010) An enhanced system for unnatural amino acid mutagenesis in E. coli, *J Mol Biol 395*, 361-374.

[94] Swanson, R., Hoben, P., Sumner-Smith, M., Uemura, H., Watson, L., and Soll, D. (1988) Accuracy of in vivo aminoacylation requires proper balance of tRNA and aminoacyl-tRNA synthetase, *Science 242*, 1548-1551.

[95] Fukunaga, J.-i., Ohno, S., Nishikawa, K., and Yokogawa, T. (2007) A base pair at the bottom of the anticodon stem is reciprocally preferred for discrimination of cognate tRNAs by Escherichia coli lysyl- and glutaminyl-tRNA synthetases *Nucleic Acids Research 34*, 3181-3188

[96] Ogle, J. M., and Ramakrishnan, V. (2005) STRUCTURAL INSIGHTS INTO TRANSLATIONAL FIDELITY, *Annual Review of Biochemistry 74*, 129-177.

[97] Rodnina, M. V. (2012) Quality control of mRNA decoding on the bacterial ribosome, In *Advances in Protein Chemistry and Structural Biology* (Assen, M., Ed.), pp 95-128, Academic Press.

[98] Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity, *Nucleic Acids Res 26*, 5017-5035.

[99] Ogle, J. M., and Ramakrishnan, V. (2005) Structural insights into translational fidelity, *Annual Review of Biochemistry 74*, 129-177.

[100] Zaher, H. S., and Green, R. (2009) Fidelity at the Molecular Level: Lessons from Protein Synthesis, *Cell 136*, 746-762.

[101] Agris, P. F., Vendeix, F. A., and Graham, W. D. (2007) tRNA's wobble decoding of the genome: 40 years of modification, *J Mol Biol 366*, 1-13.

[102] Allner, O., and Nilsson, L. (2011) Nucleotide modifications and tRNA anticodon-mRNA codon interactions on the ribosome, *RNA 17*, 2177-2188.

[103] Vold, B. S., Keith, D. E., Jr., Buck, M., McCloskey, J. A., and Pang, H. (1982) Lysine tRNAs from Bacillus subtilis 168: structural analysis, *Nucleic Acids Res 10*, 3125-3132.

[104] Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G., and Schultz, P. G. (2004) The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination, *Nat Biotechnol 22*, 1297-1301.

[105] Gribskov, M., Devereux, J., and Burgess, R. R. (1984) The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression, *Nucleic Acids Res 12*, 539-549.

[106] Sharpless, K. B., and Kolb, H. C. (1999) Click chemistry. A concept for merging process and discovery chemistry., *Abstr Pap Am Chem S 217*, U95-U95.

[107] Kolb, H. C., Finn, M. G., and Sharpless, K. B. (2001) Click chemistry: Diverse chemical function from a few good reactions, *Angew Chem Int Edit 40*, 2004-+.

[108] Wang, Q., Chan, T. R., Hilgraf, R., Fokin, V. V., Sharpless, K. B., and Finn, M. G. (2003) Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition, *J Am Chem Soc 125*, 3192-3193.

[109] Avrutina, O., Empting, M., Fabritz, S., Daneschdar, M., Frauendorf, H., Diederichsen, U., and Kolmar, H. (2009) Application of copper(I) catalyzed azide-alkyne [3+2] cycloaddition to the synthesis of template-assembled multivalent peptide conjugates, *Org Biomol Chem 7*, 4177-4185.

[110] Speers, A. E., Adam, G. C., and Cravatt, B. F. (2003) Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition, *J Am Chem Soc 125*, 4686-4687.

[111] Link, A. J., and Tirrell, D. A. (2003) Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition, *J Am Chem Soc 125*, 11164-11165.

[112] Beatty, K. E., Liu, J. C., Xie, F., Dieterich, D. C., Schuman, E. M., Wang, Q., and Tirrell, D. A. (2006) Fluorescence visualization of newly synthesized proteins in mammalian cells, *Angew Chem Int Ed Engl 45*, 7364-7367.

[113] Dedola, S., Nepogodiev, S. A., and Field, R. A. (2007) Recent applications of the Cu(I)-catalysed Huisgen azide-alkyne 1,3-dipolar cycloaddition reaction in carbohydrate chemistry, *Org Biomol Chem 5*, 1006-1017.

[114] Meldal, M., and Tornoe, C. W. (2008) Cu-catalyzed azide-alkyne cycloaddition, *Chem Rev 108*, 2952-3015.

[115] Chang, P. V., Prescher, J. A., Sletten, E. M., Baskin, J. M., Miller, I. A., Agard, N. J., Lo, A., and Bertozzi, C. R. (2010) Copper-free click chemistry in living animals, *Proc Natl Acad Sci U S A 107*, 1821-1826.

[116] Agard, N. J., Prescher, J. A., and Bertozzi, C. R. (2004) A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems, *J Am Chem Soc 126*, 15046-15047.

[117] Codelli, J. A., Baskin, J. M., Agard, N. J., and Bertozzi, C. R. (2008) Second-generation difluorinated cyclooctynes for copper-free click chemistry, *J Am Chem Soc 130*, 11486-11493.

[118] Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., Lo, A., Codelli, J. A., and Bertozzi, C. R. (2007) Copper-free click chemistry for dynamic in vivo imaging, *Proc Natl Acad Sci U S A 104*, 16793-16797.

[119] Dommerholt, J., Schmidt, S., Temming, R., Hendriks, L. J., Rutjes, F. P., van Hest, J. C., Lefeber, D. J., Friedl, P., and van Delft, F. L. (2010) Readily accessible bicyclononynes for bioorthogonal labeling and three-dimensional imaging of living cells, *Angew Chem Int Ed Engl 49*, 9422-9425.

[120] Debets, M. F., van Berkel, S. S., Schoffelen, S., Rutjes, F. P., van Hest, J. C., and van Delft, F. L. (2010) Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition, *Chem Commun (Camb) 46*, 97-99.

[121] Jewett, J. C., Sletten, E. M., and Bertozzi, C. R. (2010) Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones, *J Am Chem Soc 132*, 3688-3690.

[122] Taylor, M. T., Blackman, M. L., Dmitrenko, O., and Fox, J. M. (2011) Design and synthesis of highly reactive dienophiles for the tetrazine-trans-cyclooctene ligation, *J Am Chem Soc 133*, 9646-9649.

[123] Blackman, M. L., Royzen, M., and Fox, J. M. (2008) Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity, *J Am Chem Soc 130*, 13518-13519.

[124] Seitchik, J. L., Peeler, J. C., Taylor, M. T., Blackman, M. L., Rhoads, T. W., Cooley, R. B., Refakis, C., Fox, J. M., and Mehl, R. A. (2012) Genetically

encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes, *J Am Chem Soc 134*, 2898-2901.

[125] Song, W., Wang, Y., Qu, J., Madden, M. M., and Lin, Q. (2008) A photoinducible 1,3-dipolar cycloaddition reaction for rapid, selective modification of tetrazole-containing proteins, *Angew Chem Int Ed Engl 47*, 2832-2835.

[126] Wang, Y., Song, W., Hu, W. J., and Lin, Q. (2009) Fast alkene functionalization in vivo by Photoclick chemistry: HOMO lifting of nitrile imine dipoles, *Angew Chem Int Ed Engl 48*, 5330-5333.

[127] Liu, D. S., Tangpeerachaikul, A., Selvaraj, R., Taylor, M. T., Fox, J. M., and Ting, A. Y. (2012) Diels-Alder cycloaddition for fluorophore targeting to specific proteins inside living cells, *J Am Chem Soc 134*, 792-795.

[128] Devaraj, N. K., and Weissleder, R. (2011) Biomedical applications of tetrazine cycloadditions, *Acc Chem Res 44*, 816-827.

[129] Liu, S., Hassink, M., Selvaraj, R., Yap, L. P., Park, R., Wang, H., Chen, X., Fox, J. M., Li, Z., and Conti, P. S. (2013) Efficient 18F labeling of cysteine-containing peptides and proteins using tetrazine-trans-cyclooctene ligation, *Molecular imaging 12*, 121-128.

[130] Blackman, M. L., Royzen, M., and Fox, J. M. (2008) Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels−Alder Reactivity, *Journal of the American Chemical Society 130*, 13518-13519.

[131] Plass, T., Milles, S., Koehler, C., Schultz, C., and Lemke, E. A. (2011) Genetically Encoded Copper-Free Click Chemistry, *Angewandte Chemie (International Ed. in English) 50*, 3878-3881.

[132] Wang, Y.-S., Fang, X., Wallace, A. L., Wu, B., and Liu, W. R. (2012) A Rationally Designed Pyrrolysyl-tRNA Synthetase Mutant Has a Broad Substrate Spectrum, *Journal of the American Chemical Society 134*, 2950-2953.

[133] Karver, M. R., Weissleder, R., and Hilderbrand, S. A. (2011) Synthesis and evaluation of a series of 1,2,4,5-tetrazines for bioorthogonal conjugation, *Bioconjug Chem 22*, 2263-2270.

[134] Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D. J., Blackman, M. L., Fox, J. M., and Chin, J. W. (2012) Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions, *J Am Chem Soc 134*, 10317-10320.

[135] Devaraj, N. K., Weissleder, R., and Hilderbrand, S. A. (2008) Tetrazine-based cycloadditions: application to pretargeted live cell imaging, *Bioconjug Chem 19*, 2297-2299.

[136] Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A., and Chin, J. W. (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction, *Nat Chem 4*, 298-304.

[137] Plass, T., Milles, S., Koehler, C., Szymanski, J., Mueller, R., Wiessler, M., Schultz, C., and Lemke, E. A. (2012) Amino acids for Diels-Alder reactions in living cells, *Angew Chem Int Ed Engl 51*, 4166-4170.

[138] Rossin, R., van den Bosch, S. M., Ten Hoeve, W., Carvelli, M., Versteegen, R. M., Lub, J., and Robillard, M. S. (2013) Highly Reactive trans-Cyclooctene Tags with Improved Stability for Diels-Alder Chemistry in Living Systems, *Bioconjug Chem*.

[139] Schoch, J., Staudt, M., Samanta, A., Wiessler, M., and Jaschke, A. (2012) Site-specific one-pot dual labeling of DNA by orthogonal cycloaddition chemistry, *Bioconjug Chem 23*, 1382-1386.

[140] Zhao, Y., and Truhlar, D. G. (2008) The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06 Functionals and Twelve Other Functionals, *Theo. Chem. Acc. 120*, 215-241.

[141] Barone, V., and Cossi, M. (1998) Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model, *J. Phys. Chem. A 102*, 1995-2001.

[142] Cossi, M., Rega, N., Scalmani, G., and Barone, V. (2003) Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model, *J. Comp. Chem. 24*, 669-681.

[143] Liang, Y., Mackey, J. L., Lopez, S. A., Liu, F., and Houk, K. N. (2012) Control and Design of Mutual Orthogonality in Bioorthogonal Cycloadditions, *J. Am. Chem. Soc. 134*, 17904-17907.

[144] Huisgen, R. (1963) 1,3-Dipolar Cycloaddition. Past and Future, *Angew Chem Int Ed Engl 2*, 565-598.

[145] Huisgen, R. (1963) Kinetics and Mechanism of 1,3-Diploar Cycloadditions, *Angew Chem Int Ed Engl 2*, 633-696.

[146] Yu, Z., Ho, L. Y., and Lin, Q. (2011) Rapid, Photoactivatable Turn-On Fluorescent Probes Based On an Intramolecular Photoclick Reaction, *Journal of the American Chemical Society 133*, 11912-11915.

[147] Wang, J., Zhang, W., Song, W., Wang, Y., Yu, Z., Li, J., Wu, M., Wang, L., Zang, J., and Lin, Q. (2010) A Biosynthetic Route to Photoclick Chemistry on Proteins, *Journal of the American Chemical Society 132*, 14812-14818.

[148] Yu, Z., Pan, Y., Wang, Z., Wang, J., and Lin, Q. (2012) Genetically Encoded Cyclopropene Directs Rapid, Photoclick-Chemistry-Mediated Protein Labeling in Mammalian Cells, *Angewandte Chemie International Edition 51*, 10600-10604.

[149] Song, W., Wang, Y., Qu, J., and Lin, Q. (2008) Selective Functionalization of a Genetically Encoded Alkene-Containing Protein via "Photoclick Chemistry" in Bacterial Cells, *Journal of the American Chemical Society 130*, 9654-9655.

[150] Kaya, E., Vrabel, M., Deiml, C., Prill, S., Fluxa, V. S., and Carell, T. (2012) A Genetically Encoded Norbornene Amino Acid for the Mild and Selective

Modification of Proteins in a Copper-Free Click Reaction, *Angewandte Chemie International Edition 51*, 4466-4469.

[151] Lee, Y. J., Wu, B., Raymond, J. E., Zeng, Y., Fang, X., Wooley, K. L., and Liu, W. R. (2013) A Genetically Encoded Acrylamide Functionality, *ACS Chem Biol*.

[152] Borrmann, A., Milles, S., Plass, T., Dommerholt, J., Verkade, J. M., Wiessler, M., Schultz, C., van Hest, J. C., van Delft, F. L., and Lemke, E. A. (2012) Genetic encoding of a bicyclo[6.1.0]nonyne-charged amino acid enables fast cellular protein imaging by metal-free ligation, *Chembiochem 13*, 2094-2099.

[153] Plass, T., Milles, S., Koehler, C., Schultz, C., and Lemke, E. A. (2011) Genetically encoded copper-free click chemistry, *Angew Chem Int Ed Engl 50*, 3878-3881.

[154] Wang, Y. S., Fang, X., Chen, H. Y., Wu, B., Wang, Z. U., Hilty, C., and Liu, W. R. (2013) Genetic Incorporation of Twelve meta-Substituted Phenylalanine Derivatives Using a Single Pyrrolysyl-tRNA Synthetase Mutant, *ACS Chem Biol 8*, 405-415.

[155] Taylor, M. T., Blackman, M. L., Dmitrenko, O., and Fox, J. M. (2011) Design and Synthesis of Highly Reactive Dienophiles for the Tetrazine–trans-Cyclooctene Ligation, *Journal of the American Chemical Society 133*, 9646-9649.

[156] Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8[thinsp]A resolution, *Nature 389*, 251-260.

[157] Bhaumik, S. R., Smith, E., and Shilatifard, A. (2007) Covalent modifications of histones during development and disease pathogenesis, *Nat Struct Mol Biol 14*, 1008 - 1016.

[158] Chi, P., Allis, C. D., and Wang, G. G. (2010) Covalent histone modifications-miswritten, misinterpreted and mis-erased in human cancers, *Nature Reviews Cancer 10*, 457-469.

[159] Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., Yao, J., Zhou, L., Zeng, Y., Li, H., Li, Y., Shi, J., An, W., Hancock, S. M., He, F., Qin, L., Chin, J., Yang, P., Chen, X., Lei, Q., Xiong, Y., and Guan, K.-L. (2010) Regulation of Cellular Metabolism by Protein Lysine Acetylation, *Science 327*, 1000-1004.

[160] Weiwei, D., Kristan, K. S., Rocco, P., Jean, A. D., Johnson, F. B., Ali, S., Matt, K., Brian, K. K., and Shelley, L. B. (2009) Histone H4 lysine 16 acetylation regulates cellular lifespan, *Nature 459*, 802-807.

[161] Tsubota, T., Berndsen, C. E., Erkmann, J. A., Smith, C. L., Yang, L., Freitas, M. A., Denu, J. M., and Kaufman, P. D. Histone H3-K56 Acetylation Is Catalyzed by Histone Chaperone-Dependent Complexes, *Molecular Cell 25*, 703-712.

[162] Cheung, P., and Lau, P. (2005) Epigenetic Regulation by Histone Methylation and Histone Variants, *Molecular Endocrinology 19*, 563-573.

[163] Martin, C., and Zhang, Y. (2005) The diverse functions of histone lysine methylation, *Nat Rev Mol Cell Biol 6*, 838-849.

[164] Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009) Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation, *Cell 138*, 1122-1136.

[165] Ellis, L., Atadja, P. W., and Johnstone, R. W. (2009) Epigenetics in cancer: Targeting chromatin modifications, *Molecular Cancer Therapeutics 8*, 1409-1420.

[166] Portela, A., and Esteller, M. (2010) Epigenetic modifications and human disease, *Nature Biotechnology 28*, 1057-1068.

[167] Dawson, P. E., and Kent, S. B. H. (2000) SYNTHESIS OF NATIVE PROTEINS BY CHEMICAL LIGATION 1, *Annual Review of Biochemistry 69*, 923-960.

[168] Xu, R., Ayers, B., Cowburn, D., and Muir, T. W. (1999) Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies, *Proceedings of the National Academy of Sciences 96*, 388-393.

[169] Klose, R. J., and Zhang, Y. (2007) Regulation of histone methylation by demethylimination and demethylation, *Nat Rev Mol Cell Biol 8*, 307-318.

[170] Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-Resolution Profiling of Histone Methylations in the Human Genome, *Cell 129*, 823-837.

[171] Hojfeldt, J. W., Agger, K., and Helin, K. (2013) Histone lysine demethylases as targets for anticancer therapy, *Nat Rev Drug Discov 12*, 917-930.

[172] Tsukada, Y.-i., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins, *Nature 439*, 811-816.

[173] Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H., and Chin, J. W. (2009) Genetically Encoding Nϵ-Methyl-l-lysine in Recombinant Histones, *Journal of the American Chemical Society 131*, 14194-14195.

[174] Groff , D., Chen , P. R., Peters , F. B., and Schultz, P. G. (2010) A Genetically Encoded ε-N-Methyl Lysine in Mammalian Cells, *ChemBioChem 11*, 1066-1068.

[175] Lee, M., Simpson Jr, J. E., Woo, S., Kaenzig, C., Anlezark, G. M., Eno-Amooquaye, E., and Burke, P. J. (1997) Synthesis of an aminopropyl analog of the experimental anticancer drug tallimustine, and activation of its 4-nitrobenzylcarbamoyl prodrug by nitroreductase and NADH, *Bioorganic & Medicinal Chemistry Letters 7*, 1065-1070.

[176] Virdee, S., Kapadnis, P. B., Elliott, T., Lang, K., Madrzak, J., Nguyen, D. P., Riechmann, L., and Chin, J. W. (2011) Traceless and Site-Specific Ubiquitination of Recombinant Proteins, *Journal of the American Chemical Society 133*, 10708-10711.

[177] Hsu, W. W., Wu, B., and Liu, W. R. (2016) Sirtuins 1 and 2 Are Universal Histone Deacetylases, *ACS Chemical Biology 11*, 792-799.

[178] Szewczuk, L. M., Culhane, J. C., Yang, M., Majumdar, A., Yu, H., and Cole, P. A. (2007) Mechanistic Analysis of a Suicide Inactivator of Histone Demethylase LSD1, *Biochemistry 46*, 6892-6902.

[179] Kim, S.-A., Chatterjee, N., Jennings, M. J., Bartholomew, B., and Tan, S. (2015) Extranucleosomal DNA enhances the activity of the LSD1/CoREST histone demethylase complex, *Nucleic Acids Research 43*, 4868-4880.

[180] Grove, J. I., Lovering, A. L., Guise, C., Race, P. R., Wrighton, C. J., White, S. A., Hyde, E. I., and Searle, P. F. (2003) Generation of Escherichia Coli Nitroreductase Mutants Conferring Improved Cell Sensitization to the Prodrug CB1954, *Cancer Research 63*, 5532-5537.

[181] Race, P. R., Lovering, A. L., White, S. A., Grove, J. I., Searle, P. F., Wrighton, C. W., and Hyde, E. (2007) Kinetic and Structural Characterisation of Escherichia coli Nitroreductase Mutants Showing Improved Efficacy for the Prodrug Substrate CB1954, *Journal of Molecular Biology 368*, 481-492.

[182] Sokolovsky, M., Riordan, J. F., and Vallee, B. L. (1967) Conversion of 3-nitrotyrosine to 3-aminotyrosine in peptides and proteins, *Biochemical and Biophysical Research Communications 27*, 20-25.