SYNTHESIS OF PROTEINS WITH HOMOGENOUS CHEMICAL AND POSTTRANSLATIONAL MODIFICATIONS

A Dissertation

by

BO WU

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,	Tadhg Begley
	Rene Garcia
	Christian Hilty
Head of Department,	David Russell

August 2014

Major Subject: Chemistry

Copyright 2014 Bo Wu

ABSTRACT

Genetic encoding non-canonical amino acids (NCAAs) is a facile approach to synthesize proteins with homogenous modifications. In my graduate study, I demonstrated the application of this approach in the synthesis of a variety of proteins with site-specific chemical and post-translational modifications.

My first effort was to synthesize and demonstrate the genetic installation of N^{ε}-methyl-L-lysine (Kme1) into proteins indirectly via the incorporation of N^{ε}-(2-Nitrobenzyl)oxycarbonyl-N^{ε}-methyl-L-lysine and its following photolytic deprotection to recover Kme1. This novel method has been applied in the Liu group at Texas A&M Department of Chemistry for understanding epigenetic roles of nucleosomal posttranslational lysine modifications.

In order to develop an approach for easy access to proteins with defined two-color labeling, the next endeavor in my graduate study was to design and demonstrate a novel method for the genetic encoding of two different NCAAs in one protein that can be further modified to introduce two dyes with different colors. I showed that NCAAs with two different functionalities could be successfully incorporated into one protein and using a model protein glutamine binding protein (QBP), I demonstrated two dyes that form a Fº rster resonance energy transfer (FRET) pair could be successfully covalently linked to two NCAAs in a one-pot and catalyst-free fashion. The synthesized two-dye labeled QBP was successfully applied to undergo the folding/unfolding analysis of QBP in the presence of guanidinium.

I was also involved in the synthesis of proteins with hydroxylamine and hydrazine groups for their efficient labeling with ketos. I synthesized N^{ε}-allyl protected hydroxylamine- and hydrazine-containing NCAAs and devised a method for their incorporation into superfolder green fluorescent protein (sfGFP).

Similarly to the genetic encoding of Kme1, the direct incorporation of N^{ε},N^{ε}-dimethyl-L-lysine (Kme2) also suffers from its structural similarity with lysine. In order to genetically install Kme2 in proteins, we aimed to genetically incorporate allysine into proteins followed by reductive amination with dimethylamine to make Kme2. Several allysine precursors were synthesized. I demonstrated that they could be site-selectively incorporated into proteins and deprotected to recover allysine. The Liu group is now in the process to carry out the reductive amination reaction to synthesize proteins with Kme2.

An important biological study I was involved in was to synthesize histone H3 with different acylation types at its multiple lysines and then I used these proteins to probe substrate and modification type specificities of one group of histone deacetylases, sirtuins. H3 variants with four modification types, acetylation, propionylation, butyrylation, and crotonylation at about 10 lysine sites were recombinantly expressed. They were used to probe sirtuin enzymes. Based on our results, we conclude that Sirt1 and Sirt2 act as universal histone deacylases regardless of lysine sites or modification types. In contrast, Sirt6 and Sirt7 barely show any reactivity toward any modification types or lysine sites we tested.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Liu, and my committee members, Dr. Begley, Dr. Garcia, Dr. Hilty, for their guidance and support throughout my graduate research. I want to show my great appreciation specifically to my advisor Dr. Liu for showing me the genetic incorporation of non-canonical amino acids, which is not only very interesting in the manner of pure science, but also shows great potential for the application of drug discovery.

Thanks also go to my lab mates for their help with both my research and life. Dr. Zhiyong Wang was a great mentor and taught me organic synthesis. Other postdocs and students also helped me to overcome some obstacles at certain point in the course of my research. Faculty and staff in the chemistry department also helped me a great amount to ensure my research to proceed and make my time at Texas A&M University a great experience. In my spare time, I enjoyed my spare time with hanging out with my buddies from the lab. Thanks for making my life joyful.

Finally, thanks to my family for their encouragement and support. I want to show my special gratitude to my aunt, for helping me to settle down in the United States and giving me endless support in my life.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Protein Posttranslational Modifications

It is well known that twenty canonical amino acids are the building blocks for proteins. However, in most eukaryotic cells, proteins undergo many types of posttranslational modifications (PTMs) for their proper functions. A large variety of PTMs have been discovered in all compartments inside cells, from nucleus to mitochondria.[1-3] Among all PTMs, the ones on histones have drawn more and more attention of scientists since there is evidence showing that these PTMs are closely related to epigenetic regulation of cell functions (**Figure 1**). Lysine/arginine methylation, lysine acylations, serine/threonine/tyrosine phosphorylation, lysine ubiquitination, and lysine sumoylation have been identified on histone proteins and they dynamically regulate cellular activities.[4] Among all the modifications, lysine methylation and lysine acetylation are two most universal modifications and it has been found that they regulate chromatin conformations.[5-7] Histone lysine methylation retains the positive charge on the lysine ε-amine which results in a tighter binding between DNA and histone proteins since DNA bears negative charges on the phosphates. On the other hand, histone acetylation removes the positive charge on the ε-amine so that it loosens the binding between DNA and histone proteins, resulting in the formation of euchromatin.[8] In euchromatin, DNA binds to histone with a more relaxed structure so that transcriptional factors are recruited, leading to a higher gene transcription level.[9]

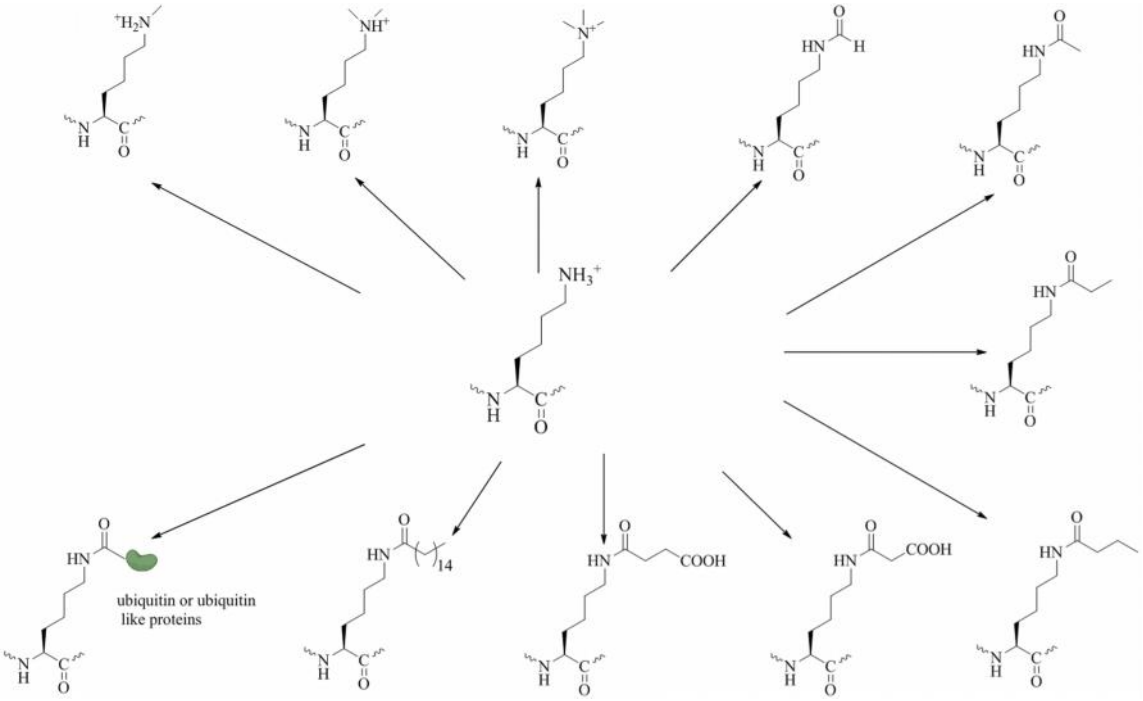


Figure 1. Lysine posttranslational modifications on proteins.

To fully understand how PTMs trigger the following cellular responses, it is pivotal to obtain proteins with homogenous PTMs. However, it is a considerable challenge since most of PTMs are dynamic and transient. It is difficult to extract a protein from eukaryotic cell lysate with a single desired PTM. Therefore, the synthesis of proteins with specific PTMs using chemical or biological methods is the preferential solution. With the advance of solid phase peptide synthesis (SPPS), a great number of important contributions have been made to biomedical research.[10-15] This approach is straightforward and can be applied to synthesize proteins with all kinds of PTMs. However, given the fact that this approach lies in the mechanism of the formation of amide bond with chemical reactions, it suffers from low yield when the protein of interest is larger than 50 residues.[16] Even with 99% yield at each step of amide formation, the overall yield of synthesis of proteins with 50 residues barely passes 60%. That is to say, SPPS is a convenient tool to synthesis peptides shorter than 30 amino acids but not applicable with larger proteins. Due to the fact that most of proteins of interest with PTMs consist more than 100 residues, SPPS is not an optimal way to synthesize proteins with PTMs.

Besides PTMs, proteins with NCAAs bearing unique functional groups are also in great demand. Fluorine NMR provides a unique prospective to study protein folding and unfolding, enzymatic actions and protein-protein interactions.[17-19] It requires to incorporate a NCAA with one or several fluorine atoms at certain position on the protein of interest. Click-chemistry, Copper-catalyzed azide–alkyne cycloaddition in particular, is ideal for protein labeling owing to the fact that it can be done under physiological

conditions and it is orthogonal to all canonical amino acids so that only the residue of interest is labeled.[20-22] Since there is no canonical amino acids bearing an azide or an alkyne group, an installation of a NCAA with those groups is a prerequisite. Moreover, alkene, aldehyde, ketone, hydroxylamine and hydrazine groups all display their unique reactivities which will be greatly beneficial to protein labeling.[23,24] Although SPPS provide a solution for some small-size proteins and peptides, a technique that enables the incorporation of NCAAs into proteins in a more efficient manner with high yield is still desired.

Chemical Synthesis of Proteins With PTMs

To solve the problem with the low yield of SPPS, researchers have shifted their focus to a more efficient way to synthesize proteins. In the past 30 years, recombinant DNA-based expression of proteins in genetically engineered cells has been the most powerful approach to synthesize proteins. Nonetheless, cells only carry the machinery to synthesize proteins with 20 canonical amino acids so there has to be a way to derive amino acids with PTMs from the native ones. Cysteine is widely chosen to be the reactive head due to its good nucleophilicity towards halogenated alkyl chains. Therefore, with the mutation of residue of interest to cysteine and the removal of all native cysteines in the protein of interest, specific mimics of amino acids with PTMs are produced. All three kinds of methyl lysine mimics have been made with this cysteine mediated approach via the reaction between cysteine and alkyl bromide (**Figure 2**).[25]

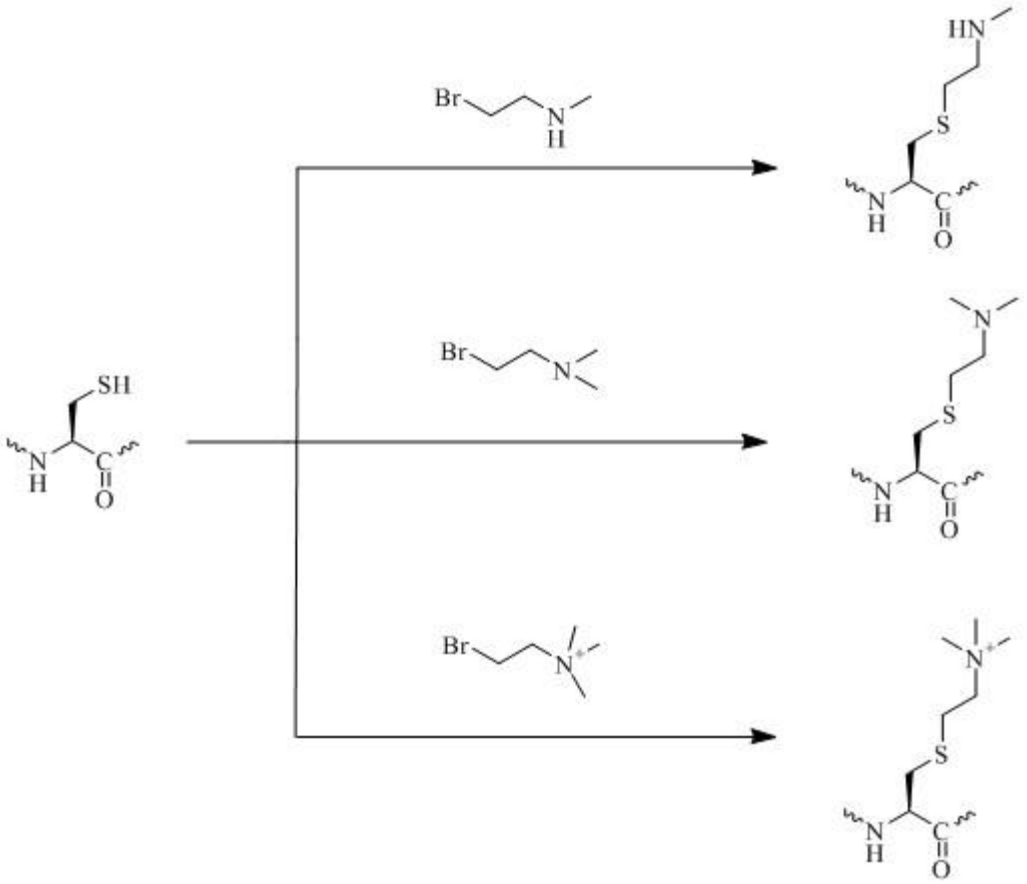


Figure 2. Cysteine-mediated methyl lysine installation.

The methylene at γ position is replaced by a size-comparable sulfur atom. Despite that this approach overcomes the low yield problem with SPPS, it has a strict requirement that the protein of interest cannot have any cysteine. All the native cysteines in the protein of interest have to be mutated to other canonical amino acids. Given the fact that cysteines serve important catalytic roles in a lot of enzymes and sometimes disulfide bonds are essential to fold proteins, this approach cannot be widely applied to install PTMs into proteins.

Native Chemical Ligation

Native chemical ligation (NCL) was a great breakthrough in the field of the synthesis of proteins with PTMs.[16] It relies on the reaction between the C-terminal thioester of one synthesized peptide and the N-terminal cysteine residue of the other. The thiol group in the cysteine reacts to the thioester by the means of reversible thiol/thioester exchange to afford the new thioester-linked conjugated peptide, and the amine in cysteine performs irreversible intramolecular rearrangement to give the full length protein (**Figure 3**). Since two peptide fragments are synthesized separately, this approach extends the size of proteins available for SPPS to more than 100 residues. In particular cases, when the NCAA locates in the N-terminal peptide fragment, the C-terminal peptide can be synthesized by recombinant DNA-based protein expression[26]. This method offers a solution to synthesize larger proteins with NCAAs incorporated close to N-termini.

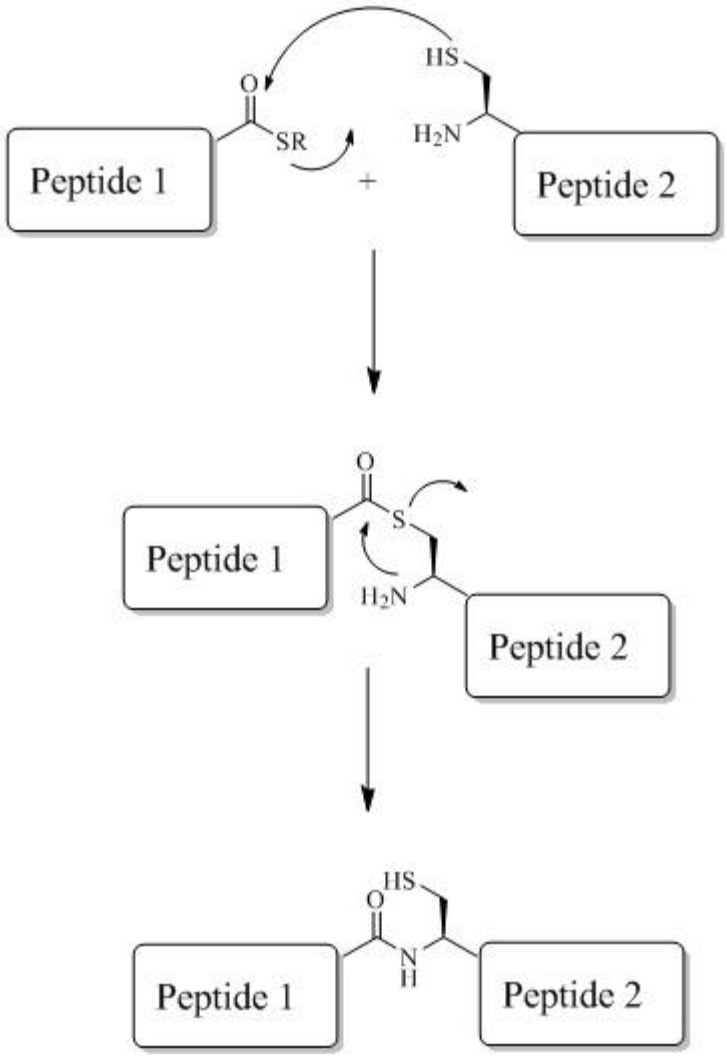


Figure 3. Native chemical ligation.

Expressed Protein Ligation

While the installation of NCAAs at N-termini of proteins was realized, the synthesis of proteins with NCAAs incorporated at C-termini became the remaining challenge. Unlike a C-terminal peptide fragment, an N-terminal fragment that bears a thioester at its C-terminus cannot be readily recombinantly produced. Therefore, a special method to express peptide with a C-terminal thioester with the recombinant DNA technique was demanded. Protein-splicing is a process involving intramolecular rearrangement to extrude the internal sequence (intein) and join the remaining sequences (extein) via reversible transesterification and an S to N acyl shift to a form an amide linkage.[27,28] A mutated system that fuses a protein of interest (POI) to intein results in the thioester formation between the POI and intein without further excision of the intein, yielding the POI with a C-terminal thioester. With the treatment of other thiol bearing nucleophiles, such as DTT, β-mercaptoethanol or thiophenol, proteins with a C-terminal thioester can then undergo 'native chemical ligation' like chemistry with peptide fragments with a N-terminal cysteine and unique NCAAs that are synthesized by SPPS. (**Figure 4**)[29] This so-called 'expressed chemical ligation' offers an optimal solution for the synthesis of proteins with NCAAs at their C-termini. With two sequential steps of native chemical ligation and expressed chemical ligation, proteins with NCAAs in the middle were also successfully synthesized.[30]

Although NCL and expressed protein ligation require the insertion of a cysteine for the ligation purpose, they don't require the removal of the rest endogenous cysteines such as in the cysteine-modification approach. Indeed, NCL and expressed protein

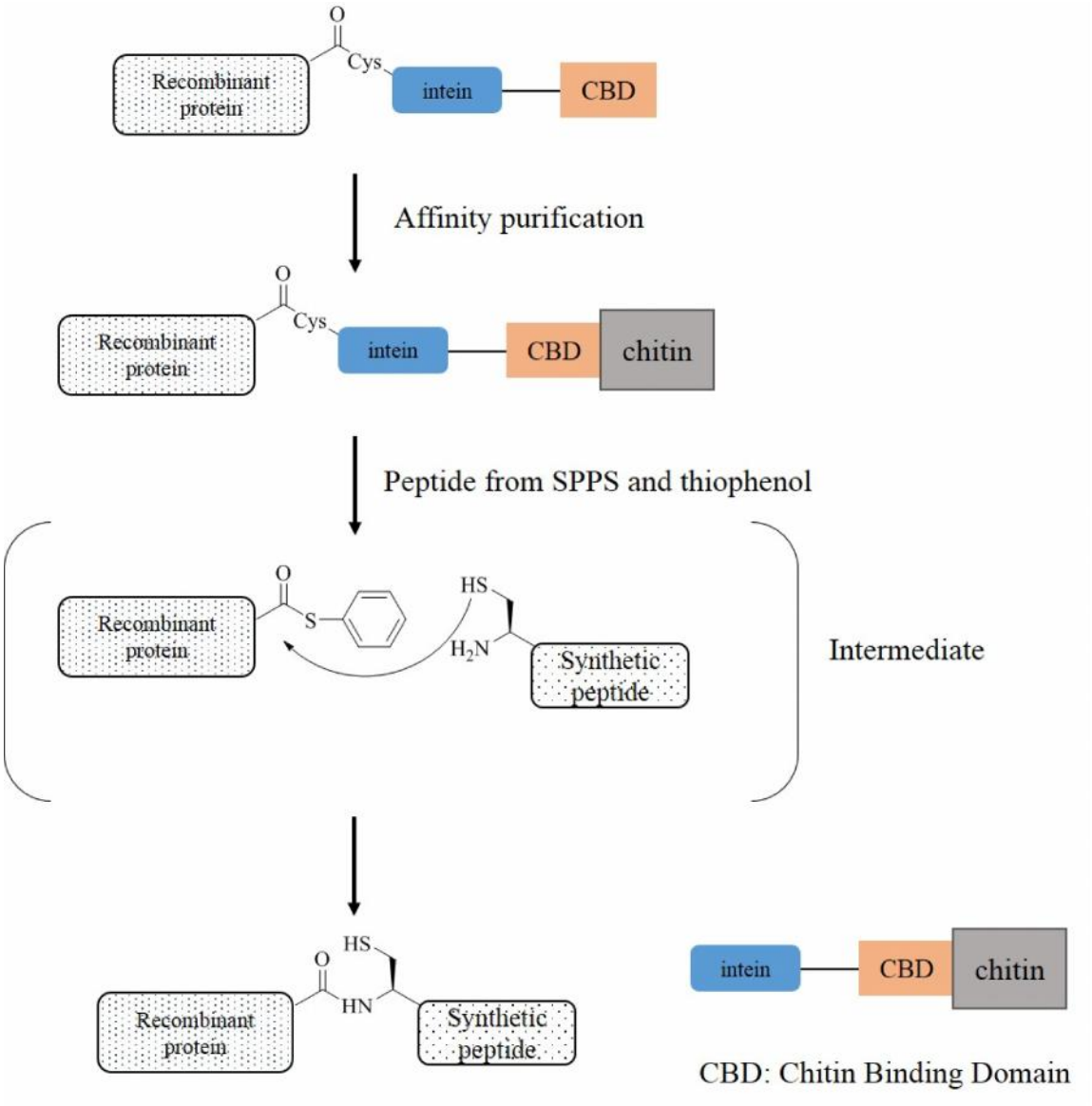


Figure 4. Expressed protein ligation.

ligation have greatly expanded the number of proteins with PTMs available, yet its drawbacks impede the application of the approach to all proteins with PTMs. For one thing, NCL may induce a cysteine mutant if there is no cysteine available as a 'stitch' close to the PTM site. For another, despite that a two-step consecutively expressed protein ligation allows to synthesize proteins with PTMs in the center, the overall experiment procedures could be very complicate and not applicable to most of the biochemistry labs.

The Genetic NCAA Encoding Approach

Considering the high efficiency of protein synthesis with the ribosomal translation system, it is an optimal solution to synthesize proteins with NCAAs. The first endeavor made to utilize the ribosomal translation system was to apply the ribosomal protein synthesis mechanism with chemically synthesized aminoacyl-tRNAs that recognized nonsense or frameshift codons.[31-33] Despite the successful incorporation of over 80 NCAAs into proteins via this approach, it suffers from the low yield and the low general application potential of the technique due to the challenge to chemically synthesize aminoacyl-tRNAs.[34]

When selenocysteine, the twenty-first proteinogenic amino acid was discovered, its astonishing encoding mechanism surprised the field. Unlike twenty canonical amino acids, selenocysteine utilizes UGA, a typical stop codon, to achieve the incorporation into proteins.[35,36] At the beginning of the 21st century, it came to the fact that the twenty-second proteinogenic amino acid, pyrrolysine, adopts another stop codon UAG for its encoding.[37,38] To encode an amino acid into proteins, a tRNA which exclusively bind to

the amino acid is required as well as an aminoacyl-tRNA synthetase (aaRS) which catalyzes the reaction linking the 3'-end of the tRNA to the carboxylate of the amino acid with the consumption of one molecule of ATP. The anticodon part of the tRNA specifically binds to the codon cognate for the amino acid on the mRNA[39]. Inspired by the fact that both selenocysteine and pyrrolysine are coded by a stop codon, Schultz and co-workers developed a method that uses stop codons and frameshift codons to encode NCAAs into proteins (**Figure 5**)[40-44]. The technique relies on the introduction of an orthogonal tRNA/aaRS into cells. That is to say, neither does aaRS charge any endogenous tRNAs nor is the tRNA recognized by any endogenous aaRSs. Moreover, the exogenous aaRS must be able to install an NCAA to the tRNA efficiently which requires no editing mechanism for the aaRS. One of the most successful case was the introduction of tyrosyl-tRNA synthetase (*Mj*TyrRS)/ $tRNA_{CUA}^{Tyr}$ derived from an archaea, *Methanocaldococcus jannaschii* into *Escherichia coli.* Not only does the *Mj*TyrRS/ $tRNA_{CUA}^{Tyr}$ show high orthogonality to all the endogenous bacterial aaRS/tRNA pairs, but also they can be well expressed in *Escherichia coli.* With the evolution of *Mj*TyrRS, a variety of NCAAs have been incorporated into proteins by this system.[45-48] However, In spite of the success of incorporation of NCAAs into proteins in *Escherichia coli,* this method is not applicable to eukaryotes since the *Mj*TyrRS/ $tRNA_{CUA}^{Tyr}$ fails to achieve orthogonality with eukaryotic systems. Therefore, a new aaRS/tRNA pair that is orthogonal to eukaryotic aaRS/tRNA pairs is needed. Several bacterial aaRS/tRNA pairs have been successfully applied to the incorporation of a number of NCAAs in both yeast and mammalian cells. Since they are derived from bacteria, they are not applicable with

Escherichia coli. Indeed TyrRS/tRNA pairs from different origins demonstrate their abilities to incorporate tyrosine derivatives into protein *in vivo* in both eukaryotic and prokaryotic systems, however, the PTMs on tyrosine only contribute to a small portion of the whole spectrum of protein PTMs and the rest of the posttranslational modified amino acids couldn't be encoded via this method. Posttranslational modified lysines are far more abundant and regulate DNA and protein functions.[49-51] The pyrrolysyl-tRNA synthetase (PylRS)-$tRNA^{Pyl}_{CUA}$ pair, an aaRS/tRNA pair naturally exists in certain methanogenic archaea to incorporate pyrrolysine (**Figure 5**), doesn't cross-react with any endogenous aaRS/tRNA pairs in either bacterial or eukaryotic systems. Consequently, it is optimal for the incorporation of lysine derivatives in both bacterial and eukaryotic cells. Interestingly, unlike the *Mj*TyrRS/tRNA pair, the anticodon region of $tRNA^{Pyl}_{CUA}$ barely affects its binding to PylRS, which allows mutations on the anticodon to make an alternative suppressor other than amber codon. It not only enables an opportunity to use a sense codon to encode NCAAs, which can boost up the suppression efficiency, but also makes it possible to incorporation two different NCAAs into proteins with the combination of the *Mj*TyrRS/ $tRNA^{Tyr}_{CUA}$ pair.[52] Moreover, with the large cavity of the catalytic site of PylRS, it can be evolved to accommodate other canonical amino acid derivatives which greatly expanded the spectrum of proteins with PTMs available via this approach.

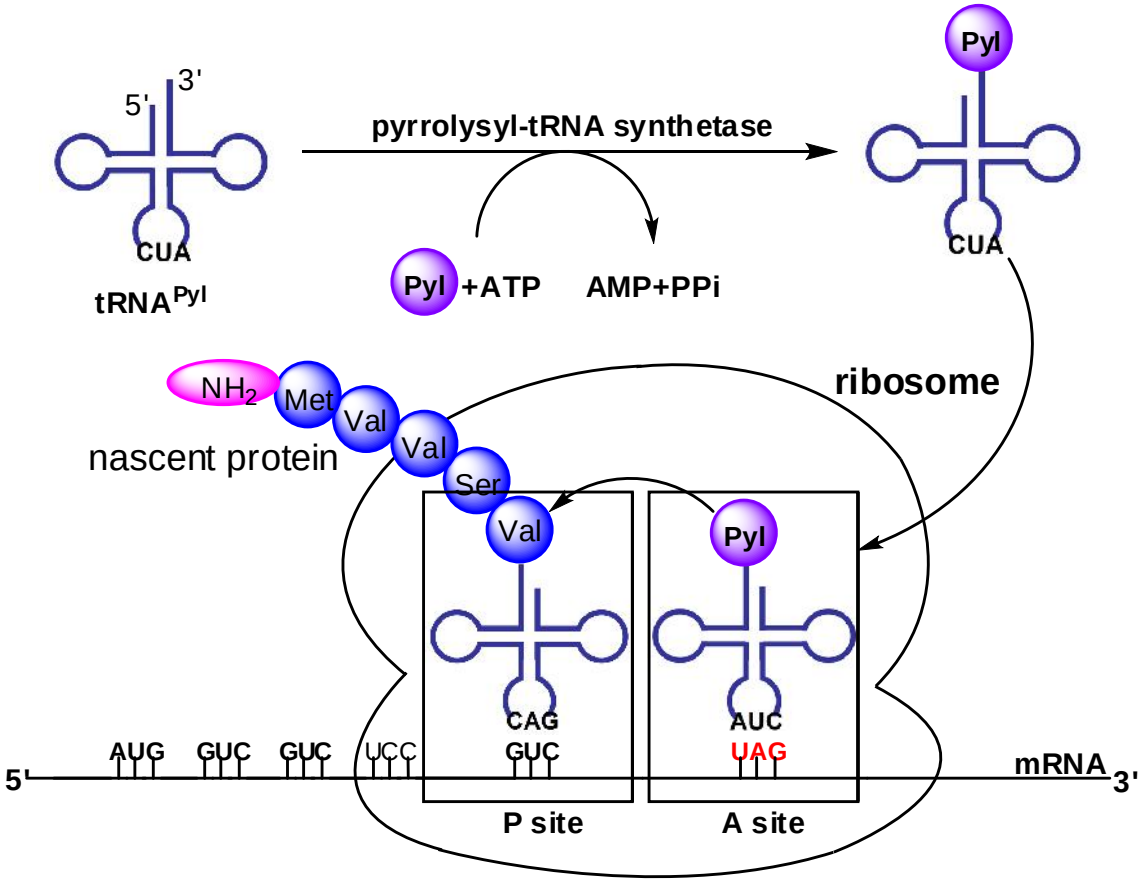


Figure 5. The genetic pyrrolysine incorporation machinery.

Histone Proteins and Their Methylations and Acylations

Histone octamer is the protein component of the nucleosome core. Histone octamer contains two copies of the four core histone proteins, H2A, H2B, H3 and H4, respectively[53]. All four proteins have a lysine-rich free loop region at the N-termini, which are widely recognized as the parts that has intensive interactions with DNA. Lysine, arginine and serine residues undergo a variety of PTMs which are believed to dynamically control cellular activities[54-58]. In particular, lysine modifications play a pivotal role in DNA transcription and repair which are closely related to a variety of diseases, especially cancers. Given the fact that these PTMs are transient and dynamic, it is important to elucidate how these PTMs are regulated. To study mechanisms of their cellular activities in detail, homogenous histones with these PTMs need to be obtained as well. Efforts have been made to study how these PTMs are regulated using peptide substrates made by SPPS. However, since these peptides are only parts of the full-length histone proteins, conclusions drawn from these studies may be incomplete or even controversial. With the genetic encoding approach with PylRS-tRNA$^{Pyl}_{CUA}$ system, we are able to synthesize homogenous full-length histones with PTMs at specific sites, which are natural substrates for a variety of enzymes for lysine modifications.

Among all four histone proteins, we chose histone H3 as the target for our study since H3 has 11 lysine sites that are reported to undergo PTMs which is the most abundant among all four histones[59-61]. H3 starts with an N-terminal free loop followed by an α-helix-loop-α-helix structure where K56, K79, K115 and K122 locates (**Figure 6**). It is well adopted that the acetylation of the lysines in the free loop region regulates

the interactions between H3 and DNA. We found that the amino groups of K56 and K115 are close to the phosphate groups on DNA so even that they are located in the core region of H3, they may still get involved with the interactions to DNA[59,62]. In this dissertation, we studied the substrate specificity of all four nuclear sirtuins towards different sites of lysine acylations in H3.

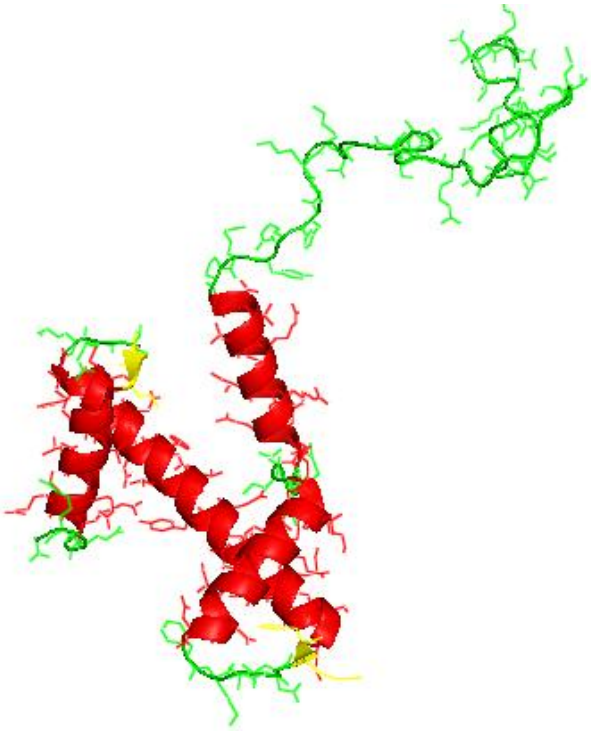


Figure 6. Histone H3 crystal structure. PDB entry: 1KX5. Red parts are α-helices, yellow parts are β-strands and green parts represents free loops.

CHAPTER II*

A GENETICALLY ENCODED PHOTOCAGED N^{ε}-METHYL-L-LYSINE

Introduction

Protein lysine methylation is an enzymatic process that involves the transfer of a methyl group from *S*-adenosyl-L-methionine to a lysine side-chain amine in a protein. It represents one the most important posttranslational modification and is crucial in modulating chromatin-based transcriptional control and shaping inheritable epigenetic programs in the eukaryotic cells.[63,64] There are three lysine methylation patterns, mono-, di- and trimethylation, which may serve different regulatory roles.[65] The study of protein lysine methylation is critical to understanding chromatin epigenetics and transcription factor regulation but has long been impeded by the challenge of synthesizing site-specifically and quantitatively methylated proteins.[66] Several methods including enzymatic protein methylation, native chemical ligation, and chemical modification of cysteines have been introduced to synthesize proteins with a defined lysine methylation.[67-69] However, they all suffer limitations. Enzymatic protein methylation hardly reaches completeness and the separation of the modified protein from the original one is difficult. Native chemical ligation requires a cysteine to mediate the ligation and the installation of a methylated lysine in the middle of a protein is problematic. The

---

*Reprinted with permission from "A genetically encoded photocaged N^{ε}-methyl-L-lysine" by Wang, Y. S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. *Mol Biosyst* 2010, 6, 1557., Copyright 2010 by Royal Society of Chemistry. Reproduced by permission of The Royal Society of Chemistry.

cysteine modification method can only install a methylated lysine analog into a protein. We sought to resolve limitations associated with the methods discussed above by applying the genetic non-canonical amino acid (NCAA) incorporation method to synthesize methylated proteins. Originally developed by Schultz *et al.*,[41,70] the genetic NCAA incorporation method relies on the read-through of an in-frame amber UAG codon in mRNA by an amber suppressor tRNA that is specifically acylated with a NCAA by an evolved aminoacyl-tRNA synthetase (aaRS). This method is equivalent to the naturally occurring pyrrolysine (Pyl) incorporation machinery discovered in methanogenic archaea and some bacteria.[38,71] In these organisms, Pyl is cotranslationally inserted into proteins and coded by an in-frame UAG codon. Suppression of this UAG codon is mediated by a suppressor tRNA, pylT that has a CUA anticodon and is acylated with Pyl by pyrrolysyl-tRNA synthetase (PylRS). Similarly, the PylRS-pylT pair can be directly applied to incorporate Pyl and several other lysine derivatives into proteins at amber mutation sites in *E. coli*.[72-75] Together with pylT, evolved PylRSs have also been applied to incorporate N^{ε}-acetyl-L-lysine into proteins in both *E. coli* and mammalian cells.[74,76,77] Inspired by these advances, we thought it might be possible to incorporate either N^{ε}-methyl-L-lysine or its derivatives into proteins at amber mutation sites using an evolved PylRS-pylT pair. This may allow the synthesis of proteins with monomethylated lysines for their functional investigations. Herein, we wish to show that a N^{ε}-methyl-L-lysine can be site-specifically installed into a protein by the genetic incorporation of a photocaged N^{ε}-methyl-L-lysine using an evolved PylRS-pylT pair and the following deprotection under UV irradiation.

Experimental Section

General Experimental

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was preformed on Whatman SiO_2 60 F-254 plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Specific rotations of chiral compounds were obtained at the designated concentration and temperature on a Rudolph Research Analytical Autopol II polarimeter using a 0.5 dm cell. Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.80 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, and a minimal amount of sodium hydroxide pellet was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral or acidic conditions. ^{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 obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

H-Lys(Z)-OH (**5**) and H-Lys(Me)-OH·HCl (**17**) was obtained from Chem-Impex International, Inc. (Wood Dale, IL). Compound **7** was synthesized as reported[78]. All other reagents were obtained from commercial suppliers and used as received.

Chemical Synthesis

Compounds **6** and **8** were synthesized from **12**[79] in a scalable route (**Scheme II-1**). For comparison, a shorter synthesis from the relatively expensive N^{ε}-methyl-L-lysine hydrochloride (**17**) following the standard protocol of copper complexation, reaction with appropriate chloroformate, and decomplexation with 8-hydroxyquinoline[80] was also developed.

2-Nitrobenzyl chloroformate (10) and 2-nitrobenzyl trichloromethyl carbonate (11)

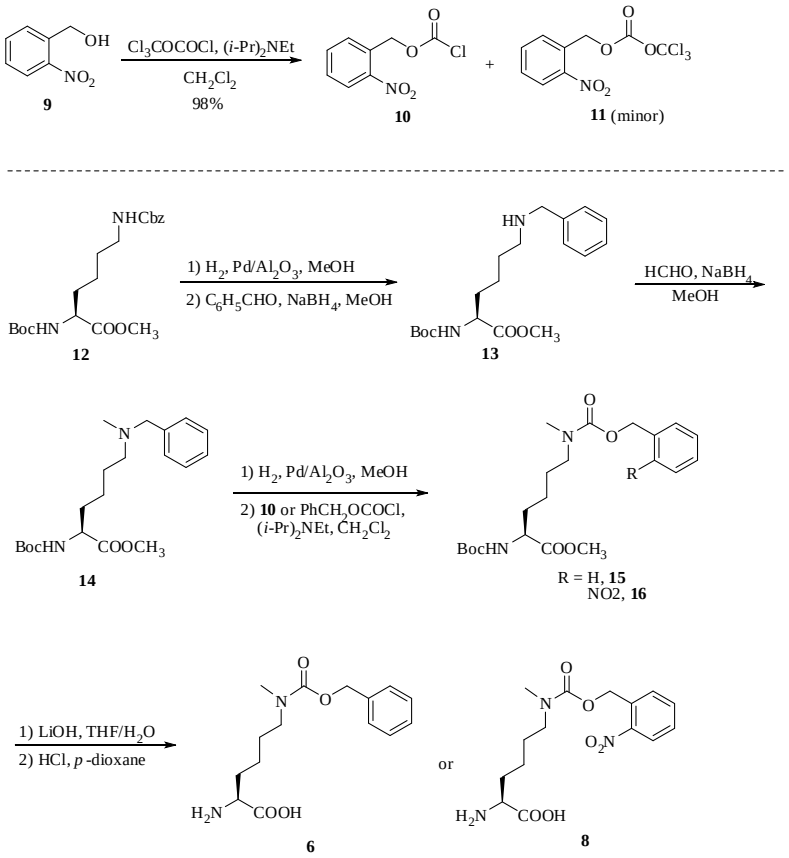
To a solution of 2-nitrobenzyl alcohol (**9**, 1.97 g, 12.9 mmol) in anhydrous dichloromethane (55 mL) cooled in an ice bath was added diphosgene (1.71 mL, 14.2 mmol) in dichloromethane (10 mL) dropwise over 10 min followed by diisopropylethylamine (2.25 mL, 12.9 mmol) in dichloromethane (10 mL) dropwise over 10 min. The reaction mixture was then stirred at room temperature for 2 h, and sodium hydroxide (1 *N*, 20 mL) was added and stirred further at room temperature for 30 min. The mixture was washed with water (30 mL), saturated sodium bicarbonate (30 mL x 2) and brine (30 mL), dried (Na_2SO_4), and flash chromatographed (EtOAc/hexanes, 1:20) to give a mixture of **10** and **11** (2.71 g, 98%) as a yellow oil. A minor fraction of impurity, presumably **11**, was evident from NMR analysis but did not interfere with the

next step reaction. No further purification was performed. For **10**[81]: ^{1}H NMR ($CDCl_3$, 500 MHz) δ 8.19 (d, 1 H, *J* = 8.0 Hz), 7.74 (t, 1 H, *J* = 7.8 Hz), 7.65 (d, 1 H, *J* = 8.0 Hz), 7.58 (t, 1 H, *J* = 7.8 Hz), 5.75 (s, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 150.6, 147.7, 134.4, 129.9, 129.8, 129.2, 125.6, 69.7. For **11**: ^{1}H NMR ($CDCl_3$, 500 MHz) δ 8.20 (d, 1 H, *J* = 6.5 Hz), 7.74 (t, 1 H, *J* = 7.7 Hz), 7.68 (d, 1 H, *J* = 8.0 Hz), 7.58 (t, 1 H, *J* = 7.7 Hz), 5.75 (s, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 147.3, 147.2, 134.5, 130.2, 129.8, 129.0, 108.0, 68.0.

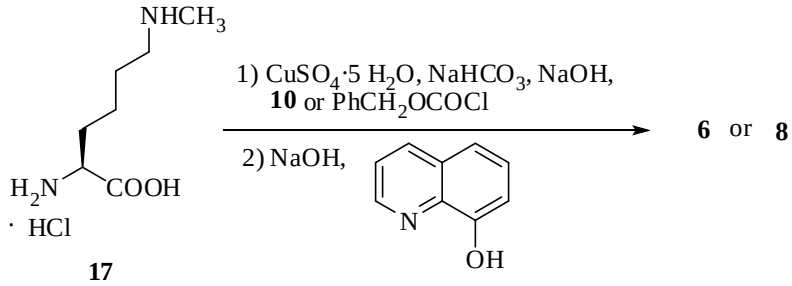
***N*˜-Boc-*N*ε-benzyl-L-lysine methyl ester (13)**

A solution of **12** (4.20 g, 10.6 mmol) in methanol (100 mL) was hydrogenated under a H_2 balloon in the presence of palladium on alumina (10 wt.% Pd, 0.71 g, 0.67 mmol) at room temperature for 3 h, and TLC analysis showed a complete conversion. The mixture was then filtered over a pad of Celite and the solution was directly used for the next step reaction. The material should be immediately used without purification since prolonged storage at room temperature or flash chromatography would contribute to lactam formation.

To a solution of the above amine (~10.6 mmol) in methanol was added benzaldehyde (4.00 mL, 39.4 mmol), and the reaction mixture was stirred at room temperature for 30 min. The mixture was then cooled in an ice bath, and sodium borohydride (0.75 g, 19.8 mmol) was added portionwise. The mixture was then stirred at room temperature overnight, and water (10 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), saturated sodium



Scheme II-1. Longer synthesis of **6** and **8**.



Scheme II-2. Shorter synthesis of **6** and **8**.

bicarbonate (30 mL) and brine (30 mL), dried (Na_2SO_4), and evaporated to give the crude **13** as a yellow oil, which was used in the next step reaction without further purification. A small fraction of pure **13** was obtained by flash chromatography (10% methanol with 5% triethylamine in dichloromethane) for characterization. $[\alpha]_D^{19}$ +7.4 (*c* 4.85, CH_2Cl_2); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.35-7.30 (m, 4 H), 7.27-7.24 (m, 1 H), 5.05 (d, 1 H, *J* = 7.0 Hz), 4.32-4.28 (m, 1 H), 3.78 (s, 2 H), 3.74 (s, 3 H), 2.63 (t, 2 H, *J* =7.0 Hz), 1.83-1.78 (m, 1 H), 1.67-1.60 (m, 1 H), 1.58-1.48 (m, 3 H), 1.44 (s, 9 H), 1.41-1.36 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.5, 155.5, 140.5, 128.5, 128.2, 127.0, 79.9, 54.1, 53.5, 52.3, 49.1, 32.7, 29.7, 28.4, 23.2; HRMS (ESI) calcd for $C_{19}H_{31}N_2O_4$ ($[M + H]^+$) 351.2284, found 351.2282.

***N*˜-Boc-*N*£-benzyl-*N*£-methyl-L-lysine methyl ester (14)**

To a solution of crude **13** (~10.6 mmol) in methanol (100 mL) in methanol 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 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 h, 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 (Na_2SO_4), evaporated, and flash chromatographed (EtOAc/hexanes, 1:1 then 5% to 10% methanol in dichloromethane) to give **14** (3.32 g, 86% yield for three steps) as a yellow oil. $[\alpha]_D^{20}$ +7.7 (*c* 2.37, CH_2Cl_2); ^{1}H NMR ($CDCl_3$, 500 MHz) δ

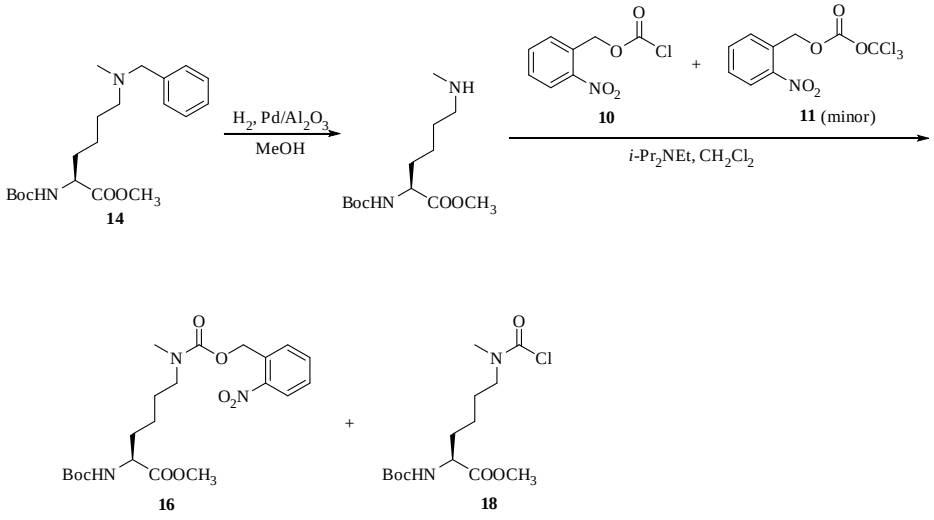
7.33-7.29 (m, 4 H), 7.26-7.23 (m, 1 H), 5.08 (d, 1 H, *J* = 8.5 Hz), 4.31-4.27 (m, 1 H), 3.73 (s, 3 H), 3.46 (s, 2 H), 2.35 (t, 2 H, *J* =7.2 Hz), 2.17 (s, 3 H), 1.82-1.76 (m, 1 H), 1.66-1.47 (m, 3 H), 1.44 (s, 9 H), 1.40-1.32 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.6, 155.6, 139.3, 129.2, 128.4, 127.1, 80.0, 62.5, 57.1, 53.6, 52.4, 42.4, 32.7, 28.5, 27.0, 23.3; HRMS (ESI) calcd for $C_{20}H_{33}N_2O_4$ ([+ H]$^+$) 365.2440, found 365.2437.

***N*˜-Boc-*N*£-Cbz-*N*£-methyl-L-lysine methyl ester (15)**

A solution of **14** (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 h. 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. A small fraction of pure amine was obtained by flash chromatography (10% methanol with 5% triethylamine in dichloromethane) for characterization. $[\alpha]_D^{19}$ +6.9 (*c* 1.90, CH_2Cl_2); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.38 (d, 1 H, *J* = 7.0 Hz), 4.14-4.12 (m, 1 H), 3.59 (s, 3 H), 2.95 (s, 1 H), 2.46 (t, 2 H, *J* =7.2 Hz), 2.29 (s, 3 H), 1.69-1.64 (m, 1 H), 1.55-1.48 (m, 1 H), 1.43-1.33 (m, 2 H), 1.30 (s, 9 H), 1.27-1.22 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.4, 155.5, 79.6, 53.3, 52.1, 51.2, 35.8, 32.2, 28.8, 28.2, 23.0; HRMS (ESI) calcd for $C_{13}H_{27}N_2O_4$ ([M + H]$^+$) 275.1971, found 275.1968.

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 benzyl chloroformate (95%, 1.50 mL, 10.5 mmol) dropwise over 10 min, and the mixture was stirred at room temperature for 12 h. 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_2SO_4), evaporated, and flash chromatographed (EtOAc/hexanes, 1:3) to give **15** (2.59 g, 90% for two steps) as a colorless oil. $[\alpha]_D^{19}$ +1.2 (*c* 1.51, CH_2Cl_2); ^1H NMR analysis showed a 1.5:1 mixture of rotamers at room temperature. Major rotamer: ^1H NMR ($CDCl_3$, 500 MHz) δ 7.33-7.32 (m, 4 H), 7.29-7.26 (m, 1 H), 5.26 (d, 1 H, *J* = 6.0 Hz), 5.10 (s, 2 H), 4.24 (m, 1 H), 3.69 (s, 3 H), 3.32-3.19 (m, 2 H), 2.87 (s, 3 H), 1.78-1.66 (m, 2 H), 1.57-1.51 (m, 2 H), 1.41 (s, 9 H), 1.36-1.28 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.4, 156.5, 155.6, 137.0, 128.5, 128.0, 127.9, 79.7, 67.0, 53.4, 52.2, 48.4, 33.9, 32.0, 28.4, 26.9, 22.2; Characteristic peaks of the minor rotamer: ^1H NMR ($CDCl_3$, 500 MHz) δ 5.05 (d, 1 H, *J* = 6.0 Hz); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 156.3, 155.4, 79.9, 53.3, 34.6, 32.4, 27.5, 22.5; HRMS (ESI) calcd for $C_{21}H_{33}N_2O_6$ ($[M + H]^+$) 409.2339, found 409.2332.



Scheme II-3. Synthesis for compound **18**.

$N^{\sim}$-Boc-$N^{£}$-(2-nitrobenzyl)oxycarbonyl-$N^{£}$-methyl-L-lysine methyl ester (16) and $N^{\sim}$-Boc-$N^{£}$-chlorocarbonyl-$N^{£}$-methyl-L-lysine methyl ester (18)

Compound **14** (1.06 g, 2.91 mmol) was converted into the corresponding amine by hydrogenolysis, which was then treated with crude **10** (0.94 g, 4.38 mmol) according to the procedure for **15** to give **16** (0.91g, 69% for two steps) as a yellow oil. A small amount of **18** (yield not determined), the structure of which was assigned based on NMR and MS analysis data, was obtained as a colorless oil. Presumably Boc-Lys(Me)-OMe reacts with **11** to give **16** and generates one molecule of phosgene at the same time, which then acylates the residual Boc-Lys(Me)-OMe to afford **18**.

For **16**: $[\alpha]_D^{20}$ +8.9 (*c* 1.70, CH_2Cl_2); R_f = 0.46 (EtOAc/hexanes, 1:1); ^{1}H NMR analysis showed a 1.1:1 mixture of rotamers at room temperature. Major rotamer: ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.96 (d, 1 H, *J* = 8.0 Hz), 7.58 (t, 1 H, *J* = 7.5 Hz), 7.50-7.47 (m, 1 H), 7.39 (t, 1 H, *J* = 8.2 Hz), 5.42 (s, 2 H), 5.21 (d, 1 H, *J* = 7.5 Hz), 4.20-4.16 (m, 1 H), 3.63 (s, 3 H), 3.21 (appar. nonet, 2 H, *J* = 7.1 Hz), 2.86 (s, 3 H), 1.74-1.44 (m, 4 H), 1.34 (s, 9 H), 1.30-1.24 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.2, 155.6, 155.4, 147.5, 133.6, 133.1, 128.7, 128.5, 124.8, 79.6, 63.7, 53.2, 48.5, 33.8, 31.9, 28.2, 26.7, 22.2; Characteristic peaks of the minor rotamer: ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.13 (d, 1 H, *J* = 7.5 Hz), 2.83 (s, 3 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.1, 155.5, 155.4, 147.5, 128.9, 128.5, 79.7, 52.1, 48.4, 34.6, 32.3, 27.4, 22.4.

For **18**: $[\alpha]_D^{20}$ +6.6 (*c* 2.20, CH_2Cl_2); R_f = 0.61 (EtOAc/hexanes, 1:1); ^{1}H NMR analysis showed a 1.2:1 mixture of rotamers at room temperature. Major rotamer: ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.06 (d, 1 H, *J* = 8.0 Hz), 4.30 (m, 1 H), 4.20-3.73 (s, 3 H),

3.44 (t, 1 H, J = 7.5 Hz), 3.40-3.36 (m, 1 H), 3.10 (s, 3 H), 1.85-1.56 (m, 4 H), 1.43 (s, 9 H), 1.40-1.31 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.3, 155.5, 149.8, 80.1, 53.2, 52.8, 51.2, 38.6, 32.5, 28.5, 26.7, 22.4; Characteristic peaks of the minor rotamer: ^{1}H NMR ($CDCl_3$, 500 MHz) δ 3.74 (s, 3 H), 3.02 (s, 3 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 149.3, 80.2, 52.5, 36.8, 32.7, 27.4, 22.4; HRMS (ESI) calcd for $C_{14}H_{25}ClN_2O_5Na$ ([M + Na]$^+$) 361.1320 (^{37}Cl)/359.1350 (^{35}Cl), found 361.1348/ 359.1359; calcd for $C_{14}H_{26}ClN_2O_5$ ([M + H]$^+$) 339.1505/337.1530, found 339.1585/ 337.1557.

N^{ε}-Benzyloxycarbonyl-N^{ε}-methyl-L-lysine (6)

To a solution of **15** (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 h. 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 (Na_2SO_4), 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 **6** (1.51 g, 81% for two steps) as a white powder. $[\alpha]_D^{20}$ +14.1 (*c* 1.07, 3 *N* HCl) (lit.[82] $[\alpha]_D^{25}$ +14.0 (*c* 0.5, acetic acid)); ^{1}H NMR analysis showed a 1:1 mixture of rotamers at room temperature. Major rotamer: ^{1}H NMR (D_2O, 500 MHz) δ 7.46-7.42 (m, 5 H), 5.16 (s, 2 H), 3.68 (m, 1 H), 3.34 (m, 2 H), 2.90 (s, 3 H), 1.83 (m, 2 H), 1.60 (quintet, 2 H, *J* = 7.3 Hz), 1.34 (m, 2 H); ^{13}C NMR (D_2O, 75 MHz) δ 184.1, 158.4, 137.1, 129.4, 129.0, 128.5, 68.0, 56.5, 49.0, 35.1, 34.8, 27.6, 22.8; Characteristic peaks of the minor rotamer: ^{1}H NMR (D_2O, 500 MHz, pH = 14) δ 2.95 (s, 3 H); ^{13}C NMR (D_2O, 75 MHz) δ 128.2, 67.2, 49.2, 34.2, 27.3. HRMS (ESI) calcd for $C_{15}H_{23}N_2O_4$ ($[M + H]^+$) 295.1658, found 295.1656.

N^{ε}-(2-Nitrobenzyl)oxycarbonyl-N^{ε}-methyl-L-lysine (8)

According to the same procedure for **6**, **16** (0.914 g, 2.02 mmol) afforded **8** (0.514 g, 75% for two steps) as a pale yellow solid. $[\alpha]_D^{20}$ +14.2 (*c* 1.16, 3 *N* HCl); ^{1}H NMR analysis showed a 1.1:1 mixture of rotamers at room temperature. Major rotamer: ^{1}H NMR (D_2O, 500 MHz, pH = 14) δ 8.00 (d, 1 H, *J* = 8.0 Hz), 7.64 (t, 1 H, *J* = 7.7 Hz), 7.50 (d, 1 H, *J* = 8.5 Hz), 7.47 (t, 1 H, *J* = 8.2 Hz), 5.32 (s, 2 H), 3.95 (m, 1 H), 3.20 (m, 2 H), 2.77 (s, 3 H), 1.84 (m, 2 H), 1.49 (m, 2 H), 1.31-1.27 (m, 2 H); ^{13}C NMR (D_2O, 75 MHz) δ 184.2, 158.0, 147.6, 135.1, 132.7, 129.8, 129.4, 125.7, 65.0, 56.5, 49.2, 35.1, 34.9, 27.6, 22.8; Characteristic peaks of the minor rotamer: ^{1}H NMR (D_2O, 500 MHz, pH = 14) δ 2.82 (s, 3 H); ^{13}C NMR (D_2O, 75 MHz) δ 133.0, 129.7, 64.9, 49.3, 34.2, 27.2; HRMS (ESI) calcd for $C_{15}H_{22}N_3O_6$ ($[M + H]^+$) 340.1509, found 340.1513.

Synthesis of 6 from the shorter pathway

To a solution of **17** (1.50 g, 7.63 mmol) in water (30 mL) was added cupric sulfate pentahydrate (1.00 g, 4.00 mmol), followed by sodium bicarbonate (1.42 g, 16.90 mmol) in small portions to prevent excessive bubble formation. Benzyl 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 h, filtered, washed with water (100 mL), ethanol (50 mL) and diethyl ether (50 mL), and dried in the open air for 1 h 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. $[\alpha]_D^{22}$ +15.3 (*c* 1.02, 3 *N* HCl). All other characterization data were identical to that of **6** from the longer route.

Synthesis of 8 from the shorter pathway

According to the same procedure for **6**, **17** (0.50 g, 2.54 mmol) afforded **8** (0.39 g, 45% yield for two steps). The compound was identical to **8** from the longer route in all aspects.

DNA and Protein Sequences

DNA Sequences

Z Domain

atgactagtgtagacaactagatcaacaaagaacaacaaaacgccttctatgagatcttacatttacctaacctgaatgaggagc
agcgtgatgccttcatccaaagtttaaaagatgacccaagccaaagcgctaacctttagcagaagctaaaaagctaaatgatgc
tcaggcgcctaagggatctgagctccatcaccatcaccatcactaa

GFP_{UV}

atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtca
gtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggcca
acacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatg
cccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggt
gataccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaatacaactat
aactcacacaatgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagat
ggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtc
gacatagtctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacac
atggcatggatgaactctacaaagagctccatcaccatcaccatcactaa

pylT

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

Methanosarcina mazei PylRS

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacac
cacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggact
gcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcac
aaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatcc
gttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgataccg
gtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactg
gtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactga
caggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctct
ctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccag
gttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgat
accgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagacccatgcttgctccaaacctttacaactacctgcg
caagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgctacagaaaagagtccgacggcaaagaac
acctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaaatcttgaaagcataattacggactt
cctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctatggggatacccttgatgtaatgcacggagac
ctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttc
gggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggattt
ctaccaacctgtaa

Proteins Sequences

Z Domain

MTSVDNXINKEQQNAFYEILHLPNLNEEQRDAFIQSLKDDPSQSANLLAEAKKL

NDAQAPKGSELHHHHHH

X represents a noncanonical amino acid.

GFP_{UV}

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP

VPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYK

TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGI

KANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTXSALSKDPNEKR

DHMVLLEFVTAAGITHGMDELYKELHHHHHH

X represents a noncanonical amino acid.

Methanosarcina mazei PylRS

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSS

RTARALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKK

AMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTG

ATASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFR

ELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIER

MGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYR

KESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYG

DTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKR AARSESYYNGISTNL

Construction of Plasmids

Constructions of pY+ and pY-

The plasmid pY+ was derived from the pRep plasmid by replacing the suppressor tRNA in pRep by pylT.[83] The gene of pylT flanked by the lpp promoter at the 5' end and the *rrnC* terminator at the 3' end was amplified from pBK-AcKRS-pylT.[77,84] The plasmid pY- was derived from the pNeg plasmid by replacing the suppressor tRNA with pylT.[83] Similarly, the gene of pylT flanked by the lpp promoter at the 5' end and the *rrnC* terminator at the 3' end was amplified from pBK-AcKRS-pylT.[23,85] pY+ has a tetracycline selection marker, a chloramphenicol acetyltransferase gene with an amber mutation at D112. pY- has an ampicillin selection marker and a barnase gene with two amber mutations at Q2 and D44. The barnase gene is under control of a pBad promoter.

Construction of pET-pylT-GFP

Plasmid pET-pylT-GFP was derived from the plasmid pAcKRS-pylT-GFP1Amber in which GFP_{UV} has an amber mutation at Q204.[23,85] The restriction enzyme *BglII* was used to cut off the ACKRS gene. The digested pAcKRS-pylT-GFP1Amber plasmid was ligated to form pET-pylT-GFP.

Construction of pET-pylT-Z

The pET-pylT-Z plasmid was derived from pET-pylT-GFP. This gene was amplified from the pLeiZ plasmid[45]. Two restriction sites, *NdeI* at the 5' end and *SacI* at

the 3’ end, were introduced in the PCR product which was subsequently digested and used to replace GFP_{UV} in pET-pylT-GFP.

Construction of the pRS1 Library

The plasmid pBK-mmPylRS that encodes wild-type *Methanosarcina mazei* PylRS was derived from a pBK plasmid containing *p*-iodophenylalanyl-tRNA synthetase.[86] The pylRS gene is under the control of *E. coli glnS* promoter and terminator. It was amplified from genomic DNA of *Methanosarcina mazei* strain DSM 3647 (ATCC) by flanking primers, pBK-mmPylRS-NdeI-F and pBK-mmPylRS-PstI~NsiI-R. To construct the pRS1 library, NNK (N=A or C or G or T, K=G or T) mutations were introduced at six sites by overlap extension PCR.[87] The following pairs of primers were used to generate a PylRS gene library with randomization at six sites: (1) pBK-mmPylRS-NdeI-F (5'-gaatcccatatggataaaaaaccactaaacactctg-3') and mmPylRS-Mutlib01-R (5'-ggccctgtcaagcttgcgmnngtagttmnnmnngtttggagcaagca tggg-3'); (2) mmPylRS-Mutlib02-F (5'-cgcaagcttgacagggccctgcctgatcc-3') and mmPylRS-Mutlib03-R (5'-gcatcccgatcccatctgmnngaamnncagcatggtaaactcttc-3'); (3) mmPylRS-Mutlib04-F (5'-cagatgggatcgggatgcacacg-3') and mmPylRS-Mutlib05-R (5'-ccgaaacctgcccctatmnngggtttatcaatacccca-3'); (4) mmPylRS-Mutlib06-F (5'-ataggggcaggtttcgggctcgaacgcc-3') and pBK-mmPylRS-PstI~NsiI-R (5'-gtttgaaaatgcatttacaggttggtagaaatccc-3'). The gene library was digested with the restriction enzymes *NdeI* and *NsiI*, gel-purified, and ligated back into the pBK vector digested by *NdeI* and *PstI* to afford plasmid the pRS1 library. 1 μg of the ligation products were then electroporated into *E. coli* Top10 cells. Electroporated cells were

recovered in SOC medium for 60 min at 37 °C, transferred into a 2 L 2YT medium with kanamycin (25 μg/mL) and were incubated at 37°C to OD_{600} at 1.0. To calculate the library size, 1 μL recovered SOC culture was subjected to serial dilutions in 2YT, then plated on LB agar plates with kanamycin (25 μg/mL), and grown overnight in a 37°C incubator. Based on the colony numbers on these plates, the pRS1 library contains approximately 1.01×10^9 independent transformants. Sequencing pylRS variants in 20 clones did not reveal any significant bias at the randomization sites.

Selection Procedure for Evolving Pyrrolysyl-tRNA Synthetase

The selections followed the scheme shown in **Scheme II-4**. For the positive selection, the pRS1 library was used to transform *E. coli* TOP10 competent cells harboring pY+ to yield a cell library greater than 1×10^9 cfu, ensuring complete coverage of the pRS1 library. Cells were plated on LB agar plates containing 12 μg/mL tetracycline (Tet), 25 μg/mL kanamycin (Kan), 68 μg/mL chloramphenicol (Cm) and 1 mM **5**. After incubation at 37°C for 72 h, colonies on the plates were collected. Total plasmids were isolated and separated by 1 % agarose gel electrophoresis. pRS1 plasmids were extracted using a Gel-extraction kit (QIAGEN). The extracted pRS1 plasmids from the positive selection were used to transform *E. coli* TOP10 harboring pY− for the negative selection. After electroporation, the cells were allowed to recover for 1 h at 37°C in SOC media before being plated on LB agar plates containing 50 μg/mL Kan, 200 μg/mL ampicillin (Amp) and 0.2% arabinose. The plates were incubated for 16 h at 37°C. Survived cells were then pooled and pRS1 plasmids were extracted. The selection power to exclude out the mutants that also took endogenous amino acids was tested on

LB agar plates containing 50 μg/mL Kan, 200 μg/mL Amp, 0.2% arabinose and 1mM **5**. The plate contains 1 mM **5** showed much fewer colony numbers as times of negative selections increased. Five alternative selections (three positive + two negative) finally yielded many colonies. 22 single colonies after the third positive selection were selected and the plasmids were isolated for sequencing. 96 single colonies from the third positive selection were also chosen for testing their ability to grow on plates with 102 μg/mL chloramphenicol, 25 μg/mL Kan, 12 μg/mL Tet, and 1 mM of **5**, **6**, **7** or **8**. A plate without NCAA supplementary was used as a control. Images of colonies growing on different plates were shown in **Figure 6**. Sequences of PylRS variants that charge pylT with different NCAAs are presented in **Table 1&2**.

Protein Purification and Photolysis to Form the Monomethylated Protein

To express GFP_{UV} incorporated with a NCAA, we cotransformed *E. Coli* BL21(DE3) cells with pBK-mKRS1 and pET-pylT-GFP. Cells were recovered in 1 mL of LB medium for 1 h at 37 ºC before being plated on LB agar plate containing Kan (25 μg/mL) and Amp (100 μg/mL). A single colony was then selected and grown overnight in a 10 mL culture. This overnight culture was used to inoculate 100 mL of M9 minimal media supplemented with 1% glycerol, 300 μM leucine, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.2% NaCl, 25 μg/mL Kan and 100 μg/mL Amp. Cells were grown at 37ºC in an incubator (300 r.p.m.) and protein expression was induced when OD_{600} was 0.7 by adding IPTG to a final concentration of 1 mM and **5** to a final concentration of 1 mM. After 6 h induction, cells were harvested, resuspended in a lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and sonicated. The cell lysate was

clarified by centrifugation (60 min, 11,000 g, 4ºC). The supernatant was injected into a 30 mL Ni^{2+}-NTA column (Qiagen) on FPLC (ÄKTApurifier™, GE Healthcare Bio-Sciences Corp) and washed with 45 mL lysis buffer and 45 mL wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM imidazole, pH 8.0). Protein was finally eluted out by running an imidazole gradient from 40 mM to 250 mM in lysis buffer. Pure fractions were collected and concentrated. The buffer was later changed to 1 mM ammonium bicarbonate using an Amicon Ultra -15 Centrifugal Filter Devices (10,000 MWCO cut) (Millopore). The purified proteins were analyzed by 15% SDS-PAGE. GFP_{UV} proteins incorporated with other NCAAs were expressed and purified similarly except the supplemented NCAA was changed. For all NCAAs, 1 mM final concentration was used.

Z domain proteins incorporated with different NCAAs were expressed and purified as same as the expression of GFP_{UV} proteins except pET-pylT-Z was used to cotransform *E. coli* BL21 (DE3) together with pBK-mKRS1. 1 mM final concentration was used for all four NCAAs. The buffer of the finally purified Z domain proteins was changed to 1 mM ammonium bicarbonate using an Amicon centriplus YM-3 (3,000 MWCO cut) (Millopore). The purified proteins were analyzed by 15% SDS-PAGE.

Photolysis to Form the Monomethylated Protein

Z-7, **Z-8**, and **GFP-8** (1 mg/mL) in 1 mM ammonium bicarbonate solution were treated with 365 nm UV light form hand-held UV light source for one hour.

Protein LC-ESI-MS Analysis

An Agilent (Santa Clara, CA) 1200 capillary HPLC system was interfaced to an API QSTAR Pulsar Hybrid QTOF mass spectrometer (Applied Biosystems/MDS Sciex, Framingham, MA) equipped with an electrospray ionization (ESI) source. Liquid chromatography (LC) separation was achieved using a Phenomenex Jupiter C4 microbore column (150 × 0.50 mm, 300 Å) (Torrance, CA) at a flow rate of 10 µL per min. The proteins were eluted using a gradient of (A) 0.1% formic acid versus (B) 0.1% formic acid in acetonitrile. The gradient timetable was as follows: 2% B for 5 min, 2-30% in 3 min, 30-60% in 44 min, 60-95% in 8 min, followed by holding the gradient at 95% for 5 min, for a total run time of 65 min. The MS data were acquired in positive ion mode (500-1800 Da) using spray voltage of +5000 V. BioAnalyst software (Applied Biosystems) was used for spectral deconvolution. For the GFPuv protein analysis, a mass range of m/z 500-1800 was used for deconvolution and the output range was 10000-50000 Da using a step mass of 0.1 Da and an S/N threshold of 20. For the Z-Domain protein analysis, a mass range of m/z 500-2000 was used for deconvolution and the output range was 5000-15000 Da for Z-domain-His6X using a step mass of 0.1 Da and an S/N threshold of 20.

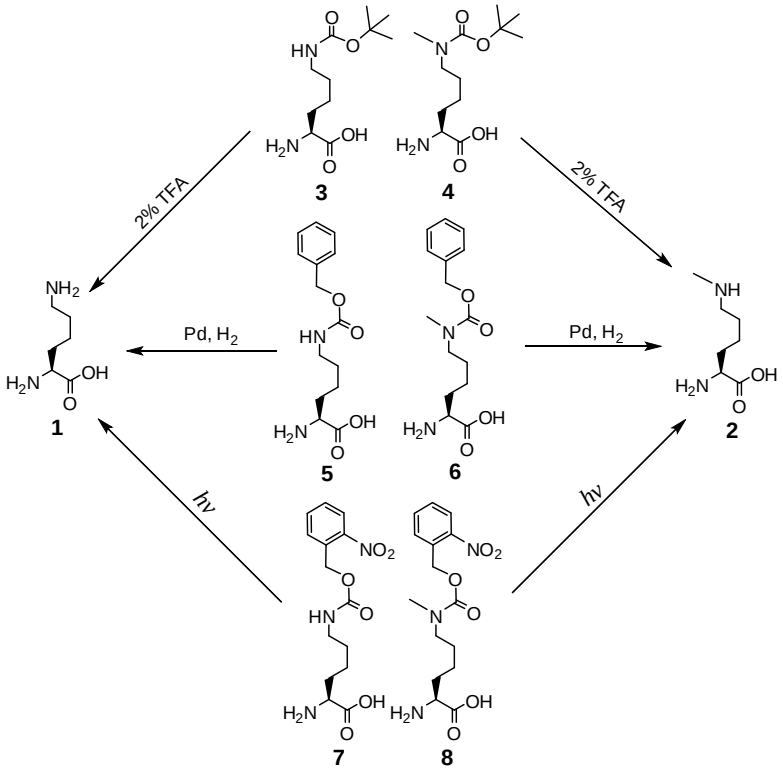
Result and Discussion

We chose to work on the genetic incorporation of protected N^{ε}-methyl-L-lysines instead of N^{ε}-methyl-L-lysine itself because it would be difficult to identify an evolved PylRS that specifically recognizes N^{ε}-methyl-L-lysine but not cellularly abundant lysine.

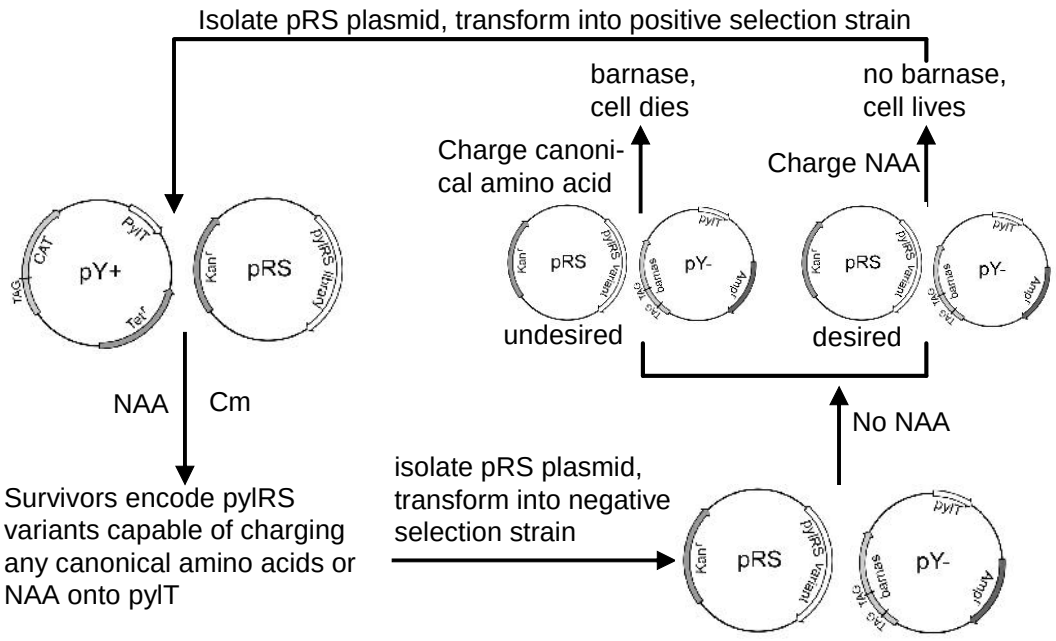
Three protected N^{ε}-methyl-L-lysines (**4**, **6** and **8** in **Scheme II-4**) were initially considered but eventually we chose **6** and **8**. The cleavage of the Boc protection group from **4** to recover N^{ε}-methyl-L-lysine needs to be carried out under a strong acidic condition. Although it has been demonstrated that wild type PylRS-pylT pair can genetically incorporate **3**[74] that is structurally close to **4** into proteins at amber codons and it is highly possible the same pair will also incorporate **4** into proteins, we thought the harsh condition for the deprotection is not suitable for many proteins (the incorporation of **4** into proteins using wild type PylRS-pylT pair in *E. coli* was published by Chin *et al*.[88]). The strong acidic condition for the deprotection denatures proteins that have to be refolded later. However, refolding is problematic for most large size proteins. On the contrary, the deprotection of **6** by catalytic hydrogenation and **8**, a photocaged N^{ε}-methyl-L-lysine by UV photolysis can be achievable under mild conditions that are suitable for most proteins.[89] In addition, the photolysis of **8** after its incorporation into proteins may also be carried out in living cells. This may allow the synthesis of methylated proteins directly in cells for their functional investigations.

The synthesis of both **6** and **8** started from lysine (S**cheme II-4**) and finished in gram quantities. To expand the substrate scope of PylRS to accommodate **6** or **8**, we constructed an active-site mutant library of the *Methanosarcina mazei* PylRS gene with randomization at six active site residues (L305, Y306, L309, N346, C348, and W417) (**Figure 7**) according to a standard protocol.[90] This gene library was then cloned into a pBK plasmid to form a pRS1 plasmid library in which mutant PylRS variants are under control of a constitutive *glnS* promoter.[90] Together with two selection plasmids, pY+ for

positive selection and pY- for negative selection, this plasmid library was then subjected to alternative positive and negative selections to identify PylRS variants specific for **6** or **8** (**Scheme II-5, Figure 8**).[76,89] The positive selection plasmid, pY+ contains genes encoding pylT and type I chloramphenicol (Cm) acetyltransferase with an amber mutation at D112.



Scheme II-4: L-Lysine and N^{ε}-methyl-L-lysine derivatives and their deprotections



Scheme II-5. The selection scheme to identify PylRS variants specific for a noncanonical amino acid.

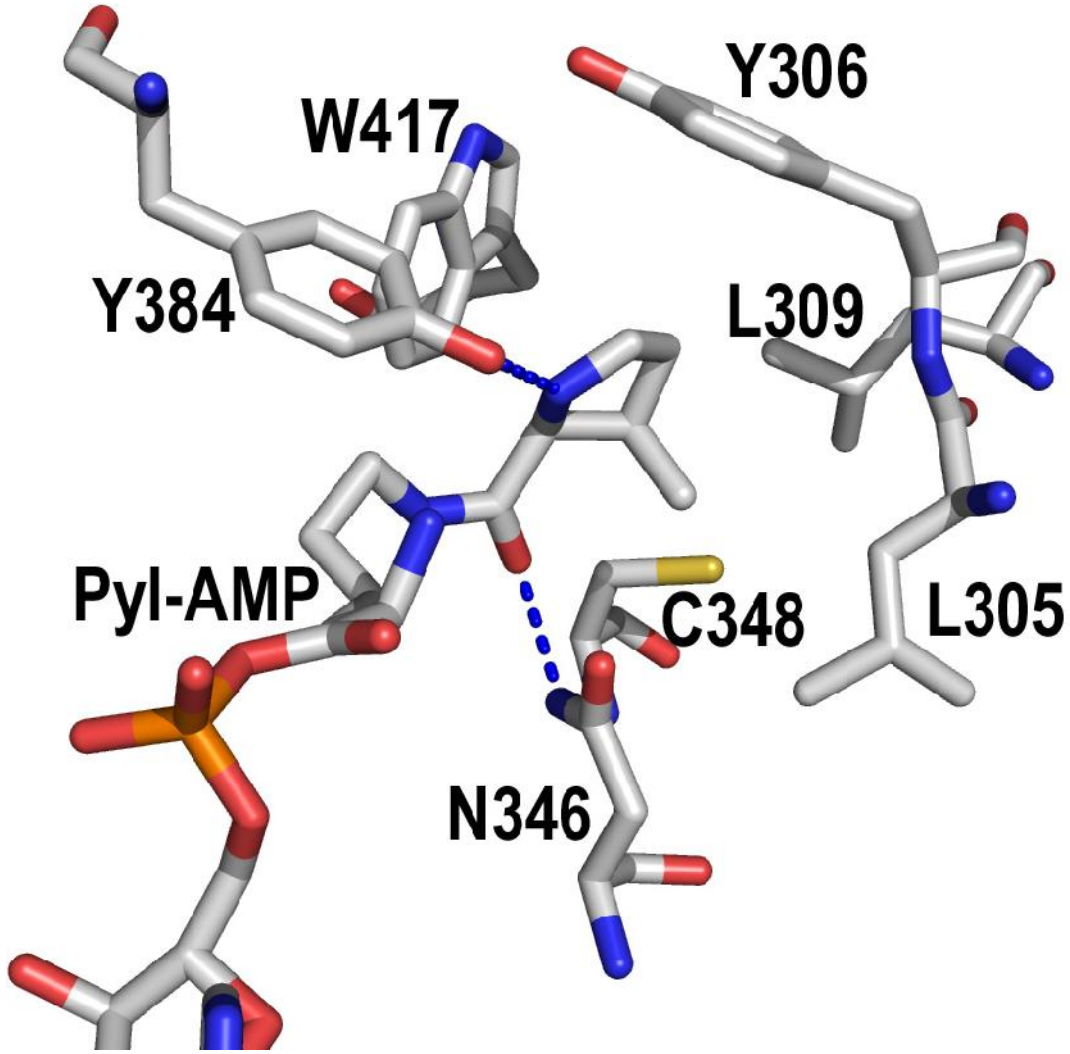


Figure 7. The active site of PylRS. The structure was derived from pdb entry: 2Q7E.[91]

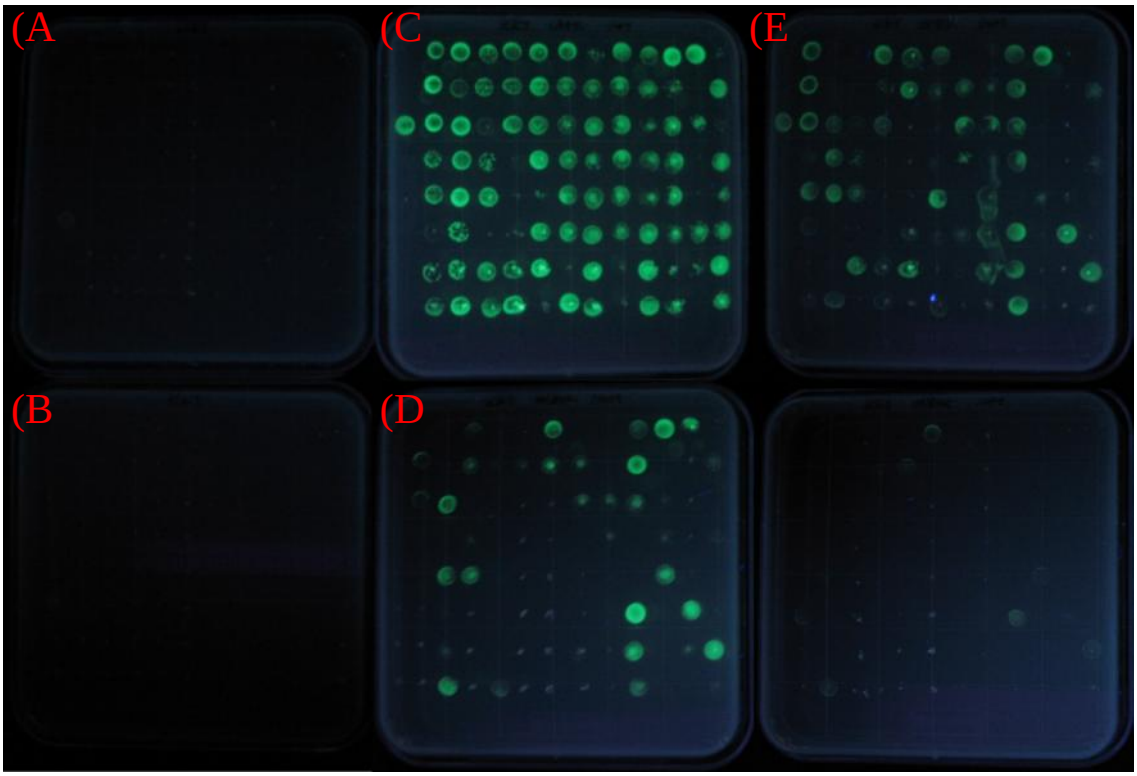


Figure 8. Growth of 96 single colonies from the third positive selection of 5 on LB plates with different supplements. (A) 68 g/mL Cm, 25 g/mL Kan and 12 g/mL Tet; (B) 102 g/mL Cm, 25 g/mL Kan and 12 g/mL Tet; (C) 1 mM 5, 102 g/mL Cm, 25 g/mL Kan and 12 g/mL Tet; (D) 1 mM 6, 102 g/mL Cm, 25 g/mL Kan and 12 g/mL Tet; (E1 mM 7, 102 g/mL Cm, 25 g/mL Kan and 12 g/mL Tet; (F) 1 mM 8, 102 g/mL Cm, 25 g/mL Kan and 12 g/mL Tet. The pY+ plasmid has a GFPUV gene under control of a T7 promoter. Its expression is promoted by the suppression of two amber mutations at positions 1 and 107 of a T7 RNA polymerase gene in pREP. The fluorescent intensity of the expression of GFPUV roughly represents the suppression efficiency at amber codons.

cotransforming *E. coli* with pY+ and pRS1 and then growing cells in Cm and NCAA-containing plates conferred the selection of PylRS variants that charge pylT with native amino acids or the supplied NCAA. The negative selection plasmid, pY- contains genes encoding pylT and toxic barnase that has two amber mutations at Q2 and D44. Cotransforming *E. coli* with pY- and pRS1 and then growing cells in plates without NCAA conferred the selection of PylRS variants that do not charge pylT with a native amino acid. Only PylRS variants that charge pylT with the provided NCAA but not any native amino acid will survive from both positive and negative selections. However, a series of selections (three positive selections + two negative selections) yielded no viable clones. We then chose an indirect route to identify clones specific for **6** or **8**. Since **5** and **7** are structurally close to **6** and **8**, respectively. We thought PylRS variants selected for **5** or **7** might also charge pylT with **6** or **8**, respectively. We carried out the selections of the pRS1 library to identify PylRS variants that are specific for **5**. After a series of selections, many colonies survived and most of them converge to two specific clones, in which mKRS1 shows the highest suppression efficiency and has mutations Y306M/L309A/C348T/T364K (**Table 1&2**). We then tested the efficiency of the evolved mKRS1-pylT pair to suppress an amber mutation at Q204 of GFP_{UV} in *E. coli*. A plasmid pET-pylT-GFP was constructed. It contains genes encoding pylT and GFP_{UV} with an amber mutation at Q204.

Table 1. Sequences of selected PylRS variants that charge pylT with all four NCAAs[1]

PylRS	Frequency	L305	Y306	L309	N346	C348	W417
mKRS1[1]	7/22	L	M	A	N	T	W
mKRS3	6/22	L	M	A	N	C	W
mKRS5	1/22	L	M	P	N	C	W

[1]Other PylRS clones are presented in **Table 2.** [2]This clone has an additional mutation T364K.

Table 2. Evolved PylRS variants that Charge pylT with different NCAAs.

Position	305	306	309	346	348	417	Remark[1]
WT	L	Y	L	N	C	W	
mKRS1[2]	L	M	A	N	T	W	**5, 7, 6, 8**
mKRS2	L	V	A	N	A	W	**5, 7, 6**
mKRS3[3]	L	M	A	N	C	W	**5, 7, 6, 8**
mKRS4	L	A	A	H	L	W	**5, 6**
mKRS5	L	M	P	N	C	W	**5, 6, 7, 8**
mKRS6	L	M	A	N	S	W	**5**
mKRS7	L	Y	A	N	A	W	**5**
mKRS8[4]	L	M	T	N	A	W	**5**
mKRS9	L	A	A	N	A	W	**5**
mKRS10	L	A	L	N	A	W	**5, 7**
mKRS11[5]	L	A	L	N	C	W	**5, 7**

[1] This column represents the NCAAs that can be taken by mutant PylRS variants. The order of compounds also indicates the decreasing encoding efficiency based on the screening results.

[2] The mutant was found seven times from 22 sequenced mutants and has extra mutation on T364K.

[3] The mutant was found six times from 22 sequenced mutants.

[4]The mutant has extra mutation on P297S.

[5] The mutant was found twice from 22 sequenced mutants.

The GFP_{UV} gene is under control of a T7 promoter. Cotransforming *E. coli* BL21 (DE3) cells with the selected pBK-mKRS1 and pET-pylT-GFP and growing cells in minimal medium supplemented with **5** afforded full-length GFP_{UV}. No full-length GFP_{UV} was expressed when **5** was excluded from the medium (**Figure 9**). As what we expected, the evolved mKRS1-pylT pair could also incorporate **6** at an amber codon position. When **6** instead of **5** was provided in the medium, full-length GFP_{UV} was also expressed (**Figure 2**). To our surprise, the evolved mKRS1 can also charges pylT with **7** and **8**. When **7** or **8** was provided in the medium, growing cells transformed with pBK-mKRS1 and pET-pylT-GFP also afforded full-length GFP_{UV} (**Figure 9**). We also tested the efficiency of the mKRS1-pylT pair to suppress an amber mutation at K7 of Z domain. A plasmid pET-pylT-Z was constructed by replacing the GFP_{UV} gene in pET-pylT-GFP with Z domain that contains an amber mutation at K7. Cotransforming *E. coli* BL21(DE3) cells with pBK-mKRS1 and pET-pylT-Z and then growing cells in the medium supplemented with either of **5**, **6**, **7**, and **8** afforded full-length Z domain. A trace amount of Z domain was expressed when no NCAA was provided in the medium (**Figure 9**). To prove the incorporation of NCAAs, purified Z domain proteins were then analyzed by electrospray ionization mass spectrometry (ESI-MS) (**Figure 10-14**). Before analysis, Z-domain containing **7** (**Z-7**) and Z domain containing **8** (**Z-8**) were photolysed under 365 nm UV light for an hour. As demonstrated previously, this treatment should efficiently cleave the photocaging group.[89] The ESI-MS analysis confirmed the expected mass for all four

Z domain proteins (**Table 3**). For both **Z-7** and **Z-8**, the photocaging group was efficiently cleaved off. Only peaks corresponding to Z domain that contains lysine or N^{ε}-methyl-L-lysine at K7 could be detected. The ESI-MS analysis of the deprotected **Z-8** also revealed two additional peaks at 8005 Da and 8136 Da.These two peaks match mass of Z domain that contains lysine, glutamate or glutamine at K7. Since mKRS1 was evolved against endogenous native amino acids and these two additional peaks are significant in the ESI-MS spectrum, direct incorporation of lysine, glutamate, or glutamine at K7 is not likely. We thought the additional mass peak was due to the incorporation of **7** that happened to be a contaminant in **8**. This is highly possible since we synthesized **8** from lysine. A very small amount of **7** might end up in **8**. We carefully examined all the NMR and MS data of **8** and concluded the contaminant was lower than 1% and insignificant. However, **7** could become significant after its incorporation into Z domain because the evolved mKRS1 has a higher affinity to **7** than **8**. Similarly, two additional mass peaks at 8139 Da and 8181 Da corresponding to Z domain containing **5** (**Z-5**) was also present in the ESI-MS spectrum of Z domain incorporated with **6** (**Z-6**). These two peaks are clearly the mass peaks of **Z-5**. This indicates the existance of **5** as a contaminant in **6**, though the contaminant was also lower than 1%. To eliminate **7** from **8**, we synthesized **8** from commercially available N^{ε}-methyl-L-lysine following a route presented in **Scheme II-2**. The purified **8** was then used to express GFP_{UV} containing **8**

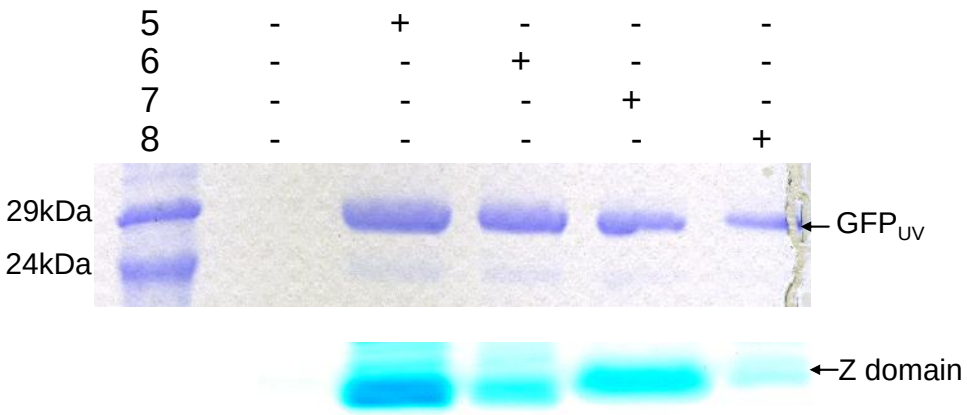


Figure 9. The expression of GFP_{UV} and Z domain at different conditions

Table 3. Z domain expression yields and MS characterization

Proteins	Yield[1] (mg/L)	Calculated Mass (Da)	Detected Mass (Da)
Z-5	9.3	8270[2]	8270
		8181[2]	8180
		8139[4]	8138
Z-6	4.0	8284[2]	
		8195[3]	8195
		8153[4]	8153
Z-7[5]	5.3	8136[2]	8136
		8047[3]	8047
		8005[4]	8004
Z-8[5]	1.2	8150[2]	8150
		8061[3]	8061
		8019[4]	8019

[1]Proteins were expressed in minimal media supplemented with 1% glycerol and 1 mM NCAA. [2]Full-legnth Z domain proteins. [3]Full-length Z domain without N-terminal methionine but with an N-terminal acetylation. [4]Full-length protein without N-terminal methionine. [5]Both proteins were deprotected by UV irradiation.

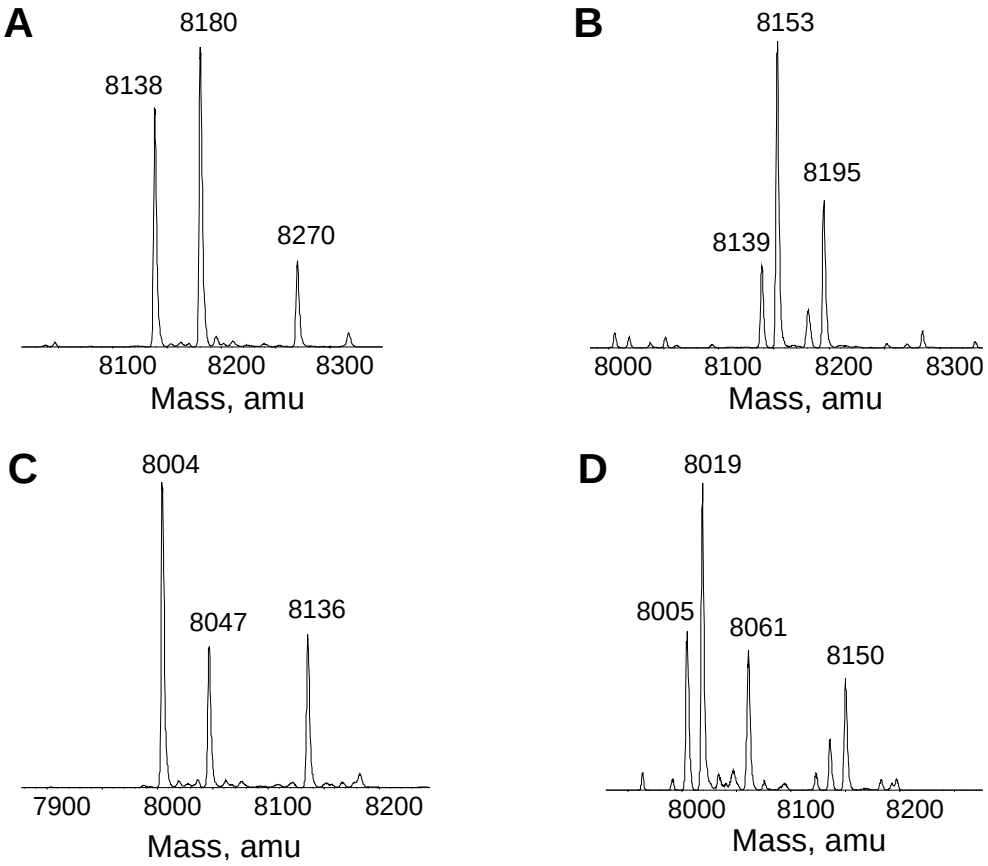


Figure 10. Deconvoluted ESI-MS of **Z-5, Z-6**, UV deprotected **Z-7**, and UV deprotected **Z-8**.

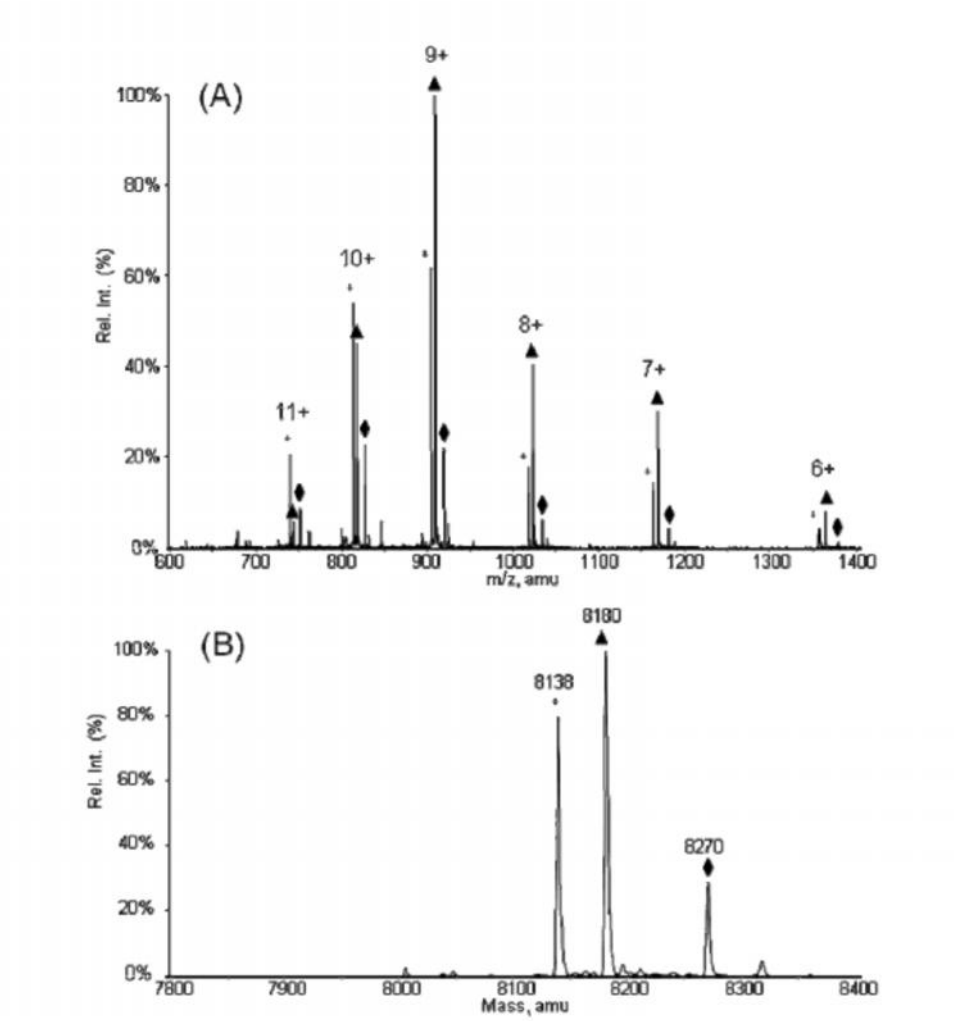


Figure 11. Mass determination of Protein **Z-5** (A) ESI-MS spectrum of **Z-5** (B). The deconvoluted ESI-MS spectrum of **Z-5**.

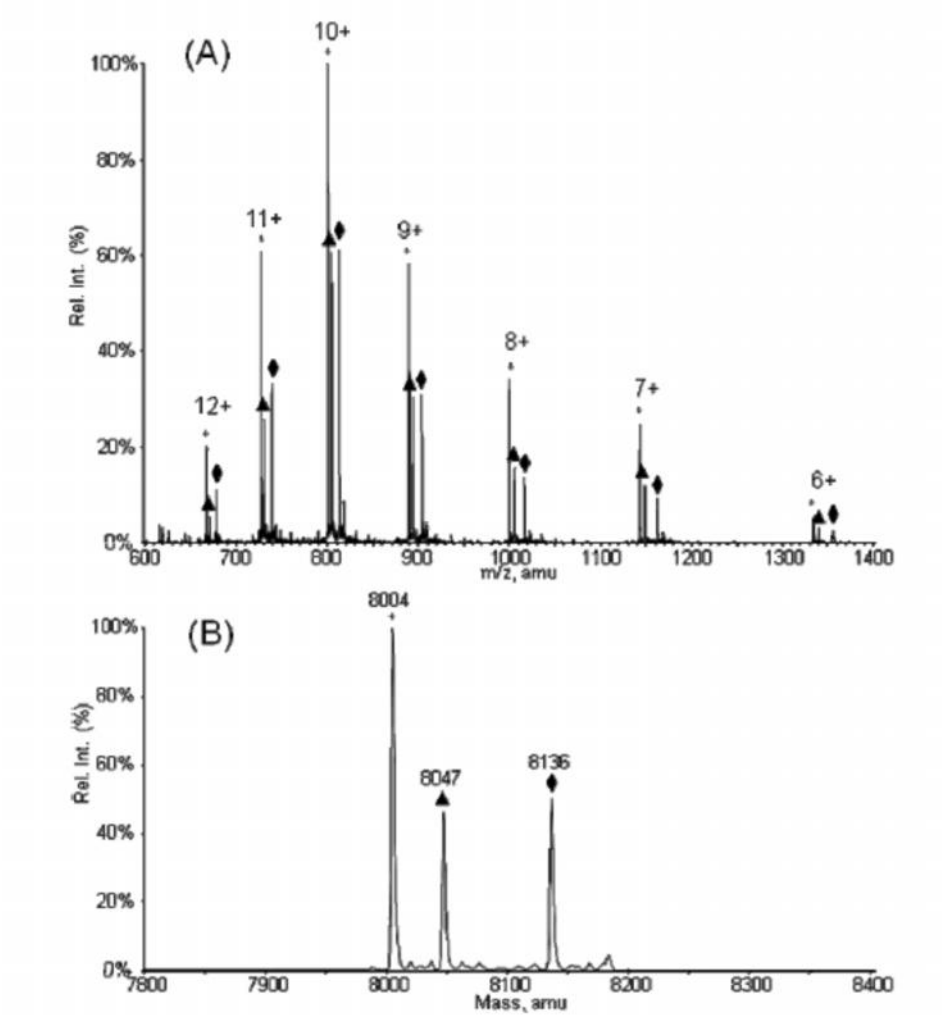


Figure 12: Mass determination of Protein **Z-6** (A) ESI-MS spectrum of **Z-6** (B). The deconvoluted ESI-MS spectrum of **Z-6**.

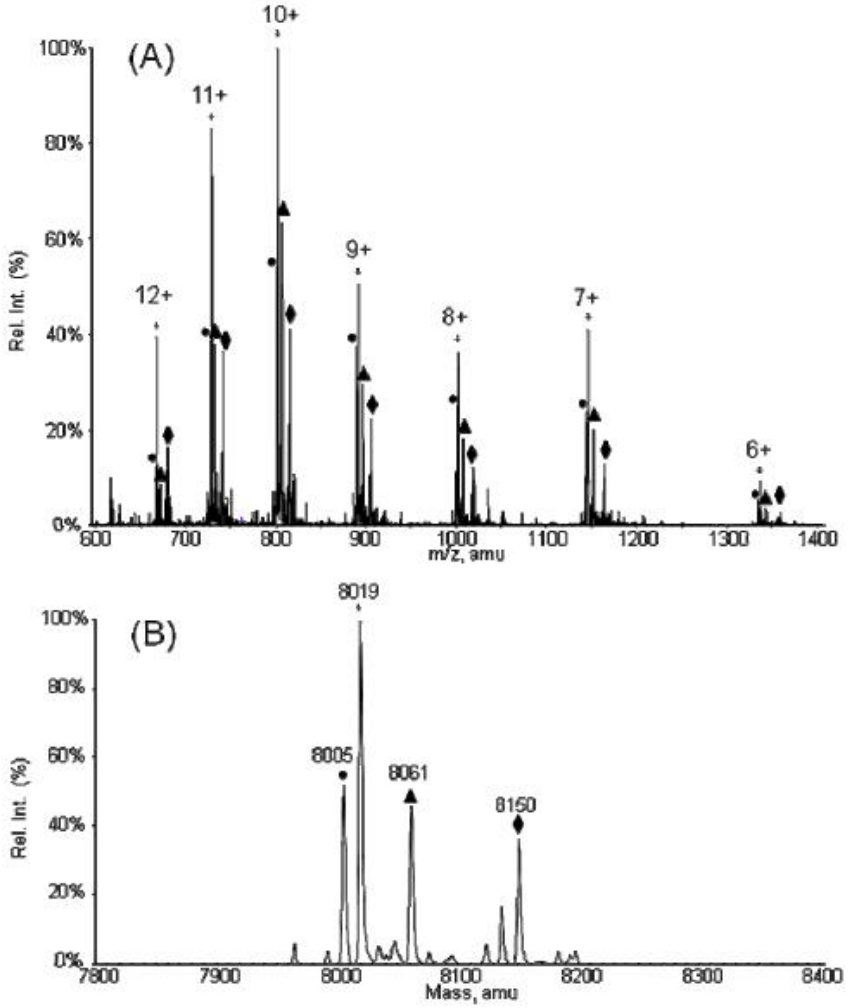


Figure 13. Mass determination of Protein **Z-7** (A) ESI-MS spectrum of **Z-7** (B). The deconvoluted ESI-MS spectrum of **Z-7**.

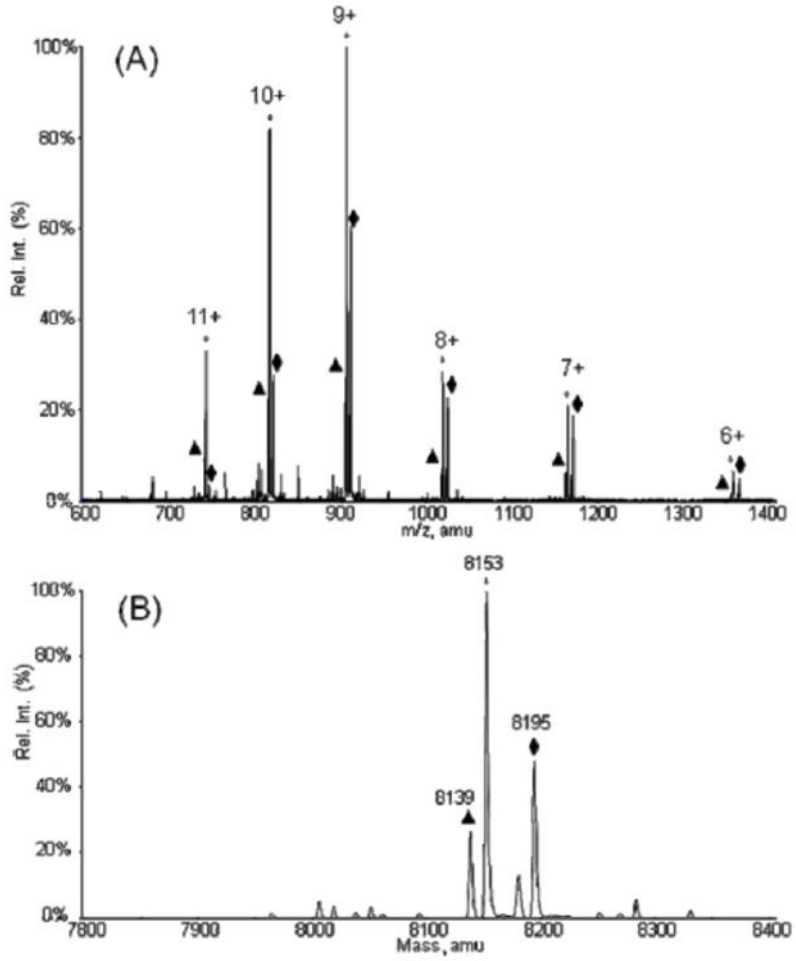


Figure 14. Mass determination of Protein **Z-8** (A) ESI-MS spectrum of **Z-8** (B). The deconvoluted ESI-MS spectrum of **Z-8**.

at Q204 (**GFP-8**). The purified **GFP-8** before and after 1 hour photolysis was then analyzed by ESI-MS (**Figure 15, 16**). The spectrum of **GFP-8** before photolysis showed one major peak at 27904 Da corresponding to **GFP-8** without N-terminal methionine (calculated mass: 27903 Da). No mass peak corresponding to GFP_{UV} incorporated with **7** (**GFP-7**) was identified. Similarly, the spectrum of **GFP-8** after photolysis showed one major peak at 27725 Da. It matches the calculated mass (27724 Da) of GFP_{UV} incorporated with N^{ε}-methyl-L-lysine at Q204. No mass peak corresponding to GFP_{UV} incorporated with lysine at Q204 was present.

From the ESI-MS spectral data of deprotected **Z-7**, **Z-8**, and **GFP-8** (**Figure 13, 14, 16)**, it is clear that deprotecting the photocaging group under UV light is very efficient. No additional treatment is necessary. We have also tried palladium black catalyzed hydrogenation to deprotect **Z-6**. However, the protein either aggregated or did not show any detectable deprotection. We are currently searching for homogenous hydrogenation catalysts that can efficient deprotect Cbz group from **Z-6**.

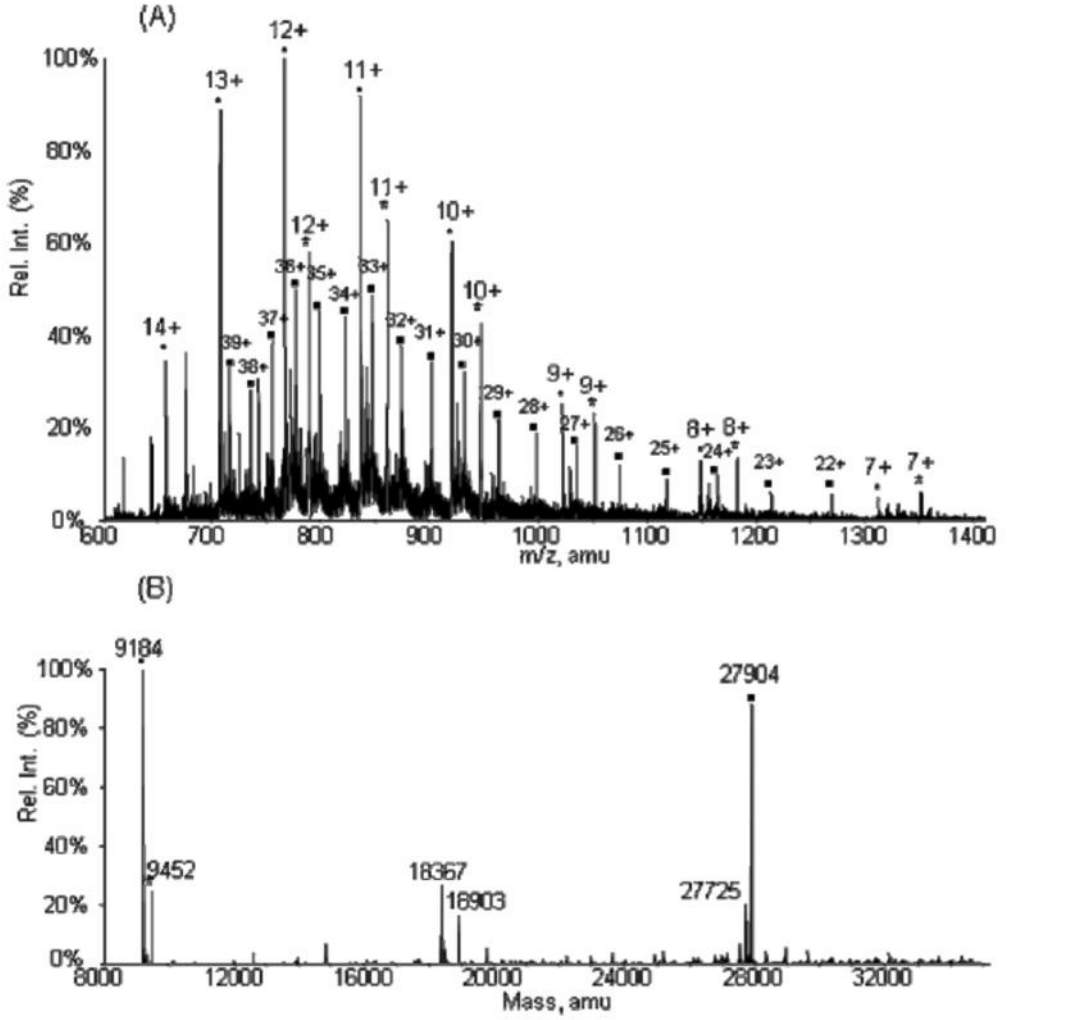


Figure 15. Mass determination of Protein GFP-**8** (A) ESI-MS spectrum of GFP-**8** (B). The deconvoluted ESI-MS spectrum of GFP-**8**.

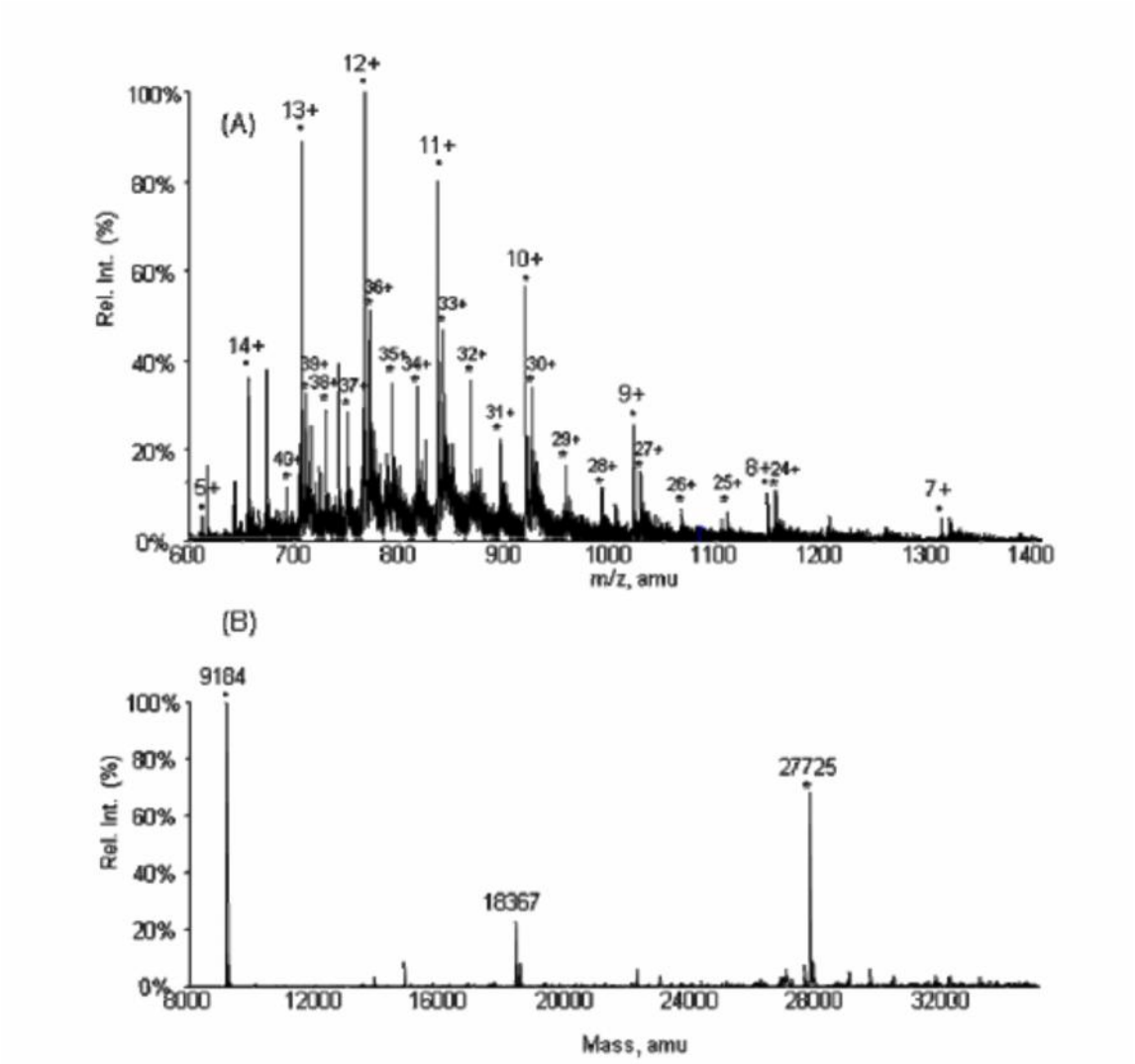


Figure 16. Mass determination of Protein GFP-**8** after photolysis. (A) ESI-MS spectrum of GFP-**8** (B). The deconvoluted ESI-MS spectrum of GFP-**8**.

Conclusion

In summary, we have demonstrated the genetic incorporation of a photocaged N^{ϵ}-methyl-L-lysine into proteins in *E. coli*. Since deprotecting the photocaging group to recover N^{ϵ}-methyl-L-lysine only requires UV irradiation, this method is suitable to directly synthesize many proteins with monomethylated lysines. Given the fact that protein lysine methylation has a fundamental role in regulating functions of chromatin and many transcription factors, a broad application of this developed method is anticipated. Since wild type and evovled PylRS-pylT pairs have been directly used to incorporate NCAAs into proteins in mammalian cells,[74,89] the evolved mKRS1-pylT pair might also be applied to synthesize proteins with monomethylated lysines directly in mammalian cells. This will allow the functional analysis of lysine monomethylations directly *in vivo.*

CHAPTER III*

CATALYST-FREE AND SITE-SPECIFIC ONE-POT DUAL-LABELING OF A PROTEIN DIRECTED BY TWO GENETICALLY INCORPORATED NONCANONICAL AMINO ACIDS

Introduction

Förster resonance energy transfer (FRET) between a pair of donor and acceptor dyes is an invaluable tool to study dynamic protein conformational changes such as conformation rearrangement and folding/unfolding.[92-94] The efficiency of energy transfer that is dependent on the distance between the two dyes not only represents the conformational distributions and also reflects their change upon time. Two methods are usually applied to achieve dual labeling of a protein with a FRET pair. One is to genetically fuse two green fluorescent protein (GFP) variants at the N- and C-termini of a protein.[95,96] The other is to modify two cysteine residues in a protein with a small-molecule FRET pair.[93] The GFP labeling approach has important advantages such as high labeling specificity and simplicity.[97,98] However, it has its intrinsic limitations. Labeling a protein with two GFP variants is, in general, restricted at two termini. The size of GFP (~27 kDa) is also large enough to potentially interfere with the structures

*Reprinted with permission from "Catalyst-free and site-specific one-pot dual-labeling of a protein directed by two genetically incorporated noncanonical amino acids" by Wu, B.; Wang, Z. Y.; Huang, Y.; Liu, W. S. R. *Chembiochem* **2012**, *13*, 1405, Copyright 2012 by John Wiley & Sons, Inc..

and functions of proteins to which they are fused. The cysteine labeling approach resolves issues such as site and size restrictions associated with the GFP labeling approach. It also has advantages such as the flexibility in choosing small molecule fluorophores and achieving labeling at the single-residue level. The straightforward labeling reactions and the commercial availability of many thiol-reactive dyes have made the cysteine labeling approach a favourable choice especially for the single-molecule FRET analysis.[99,100] However, the cysteine labeling approach also has its drawbacks. It requires mutating all non-targeted cysteine residues, therefore not applicable for proteins in which cysteine residues serve as critical structural and/or functional roles. In addition, dual labeling of two cysteine residues, in general, leads to heterogeneously labeled proteins due to the lack of selectivity between two cysteine residues. To achieve site-selective labeling of a protein with a small-molecule FRET pair, several other methods have been developed, including total chemical synthesis of a modified protein, specifically modifying a protein N-terminus and a cysteine residue, the use of the protein-protein interaction to protect a cysteine to achieve sequential labeling of two cysteine residues, and the combination of chemical and enzymatic modifications.[101-104] These methods either work for specific proteins or are complicated so that their general applications are difficult. An alternative dual labeling approach was developed by Schultz *et al.*, in which a cysteine residue and a noncanonical amino acid (NCAA), *p*-acetyl-L-phenylalanine coded by an amber codon were used to site-selectively label a protein.[105] Although elegant, this approach still suffers limitations

such as the requirement to mutate non-targeted cysteine residues and is not applicable to many mammalian proteins.

An ideal dual labeling approach that resolves issue related to methods discussed above is to install two different bio-orthogonal and chemically reactive groups into a protein followed by their selective reactions with corresponding dyes. One way to achieve this goal is to incorporate two different NNAs into a target protein. Recently Chin and our groups independently developed two methods for the genetic incorporation of two different NCAAs into one protein.[106,107] Our method relies on the suppression of two stop codons, namely amber UAG codon and ochre UAA codon, which is achieved by genetically encoding two orthogonal pairs, an evolved M. jannaschii tyrosyl-tRNA synthetase (MjTyrRS)-$tRNA^{Tyr}_{CUA}$ pair[41,70,108,109] and a wild type or evolved pyrrolysyl-tRNA synthetase (PylRS)- $tRNA^{Pyl}_{UUA}$ pair.[38,76,91,110,111] Herein, we show that this double NCAA incorporation method can be applied to genetically install two bio-orthogonal functional groups into one protein, allowing the catalyst-free and site-specific one-pot labeling of the protein with a FRET pair.

Experimental Section

General Experimental

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was performed on Whatman SiO_2 60 F-254 plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-

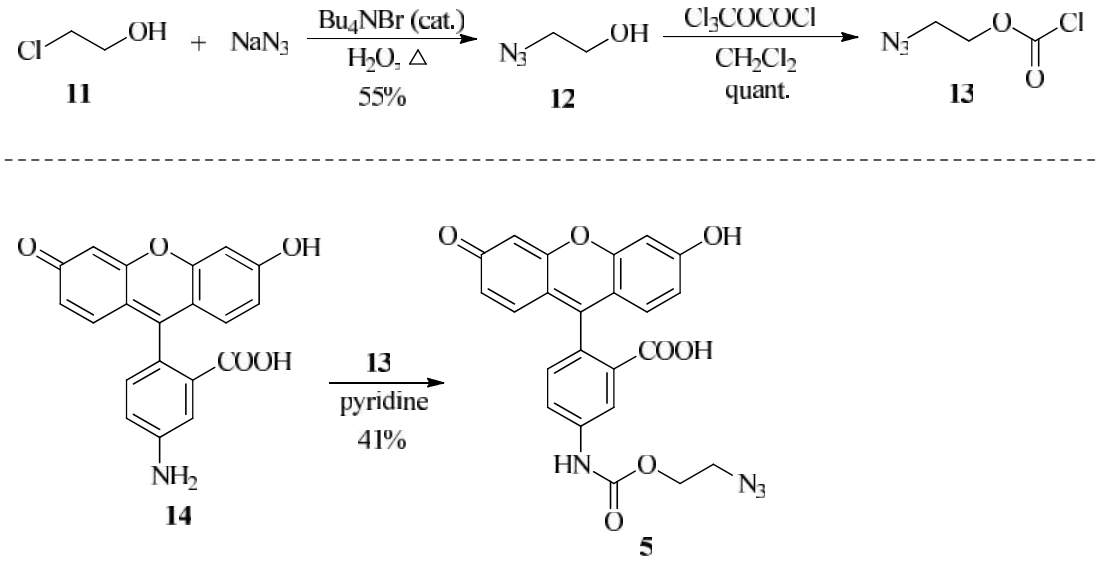
butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Specific rotations of chiral compounds were obtained at the designated concentration and temperature on a Rudolph Research Analytical Autopol II polarimeter using a 0.5 dm cell. Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C), methanol (3.31 ppm for ^{1}H and 49.15 ppm for ^{13}C), DMSO (2.50 ppm for ^{1}H and 39.51 ppm for ^{13}C), or water (4.80 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, and a minimal amount of sodium hydroxide pellet or concentrated hydrochloric acid was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral conditions. ^{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 obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

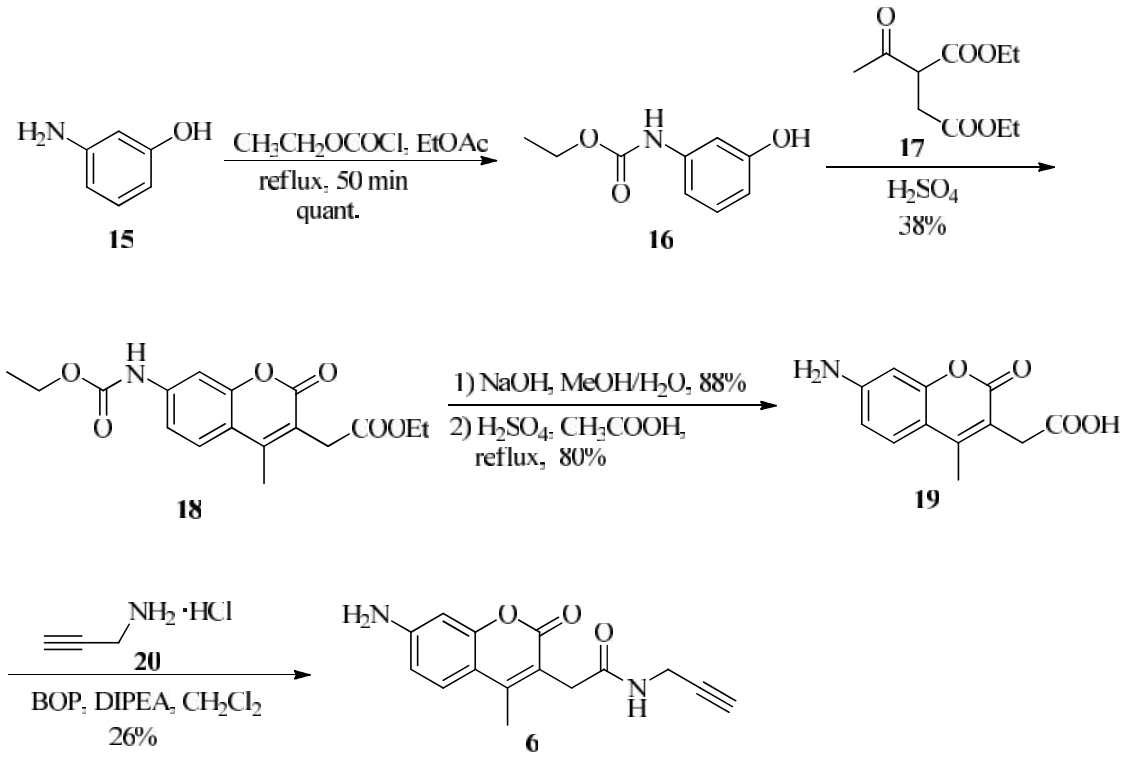
Compounds **1**,[112] **2** and **3**,[23] **4**,[52] and **7** and **8**[23] were prepared according to literature procedures as previously described. Compound **9** was purchased from Click Chemistry Tools (Scottsdale, AZ). All other reagents were obtained from commercial suppliers and used as received.

Chemical Synthesis

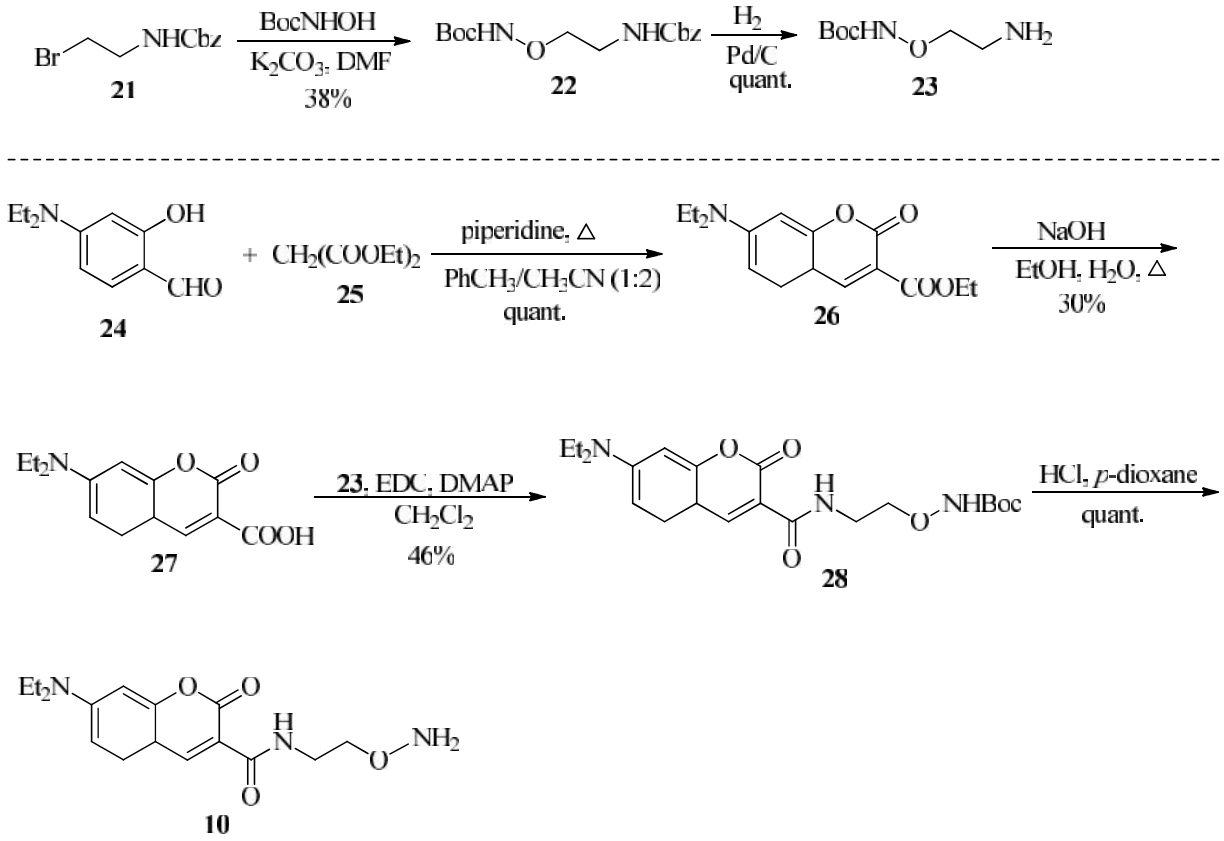
Compounds **5** was synthesized from fluorescein amine **14** through carbamate formation (**Scheme III-1**). Compounds **6** was synthesized by amide coupling between and **19** and **20** (**Scheme III-2**). Compound **10** was similarly obtained by coupling between **27** and **23** followed by deprotection (**Scheme III-3**).



Scheme III-1. Synthesis of **5**.



Scheme III-2. Synthesis of **6**.



Scheme III-3. Synthesis of **10**.

2-Azidoethanol (12)[113]

To a solution of 2-chloroethanol (**11**, 25.2 g, 0.31 mol) in water (80 mL) was added sodium azide (26.3 g, 0.40 mol) and tetrabutylammonium bromide (2.0 g, 6.2 mmol), and the mixture was stirred at room temperature for 2 h before being heated at 120 ^{o}C for 2 h. Sodium chloride was added to the cooled yellow solution until saturation, and the mixture was extracted with ethyl acetate (60 mL x 3). The combined organics were dried ($MgSO_4$), filtered, and evaporated to give a crude yellow oil (31 g). Distillation under an oil pump-generated vacuum (~0.1 mmHg, bp 87-90 ^{o}C) afforded **12**

(14.9 g, 55%) as a colorless oil. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 3.80-3.78 (m, 2 H), 3.46 (t, 2 H, *J* = 5.0 Hz), 1.88 (bs, 1 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 61.7, 53.7.

2-Azidoethyl chloroformate (13)

To a solution of **12** (82 mg, 0.94 mmol) in anhydrous dichloromethane (0.5 mL) was added trichloromethyl chloroformate (0.13 mmol, 1.1 mmol), and the mixture was stirred at room temperature for 12 h. The volatiles were evaporated to leave crude **13** (0.14 g, quant.) as a yellow oil which was used without further purification. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 4.43 (t, 2 H, *J* = 5.2 Hz), 3.63 (t, 2 H, *J* = 5.0 Hz).

5-((2'-Azidoethyl)oxycarbonylamino)fluorescein (5)

To a solution of fluorescein amine isomer I (Aldrich, 0.20 g, 0.58 mmol) in pyridine (2.0 mL, 24.7 mmol) cooled in an ice/water bath was added **13** (0.14 g, 0.94 mmol) dropwise over 5 min with the aid of a small amount of anhydrous dichloromethane (0.1 mL). The mixture was then stirred at room temperature for 48 h, and water (5 mL) was added. The mixture was diluted in ethyl acetate (50 mL), washed with water (10 mL), hydrochloric acid (0.1 *N*, 10 mL) and brine (10 mL), dried (Na_2SO_4), evaporated, chromatographed (EtOAc/hexanes, 1:1), and crystallized in ethyl acetate/hexanes to give **5** (0.11 g, 41%) as an orange solid. ^{1}H NMR (DMSO-d_6, 500 MHz) δ 10.32 (s, 1 H), 10.12 (s, 2 H), 8.12 (s, 1 H), 7.78 (dd, 1 H, *J* = 8.5, 1.5 Hz), 7.20 (d, 1 H, *J* = 8.5 Hz), 6.66 (d, 2 H, *J* = 2.0 Hz), 6.59 (d, 2 H, *J* = 8.5 Hz), 6.54 (dd, 2 H, *J* = 8.7, 2.2 Hz), 4.31 (t, 2 H, *J* = 4.7 Hz), 3.65 (t, 2 H, *J* = 5.2 Hz); ^{13}C NMR (DMSO-d_6, 75 MHz) δ 168.6, 159.5, 153.3, 151.9, 146.2, 140.7, 129.1, 127.1, 125.7, 124.6, 112.6,

112.4, 109.7, 102.2, 83.3, 63.1, 49.7; HRMS (ESI) calcd for $C_{23}H_{17}N_4O_7$ ($[M+H]^+$) 461.1097, found 461.1094.

3-*N*-(Carbethoxy) aminophenol (16)[114]

To a 500 mL round-bottomed flask fitted with a condenser, an addition funnel with pressure-equilibrating side arm, and a magnetic stir bar was charged 3-aminophenol (**15**, 50.0 g, 0.45 mol) and ethyl acetate (170 mL), which was dried over $MgSO_4$ and filtered prior to use. The mixture was heated at reflux to give a grey solution, and ethyl chloroformate (22.2 mL, 0.22 mol) was added dropwise over 50 min. The mixture was cooled to room temperature, filtered, washed with ethyl acetate (100 mL x 3) and hexanes (100 mL x 3), and dried in vacuo to give the recovered hydrochloride salt of 3-aminophenol (32.9 g, quant.) as an off-white powder. The combined filtrate was evaporated to remove most of the solvents, and hexanes (50 mL) was added. The mixture was frozen at -20 ^oC for 1 h, crushed, filtered, washed with hexanes (100 mL), and dried in vacuo to give pure **16** (37.0 g) as an off-white solid. The filtrate was again concentrated, mixed with hexanes (50 mL) and frozen, and filtered and treated as above to give more **16** (4.5 g, quant.) as a brown solid (slightly impure). ^1H NMR ($CDCl_3$, 500 MHz) δ 7.42 (s, 1 H), 7.14 (t, 1 H, J = 8.0 Hz), 6.96 (bs, 1 H), 6.78 (s, 1 H), 6.63 (d, 1 H, J = 8.0 Hz), 6.59 (dd, 1 H, J = 9.0, 2.5 Hz), 4.24 (q, 2 H, J = 7.0 Hz), 1.32 (t, 3 H, J = 7.0 Hz); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 157.2, 154.3, 139.0, 130.1, 110.9, 110.5, 106.1, 61.9, 14.6.

Ethyl 7-carbethoxyamido-4-methylcoumarin-3-acetate (18)[115]

To a brown solution of **16** (4.56 g, 25.2 mmol) in 71% sulfuric acid (18 mL) was added diethyl acetylsuccinate (**17**, 5.7 mL, 27.6 mmol) dropwise through an addition funnel over 30 min. After 4 h stirring at room temperature, the mixture was poured into a mixture of ice and water (~100 g). Upon stirring and trituration the initial white gel solidified, which was filtered and dried to give an off-white solid (6.6 g). The material was suspended in sodium hydroxide (1.0 *N*, 100 mL) and extracted with dichloromethane (120 mL x 3). The combined organics were washed with sodium hydroxide (0.5 *N*, 50 mL x 2) and brine (50 mL), dried ($MgSO_4$), evaporated, suspended in ether (50 mL), filtered, and dried in vacuo to give **18** (3.2 g, 38%) as a white solid. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.52 (d, 1 H, *J* = 9.0 Hz), 7.41 (d, 1 H, *J* = 2.0 Hz), 7.34 (d, 1 H, *J* = 8.5 Hz), 7.04 (s, 1 H), 4.26 (q, 2 H, *J* = 7.2 Hz), 4.20 (q, 2 H, *J* = 7.2 Hz), 3.72 (s, 2 H), 2.38 (s, 3 H), 1.33 (t, 3 H, *J* = 7.2 Hz), 1.28 (t, 3 H, *J* = 7.0 Hz); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 170.7, 161.9, 153.4, 153.2, 149.0, 141.3, 125.7, 117.8, 115.8, 114.6, 105.8, 61.9, 61.5, 33.2, 15.5, 14.7, 14.4; HRMS (ESI) calcd for $C_{17}H_{19}NO_6Li$ ($[M+Li]^+$) 340.1372, found 340.1373.

7-Carbethoxyamido-4-methylcoumarin-3-acetic acid (19)[115]

To a suspension of **18** (1.37 g, 4.1 mmol) in a mixed solvent of methanol (20 mL) and water (20 mL) was added sodium hydroxide (1.0 g, 25.0 mmol), and the mixture was stirred at room temperature for 36 h to give a milky solution. Most of the methanol was evaporated, and the remaining solution was diluted in water (10 mL) and extracted with ethyl acetate (30 mL). The separated aqueous phase was adjusted to pH 1

with concentrated hydrochloric acid, and the precipitate was filtered, washed with water (50 mL) and dichloromethane (50 mL), and dried to give a 7:1 mixture of the corresponding acid of **18** with **19** (1.11 g, 88%) as a pink solid. ^{1}H NMR (DMSO-d_6, 500 MHz) δ 10.17 (s, 1 H, minor product), 10.13 (s, 1 H), 7.72 (d, 1 H, *J* = 8.0 Hz), 7.54 (s, 1 H), 7.39 (d, 1 H, *J* = 8.5 Hz), 7.44 (d, 1 H, *J* = 8.5 Hz, minor product), 6.57 (d, 1 H, *J* = 7.5 Hz, minor product), 6.41 (s, 1 H, minor product), 6.05 (s, 2 H, minor product), 4.16 (q, 2 H, *J* = 6.8 Hz), 3.70 (s, 2 H, minor product), 3.56 (s, 2 H), 2.34 (s, 3 H), 2.25 (s, 3 H, minor product), 1.25 (t, 3 H, *J* = 7.0 Hz); ^{13}C NMR (DMSO-d_6, 125 MHz) δ 171.7, 161.0, 153.5, 152.6, 149.0, 142.4, 126.3, 117.4, 114.6, 114.5, 104.3, 60.8, 32.8, 15.0, 14.5.

The above crude mixture of products (1.11 g, 3.6 mmol) was dissolved in conc. sulfuric acid (1.5 mL) and acetic acid (1.5 mL) in a 100 mL flask and heated at reflux for 3 h to give a dark solution. Upon cooling down to room temperature ice/water (15 g) was added followed by another 25 mL of water. Charcoal (0.5 g) and Celite 545 (0.5 g) were then added, and the mixture was heated at reflux for 10 min. The hot mixture was carefully filtered to give a red filtrate, which was evaporated to a volume of ~15 mL. The solution was allowed to stand at room temperature overnight, and further cooling at -20 ^{o}C did not yield more precipitate. The mixture was filtered, washed with cold water (4 mL) and cold ethanol (10 mL), and dried to give **19** (0.71 g, 80%) as a brown solid. ^{1}H NMR (DMSO-d_6, 500 MHz) δ 7.50 (d, 1 H, *J* = 9.0 Hz), 6.65 (dd, 1 H, *J* = 8.7, 1.8 Hz), 6.51 (d, 1 H, *J* = 2.0 Hz), 3.50 (s, 2 H), 2.27 (s, 3 H); ^{13}C NMR (DMSO-d_6, 125 MHz) δ 172.0, 161.5, 153.9, 150.8, 149.7, 126.6, 113.6, 112.3, 110.2, 99.8, 32.6, 14.9.

2-(7-Amino-4-methyl-2-oxo-2*H*-chromen-3-yl)-*N*-(prop-2-yn-1-yl)acetamide (6)

To a solution of **19** (10 mg, 43 μmol), (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate (BOP, 26 mg, 62 μmol), *N*,*N*-diisopropylethylamine (20 μL, 0.12 mmol) in anhydrous DMF (0.2 mL) was added propargylamine hydrochloride (**20**, 7.5 mg, 78 μmol), and the mixture was stirred at room temperature for 14 h. The mixture was diluted in ethyl acetate (40 mL), washed with sodium hydroxide (0.5 *N*, 10 mL), hydrochloric acid (0.5 *N*, 10 mL) and brine (10 mL), dried (Na_2SO_4), evaporated, and chromatographed (EtOAc/hexanes, 2:1 followed by 5% MeOH in CH_2Cl_2) to give **6** (3.0 mg, 26%) as a yellow solid. ^{1}H NMR (DMSO-d_6, 500 MHz) δ 8.31 (t, 1 H, *J* = 5.7 Hz), 7.45 (d, 1 H, *J* = 8.5 Hz), 6.57 (dd, 1 H, *J* = 8.7, 2.2 Hz), 6.40 (d, 1 H, *J* = 2.0 Hz), 6.04 (s, 2 H), 3.83 (dd, 2 H, *J* = 5.5, 2.5 Hz), 3.40 (s, 2 H), 3.09 (s, 2 H), 2.24 (s, 3 H); ^{13}C NMR (DMSO-d_6, 75 MHz) δ 169.1, 161.6, 154.1, 152.4, 150.0, 126.3, 113.1, 111.3, 109.4, 99.4, 81.3, 72.8, 33.5, 28.0, 14.9; HRMS (ESI) calcd for $C_{15}H_{14}N_2O_3Na$ ($[M+Na]^+$) 293.0902, found 293.0914.

***N*-Boc-*O*-(2-(benzyloxycarbonylamino)ethyl)hydroxylamine (22)**

To a solution of **21**[116] (0.30 g, 1.2 mmol) and *N*-Boc-hydroxylamine (0.16 g, 1.2 mmol) in anhydrous DMF (1.5 mL) was added potassium carbonate (0.44 g, 3.2 mmol), and mixture was stirred at room temperature for 24 h. Water (10 mL) was then added, and the mixture was extracted with ether (50 mL). The organic phase was washed with sodium hydroxide (0.5 *N*, 10 mL), hydrochloric acid (0.5 *N*, 10 mL) and brine (10 mL), dried (Na_2SO_4), evaporated, and chromatographed (EtOAc/hexanes, 1:5) to give **22** (0.14 g, 38%) as a colorless oil. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.47 (bs, 1 H), 7.38-7.29

(m, 5 H), 5.80 (bs, 1 H), 5.12 (s, 2 H), 3.86 (t, 2 H, J = 4.5 Hz), 3.44-3.43 (m, 2 H), 1.47 (s, 9 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 158.0, 157.4, 137.2, 129.1, 128.9, 128.6, 82.7, 76.2, 67.3, 39.7, 28.8; HRMS (ESI) calcd for $C_{15}H_{22}N_2O_5Na$ ($[M+Na]^+$) 333.1426, found 333.1418.

***tert*-Butyl 2-aminoethoxycarbamate (23)**

To a solution of **22** (0.33 g, 1.1 mmol) in methanol (20 mL) was added palladium on activated carbon (10% Pd, 0.11 g, 0.1 mmol), and the mixture was hydrogenated under a H_2 balloon at room temperature for 5 h. The mixture was filtered over Celite and evaporated to give **23** (0.20 g, quant.) as a grey oil. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 3.93 (t, 2 H, J = 5.0 Hz), 2.99 (t, 2 H, J = 5.2 Hz), 1.48 (s, 9 H), 1.46 (s, 2 H). The material was used without further purification.

Ethyl 7-(diethylamino)coumarin-3-carboxylate (26)[117]

To a solution of 4-(diethylamino)salicylaldehyde (**24**, 5.93 g, 30.1 mmol) and diethyl malonate (7.6 g, 47.0 mmol) in a mixed solvent of toluene and acetonitrile (1:2, 210 mL) was added piperidine (8.9 mL, 90.1 mmol), and the red solution was heated at reflux for 10 h. The solvent was evaporated under reduced pressure, and the residue was directly chromatographed (EtOAc/hexanes, 1:3) to give **26** (9.2 g, quant.) as a yellow oil. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 8.34 (s, 1 H), 7.29 (d, 1 H, J = 8.5 Hz), 6.54 (dd, 1 H, J = 8.7, 2.2 Hz), 6.35 (d, 1 H, J = 2.5 Hz), 4.29 (q, 2 H, J = 7.2 Hz), 3.37 (q, 2 H, J = 7.2 Hz), 1.32 (t, 3 H, J = 7.2 Hz), 1.16 (t, 6 H, J = 7.2 Hz); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 164.2, 158.4, 152.9, 149.2, 131.1, 109.6, 108.8, 107.6, 96.6, 61.1, 45.1, 14.4, 12.5.

7-(Diethylamino)coumarin-3-carboxylic acid (27)[117]

To a solution of **26** (3.34 g, 11.5 mmol) in ethanol (30 mL) was added sodium hydroxide (1.0 *N*, 20.0 mL, 20.0 mmol), and a yellow precipitate quickly formed. The mixture was heated at reflux for 3 h to give a clear red solution, which was cooled to room temperature and filtered. The filtrate was adjusted to pH 3 with hydrochloric acid (2 *N*, ~18 mL), filtered, washed with water (30 mL), ethanol (10 mL) and ether (30 mL), and dried to give **27** (0.90 g, 30%) as an orange solid. ^{1}H NMR (CD_3OD, 500 MHz) δ 8.62 (s, 1 H), 7.58 (d, 1 H, *J* = 9.0 Hz), 6.86 (dd, 1 H, *J* = 9.0, 2.5 Hz), 6.62 (d, 1 H, *J* = 2.5 Hz), 3.56 (q, 2 H, *J* = 7.0 Hz), 1.25 (t, 6 H, *J* = 7.2 Hz); HRMS (ESI) calcd for $C_{14}H_{16}NO_4$ ($[M+H]^+$) 262.1079, found 262.1073.

***tert*-Butyl 2-(7-(diethylamino)-2-oxo-4a,5-dihydro-2*H*-chromene-3-carboxamido)ethoxy-carbamate (28)**

To a solution of **27** (0.146 g, 0.56 mmol), 4-(dimethylamino)pyridine (44 mg, 0.36 mmol) and **23** (74 mg, 0.42 mmol) in anhydrous dichloromethane (2 mL) was added *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC hydrochloride, 0.136 g, 0.71 mmol), and the mixture was stirred at room temperature for 12 h. The mixture was diluted in ethyl acetate (50 mL), washed with sodium hydroxide (0.5 *N*, 10 mL), hydrochloric acid (0.5 *N*, 10 mL) and brine (10 mL), dried (Na_2SO_4), evaporated, and chromatographed (EtOAc/hexanes, 1:3 to 1:1) to give **28** (81 mg, 46%) as a yellow oil which solidified upon standing. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 9.06 (t, 1 H, *J* = 6.0 Hz), 8.67 (s, 1 H), 8.07 (s, 1 H), 7.40 (d, 1 H, *J* = 9.0 Hz), 6.63 (dd, 1 H, *J* = 8.8, 2.5 Hz), 6.48 (d, 1 H, *J* = 2.4 Hz), 3.96 (t, 2 H, *J* = 5.1 Hz), 3.70 (dt, 2 H, *J* = 5.4,

5.4 Hz), 3.44 (q, 4 H, *J* = 7.2 Hz), 1.47 (s, 9 H), 1.22 (t, 6 H, *J* = 7.0 Hz); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 164.2, 162.8, 157.8, 156.9, 152.6, 148.4, 131.3, 110.2, 108.6, 96.9, 81.5, 74.9, 45.3, 37.8, 28.4, 12.5; HRMS (ESI) calcd for $C_{21}H_{30}N_3O_6$ ($[M+H]^+$) 420.2135, found 420.2129; calcd for $C_{21}H_{29}N_3O_6Li$ ($[M+Li]^+$) 426.2216, found 426.2214; calcd for $C_{21}H_{29}N_3O_6Na$ ($[M+Na]^+$) 442.1954, found 442.2010.

_N_-(2-(Aminooxy)ethyl)-7-(diethylamino)-2-oxo-4a,5-dihydro-2_H_-chromene-3-carboxamide (10)

To a solution of **28** (62 mg, 0.15 mmol) in 1,4-dioxane (1.0 mL) was added hydrogen chloride in dioxane (4.0 M, 0.3 mL, 1.2 mmol), and the mixture was stirred at room temperature for 4 h. Water (10 mL) was added followed by saturated sodium bicarbonate (20 mL), and the mixture was extracted with chloroform (50 mL). The separated organic phase was washed with brine (10 mL), dried (Na_2SO_4), and evaporated to afford **10** (48 mg, quant.) as a yellow oil. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 9.03 (s, 1 H), 8.69 (s, 1 H), 7.42 (d, 1 H, *J* = 8.5 Hz), 6.63 (dd, 1 H, *J* = 9.2, 1.7 Hz), 6.48 (d, 1 H, *J* = 2.0 Hz), 3.88 (bs, 2 H), 3.69 (bs, 2 H), 3.44 (q, 4 H, *J* = 7.2 Hz), 1.23 (t, 6 H, *J* = 7.0 Hz); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 163.9, 162.9, 157.8, 152.7, 148.4, 131.3, 110.2, 110.1, 108.5, 96.7, 45.2, 38.5, 12.6; HRMS (ESI) calcd for $C_{16}H_{22}N_3O_6$ ($[M+H]^+$) 320.1610, found 320.1616.

Primer and Gene Sequences

Primer Sequences

Forward primer NP2 agcgcggccgcgtcgacggtaccctcgagtctggtaaag

Reverse primer NP2 attgcggccgcccatggtatatctccttcttatacttaac

F1-QBP3TAG tatacatatggcctaggattaaaaattagttgtcgc

R1-QBP141TAA cagttccatataggcttaatcgatgttcgggaa

F2-QBP141TAA ttcccgaagaatcgattaagcctatatggaactg

R2-QBP gagggtacctcagtgatggtgatggtgatgtttcggttcagtacc

PylRS->AcKRS acctgcgataccggtttccacccaag

PylRS->AcKRS R cttaagttacaggttggtagaaatccc

Genes Sequences

QBP2m

atggcctaggataaaaaattagttgtcgcgacggataccgccttcgttccgtttgaatttaaacagggcgataaatatgtgggcttt

gacgttgatctgtgggctgccatcgctaaagagctgaagctggattacgaactgaagccgatggatttcagtgggatcattccg

gcactgcaaaccaaaaacgtcgatctggcgctggcgggcattaccatcaccgacgagcgtaaaaaagcgatcgatttctctga

cggctactacaaaagcggcctgttagtgatggtgaaagctaacaataacgatgtgaaaagcgtgaaagatctcgacgggaaag

tggttgcggtgaagagcggcactggctccgttgattacgcgaaagcaaacatcaaaactaaagatctgcgtcagttcccgaaca

tcgattaagcctatatggaactgggcaccaaccgcgcagacgccgttctgcacgatacgccaaacattctgtacttcatcaaaac

cgccggtaacggtcagttcaaagcggtaggcgactctctggaagcgcagcaatacggtattgcgttcccgaaaggtagcgac

gagctgcgtgacaaagtcaacggcgcgttgaaaaccctgcgcgagaacggaacttacaacgaaatctacaaaaaatggttcg

gtactgaaccgaaacatcaccatcaccatcactga

MmAcKRS

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacac

cacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggact

gcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcac

aaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatcc

gttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgataccg

gtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactg

gtaaaagggaatacgaatcccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactgac

aggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctctc

tcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccag

gttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgat

accgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagacccatgatggctccaaacctgctgaactacgccc

gcaagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgctacagaaaagagtccgacggcaaagaa

cacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgtacacgggaaaatcttgaaagcataattacggact

tcctgaaccacctgggaattgatttcaagatcgtaggcgattccttcatggtctatggggatacccttgatgtaatgcacggagac

ctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttc

gggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggattt

ctaccaacctgtaa

MmPylRS

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacac

cacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggact

gcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcac

aaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatcc

gttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgataccg

gtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactg

gtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactga

caggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctct

ctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccag
gttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgat
accgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagacccatgcttgctccaaacctttacaactacctgcg
caagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgctacagaaaagagtccgacggcaaagaac
acctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaaatcttgaaagcataattacggactt
cctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctatggggataccttgatgtaatgcacggagac
ctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttc
gggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggattt
ctaccaacctgtaa

Construction of Plasmids

All the plasmid structures were confirmed by DNA sequencing. All oligonucleotide primers were purchased from Integrated DNA Technologies, Inc. All the PCR reactions were performed with Phusion® High-Fidelity DNA Polymerase from New England Biolabs Inc. (Ipswich, MA). All the restriction enzymes, T4 DNA ligase and T4 polynucleotide kinase (T4 PNK) were purchased from New England Biolabs Inc. (Ipswich, MA).

Construction of pETtrio-PylT-PylRS-MCS

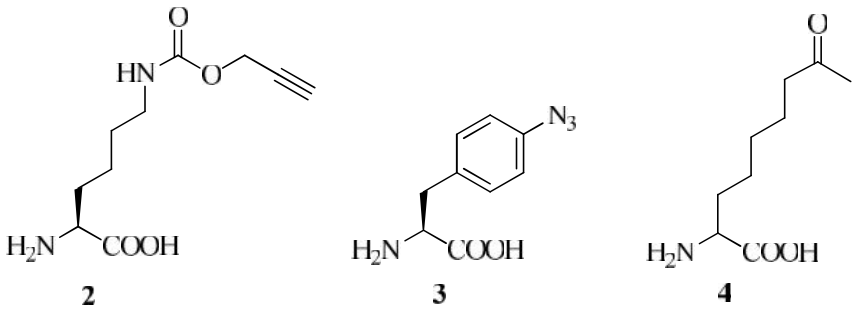
The plasmid pPylRS-PylT-MCS was derived from pPylRS-pylT-GFP1TAG149TAA.[23] Forward primer NP2 and reverse primer NP2 were used to clone the whole plasmid without the gene GFP1TAG149TAA. Meanwhile, four restriction sites, *Nco I, Not I, Sal I and Kpn I* were introduced to the positions where GFP1TAG149TAA was located in the original plasmid.

Construction of pETtrio-pylT-PylRS-QBP3TAG141TAA

F1-QBP3TAG and R1-QBP141TAA were used to clone the first part of site-mutated glutamine binding protein (QBP) from *E. coli* TOP 10 cell. The second part of QBP3TAG141TAA gene was cloned out in the same manner by using F2-QBP141TAA and R2-QBP as the two primers. Overlap PCR was performed with F1-QBP3TAG and R2-QBP as the two primers and the two fragments obtained from the PCR reactions mentioned above to afford QBP3TAG141TAA, which was inserted to pETtrio-PylT-PylRS-MCS with *Nco I* at 5' end and *Kpn I* at 3'end.

Construction of pETtrio-pylT-MmAcKRS-pylT-QBP3TAG141TAA

pKTS-MmAcKRS, the plasmid contains MmAcKRS that takes **4** efficiently evolved from MmPylRS was constructed by Prof. Dieter Söll's group (Yale University, New Haven, CT) and given to us as a gift.[118] To construct pETtrio-pylT-MmAcKRS-pylT-QBP3TAG141TAA, pETtrio-pylT-PylRS-QBP3TAG141TAA was used. Since MmAcKRS was derived from PylRS and all the mutations are beyond the 600 base pairs in the gene, PylRS->AcKRS F, a forward primer basing on the sequence of the PylRS at 450 bp was designed. Together with PylRS->AcKRS R, PylRS was converted to MmAcKRS with *Age I* at 5' end and *AflII* at 3' end.



Scheme III-4. Non-canonical amino acids used for QBP expression.

Expression of QBP(3+2)

E. coli BL21 cells co-transformed with pETtrio-pylT-MmAcKRS-pylT-QBP3TAG141TAA and pEVOL-AzFRS[118] were grown in 2TY medium (150 mL) with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol overnight. The culture was inoculated into 2YT medium (450 mL) with the same concentration of antibiotics. IPTG (500 mM), arabinose (0.2% w/v), together with **2** and **3** (both 1 mM) were added into the cell culture after the OD_{600} reached 1.2~1.4. The cell culture was incubated at 37 ^{o}C for 8 h, and the cells were harvested by centrifugation at 4000 r.p.m. for 20 min at 4 ^{o}C and re-suspended in 50 mL of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM imidazole, pH 7.8). The re-suspended cells were sonicated in an ice/water bath four

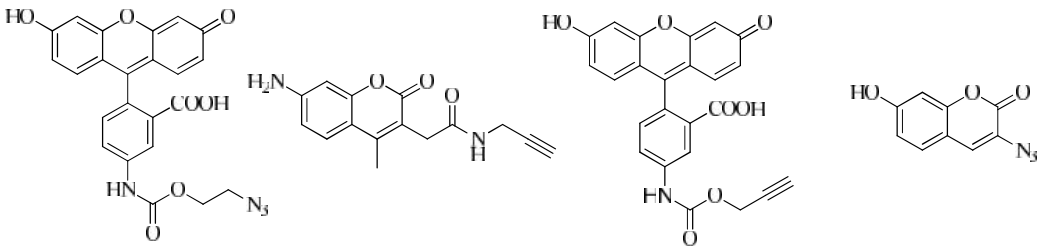
times (4 min each, 10 min interval to cool the suspension below 10 ºC before the next run) and the lysate was clarified by centrifugation at 10000 r.p.m. for 40 min at 4 °C. The supernatant was then incubated with 3 mL of Ni Sepharose^{TM} 6 Fast Flow from GE Healthcare (Little Chalfont, United Kingdom) for 1 h, and then washed with 100 mL of lysis buffer. QBP(**3+2**) was then eluted out with 12 mL of elution buffer (50 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 7.8) and concentrated by Amicon Ultra-15 Centrifugal Filter Units – 10,000 NMWL from Millipore (Billerica, MA) to 3 mL. The buffer was then changed to ammonium bicarbonate (ABC, 20 mM, pH 8.1) by dialysis. The concentration was determined by BCA protein assay kit from Thermo Fisher Scientific Inc. (Rockford, IL). According to the concentration, QBP(**3+2**) expression yield was 12 mg/L from the 2YT medium.

Expression of QBP(3+4)

E. coli BL21 cells were co-transformed with pETtrio-pylT-MmAcKRS-pylT-QBP3TAG141TAA and pEVOL-AzFRS[2] for the expression of QBP(**3+4**), which followed the same procedure of the expression of QBP(**3+2**) except **4** (2 mM) was supplemented with **2** (1 mM) instead of **3**. Purified QBP(**3+4**) was dialyzed against phosphate buffered saline (pH 6.4) for the following labeling reactions. The expression yield for QBP(**3+4**) was 11 mg/L from the 2YT medium.

Protein Labeling Procedures

Copper-Catalyzed Azide-Alkyne Cycloaddition (CuAAC)



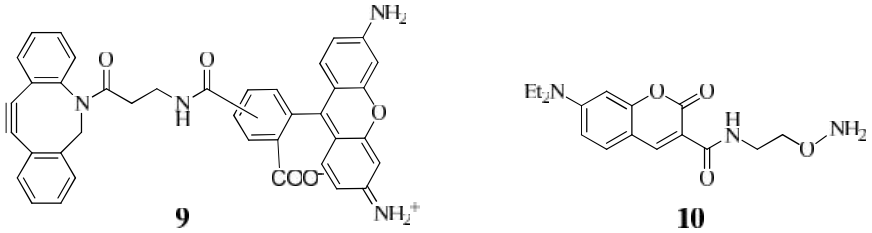
5 **6** **7** **8**

Scheme III-5. Dyes used for CuAAC.

To QBP(**3+2**) (concentration varied from 0.017 mM to 0.072 mM in 20 mM ABC buffer, 270 μL, pH 8.1) was added $CuSO_4$ (100 μM), $NiCl_2$ (1 mM), tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA, stock solution in DMSO, 500 μM) and one of the dyes (**5, 6, 7** and **8**, stock solutions in DMSO, 50 equiv. to the protein) sequentially, followed by sodium ascorbate (5 mM). The reaction was performed at room temperature for 3 h. Then ethylenediaminetetraacetic acid (EDTA, pH 8.0, 5 μL, 0.5 M final concentration) was added to the reaction mixture to chelate the two metals. The reaction product was transferred into lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM imidazole, pH 7.8, 10 mL) with Ni Sepharose™ 6 Fast Flow (1 mL) and

incubated at 4 ºC for 1 h. The resin was loaded onto an empty column and the catalysts were washed away by lysis buffer (100 mL). The labeled QBP(**3+2**) was eluted out by elution buffer (50 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 7.8, 6 mL), concentrated, dialyzed against ABC buffer (20 mM, pH 8.1) and then analyzed by mass spectrometry. The second CuAAC labeling was performed in the same manner with the appropriate dye to afford doubly labeled QBP(**3+2**).

Catalyst-free labeling



9 **10**

Scheme III-6. Dyes used for Catalyst-free labeling.

To QBP(**3+4**) (0.024 mM to 0.035 mM) was added **9** (50 equiv.) and **10** (10 equiv.). The reaction was performed at room temperature overnight. The reaction product was transferred into lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM imidazole, pH 7.8, 10 mL) with Ni Sepharose^{TM} 6 Fast Flow (1 mL) and incubated at 4

^{o}C for 1 h. The resin was loaded onto an empty column and the dye was washed away by lysis buffer (400 mL). The labeled QBP(**3+4**) was eluted out by elution buffer (50 mM HEPES, 500 mM NaCl, 250 mM imidazole, pH 7.8, 6 mL), concentrated, dialyzed against ABC buffer (20 mM, pH 8.1) and then analyzed by mass spectrometry. It is noteworthy that excessive lysis buffer must be used to completely remove **9**, which has very poor solubility in water.

Mass Spectrometry Analysis

Nanoelectrospray ionization in positive mode was performed using an Applied Biosystems QSTAR Pulsar (Concord, ON, Canada) equipped with a nanoelectrospray ion source. Solution was flowed at 700 nL/min through a 50 µm ID fused-silica capillary that was tapered at the tip. Electrospray needle voltage was held at 2100 V.

FRET Assay

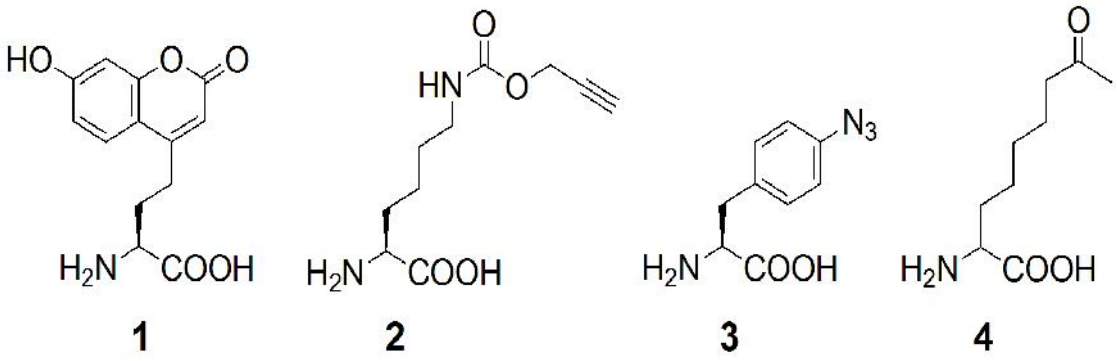
QBP(**3+2**)-**6**-**5** and QBP(**3+4**)-**9**-**10** were diluted with various concentrations of guanidine hydrochloride (GndCl, 0 M, 1 M, 2 M, 3 M, 4 M, 5 M and 6 M, respectively) in PBS buffer (pH 7.8). The fluorescent emission of those solutions was tested by QuantaMaster™ 40 Intensity Based Spectrofluorometer from Photon Technology International Inc. (Birmingham, NJ) with excitation at 350 nm (QBP(**3**+**2**)) or 430 nm (QBP(**3**+**4**)). Emission change based on the concentration of GndCl was plotted. All measurements were taken on freshly made samples and the data was collected every 0.2 second with 0.5 nm intervals.

Result and Discussion

We initially attempted to incorporate one fluorescent NCAA, L-(7-hydroxycoumarin-4yl) ethylglycine (**1** in **Scheme III-7**)[119] and *N*$^{\varepsilon}$-propargyloxycarbonyl-L-lysine (**2** in **Scheme III-7**)[72] into glutamine binding protein (QBP) at 3 and 141 positions respectively followed by the copper(I)-catalyzed azide-alkyne Huisgen cycloaddition (CuAAC) click reaction[120,121] on **2** to generate a dual-labeled QBP. **1** is blue fluorescent itself. Its incorporation into proteins at amber mutation sites using an evolved *Mj*TyRS (CouRS)-tRNA$^{Tyr}_{CUA}$ pair enables their quantitative fluorescent labeling. The incorporation of both **1** and **2** into QBP will only require one following labeling reaction to install a FRET pair, therefore simplifying the FRET installation procedure. To incorporate **1** and **2** into QBP, the QBP gene with an amber mutation at its amino acid position 3 and an ochre mutation at its amino acid position 141 was cloned into plasmid pETtrio-pylT-PylRS-MCS, which was derived from pPylRS-pylT-GFP1TAG149TAA[107] and carries genes coding the PylRS-tRNA$^{Pyl}_{UUA}$ pair, to afford pETtrio-pylT-PylRS-QBP3TAG141TAA. This plasmid was coupled together with pEVOL-CouRS[122] that carries genes coding the CouRS-tRNA$^{Tyr}_{CUA}$ pair to cotransform *E. coli* BL21 cells. However, growing the transformed cells in the presence of 1 mM **1** and 1 mM **2** resulted in negligible QBP expression. This low QBP expression level is probably due to the low incorporation efficiency of **1** mediated by CouRS.[119]

We next sought other NCAAs that can be coupled together with **2** to be incorporated into QBP for a FRET pair installation. Since *p*-acetyl-L-phenylalanine needs to react with hydroxylamine dyes at pH ≤ 4, a condition that denatures most

proteins,[105] we focused on *p*-azido-L-phenylalanine (**3** in **Scheme III-7**)[123] that can undergo the CuAAC reaction. To incorporate **2** and **3** into QBP, pETtrio-pylT-PylRS-QBP3TAG141TAA and pEVOL-AzFRS were used to cotransform *E. coli* BL21 cells. Plasmid pEVOL-AzFRS carries genes coding an evolved **3**-specific *Mj*TyRS (AzFRS) and $tRNA_{CUA}^{Tyr}$. The transformed cells were then used to express QBP with **3** and **2** incorporated at positions 3 and 141 respectively (QBP(**3**+**2**)). QBP was overexpressed when 1 mM **2** and 1 mM **3** were provided in the growth medium. The expression yield in 2YT medium was 14 mg/L. Providing one NCAA or No NCAA in the medium only gave negligible QBP expression (**Figure 17**).



Scheme III-7. NCAAs used in the study.

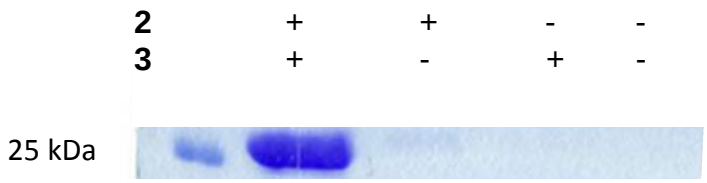


Figure 17. The expression of QBP(3+2) supplemented with different NCAAs. Visualization was effected by Coomassie blue staining.

Since both **2** and **3** can undergo the CuAAC reaction, we carried out labeling of QBP(**3**+**2**) in two separate steps to avoid cross-reactions of dyes. Our initial trials of labeling **3** of QBP(**3**+**2**) with a coumarin alkyne (**6** in **Scheme III-8**) in the presence of a Cu(I):tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) complex and 5 mM ascorbate met problems due to several reasons. QBP(**3**+**2**) has a 6×His tag at its C-terminus for its affinity purification using Ni-NTA resins. Imidazole that was used to elute QBP(**3**+**2**) from Ni-NTA resins severely influenced the CuAAC labeling process. Directly performing the CuAAC reaction between QBP(**3**+**2**) and **6** in a buffer containing 250 mM imidazole gave poor labeling efficiency. A three-hour reaction only led to an insignificant amount of labeled QBP(**3**+**2**) as proved by the electrospray ionization mass spectrometry (ESI-MS) analysis of the final reaction mixture. Thoroughly dialysing imidazole from QBP(**3**+**2**) did increase the labeling efficiency but not to a great extent. According to Finn *et al.*,[124] the 6×His tag of QBP(**3**+**2**) could potentially chelate Cu(I) and thus reduce the effective Cu(I) concentration for catalysis. To counteract the chelating effect of the 6×His tag, 1 mM Ni^{2+} was later provided in the reaction mixture. Performing the reaction in the optimized labeling condition that had 0.1 mM Cu(I):TBTA complex, 0.5 mM additional TBTA, 5 mM ascorbate, and 1 mM $NiCl_2$ for three hours afforded close to quantitatively labeling of QBP(**3**+**2**) with **6** to form QBP(**3**+**2**)-**6**, as confirmed by the ESI-MS analysis of the final product (**Figure 18C**). The ratio of dye to protein during the reaction was 50 to 1. A same reaction was also carried out to label QBP(**3**+**2**) with **5** to form QBP(**3**+**2**)-**5** (**Figure 18D**). In order to label **2** of QBP(**3**+**2**)-**6** with **5** that forms a FRET pair with **6**, QBP(**3**+**2**)-**6** was affinity

purified using Ni-NTA resins to remove unreacted **6** and then reacted with **5** in the same optimized CuAAC conditions for 3 h to form QBP(**3**+**2**)-**6**-**5**. The ESI-MS analysis of the final product suggested close to 80% conversion from QBP(**3**+**2**)-**6** to QBP(**3**+**2**)-**6**-**5** (**Figure 18E**). QBP(**3**+**2**)-**6**-**5** was further purified using Ni-NTA resins to remove unreacted **5**. To demonstrate the application of this dual labeled protein, it was used to undergo protein unfolding analysis using guanidium chloride (GndCl). With increasing GndCl concentration and excited by an 350 nm light, QBP(**3**+**2**)-**6**-**5** displayed increase of fluorescent emission from **6** at 430 nm and decrease of fluorescent emission from **5** at 520 nm (**Figure 19**), suggesting the distance increase between the two dyes during the protein unfolding process.

Although we have used two optimized consecutive CuAAC reactions to label QBP(**3**+**2**) with **6** and **5** to achieve close to quantitative dual labeling, there are downsides related to the CuAAC reaction on proteins. During labeling QBP(**3**+**2**) with **5**, we noticed a significant amount of protein aggregated. This aggregation process was time dependent. 47% of QBP(**3**+**2**) was recovered after labeling with **5** for 1 h; however, only 12% was recovered after a 5 h reaction. Since the two labeling reactions have to be performed sequentially for 3 h each and the protein has to be purified by Ni-NTA resins after each round to remove residual dyes, the protein recovery rate is low. The finally obtained product QBP(**3**+**2**)-**6**-**5** only accounted for 9% of the original QBP(**3**+**2**). This protein aggregation and low recovery problem poses a big challenge to use other less reactive dyes to label QBP(**3**+**2**). We have been trying to use dyes **7** and **8**[125] shown in **Scheme III-5.**

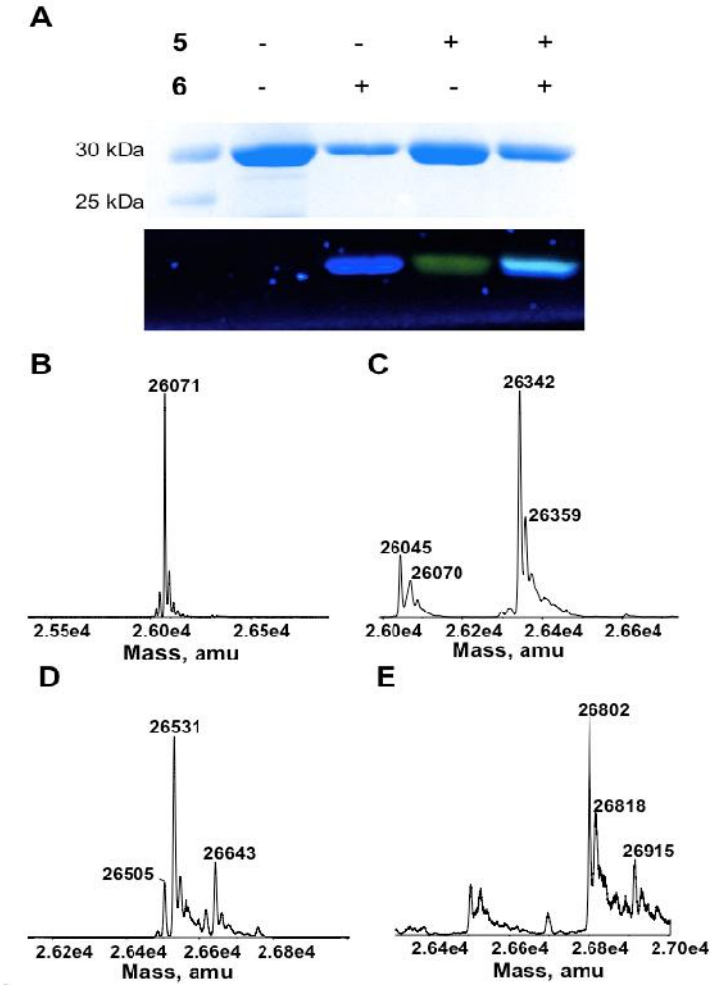


Figure 18. (A) Labeling QBP(3+2) with 6 and 5. Top panel: proteins stained with Coomassie blue in a SDS-PAGE gel; bottom panel: fluorescent imaging of the same gel irradiated by 365 nm UV light. The image shows real colors captured by a regular camera. Deconvoluted ESI-MS spectra of (B) QBP(3+2), (C) QBP(3+2)-6, (D) QBP(3+2)-5, and (E) QBP(3+2)-6-5. The theoretic molecular weights of QBP(3+2), QBP(3+2)-6, QBP(3+2)-5,and QBP(3+2)-6-5 are 26,069, 26,339, 26,529, and 26,799 Da, respectively.

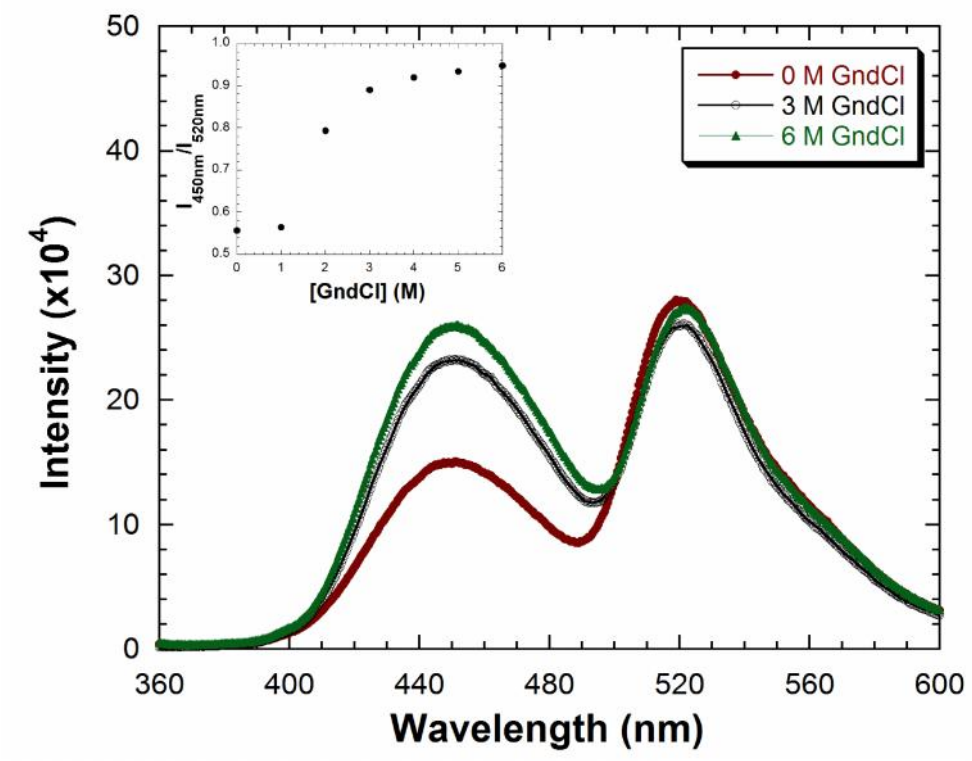


Figure 19. Fluorescent emission spectra of QBP(**3+2**)-**5**-**6** at different concentrations of GndCl. Only spectra from three GndCl concentrations are shown for clarity. The excitation wavelength was at 350 nm. The inset shows the dependence of I_{450nm}/I_{520nm} on the concentration of GndCl.

III-8 to achieve a dual-labeled QBP(**3**+**2**). Although the reaction between QBP(**3**+**2**) with **8** under the optimized CuAAC conditions could be finished in 3 h, the reaction between QBP(**3**+**2**) and **7** was less than 50% completed after 6 h (**Figure 20**), a time at which much protein aggregated. Besides the protein aggregation issue, there is another problem associated with the CuAAC reaction: protein oxidation. It has been demonstrated in the previous literature that Cu(I) can promote the formation of reactive oxygen species, leading to protein oxidation.[124] As confirmed by the ESI-MS analysis, our finally obtained QBP(**3**+**2**)-**5-6** did show a large unexpected peak at 26,818 Da that matches the molecular weights of the final production with one addition oxygen atom (**Figure 18E**). The final product also has a quite messy ESI-MS spectrum that might result from additional modifications of the protein during the labeling process. Of course, providing some reactive oxidative species scavengers and purging the reaction mixture with Argon followed by an anaerobic labeling reaction might alleviate protein oxidation. However, these additional treatments will certainly make the labeling process complicated.

A more ideal dual labeling method that resolves the issues associated with two consecutive CuAAC reactions is to run two labeling reactions at one step and in a catalyst free fashion. To run two labeling reactions in one pot, they not only need to be bio-orthogonal and also have to be orthogonal to each other. Two reactions that meet this requirement and also are catalyst free are the azide cyclooctyne click reaction and the oxime formation reaction. Although the rates of the azide cyclooctyne click reaction is slower than the CuAAC reaction,[126] presumably there is no protein aggregation so that

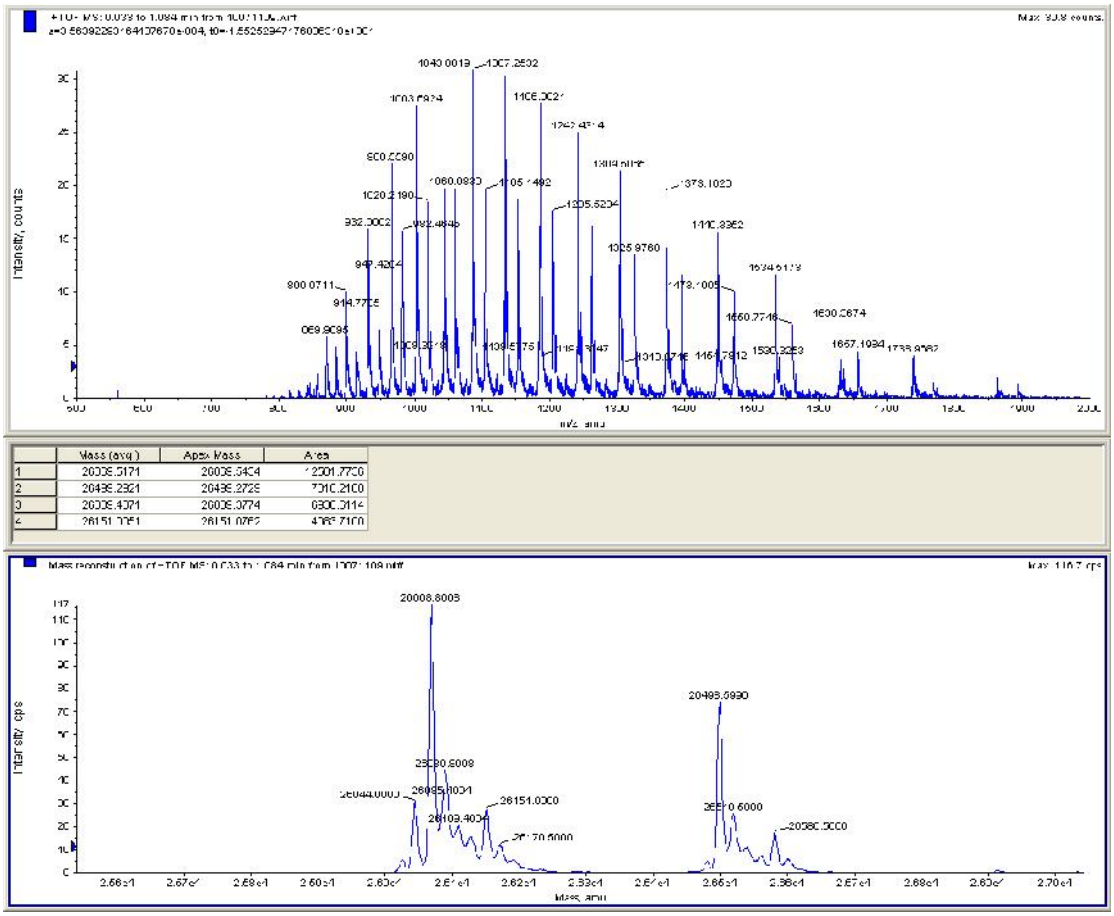


Figure 20. Labeling of QBP(3+2) with 7 for 6 h. Calculated protein mass without labeling: 26069; calculated protein mass with labeling: 26499.

the reaction time could be prolonged. We previously showed that a keto-containing NCAA, **4** in **Scheme III-7** could be genetically incorporated into proteins at amber mutation sites using an evolved *M. barkeri* PylRS (MbAcKRS1)- $tRNA^{Pyl}_{CUA}$ pair.[127] Since the labeling of **4** with a hydroxylamine dye could be achieved at a physiological pH, the incorporation of **3** and **4** into a protein will enable labeling the protein with a hydroxylamine dye and a cyclooctyne dye in a one-pot and catalyst-free fashion. Since MbAcKRS1 has a low efficiency for the incorporation of **4**, we switched to use a more efficient MmAcKRS that was evolved from *M. mazei* PylRS and has mutations L301M/Y306L/L309A/ C348F.[128] This gene was used to replace PylRS in pETtrio-pylT-PlyRS-QBP3TAG141TAA to afford pETtrio-pylT-MmAcKRS-QBP3TAG141TAA. Transforming *E. coli* BL21 cells with pEVOL-AzFRS and pETtrio-pylT-MmAcKRS-QBP3TAG141TAA and growing the transformed cells in the presence of 1 mM **3** and 2 mM **4** led to overexpression of QBP with **3** and **4** incorporated at positions 3 and 141 respectively (QBP(**3**+**4**)). The protein expression yield was 13 mg/L in 2YT medium. Providing only one NCAA or no NCAA in the medium gave a negligible QBP expression level (**Figure 21**). With QBP(**3**+**4**) in hands, we next performed a one-pot and catalyst-free dual label process of QBP(**3**+**4**) using dyes **9** and **10** shown in **Scheme III-8**. The overnight incubation at pH 6.4 conferred close to full conversion to the desired labeled QBP(**3**+**4**) (QBP(**3**+**4**)-**9**-**10**), as indicated by the ESI-MS spectrum of the final product (**Figure 22E**). Moreover, after affinity purification using Ni-NTA resins to remove unreacted dyes, the finally obtained QBP(**3**+**4**)-**9**-**10** accounted for 83% of the original QBP(**3**+**4**). In comparison to QBP(**3**+**2**)-**6**-**5** shown **Figure 22E**, QBP(**3**+**4**)-**9**-**10**

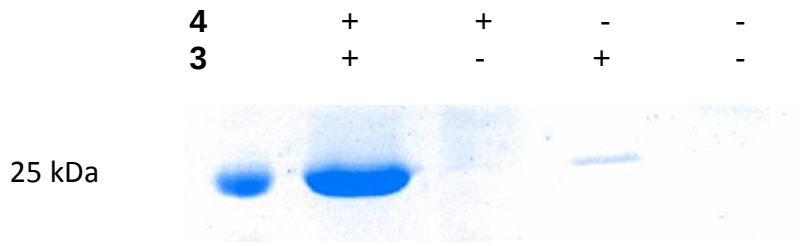


Figure 21. The expression of QBP(**3**+**4**) supplemented with different NCAAs. Visualization was effected by Coomassie blue staining.

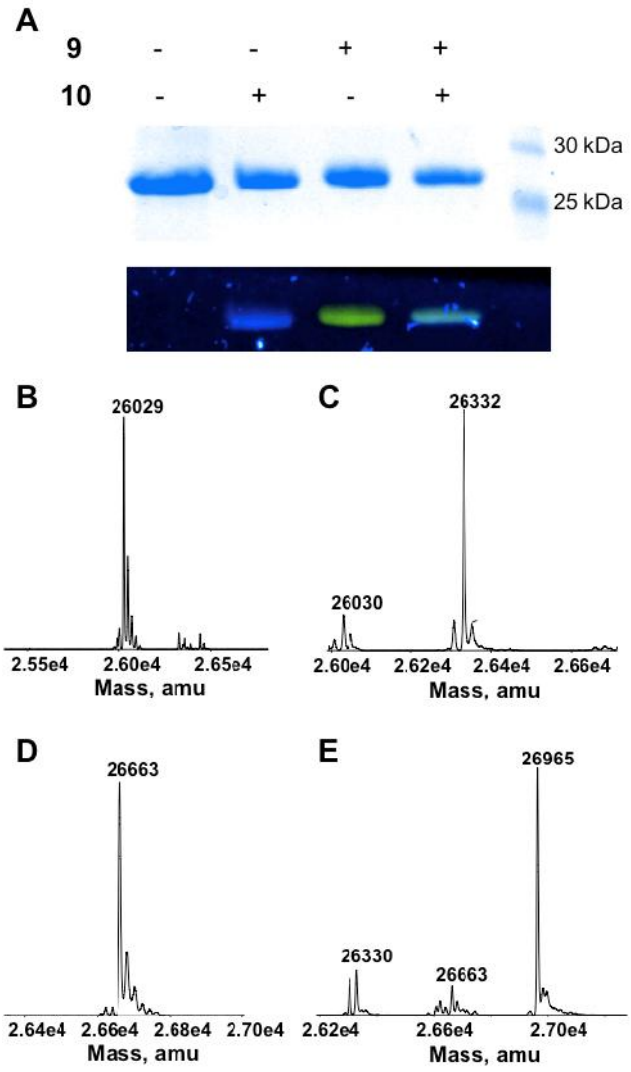


Figure 22. Labeling QBP(**3**+**4**) with **9** and **10**. Top panel: proteins stained with Coomassie blue in a SDS-PAGE gel; bottom panel: fluorescent imaging of the same gel irradiated by 365 nm UV light. The image shows real colors captured by a regular camera. Deconvoluted ESI-MS spectra of (**B**) QBP(**3**+**4**), (**C**) QBP(**3**+**4**)-**10**, (**D**) QBP(**3**+**4**)-**9**, and (**E**) QBP(**3**+**4**)-**9**-**10**. The theoretic molecular weights of QBP(**3**+**4**), QBP(**3**+**4**)-**10**, QBP(**3**+**4**)-**9**, and QBP(**3**+**2**)-**9**-**10** are 26,028, 26,329, 26,661, and 26962 Da, respectively.

displayed a much cleaner ESI-MS spectrum, suggesting no additional modification took place during the labeling process. Two separate reactions to label QBP(**3**+**4**) with **9** and **10** individually were also carried out. Each led to close to quantitative labeling (**Figure 22C**&**D**). With a much simpler labeling procedure and a much better protein recovery rate, this one-pot catalyst-free dual-labeling method is undoubtedly a more optimal choice than the dual labeling achieved by two consecutive CuAAC reactions. One thing that needs to be pointed out is the pH dependence of the oxime formation reaction. We have also attempted to run the labeling of QBP(**3**+**4**) with **10** at pH 8.1. However, overnight incubation only led to a negligible level of labeling.

We also used QBP(**3**+**4**)-**9**-**10** to carry out the protein unfolding analysis. With increasing GndCl concentration and excited by a 430 nm light, QBP(**3**+**4**)-**9**-**10** displayed increase of fluorescent emission from **10** at 470 nm and decrease of fluorescent emission from **9** at 520 nm (**Figure 23**), suggesting the distance increase between the two dyes. The inlet of **Figure 23** presents a smooth unfolding pattern determined by the change of I_{470nm}/I_{520nm}.

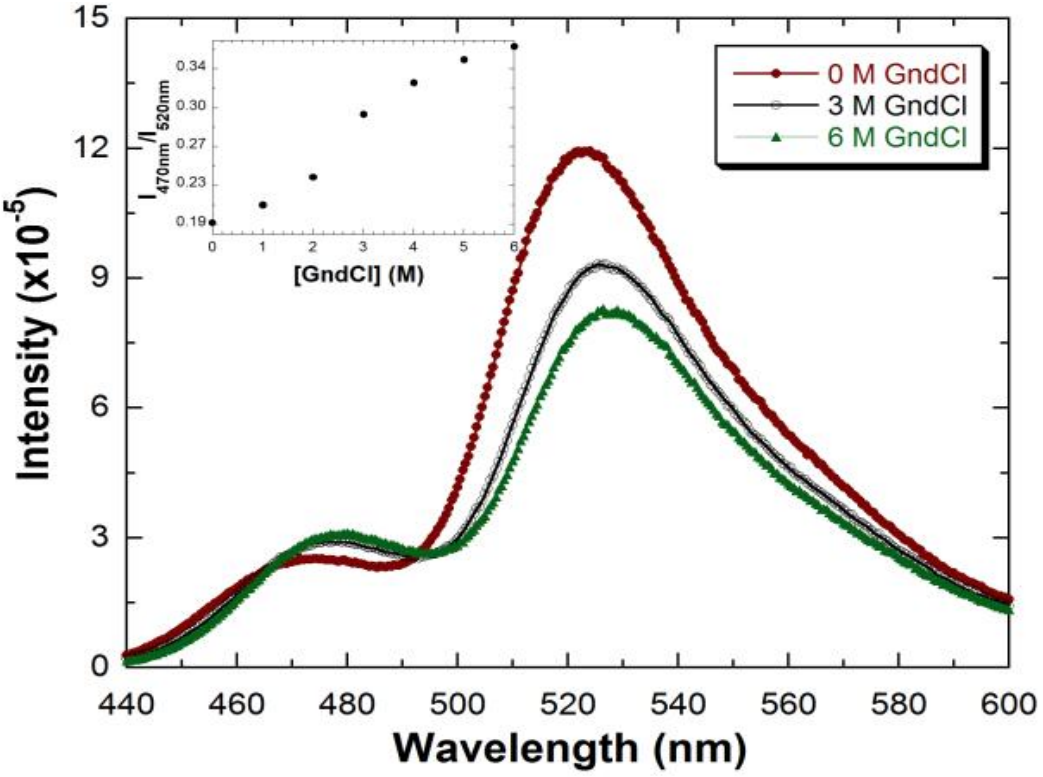


Figure 23. Fluorescent emission spectra of QBP(**3**+**4**)-**9**-**10** at different concentrations of GndCl. The excitation wavelength was 430 nm. The inlet shows the dependence of I_{470nm}/I_{520nm} on the concentration of GndCl.

Conclusion

In summary, we have developed a protein dual-labeling method that can be carried out in a one-pot and catalyst-free fashion. The two reactions for the dual labeling are both biocompatible. No treatment of proteins to avoid non-specific modifications with amino acid side chains such as cysteine thiols is necessary. The two reactions are orthogonal to each other and are directed by two genetically incorporated NCAAs, assuring the labeling specificity and selectivity. The two labeling reactions are also highly efficient, leading to almost quantitative labeling after an overnight incubation. The recovery of the finally labeled protein is also excellent. This simple and straightforward protein dual-labeling method resolves limitations associated with other current dual-labeling strategies and can be easily adopted by other research groups. Its potential applications range from single molecule FRET studies of protein dynamics, protein folding/unfolding investigations, and biosensor development.

CHAPTER IV

THE INCORPORATION OF NON-CANONICAL AMINO ACIDS WITH HYDROXYLAMINE AND HYDRAZINE MOIETY

Introduction

Lysine derivatives with hydroxylamine and hydrazine groups are non-canonical amino acids people want to install to proteins for a long time. They are suitable for protein labeling due to their high reactivity towards carbonyl groups and thioesters. Moreover, since they share similar structures to lysine, a great number of post-translational modified lysines such as long-chain acylated lysines, which are hard to be incorporated into proteins via other methods, can be obtained from **1** & **2** (**Scheme IV-1**) with further reactions. However, the direct incorporation of non-canonical amino acids bearing hydroxylamine or hydrazine groups has some obstacles. For one thing, they are too reactive to the metabolites in cells, such as ketones and aldehydes. For another, hydroxylamine and hydrazine lysine derivatives are hard to be directly genetically encoded into proteins due to their great similarity to lysine. Here we present a proposal to genetically encode protected lysine derivatives with hydroxylamine and hydrazine groups and the following deprotection after the purification of the proteins.

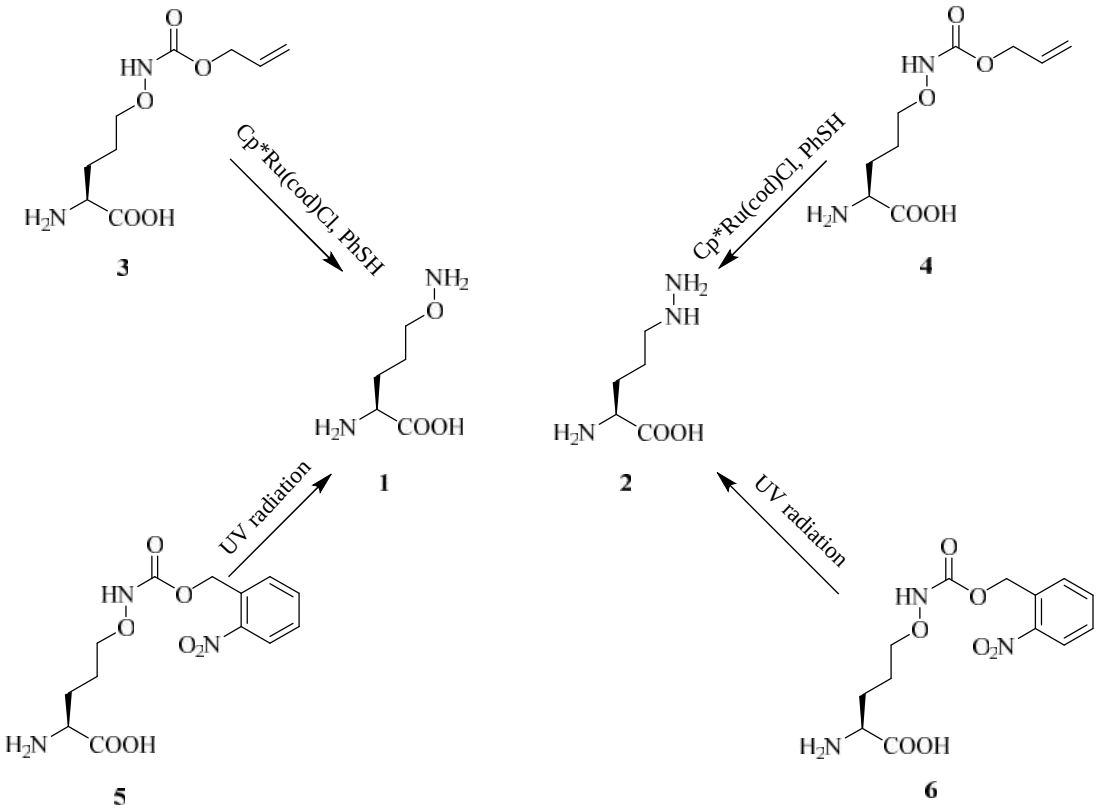
A variety of protecting groups have been applied to make proteins with modified NCAAs. Boc group is widely used in organic synthesis and N^{ε}-Boc lysine is able to be genetically encoded with the wild type PylRS/$tRNA^{Pyl}_{CUA}$ pair with a high yield. However, due to the fact that the deprotection of boc group requires strong acidic condition, it is not favored by protein chemists. Cbz group is another commonly used protecting group

and the N^{ε}-Cbz lysine has also been successfully genetically encoded with mKRS1. Nonetheless, the deprotection of the cbz group may also potentially damage proteins of interest since palladium metal, which may cause oxidation on the protein, is key catalyst for the deprotection. Alloc protecting group undergoes deprotection with chelated ruthenium ion which is considered to be a milder condition to proteins. Inspired by our previous work, we postulate that nitro-cbz protection group is another plausible protection option since it can be easily deprotected by photolysis. Although the product of the deprotection of nitro-cbz group might potentially reaction to hydroxylamine and hydrazine group, the issue can be easily resolved by adding scavengers in the reaction mixture.

Experimental Section

General Experimental

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was performed on Whatman SiO_2 60 F-254 plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).



Scheme IV-I. Encoding strategy for **1** & **2**.

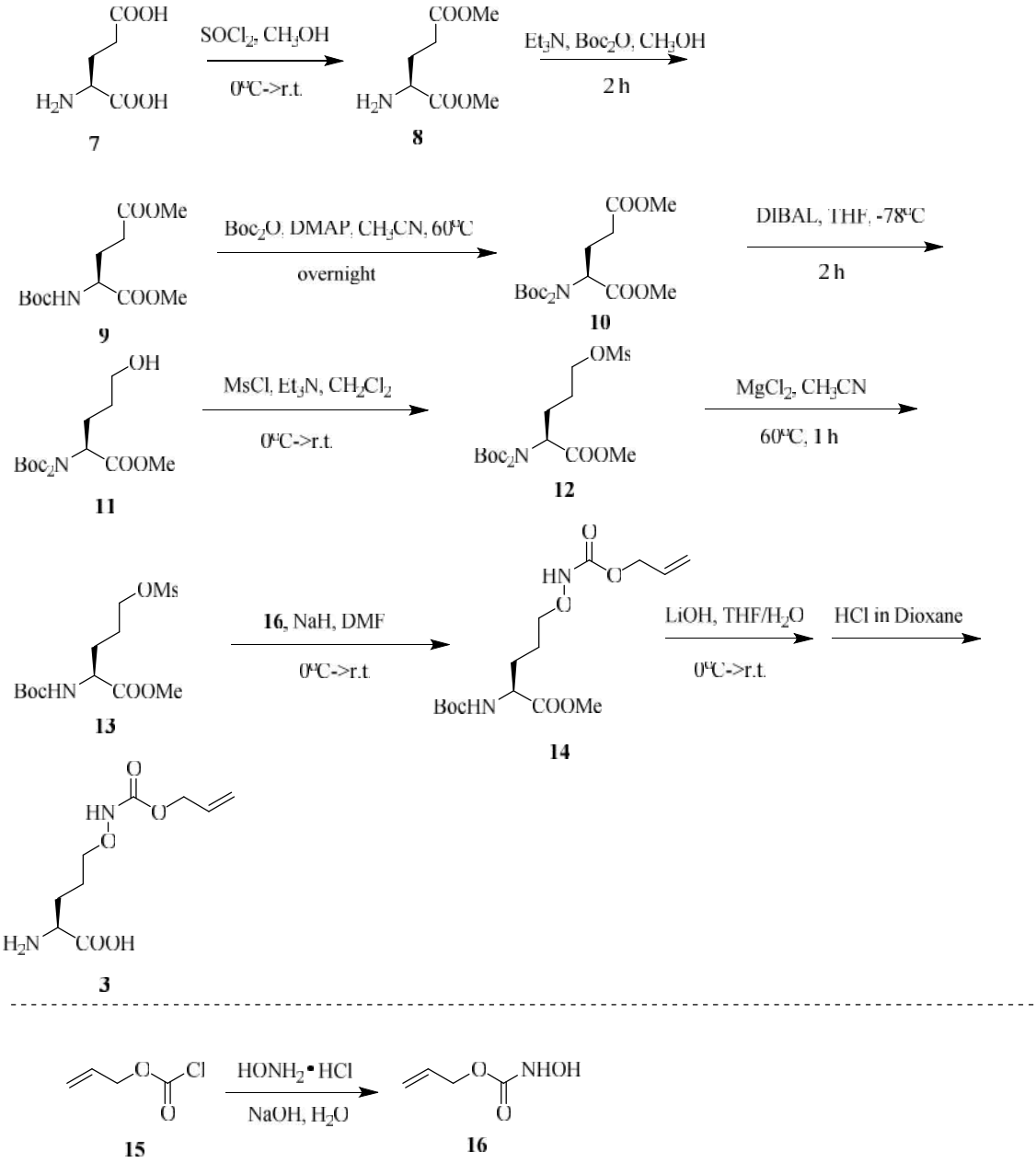
Specific rotations of chiral compounds were obtained at the designated concentration and temperature on a Rudolph Research Analytical Autopol II polarimeter using a 0.5 dm cell. Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.80 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, and a minimal amount of sodium hydroxide pellet was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral or acidic conditions. ^{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 obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

Chemical Synthesis

Dimethyl (*tert*-butoxycarbonyl)-*L*-glutamate (9)

To a suspension of **7** (5.3 g, 36.05 mmol) with ice bath in methanol (90 mL) was added thionyl chloride (8.5 mL, 116.8 mmol). The reaction mixture was stirred at 0^{o}C then room temperature overnight. Solvent was removed by water aspirator and methanol (50 mL) was added to re-dissolve the material and TEA (10 mL, 71.88 mmol) was added dropwise to the reaction mixture, followed by boc anhydride (8.738 g, 40.08 mmol) with methanol (20 mL) and the reaction mixture was stirred overnight. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to afford a slightly yellow oil. The product was purified by a silica gel chromatography with 1:5 ethyl acetate/hexanes. (9.0 g, 91%) ^{1}H NMR ($CDCl_3$, 300 MHz) δ 5.11 (d, J = 6.6 Hz), 4.35 (quartet, 1 H, J = 7.2 Hz), 3.67 (s, 6 H), 2.44-2.37 (m, 2 H), 2.22-2.15 (m, 1 H), 2.04-1.90 (m, 1 H), 1.46 (s, 9 H).



Scheme IV-2. The synthesis of **3**.

Dimethyl (di-*tert*-butoxycarbonyl)-*L*-glutamate (10)

To a solution of **9** (3.11 g, 11.31 mmol) in acetonitrile (40 mL) was added DMAP (272 mg, 2.22 mmol) and boc anhydride (4.5 g, 20.64 mmol). The reaction mixture was stirred at 60 ^{o}C overnight. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to afford a slightly yellow oil. The product was purified by a silica gel chromatography with 1:10 ethyl acetate/hexanes (4.3 g, quant.)) ^{1}H NMR ($CDCl_3$, 300 MHz) 4.94 (dd, 1 H, *J* = 5.4 Hz, 9.0 Hz), 3.70 (s, 6 H), 2.45-2.40 (m, 2 H), 2.20-2.11 (m, 1 H), 2.04-1.90 (m, 1 H), 1.46 (s, 18 H).

Methyl (*S*)-2-((di-*tert*-butoxycarbonyl)amino)-5-hydroxypentanoate (11)

To a solution of **10** (3.5g, 9.33 mmol) in anhydrous THF (100 mL) was added DIBAL (1.0 M, 24 mL, 24.0 mmoL) dropwise under dry ice acetone bath. The reaction mixture was stirred at -78^{o}C then room temperature overnight. Water (5 mL) was added to quench the excessive DIBAL. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to afford a slightly yellow oil. The product was purified by a silica gel chromatography with 1:3 ethyl acetate/hexanes. (2.66 g, 83%) ^{1}H NMR ($CDCl_3$, 300 MHz) 4.94 (dd, 1 H, *J* = 5.4 Hz, 9.0 Hz), 3.72 (s, 6 H), 3.68 (t, *J* = 6.3 Hz, 2 H) 2.31-2.18 (m, 1 H), 2.01-1.90 (m, 1 H), 1.71-1.58 (m, 2 H), 1.46 (s, 18 H).

Methyl (*S*)-2-((di-*tert*-butoxycarbonyl)amino)-5-((methylsulfonyl)oxy)pentanoate (12)

To a solution of **11** (1.445 g, 4.16 mmol) and triethylamine (1.0 mL, 7.19 mmol) in anhydrous dichloromethane (20 mL) under ice bath was added mesyl chloride (0.6 mL, 7.75 mmol) dropwise. The reaction mixture was stirred at room temperature for 2 h. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to afford a colorless oil. The product was purified by a silica gel chromatography with 1:5 ethyl acetate/hexanes. (1.684 g, 95%) ^{1}H NMR ($CDCl_3$, 300 MHz) 4.85 (dd, 1 H, J = 5.4 Hz, 9.0 Hz), 4.24 (t, J = 6.9 Hz, 3 H), 3.70 (s, 3 H), 2.99 (s, 3 H) 2.31-2.18 (m, 1 H), 2.04-1.90 (m, 1 H), 1.90-1.74 (m, 2 H), 1.46 (s, 18 H).

Methyl (*S*)-2-((*tert*-butoxycarbonyl)amino)-5-((methylsulfonyl)oxy)pentanoate (13)

To a solution of **12** (4.26 g, 10.02 mmol) in acetonitrile (20 mL) was added magnesium (0.5 g, 5.25 mmol). The reaction suspension was stirred at 50^{o}C for 1 h. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to afford a colorless oil. The product was applied for the next step without further purification.

Methyl (*S*)-5-((((allyloxy)carbonyl)amino)oxy)-2-((*tert*-utoxycarbonyl)amino)pentanoate (14)

To a solution of **16** (1.494 g, 12.77 mmol) in anhydrous DMF (20 mL) was added sodium hydride (1.076 g, 26.9 mmol) portion wise. Then **13** (2.964 g, 9.12 mmol) with anhydrous DMF (10 mL) was added into the reaction mixture dropwise. The

reaction mixture was stirred at room temperature for 1 h under argon. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to afford a colorless oil. The product was purified by a silica gel chromatography with 1:3 ethyl acetate/hexanes. (1.456 g, 47%). ^{1}H NMR ($CDCl_3$, 300 MHz) 6.00-5.84 (m, 1 H), 5.37-5.23 (m, 2 H). 5.11 (bs, 1 H), 4.61 (dt, 2 H, *J* = 1.5 Hz, 7.5 Hz), 4.16 (bs, 1 H), 3.89 (t, *J* = 6.9 Hz, 2 H), 3.71 (s, 3 H), 2.30-2.16 (m, 1 H), 2.04-1.90 (m, 1 H), 1.80-1.69 (m, 2 H), 1.46 (s, 9 H).

Allyl hydroxycarbamate (16)

To a solution of sodium hydroxide (18.2g, 0.46 mol) in water (50 mL) under ice bath was added hydroxylamine hydrochloride (21.0 g, 0.30 mol), followed by allyl chloroformate (18.0 mL, 0.17 mol) in THF (2 mL) dropwise over 20 min. The mixture was stirred at room temperature for 3 h. Then the pH was adjusted to 1 with 12 M hydrochloride acid (12 mL). The product was extracted by ethyl acetate and washed by water and brine. The product was dried by sodium sulfate and evaporated to afford a colorless oil. (19.9 g quant.) ^{1}H NMR ($CDCl_3$, 300 MHz) 5.98-5.84 (m, 1 H), 5.36-5.22 (m, 2 H), 4.64 (dt, *J* = 1.5 Hz, 6.0 Hz, 2 H).

(*S*)-5-((((allyloxy)carbonyl)amino)oxy)-2-aminopentanoic acid (3)

To a solution of **14** (1.456 g, 4.28 mmol) was in THF/water (5:1, 12 mL) was added lithium hydroxide (1 N, 6 mL, 6.0 mmol) under ice bath. The reaction mixture was stirred at 0^{o}C then room temperature for 2 h. The product was diluted in water and extracted with ether. 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, and the combined organic phases were washed once with brine, dried with sodium sulfate, and evaporated to give the crude carboxylic acid as a colorless oil, which was directly used without further purification.

The above crude acid (~4.28 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 2 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 (200 mL) and then eluted with pyridine (1 M, 250 mL) to give **3** (0.474 g, 48%). ^{1}H NMR ($CDCl_3$, 300 MHz) 5.88-5.74 (m, 1 H), 5.24-5.10 (m, 2 H), 4.49 (*J* = 1.2 Hz, 8.4 Hz, 2 H), 3.85 (t, *J* = 6.3 Hz, 2 H, major rotamer), 3.78 (t, *J* = 6.0 Hz, 2 H, minor rotamer), 2.01-1.74 (m, 2 H), 1.71-1.56 (m, 2 H) HRMS (ESI) calcd for $C_9H_{17}N_2O_5$ ($[M + H]^+$) 233.1137, found 233.1130.

Allyl (*S*)-2-(4-((di-*tert*-butoxycarbonyl)amino)-5-methoxy-5-oxopentyl)hydrazine-1-carboxylate (18)

To a solution of **10** (1.302 g, 3.47 mmol) in anhydrous THF (20 mL) was added DIBAL (1.0 M, 4 mL, 4.0 mmoL) dropwise under dry ice acetone bath. The reaction mixture was stirred at -78^{o}C then room temperature overnight. Water (5 mL) was added to quench the excessive DIBAL. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was

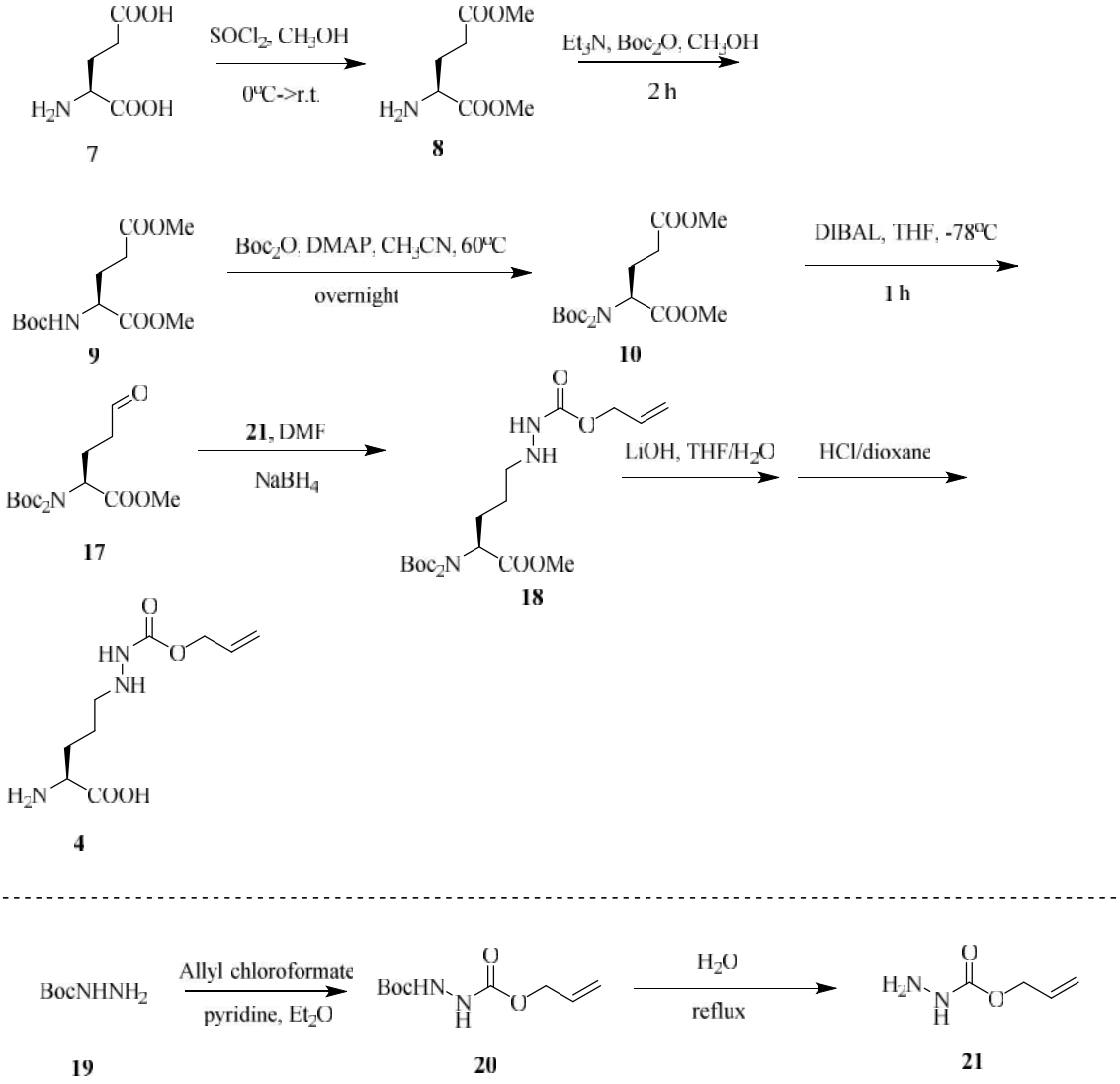
removed to give a slightly yellow oil which was applied to the next step reaction without purifications.

To the crude product of **17** in THF (20 mL) was added **21** (0.48 g, 4.07 mmol) and the reaction mixture was stirred at room temperature for 15 min. Then sodium borohydride (0.38 g, 10 mmol) was added into the reaction mixture which was further stirred vernight afterwards. The product was extracted by ethyl acetate and washed by water and brine, then dried by sodium sulfate and evaporated, further purified by a silica gel chromatography with 1:3 ethyl acetate/hexans to afford a colorless oil (1.15 g, 78%). ^{1}H NMR ($CDCl_3$, 300 MHz) 6.00-5.84 (m, 1 H), 5.37-5.23 (m, 2 H), 4.86 (dd, 1 H, *J* = 5.4 Hz, 9.0 Hz), 4.61 (dt, *J* = 1.5 Hz, 7.5 Hz, 2 H), 3.72 (s, 3 H), 2.64 (t, *J* = 6.3 Hz, 2 H) 2.31-2.18 (m, 1 H), 2.01-1.90 (m, 1 H), 1.64-1.52 (m, 2 H), 1.46 (s, 18 H).

Allyl hydrazinecarboxylate (21)

To a solution of **19** (1.33 g, 10.08 mmol) in ether (40 mL) was added pyridine (0.9 mL, 11.14 mmol), followed by allyl chloroformate (1.1 mL, 10.31 mmol) dropwise. The reaction was stirred at room temperature overnight. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to give a colorless oil **20** which was applied to the next step reaction without purifications.

A solution of **20** in water was heated to reflux overnight. The reaction mixture was adjusted to basic and the product was extracted with ethyl acetate, washed with water twice and brine, dried in sodium sulfate, and concentrated to give a yellow oil (1.16 g, quant.). The product was used to next step reaction without further purifications.



Scheme IV-2. The synthesis of **4**.

(*S*)-5-(2-((allyloxy)carbonyl)hydrazinyl)-2-aminopentanoic acid (4)

To a solution of **18** (1.15 g, 2.71 mmol) was in THF/water (5:1, 12 mL) was added lithium hydroxide (1 N, 6 mL, 6.0 mmol) under ice bath. The reaction mixture was stirred at 0^{o}C then room temperature for 2 h. The product was diluted in water and extracted with ether. 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, and the combined organic phases were washed once with brine, dried with sodium sulfate, and evaporated to give the crude carboxylic acid as a colorless oil, which was directly used without further purification.

The above crude acid (~2.71 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 2 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 (200 mL) and then eluted with pyridine (1 M, 250 mL) to give **3** (0.353 g, 56%). ^{1}H NMR ($CDCl_3$, 300 MHz) 5.88-5.74 (m, 1 H), 5.24-5.10 (m, 2 H), 4.49 (*J* = 1.2 Hz, 8.4 Hz, 2 H), 2.85 (t, *J* = 6.3 Hz, 3 H, major rotamer), 2.68 (t, *J* = 6.0 Hz, 3 H, minor rotamer), 1.90-1.74 (m, 2 H), 1.71-1.56 (m, 2 H) HRMS (ESI) calcd for $C_9H_{17}N_2O_5$ ($[M + H]^+$) 232.1297, found 232.1289.

Plasmid Construction

Both pEVOL-pylT-PylRS and pET-sfGFPS2TAG were constructed as previous reported[24].

Protein Expression and Purification

E. coli BL21(DE3) cells were transformed with pEVOL-pylT-PylRS and pET-sfGFPS2TAG and plated on LB agar plate containing chloramphenicol (Cam) (34 μg/mL) and ampicillin (Amp) (100 μg/mL). A 5 mL o/n culture was prepared from a single colony. This overnight culture was used to inoculate 500 mL of LB medium supplemented with Cam and Amp. Cells were grown at 37 ^{o}C in an incubator (250 r.p.m.) and the protein expression was induced with 1 mM IPTG, 0.2% arabinose and 5 mM **3** or **4** when OD_{600} reached 0.6. After 8 h induction, cells were harvested, resuspended in a lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and sonicated in an ice/water bath four times (4 min each, 10 min interval to cool the suspension below 10 ^{o}C between each pulse). The cell lysate was centrifuged (60 min, 11,000 g, 4 ^{o}C). The supernatant was incubated with 3 mL Ni-NTA resin (Qiagen) (2 h, 4 ^{o}C). The slurry was then loaded to a column and the protein-bound resin was washed with 30 mL of the wash buffer. Protein was finally eluted by the lysis buffer with 250 mM imidazole. Eluted fractions were collected, concentrated and buffer exchanged to 10 mM ammonium bicarbonate using an Amicon Ultra-15 Centrifugal Filter Devices (10k MWCO cut, Millipore). The protein purity was analyzed by 15% SDS-PAGE.

Deprotection of Lysine Derivatives

The deprotection of 3 and 4

3 and **4** in (50 mM, 100 µL) was mixed with [Cp*Ru(cod)Cl] (20 mM, 100 µL) and thiophenol (10 M, 20 µL). The reaction mixture was stirred at room temperature for 2 h. For TLC, a more polar compound was formed which indicated the alloc protection group was taken off.

The deprotection of sfGFPS2 with 3

sfGFP with **3** (20 µM, 100 µL) in PBS buffer with mixed with [Cp*Ru(cod)Cl] (20 mM, 1 µL) and thiophenol (200 mM, 1 µL) and the reaction mixture was stirred at 37^{o}C for 2 h. The catalyst was attempted to be removed by purifying the sfGFP with Ni-NTA resin. However, after two hours of reaction, the sfGFP was damaged and it didn't bind to the resin anymore. The same condition was applied to sfGFP with the two catalysts individually and the sfGFP could be recovered by Ni-NTA purification.

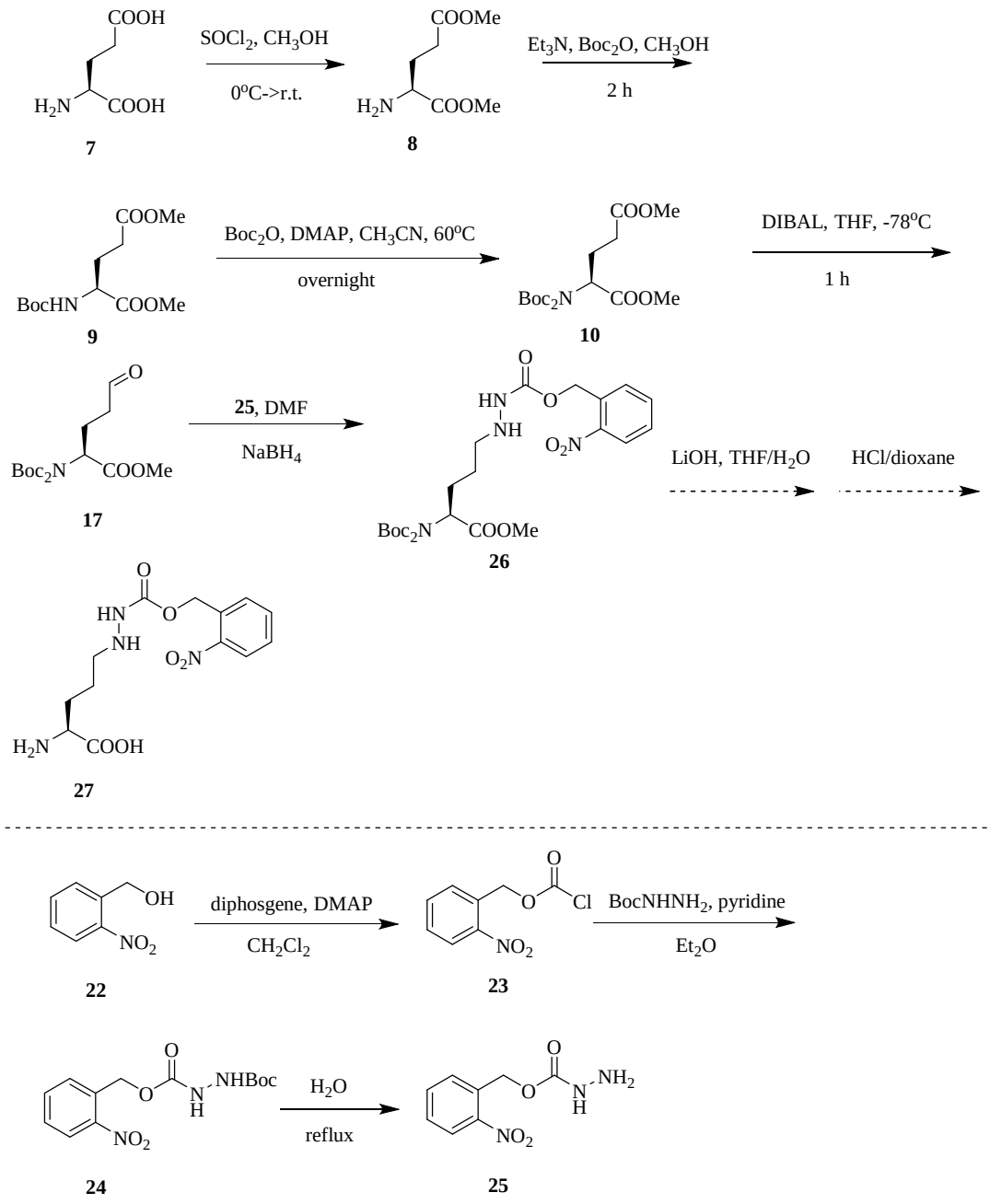
Result and Discussion

3 have been successfully genetically encoded into sfGFPS2TAG with an acceptable yield and low background (**Figure 24**). The incorporation was confirmed by mass spectrometry (**Figure 25**). However, after deprotection, sfGFP didn't bind to Ni-NTA resin anymore, which indicated that either the His-tag got cut off or there was some reaction on histidine causing the loss of its chelating ability to nickel. To investigate whether it was [Cp*Ru(cod)Cl] or thiophenol caused the damage of protein, sfGFP with **3** was treated with [Cp*Ru(cod)Cl] (10 mM) or thiophenol (100 mM) individually and incubated in the shaker at 37^{o}C with 2 h. sfGFP was recovered with Ni-NTA purification

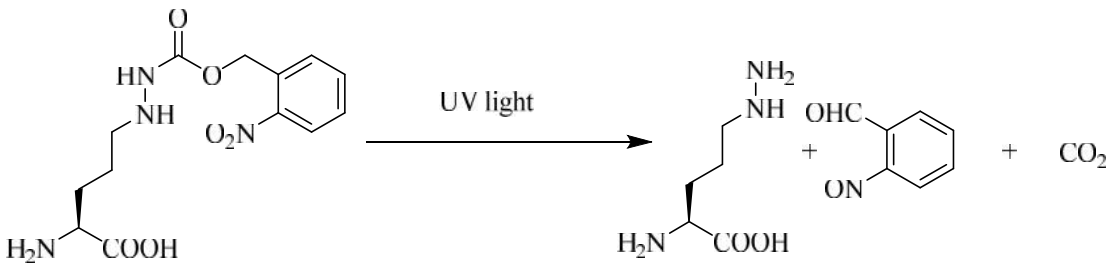
protocol in both cases. Given the fact that neither [Cp*Ru(cod)Cl] nor the thiophenol caused the cleavage individually, the damage of the protein must be related to the nature of the ruthenium deprotection reaction. Other milder conditions had also been applied with the deprotection (reduction of the catalysts, low the temperature), none of them successfully avoided the damage of the protein. Therefore we chose a different group as the protecting group which can be deprotected with a milder condition. 2-Nitrobenzyloxycarbonyl was select as the protection group since it is easily deprotected by UV radiation. Although the product of the deprotection, 6- methylenecyclohexa-2,4-dien-1-one, could potentially react with hydroxylamine/hydrazine group, it can be easily avoided by the addition of hydroxylamine as a scavenger (**Scheme IV-5**). However, the methyl ester deprotection of **26** ended up with the deprotection of alloc group as well. From the TLC tracing, it seemed that the alloc group was more labile than the methyl ester towards the base.

Conclusion

In this study, we demonstrate the synthesis and genetic incorporation of protected hydroxylamine/hydrazine-containing lysine mimics. However, the deprotection caused damage to sfGFP. An alternative synthetic route had been tried, yet the deprotection of the methyl ester also rendered the removal of the protecting group of the hydroxylamine/hydrazine. Methyl ester can be removed before coupling the hydrazine moiety to the lysine part since the presence of a carboxylic acid doesn't interfere with the reaction. It can be a solution for the genetic incorporation of lysine mimics with hydrazine group.



Scheme IV-4. The synthesis of **5**.



Scheme IV-5: Deprotection of **5**.

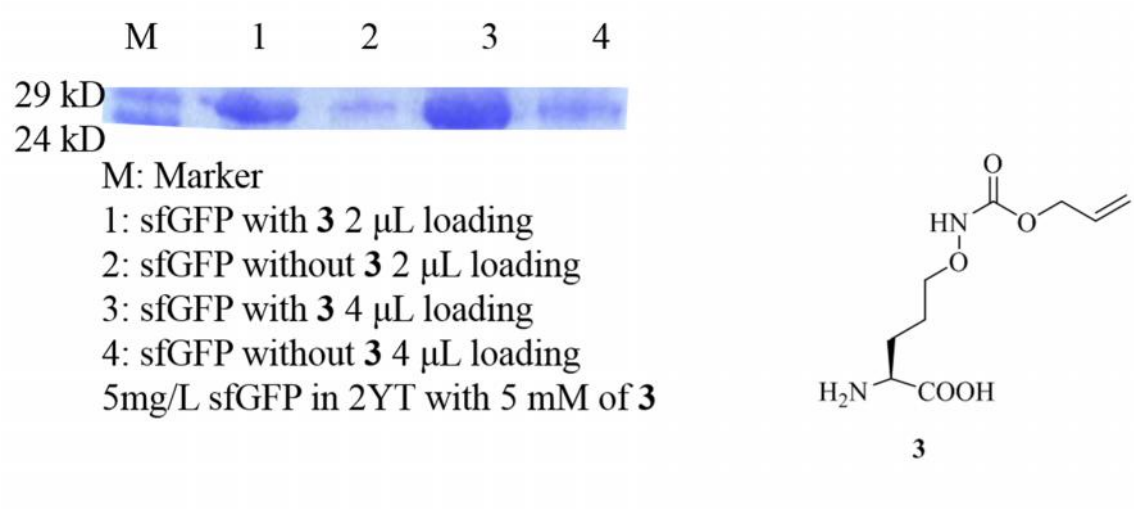


Figure 24. sfGFP expression with 5 mM of **3** in 2YT medium

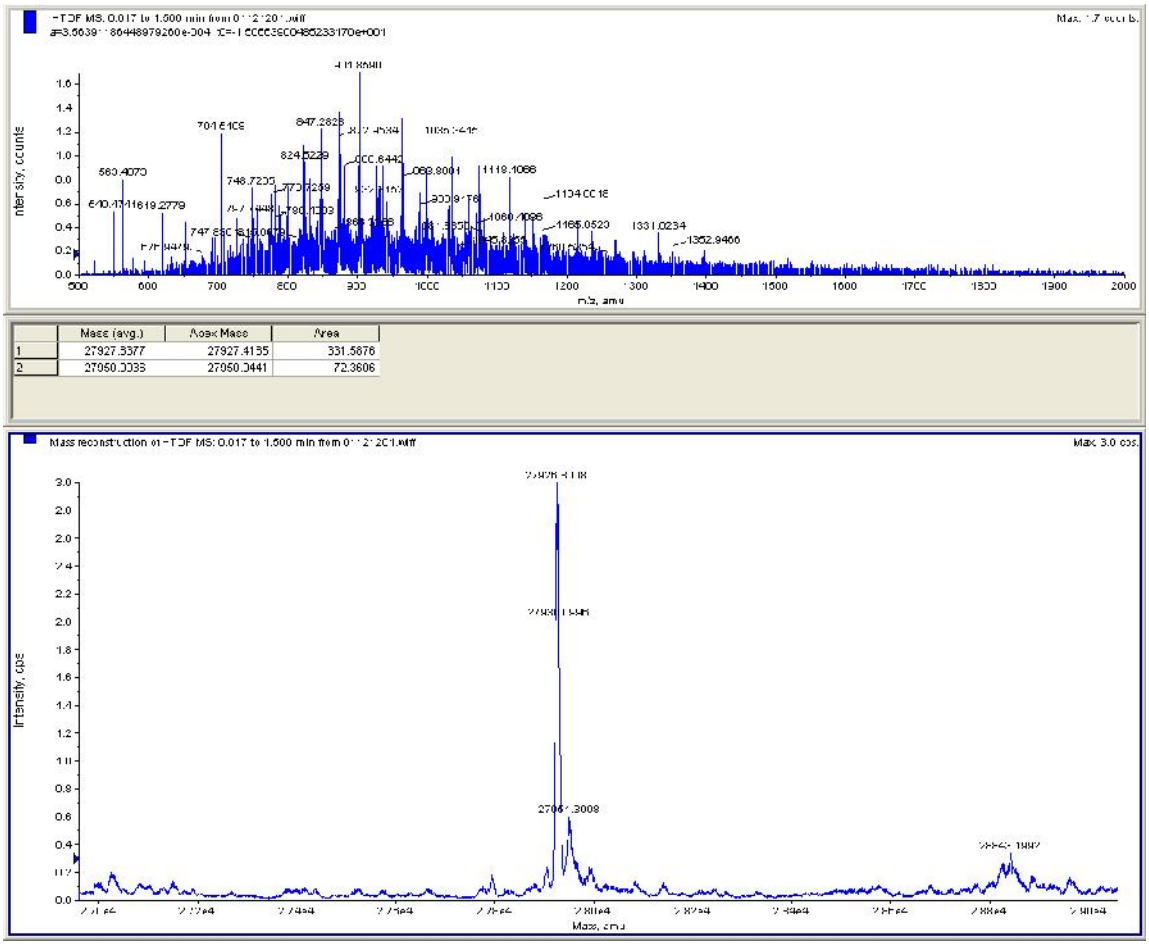


Figure 25. Mass spectrum data for sfGFPS2 with **3**.

CHAPTER V

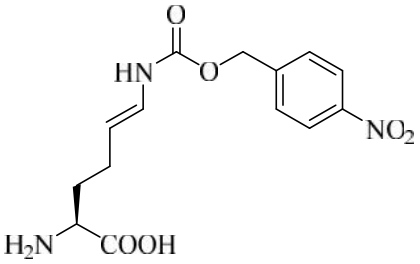
THE GENETIC ENCODING OF A 'SWISS ARMY KNIFE' INTO PROTEINS-AN ULTIMATE SOLUTION FOR THE SYNTHESIS OF PROTEINS WITH METHYLATIONS AND ACYLATIONS

Introduction

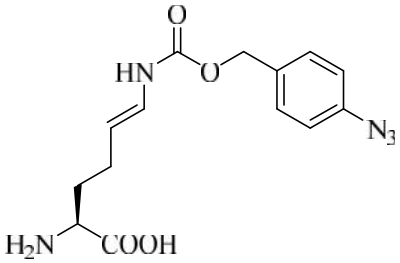
Histone methylations are widely related to cancer, aging and cell differentiation.[129-133] Interestingly, unlike histone acetylation, which is known to be a mark for the activation of gene transcription,[134,135] site-specific histone methylations recruit different proteins to activate or suppress gene transcription.[5,136,137] To study histone methylations, it is crucial to synthesize homogenous site-specific methylated histones. A number of histones with methyl lysine mimics have been made.[138,139] However, with the sulfur atom as a methylene substituent, it may interfere with the binding of histone to the protein it recruits. Previously we reported an approach for the genetic encoding of Kme1 into GFP_{UV} and Z-domain by using a photocaged protecting group.[140] Comparing with mono-methylation, dimethylation is more widely abundant in histones and it is closely related to gene transcription. Unfortunately, the same approach cannot by applied to incorporate Kme2 due to the fact that the installation of a protecting group on a tertiary amine is very unlikely and it will generate a positive charge which may impede the cell permeability of the amino acid.

Lysine myristoylation has been proven to relate to protein localization and protein-protein interactions[141-143]. However, the genetic incorporation of long-chain acylated lysine has been challenging due to the difficulty to accommodate the bulky

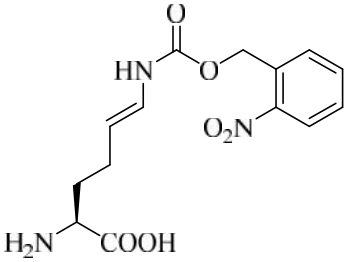
side-chain by the active site of PylRS. Given the fact that it is very likely that there are other kinds of lysine long-chain acylation yet been discovered, it is beneficial to discover a general approach for the genetic encoding instead of performing the evolution for each one of them. Here we present an approach to encode all kinds of methyl lysines and acylated lysines via a protected double bond-containing lysine (**1, 2**), which we regard as a 'swiss army knife' undergoing different reactions to afford desired modified lysines.



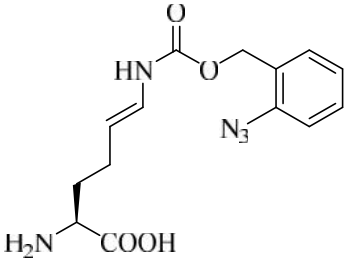
1



2



3



4

Scheme V-1. Non-canonical amino acids used in this study.

Experimental Section

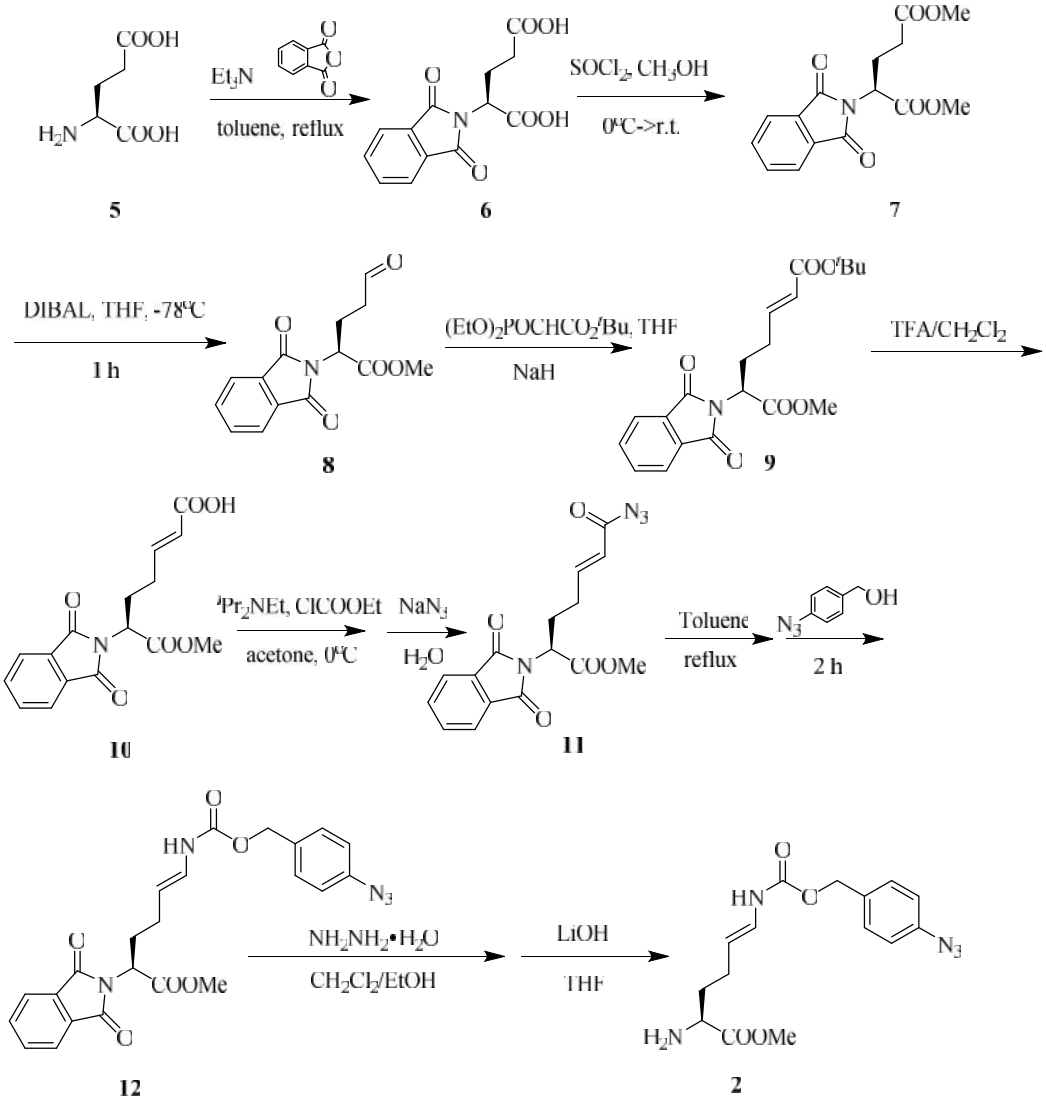
General Experimental

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was preformed on Whatman SiO_2 60 F-254 plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Specific rotations of chiral compounds were obtained at the designated concentration and temperature on a Rudolph Research Analytical Autopol II polarimeter using a 0.5 dm cell. Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.80 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, and a minimal amount of sodium hydroxide pellet was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral or acidic conditions. ^{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 obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

Chemical Synthesis



Scheme V-2. The synthesis of **2**.

(*S*)-2-(1, 3-dioxoisoindolin-2-yl)pentanedioic acid (6)

To a suspension of **5** (30 g, 0.2 mol) in methanol (300 mL) was added phthalic anhydride (29.6 g, 0.2 mol) and triethylamine (2.79 mL, 20 mmol). The reaction mixture was heated up to reflux overnight. Solvent was removed by water aspirator and ethyl acetate was added to re-dissolve the material and hydrochloride acid was added to adjust the pH to 2. The product was extracted by ethyl acetate, washed by hydrochloride acid, water and brine, dried by sodium sulfate. Solvent was removed to afford a white solid. (24.8 g, 69%)

Dimethyl (*S*)-2-(1, 3-dioxoisoindolin-2-yl)pentanedioate (7)

To a solution of **6** (24.8 g, 89.5 mmol) in MeOH (200 mL) was added thionyl chloride (19.47 mL, 0.27 mol) under ice bath. The reaction mixture was stirred at room temperature overnight. The solvent was removed by water aspirator and the product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to afford a slightly yellow oil. The product was used without purification and directly subjected to the next step reaction

Methyl (*S*)-2-(1, 3-dioxoisoindolin-2-yl)-5-oxopentanoate (8)

To a solution of **7** (5.8 g, 0.019 mol) in anhydrous ether (100 mL) was added DIBAL (1.0 M, 20.9 mL, 20.9 mmoL) dropwise under argon and dry ice acetone bath. The reaction mixture was stirred at -78^{o}C for 15 min then water (1.29 mL) was added. The reaction mixture was warmed up to room temperature and stirred for another 30 min. The product was filtrated, extracted by ethyl acetate, washed by hydrochloride acid,

sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to give a slightly yellow oil which was applied to the next step reaction without purifications.

1-(*tert*-butyl) 7-methyl (*S*,*E*)-6-(1,3-dioxoisoindolin-2-yl)hept-2-enedioate (9)

To anhydrous THF (20 mL) under ice bath was added sodium hydride (0.48 g, 60%, 12 mmol), followed by *tert*-butyl diethylphosphonoacetate (3 g, 12 mmol) dropwise. After the bubbles stopped forming, **8** (2.8 g, 10 mmol) was added in anhydrous THF (5 mL) and the reaction mixture was warmed to room temperature the stirred for an additional 2 h. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to give a slightly yellow oil (3.4 g, 92%)

(*S*,*E*)-6-(1,3-dioxoisoindolin-2-yl)-7-methoxy-7-oxohept-2-enoic acid (10)

To a solution of **10** (2.9 g, 7.8 mmol) in dichloromethane (30 mL) was added trifluoroacetic acid (5.96 mL, 78 mmol) under ice bath. The reaction mixture was stirred at room temperature for 2 h. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to give a slightly yellow oil. The product was subjected to the next step reaction without further purifications.

Methyl (*S*,*E*)-7-azido-2-(1,3-dioxoisoindolin-2-yl)-7-oxohept-5-enoate (11)

To solution of **10** (2.8 g, 8.8 mmol) in acetone (60 mL) was added *N*,*N*-Diisopropylethylamine (3.535 mL, 20 mmoL) and ethyl chloroformate (1.85 mL, 20 mmol) dropwise. The reaction mixture was stirred under ice bath for 1 h. The sodium azide (2.86 g, 44 mmol) was added with water (15 mL), then the reaction mixture was

warmed up to room temperature. The product was extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate. Solvent was removed to give a slightly yellow oil. The product was subjected to the next step reaction without further purifications. (3.0 g, quant.)

Methyl (*S,E*)-6-((((4-azidobenzyl)oxy)carbonyl)amino)-2-(1,3-dioxoisoindolin-2-yl)hex-5-enoate (12)

11 (1.2 g, 2.8 mmol) was dissolved in toluene (20 mL) and heated to reflux for 1 h. then 4-azidobenzyl alcohol (0.51 g, 3.4 mmol) was added. The reaction mixture was stirred at reflux for another 0.5 h. The product was cooled down to room temperature, extracted by ethyl acetate, washed by hydrochloride acid, sodium hydroxide and brine, dried by sodium sulfate, purified by a silica gel chromatography to afford a colorless oil. (0.74 g, 57%)

Methyl (*S,E*)-2-amino-6-((((4-azidobenzyl)oxy)carbonyl)amino)hex-5-enoate (2)

To a solution of **12** (0.4 g 0.86 mmol) in ethanol/dichloromethane (2mL/3mL) was added hydrazine monohydrate (0.066 mL, 1.36 mmol). The reaction mixture was stirred at room temperature overnight. Solvent was removed with water aspirator and the product was re-dissolved in THF (3 mL). Lithium hydroxide solution (1.0 N, 1.2 mL, 1.2 mmol) was added under ice bath. The reaction mixture was further stirred at room temperature for another 4 h. Hydrochloride acid was added to just the pH to be 7. And the water was removed by oil pump to afford a yellow solid. (0. 15 g, 52%)

Compound **1, 3 and 4** were synthesized with the same manner with the corresponding benzyl alcohols.

DNA Sequences

GFP_{UV}

Atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtca
gtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggcca
acacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatg
cccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggt
gatacccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaatacaactat
aactcacacaatgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagat
ggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtc
gacatagtctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacac
atggcatggatgaactctacaaagagctccatcaccatcaccatcactaa

pylT

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

Methanosarcina mazei PylRS

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacac
cacgaagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggact
gcaagagcgctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcac
aaaggcaaacgaagaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatcc
gttgcgagagccccgaaacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgataccg
gtttccacccaagagtcagtttctgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactg

gtaaaagggaatacgaaccccattacatccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactga
caggcttgaagtcctgttaaacccaaaagatgagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctct
ctcgcagaaaaaaagacctgcagcagatctacgcggaagaaagggagaattatctggggaaactcgagcgtgaaattaccag
gttctttgtggacaggggttttctggaaataaaatccccgatcctgatccctcttgagtatatcgaaaggatgggcattgataatgat
accgaactttcaaaacagatcttcagggttgacaagaacttctgcctgagacccatgcttgctccaaacctttacaactacctgcg
caagcttgacagggccctgcctgatccaataaaaatttttgaaataggcccatgctacagaaaagagtccgacggcaaagaac
acctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcacacgggaaaatcttgaaagcataattacggactt
cctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctatggggataccttgatgtaatgcacggagac
ctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaaccctggataggggcaggtttc
gggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtcttactataacgggattt
ctaccaacctgtaa

Construction of Plasmids

Constructions of pY+ and pY-

The plasmid pY+ was derived from the pRep plasmid by replacing the suppressor tRNA in pRep by pylT.[83] The gene of pylT flanked by the lpp promoter at the 5’ end and the *rrnC* terminator at the 3’ end was amplified from pBK-AcKRS-pylT.[77,84] The plasmid pY- was derived from the pNeg plasmid by replacing the suppressor tRNA with pylT.[83] Similarly, the gene of pylT flanked by the lpp promoter at the 5’ end and the *rrnC* terminator at the 3’ end was amplified from pBK-AcKRS-pylT.[23,85] pY+ has a tetracycline selection marker, a chloramphenicol acetyltransferase gene with an amber mutation at D112. pY- has an ampicillin selection marker and a barnase gene with two amber mutations at Q2 and D44. The barnase gene is under control of a pBad promoter.

Construction of pET-pylT-GFP

Plasmid pET-pylT-GFP was derived from the plasmid pAcKRS-pylT-GFP1Amber in which GFP_{UV} has an amber mutation at Q204.[23,85] The restriction enzyme *BglII* was used to cut off the ACKRS gene. The digested pAcKRS-pylT-GFP1Amber plasmid was ligated to form pET-pylT-GFP.

Construction of the mKRS1$^+$ Library

The plasmid pBK-mmPylRS that encodes wild-type *Methanosarcina mazei* PylRS was derived from a pBK plasmid containing *p*-iodophenylalanyl-tRNA synthetase.[86] The pylRS gene is under the control of *E. coli glnS* promoter and terminator. It was amplified from genomic DNA of *Methanosarcina mazei* strain DSM 3647 (ATCC) by flanking primers, pBK-mmPylRS-NdeI-F and pBK-mmPylRS-PstI~NsiI-R. To construct the mKRS1$^+$ library, NNK (N=A or C or G or T, K=G or T) mutations were introduced at three sites by overlap extension PCR.[87] The following pairs of primers were used to generate a PylRS gene library with randomization at three sites: (1) pBK-mmPylRS-NdeI-F (5'-gaatcccatatggataaaaaaccactaaacactctg-3') and mmPylRS-Mutlib01-R (5'-ggccctgtcaagcttgcgmnngtagttmnnmnngtttggagcaagca tggg-3'); (2) mmPylRS-Mutlib02-F (5'-cgcaagcttgacagggccctgcctgatcc-3') and mmPylRS-Mutlib03-R (5'-gcatcccgatcccatctgmnngaamnncagcatggtaaactcttc-3'); (3) pBK-mmPylRS-PstI~NsiI-R (5'-gtttgaaaatgcatttacaggttggtagaaatccc-3'). The gene library was digested with the restriction enzymes *NdeI* and *NsiI*, gel-purified, and ligated back into the pBK vector digested by *NdeI* and *PstI* to afford plasmid the *mKRS1$^+$* library. 1 μg of the ligation

products were then electroporated into *E. coli* Top10 cells. Electroporated cells were recovered in SOC medium for 60 min at 37 °C, transferred into a 2 L 2YT medium with kanamycin (25 μg/mL) and were incubated at 37°C to OD_{600} at 1.0. To calculate the library size, 1 μL recovered SOC culture was subjected to serial dilutions in 2YT, then plated on LB agar plates with kanamycin (25 μg/mL), and grown overnight in a 37°C incubator. Based on the colony numbers on these plates, the mKRS1$^+$ library contains approximately 1.01×10^9 independent transformants. Sequencing pylRS variants in 20 clones did not reveal any significant bias at the randomization sites.

Selection Procedure for Evolving Pyrrolysyl-tRNA Synthetase

The selections followed the scheme shown as of **Scheme II-4**. For the positive selection, the mKRS1$^+$ library was used to transform *E. coli* TOP10 competent cells harboring pY+ to yield a cell library greater than 1×10^9 cfu, ensuring complete coverage of the *mKRS1*$^+$ library. Cells were plated on LB agar plates containing 12 μg/mL tetracycline (Tet), 25 μg/mL kanamycin (Kan), 68 μg/mL chloramphenicol (Cm) and 1 mM **2**. After incubation at 37°C for 72 h, colonies on the plates were collected. Total plasmids were isolated and separated by 1 % agarose gel electrophoresis. mKRS1$^+$ plasmids were extracted using a gel-extraction kit (QIAGEN). The extracted mKRS1$^+$ plasmids from the positive selection were used to transform *E. coli* TOP10 harboring pY− for the negative selection. After electroporation, the cells were allowed to recover for 1 h at 37°C in SOC media before being plated on LB agar plates containing 50 μg/mL Kan, 200 μg/mL ampicillin (Amp) and 0.2% arabinose. The plates were incubated for 16 h at 37°C. Survived cells were then pooled and pRS1 plasmids were

extracted. The selection power to exclude out the mutants that also took endogenous amino acids was tested on LB agar plates containing 50 μg/mL Kan, 200 μg/mL Amp, 0.2% arabinose and 1mM **2**. The plate contains 1 mM **2** showed much fewer colony numbers as times of negative selections increased. Three alternative selections (two positive + one negative) finally yielded many colonies. 32 single colonies after the third positive selection were selected and the plasmids were isolated for sequencing. 32 single colonies from the third positive selection were also chosen for testing their ability to grow on plates with 102 μg/mL chloramphenicol, 25 μg/mL Kan, 12 μg/mL Tet, and 1 mM of **2**. A plate without NCAA supplementary was used as a control. Images of colonies growing on different plates were shown in **Figure 26**. Sequences of PylRS variants that charge pylT with different NCAAs are presented in **Table 4.**

Protein Expression and Purification

E. coli BL21(DE3) cells were transformed with pEVOL-pylT-PylRS and pET-sfGFPS2TAG and plated on LB agar plate containing chloramphenicol (Cam) (34 μg/mL) and ampicillin (Amp) (100 μg/mL). A 5 mL o/n culture was prepared from a single colony. This overnight culture was used to inoculate 500 mL of LB medium supplemented with Cam and Amp. Cells were grown at 37 ^{o}C in an incubator (250 r.p.m.) and the protein expression was induced with 1 mM IPTG, 0.2% arabinose and 5 mM **3** or **4** when OD_{600} reached 0.6. After 8 h induction, cells were harvested, resuspended in a lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and sonicated in an ice/water bath four times (4 min each, 10 min interval to cool the suspension below 10 ^{o}C between each pulse). The cell lysate was centrifuged (60

min, 11,000 g, 4 °C). The supernatant was incubated with 3 mL Ni-NTA resin (Qiagen) (2 h, 4 °C). The slurry was then loaded to a column and the protein-bound resin was washed with 30 mL of the wash buffer. Protein was finally eluted by the lysis buffer with 250 mM imidazole. Eluted fractions were collected, concentrated and buffer exchanged to 10 mM ammonium bicarbonate using an Amicon Ultra-15 Centrifugal Filter Devices (10k MWCO cut, Millipore). The protein purity was analyzed by 15% SDS-PAGE.

*Deprotection of sfGFP With **3** and **4***

To a solution of sfGFP with **3** (46 µM, 0.1 mL) was added sodium hydrosulfite (500 mM, 1 µL). The reaction was performed at room temperature for 1 h. Then the sfGFP was purified by Ni-NTA resin to remove the excessive salt.

To a solution of sfGFP with **4** (28 µM, 0.2 mL) was added TCEP (100 mM, 3 µL). The reaction was performed at room temperature for 1 h. Then the sfGFP was purified by Ni-NTA resin to remove the excessive TCEP.

Result and Discussion

We reported the genetic incorporation of N^{ε}-(2-nitrobenzyl)oxycarbonyl protected lysine with evolved mKRS1-tRNA$^{Pyl}_{CUA}$ pair[140], and we concluded that it should be able to take **3** & **4** as well. The nitro group in **3** and the azide group in **4** can be reduced to amino group with reducing reagent, which triggers the deprotection and affords the enamine undergoing rearrangement followed by hydrolysis. Allysine

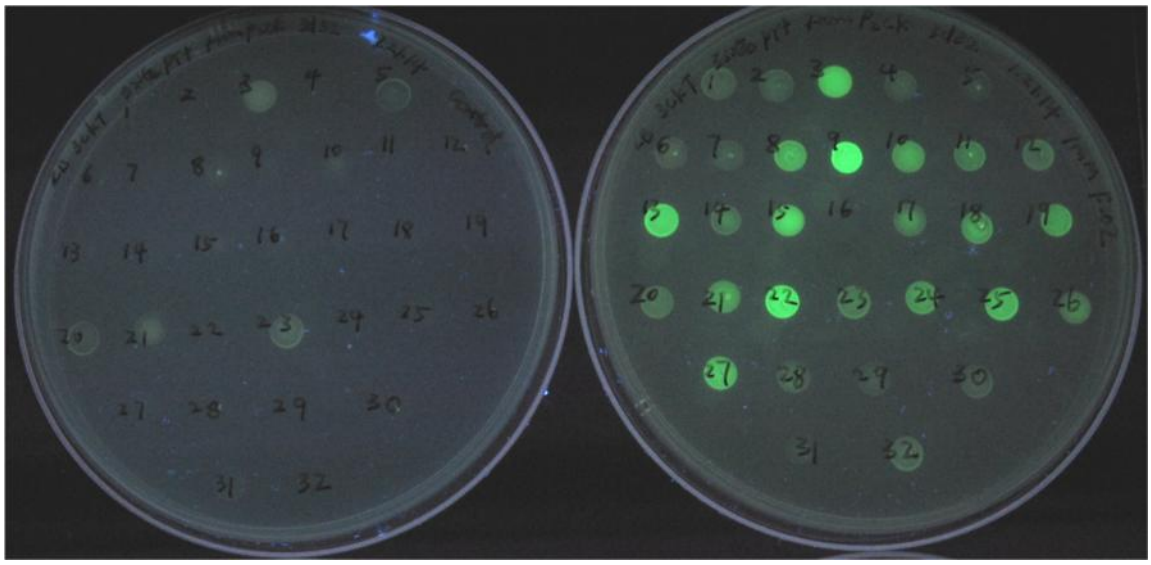


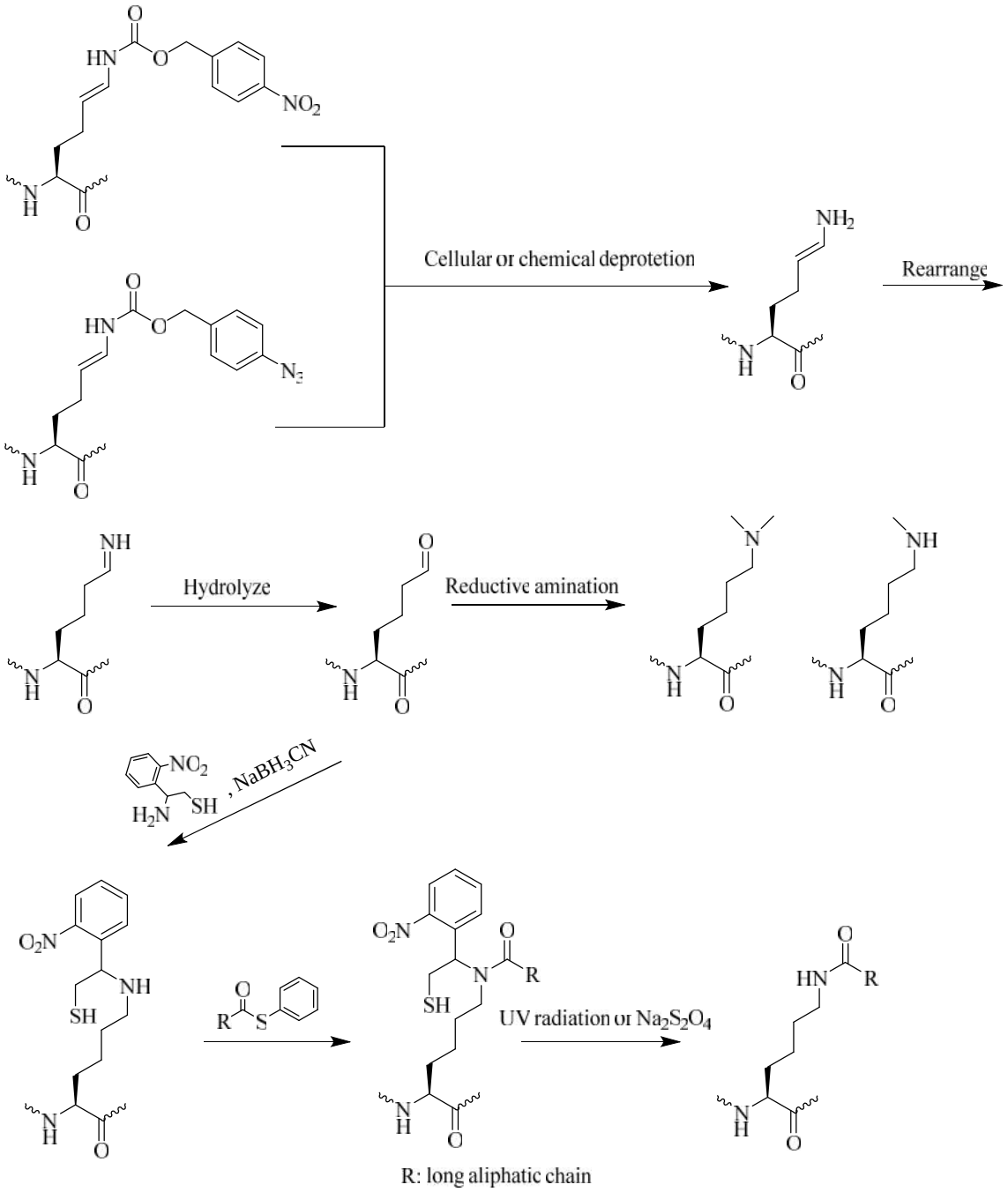
Figure 26. Screening after selection of **2**. Left : Cells grew on the plate without Non-canonical amino acids. Right Cells grew on the plate with 1 mM **2**.

Table 4. Selection result for **2**.

	306	309	348	384
No.9	Y	T	G	F
No. 13	Y	S	S	F

undergoes reactive amination with dimethyl amine to form dimethyl lysine (**Scheme V-3**). With 5 mM of either **3** or **4** supplemented in 2YT media, they were incorporated into sfGFPS2TAG with acceptable yield. Interestingly, we discovered that sfGFP was partially deprotected for both NCAAs after expression and purification. However, tryptophan was incorporated into the protein with **4** as an impurity and although **3** was encoded into sfGFP without the interference of tryptophan (**Figure27, 28**), the following deprotection with sodium hydrosulfite rendered oxidations on the protein (**Figure29, 30**). Virdee and Chin reported that (4-nitrobenzyl)oxycarbonyl group can be taken off inside of *E coli*. cell with endogenous nitro reductase[144]. Therefore, we designed and synthesized **1** & **2** as new candidates for the genetic encoding of Kme2. We created a new four-site library basing on mKRS1 for an elevated protein expression yield so that the miss-encoded of tryptophan is eliminated (**Figure 31**).

Out of the six sites picked up in our previous paper, we focused on four of them for the new library construction (**Figure 31**). Y306, L309 and Y384 were randomized to all the twenty amino acids while Y384 was mutated to phenylalanine, tyrosine and tryptophan. With selection as reported earlier[140], two mutants were selected and sequenced which showed as 306Y/309T/348G/384F and 306Y/309S/348S/384F (**Table 4**). **2** was encoded into sfGFPS2TAG and showed 85% of allysine with 15% of **2** after purification (**Figure 32**).



Scheme V-3. In-situ dimethyl lysine and acylated lysines formation.

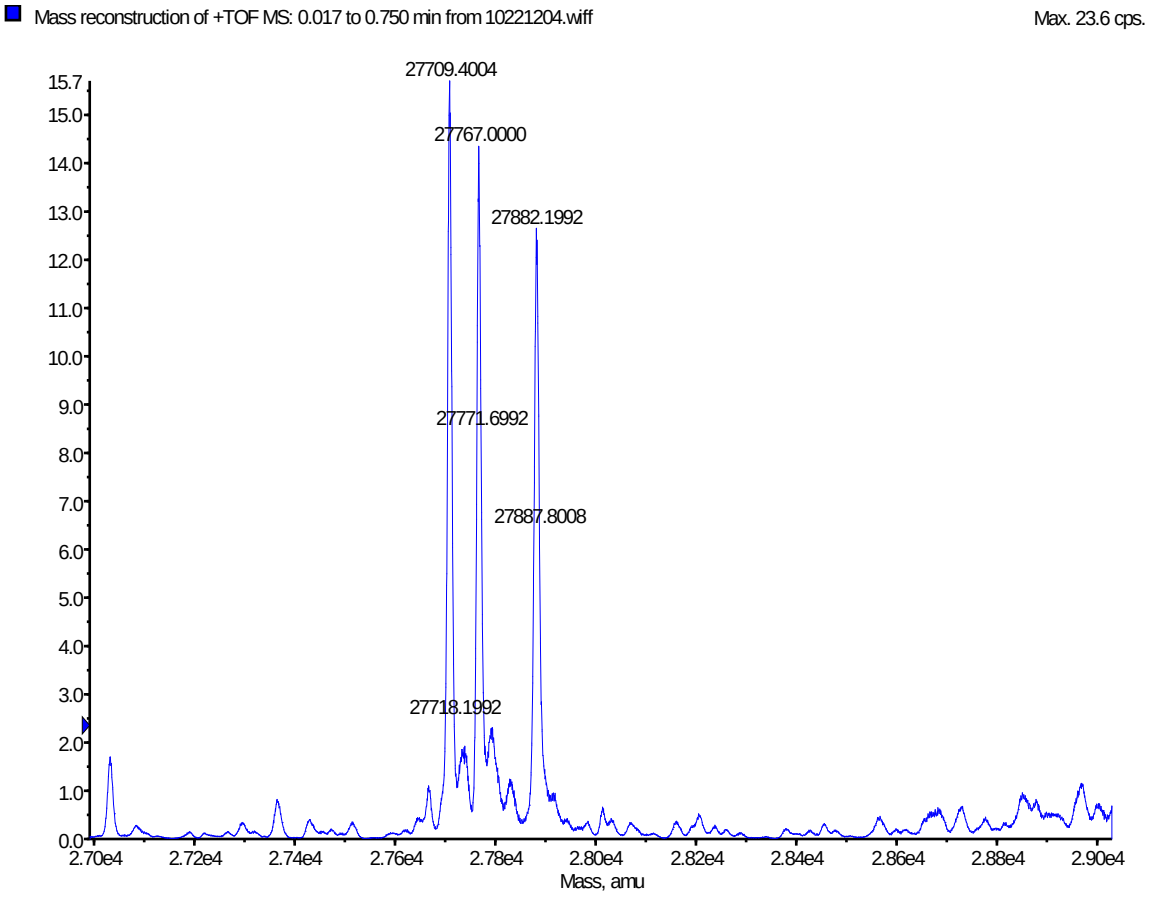


Figure 27. sfGFP with **4.** Calculated: 27882. Found: 27882. Deprotected: 27709. sfGFP with trptophan: 27767.

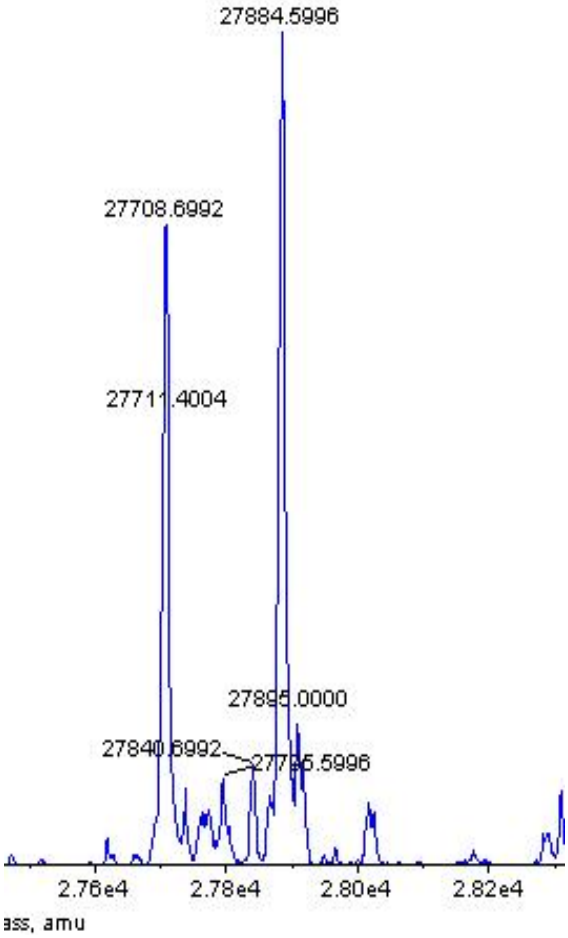


Figure 28. sfGFP with **3**. Calculated: 27886. Found: 27884. Deprotected :27709.

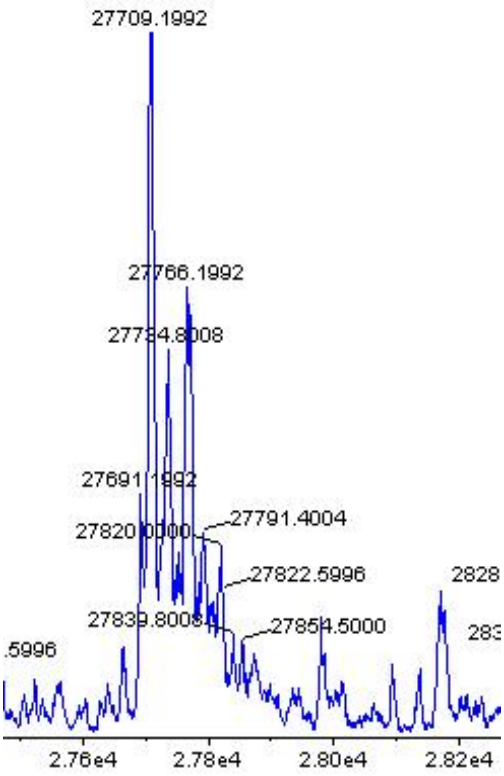


Figure 29. sfGFP with **4** after deprotection. Calculated: 27709. Found: 27709.

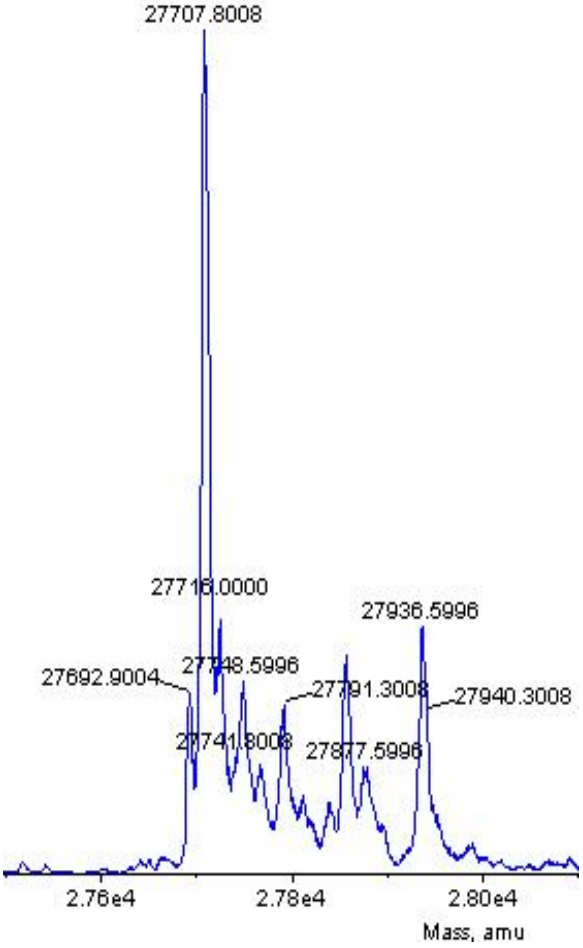


Figure 30. sfGFP with **3** after deprotection. Calculated: 27709. Found: 27708.

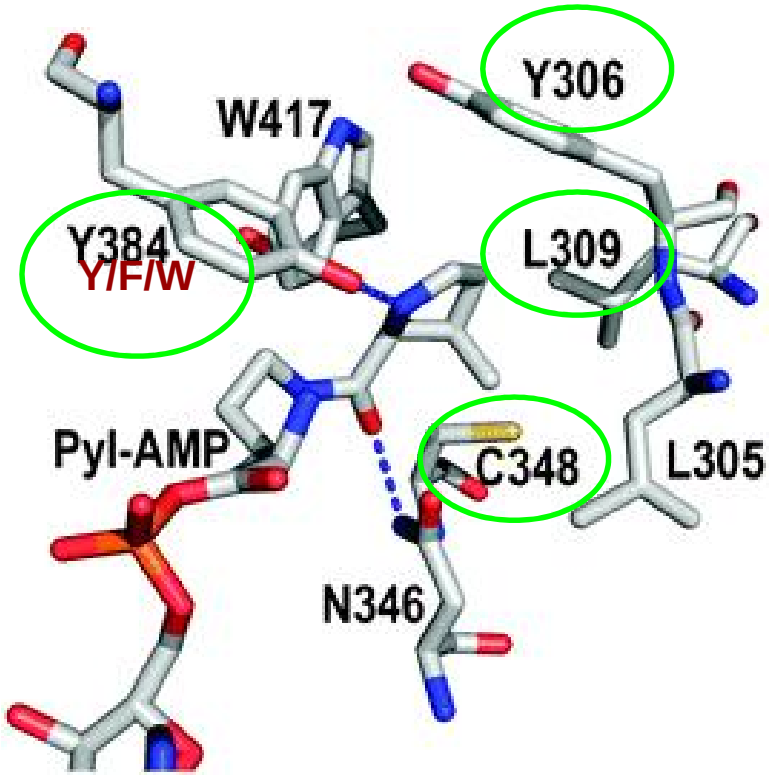


Figure 31. Four sites library for the incorporation of **1** & **2.**

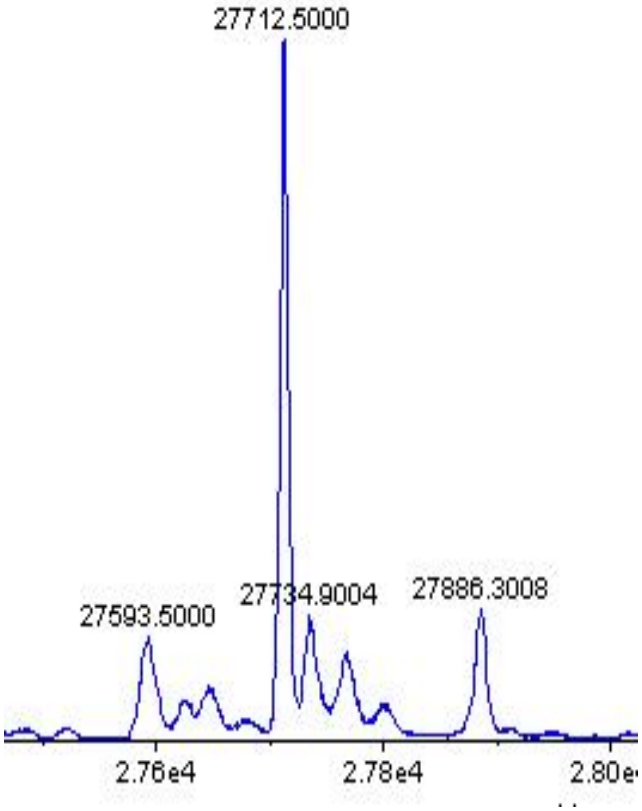


Figure 32. sfGFP with **2**. Calculated: 27709. Found: 27712.

Conclusion

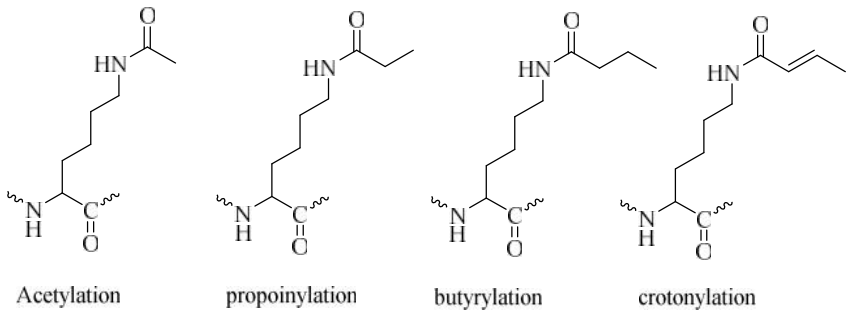
In this study, we demonstrate the synthesis of protected lysine mimics with a double bond which undergo different deprotections to afford allysine. With the recent evolved PylRS-tRNA$^{Pyl}_{CUA}$ pairs, the NCAAs are incorporated into sfGFP without canonical amino acids interference and at elevated yields. It is shown that the *p*-azidobenzyloxycarbonyl protected lysine mimic even got 85% deprotected after expression. With the deprotection with TCEP, the remain 15% of the protected lysine mimics should be able to be converted to allysine, which undergoes reductive amination and transthioesterification reactions to afford proteins with Kme2 and N^{ε}-long-chain acylated lysines.

CHAPTER VI

SIRTUIN 1 AND SIRTUIN 2 ARE UNIVERSAL H3 DEACYLASES

Introduction

More and more attention has been drawn towards histone acylations in the last decade given the fact that they dynamically control cellular activities.[145-147] Histone proteins undergo acetylations on lysine residues which neutralize the positive charge on them, loosening the binding between DNA and histone so that gene transcription gets initiated.[135,148] Sirtuins are NAD^+-dependent lysine deacetylases which control gene expression.[149] Propionylation and butyrylation are recognized as post-translational modifications (PTMs) on a variety of proteins, such as P300 and H4.[55] Furthermore, it has been discovered that propionylated H3 binds to bromodomain with a similar affinity as the same-site acetylated H3.[150] Crotonylation also has been reported to play an essential role as a mark on active sex chromosome-linked genes in postmeiotic male germ cells.[54] Despite of the blooming discovery of acylated proteins in mammalian cells, no one has yet unveiled the link between the specificity of deacylations and sirtuins, neither the site nor the modification. Here we take the benefit of our genetic NCAA encoding system to build up histone H3 with four known short-chain acylations (**Scheme VI-1**),[24] on which further we conducted deacylations with nuclear sirtuins (SIRT1, SIRT2, SIRT6 and SIRT7), so that we can obtain the deacylation level of a particular site with each sirtuin. Sirtuin from *thermotoga maritima* was introduced to the assay to investigate the differences in the substrate preferences between the enzymes from bacteria and eukaryotes.



Scheme VI-1. Short-chain acylations investigated in this study.

All nuclear sirtuins have been demonstrated that they are involved in histone deacylations. SIRT1 was found to target H3K9 and H4K16 and SIRT2 was capable to take off H4K16 acetylation.[151,152] Recently, SIRT6 and SIRT7 have also been discovered involved in H3 deacetylation at specific sites.[153-155] However, the information we obtained is still rather scattered. We decided to measure the deacylation ability of all the sirtuins in nucleus with four short-chain acylated (acetylated, propionylated, butyrylated, crotonylated) H3 with most of the lysine sites (K4, K9, K14, K18, K23, K27, K36, K56 and K79) so that we can build up a H3 deacylation map with sirtuins, which may give us a better insight about how the primary and the secondary structure of proteins affect the preference of deacylations by sirtuins.

Experimental Section

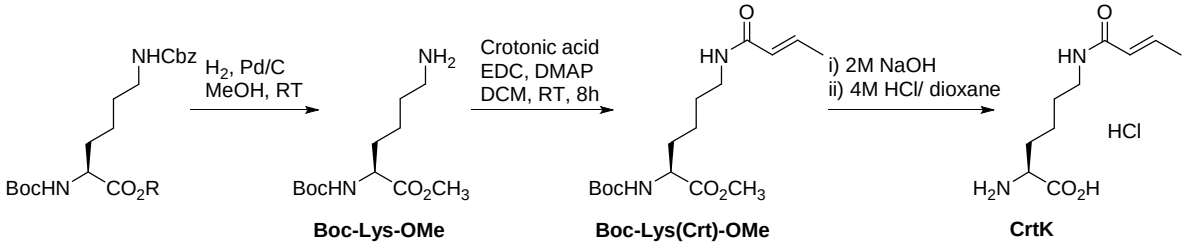
General Experimental

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was performed on Whatman SiO_2 60 F-254 plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Specific rotations of chiral compounds were obtained at the designated concentration and temperature on a Rudolph Research Analytical Autopol II polarimeter using a 0.5 dm cell. Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C), methanol (3.31 ppm for ^{1}H and 49.15 ppm for ^{13}C), DMSO (2.50 ppm for ^{1}H and 39.51 ppm for ^{13}C), or water (4.80 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, and a minimal amount of sodium hydroxide pellet or concentrated hydrochloric acid was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral conditions. ^{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 obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

Chemical Synthesis



Scheme VI-2: Synthesis of CrtK

Boc-Lys-OMe

To a solution of Boc-Lys(Cbz)-OMe (3.9 g, 10 mmol) in methanol was added palladium on activated carbon (Pd 10%, 0.6 g, 0.6 mmol) and the suspension was stirred with hydrogen bubbled through for 3 h. The product was filtered with celite and concentrated to be an almost colorless oil. The product was directly used in the next step without further purification.

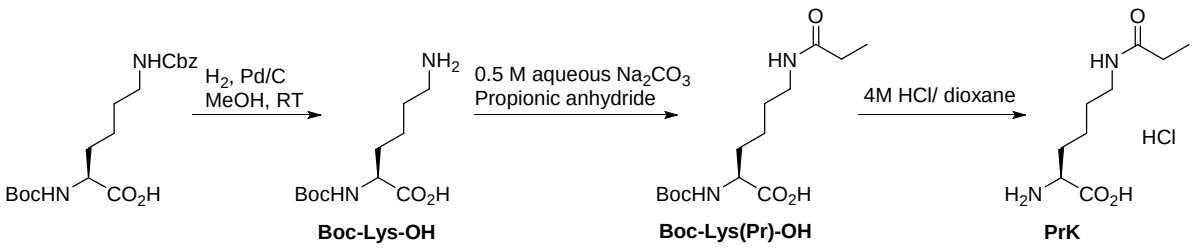
Boc-Lys(Crt)-OMe

To a solution of Boc-Lys-OMe (2.6 g, 10 mmol) in dichloromethane (40 mL, dried with calcium hydride) was added crotonic acid (1.76 g, 20 mmol), followed by 4-dimethylaminopyridine (50 mg, 0.4 mmol) and N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC, 2.5 g, 13 mmol) in dichloromethane (10 mL) dropwise. The reaction mixture was stirred at RT under a balloon of argon overnight. The product was extracted by ethyl acetate (100 mL) and washed with water (60 mL twice), brine (60 mL), dried with sodium sulfate, concentrated and chromatographed (EtOAc/hexanes, 1:3) to give a colorless oil (2.8 g, 85%).^{1}H NMR ($CDCl_3$, 300 MHz) δ 6.83 (m, 1H), 5.79 (d, 1H, J = 15.0 Hz), 5.66 (s, 1H), 5.13 (d, 1H, J = 8.1 Hz), 4.26 (m, 1H), 3.72 (s, 3H), 3.29 (quart, 2H, J = 6.3 Hz), 1.83 (dd, J = 6.9, 1.8 Hz, 3H), 1.69-1.43 (m, 4H), 1.41 (s, 9H), 1.39 (m, 2H).

CrtK

To a solution of Boc-Lys(Cr)-OMe (2.8 g, 8.5 mmol) in THF (4 mL) was added sodium hydroxide solution (2 M, 20 mL, 40 mmol) dropwise in an ice bath. The reaction mixture was stirred at 0 ^{o}C for 1h then at RT for another 2h. The organic impurities were washed by ethyl acetate (40 mL twice) and then the pH of the aqueous layer was adjusted to 3 with 3 M aqueous HCl. Then the product was extracted by ethyl acetate (60 mL then another 20 mL), washed with water (40 mL twice), brine (40 mL), dried with anhydrous sodium sulfate, and concentrated to give a sticky colorless oil, which was then dissolved in dioxane (5 mL). Hydrochloric acid in dioxane (4 M, 5 mL, 20 mmol) was then added and the reaction mixture was stirred at RT for 1h. The solvent was

removed to give an almost colorless oil (1.7g, 94%). The product was applied to biological studies without further purification. ^{1}H NMR (D_2O, 300 MHz) δ 6.78 (m, 1 H), 5.93 (d, 1 H, *J* = 15.0 Hz), 4.01 (t, 1 H, *J* = 6.3 Hz), 3.26 (t, 2 H, *J* = 6.6Hz), 1.95 (m, 2 H), 1.85 (dd, 3 H, *J* = 6.9, 1.5 Hz), 1.59 (m, 2 H), 1.46 (m, 2 H). ^{13}C NMR (75 MHz, D_2O) δ 175.8, 169.6, 142.2, 124.5, 55.3, 39.5, 30.9, 28.7, 22.4, 17.7. HRMS (ESI) calcd. for $C_{10}H_{19}N_2O_3$ $[M+H]^+$ 215.1390, found 215.1387.



Scheme VI-3. Synthesis of PrK.

Boc-Lys-OH

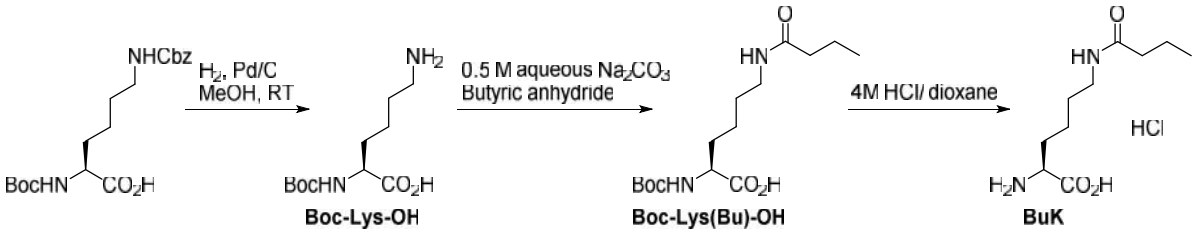
To a solution of Boc-Lys(Cbz)-OH (3.8 g, 10 mmol) in methanol (50 mL) was added palladium on activated carbon (Pd 10%, 0.6 g, 0.6 mmol) and the suspension was stirred with hydrogen bubbled through for 3 h. The product was filtered with celite and concentrated to be an almost colorless oil. The product was directly used in the next step without further purification.

Boc-Lys(Pr)-OH

Boc-Lys-OH (2.4 g, 9.8 mmol) was dissolved in the solution of sodium carbonate (0.5 M, 40 mL) and then propionic anhydride (2.0 mL, 15.7 mL) was added dropwise into the solution. The reaction mixture was stirred at RT for 2h. Sodium hydroxide (2 M, 20 mL) solution was added and the excessive propanoic anhydride was extracted by ethyl acetate. Then the pH of the aqueous phase was adjusted by hydrochloric acid to pH 3. The product was extracted by ethyl acetate (80 mL) and concentrated to give a colorless oil. The product was directly used in the next step without further purification.

PrK

To a solution of Boc-Lys(pr)-OH (2.8 g, 9.3 mmol) in dioxane (5 mL) was added hydrochloric acid in dioxane (4 M, 5 mL, 20 mmol). The reaction mixture was stirred at RT for 2h. A white sticky solid precipitated out as the reaction went. The liquid was filtered off and the solid was washed with excess ethyl acetate (200 mL). The product was then dried under oil pump overnight to afford a slightly yellow solid (1.7 g, 90%). ^{1}H NMR (D_2O, 300 MHz) δ 4.05 (t, 1 H, *J* = 6.3 Hz), 3.16 (t, 2 H, *J* = 6.9Hz), 2.20 (quart, 2 H, *J* = 7.8Hz,), 1.93 (m, 2 H), 1.52 (m, 2 H), 1.44 (m, 2 H). ^{13}C NMR (75 MHz, D_2O) δ 175.6, 173.0, 53.6, 39.3, 30.0, 29.8, 28.5, 22.1, 10.3. HRMS (ESI) calcd. for $C_9H_{19}N_2O_3$ $[M+H]^+$ 203.1390, found 203.1403.



Scheme VI-4. Synthesis of N^{ε}-butyryl-L-lysine (BuK)

BuK was prepared in the same manner as *PrK* except that a cation exchange column was applied to purify the final product. ^{1}H NMR (D_2O, 300 MHz) δ 3.72 (t, 1 H, J = 6.3 Hz), 3.20 (t, 2 H, J = 6.9 Hz), 2.21 (t, 2 H, J = 7.5 Hz,), 1.87 (m, 2 H), 1.59 (m, 4 H), 1.42 (m, 2 H), 0.90 (t, 2 H, J = 7.5 Hz). ^{13}C NMR (75 MHz, D_2O) δ 177.6, 175.3, 55.3, 39.5, 38.3, 30.7, 28.7, 22.4, 19.7, 13.3. HRMS (ESI) calcd. for $C_{10}H_{21}N_2O_3$ $[M+H]^+$ 217.1547, found 217.1550.

Primer and Gene Sequences

Primer Sequences

K4TAG-F TTCCAGGCTCGCACCTAGCAGACTGCTCGTAAG

K4TAG-R CTTACGAGCAGTCTGCTAGGTGCGAGCCTGGAA

K9TAG-F AAACAGACTGCTCGTTAGTCCACTGGCGGTAAA

K9TAG-R TTTACCGCCAGTGGACTAACGAGCAGTCTGTTT

K14TAG-F AAGTCCACTGGCGGTTAGGCGCCGCGTAAACAG

K14TAG-R CTGTTTACGCGGCGCCTAACCGCCAGTGGACTT

K18TAG-F GGTAAAGCGCCGCGTTAGCAGCTGGCAACCAAG

K18TAG-R CTTGGTTGCCAGCTGCTAACGCGGCGCTTTACC

K23TAG-F AAACAGCTGGCAACCTAGGCAGCGCGTAAAAGC

K23TAG-R GCTTTTACGCGCTGCCTAGGTTGCCAGCTGTTT

K27TAG-F ACCAAGGCAGCGCGTTAGAGCGCTCCAGCTACT

K27TAG-R AGTAGCTGGAGCGCTCTAACGCGCTGCCTTGGT

K36TAG-F GCTACTGGCGGCGTGTAGAAGCCGCACCGTTAT

K36TAG-R ATAACGGTGCGGCTTCTACACGCCGCCAGTAGC

K56TAG-F ATCCGCCGCTACCAGTAGAGCACCGAACTGCTG

K56TAG-R CAGCAGTTCGGTGCTCTACTGGTAGCGGCGGAT

K79TAG-F ATTGCTCAGGATTTCTAGACCGACCTGCGCTTC

K79TAG-R GAAGCGCAGGTCGGTCTAGAAATCCTGAGCAAT

K115TAG-F GCAGCAATCCATGCTAGGCGTGTAACCATTATG

K115TAG-R CATAATGGTTACACGCCTAGCATGGATTGCTGC

Histag-TEV-H3 opt-F

GATCCGGAAAATCTGTACTTCCAGGCTCGCACCAAACAG

Histag-TEV-H3 opt-R GCAACAAGCTTTCACGCACGCTCACCACGGATACGACG

Gene sequences

Histag-TEV-H3 opt

ATGGGCAGCAGCCATCACCATCATCACCACAGCCAGGATCCGGAAAATCTG

TACTTCCAGGCTCGCACCAAACAGACTGCTCGTAAGTCCACTGGCGGTAAAG

CGCCGCGTAAACAGCTGGCAACCAAGGCAGCGCGTAAAAGCGCTCCAGCTA

CTGGCGGCGTGAAGAAGCCGCACCGTTATCGCCCGGGTACTGTGGCTCTGCG

TGAAATCCGCCGCTACCAGAAAAGCACCGAACTGCTGATTCGCAAACTGCC

ATTTCAACGTCTGGTTCGCGAAATTGCTCAGGATTTCAAAACCGACCTGCGC

TTCCAGTCTAGCGCTGTGATGGCACTGCAAGAGGCGTCTGAGGCATATCTGG

TTGGCCTGTTCGAAGATACCAACCTGGCAGCAATCCATGCAAAGCGTGTAAC

CATTATGCCGAAAGACATCCAACTGGCTCGTCGTATCCGTGGTGAGCGTGCG

MmPrKRS

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGT

CCAGGACCGGAACAATTCATAAAATAAAACACCACGAAGTCTCTCGAAGCA

AAATCTATATTGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCCAG

GAGCAGCAGGACTGCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTG

CAAACGCTGCAGGGTTTCGGATGAGGATCTCAATAAGTTCCTCACAAAGGCA

AACGAAGACCAGACAAGCGTAAAAGTCAAGGTCGTTTCTGCCCCTACCAGA

ACGAAAAAGGCAATGCCAAAATCCGTTGCGAGAGCCCCGAAACCTCTTGAG

AATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTTTCACCTGCGA

TACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCA

ATATCAAGCATTTCTACAGGAGCAACTGCATCCGCACTGGTAAAAGGGAAT

ACGAATCCCATTACATCCATGTCTGCCCCTGTTCAGGCAAGTGCCCCCGCAC

TTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATG

AGATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCT

CTCTCGCAGAAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAA

TTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGT

TTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGA

TGGGCATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAA

GAACTTCTGCCTGAGACCCATGATGGCTCCAAACCTGCTGAACTACGCCCGC

AAGCTTGACAGGGCCCTGCCTGATCCAATAAAAATTTTTGAAATAGGCCCAT

GCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGC

TGAACTTCTGCCAGATGGGATCGGGATGTACACGGGAAAATCTTGAAAGCA

TAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGGCGA

TTCCTTCATGGTCTTGGGGGATACCCTTGATGTAATGCACGGAGACCTGGAA

CTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTG

ATAAACCCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAA

ACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAAC

GGGATTTCTACCAACCTGTAA

MmBuKRS

ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGT

CCAGGACCGGAACAATTCATAAAATAAAACACCACGAAGTCTCTCGAAGCA

AAATCTATATTGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCCAG

GAGCAGCAGGACTGCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTG

CAAACGCTGCAGGGTTTCGGATGAGGATCTCAATAAGTTCCTCACAAAGGCA

AACGAAGACCAGACAAGCGTAAAAGTCAAGGTCGTTTCTGCCCCTACCAGA

ACGAAAAAGGCAATGCCAAAATCCGTTGCGAGAGCCCCGAAACCTCTTGAG

AATACAGAAGCGGCACAGGCTCAACCTTCTGGATCTAAATTTTCACCTGCGA

TACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCATCTGTTTCAACATCA

ATATCAAGCATTTCTACAGGAGCAACTGCATCCGCACTGGTAAAAGGGAAT

ACGAACCCCATTACATCCATGTCTGCCCCTGTTCAGGCAAGTGCCCCCGCAC

TTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATG

AGATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCT

CTCTCGCAGAAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAA

TTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGT

TTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGA

TGGGCATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAA

GAACTTCTGCCTGAGACCCATGCTTGCTCCAAACCTTTACAACTACCTGCGC

AAGCTTGACAGGGCCCTGCCTGATCCAATAAAAATTTTTGAAATAGGCCCAT

GCTACAGAAAAGAGTCCGACGGCAAAGAACACCTCGAAGAGTTTACCATGC

TGAACTTCTGCCAGATGGGATCGGGATGCACACGGGAAAATCTTGAAAGCA

TAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCAAGATCGTAGGCGA

TTCCTGCATGGTCTGGGGGGATACCCTTGATGTAATGCACGGAGACCTGGAA

CTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTG

ATAAACCCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAA

ACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAAC

GGGATTTCTACCAACCTGTAA

Histag-TEV-H3 opt protein sequence

MGSSHHHHHHSQDPENLYFQARTKQTARKSTGGKAPRKQLATKAARKSAPAT

GGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSS

AVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

Construction of Plasmids

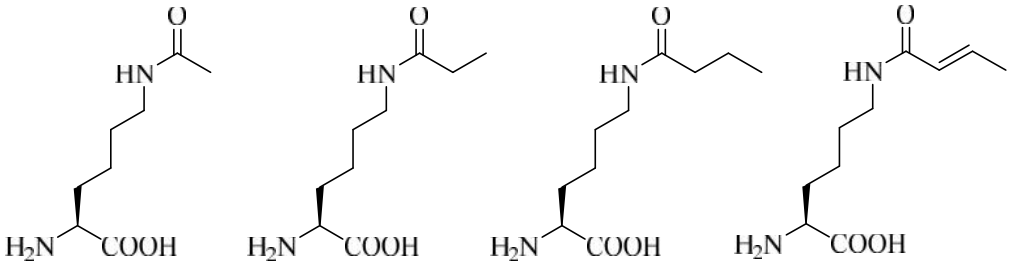
Construction of pETduet-histag-TEV-wtH3

Codon optimized human Histone H3 was purchased from Epoch Life Science, Inc. (Sugar Land, TX). Histag-TEV-H3opt-F and histag-TEV-H3opt-R were used to clone the H3 gene into pETduet-1. With BamH I and Hind III as the restriction sites.

Construction of pETduet-histag-TEV-H3mutants

All the H3 mutants were constructed with PFU-quickchange PCR with their corresponding primers.

Protein Expression of Purification



Scheme VI-5. Short-chain acylated lysines.

Expression of Histone H3 Mutants with 1

E. Coli. BL21 cells co-transformed with pETduet-histag-TEV-H3 mutants and pEVOL-mmAcKRS were grown in 2YT medium (12 mL) with 100 µg/mL ampicillin and 34 µg/mL chloramphenicol overnight. The culture was inoculated into 2YT medium (190 mL) with same concentration of antibiotics. IPTG (500 mM), together with Arabinose (0.2% w/v), **1** (5 mM) and nicotinamide (5 mM) was added into the cell culture after the OD reached to 0.6. The cell culture was incubated at 37 ^{o}C for overnight, and the cells were harvested and by centrifugation at 4000 r.p.m. for 20 min at 4 ^{o}C and re-suspended in 20 mL of lysis buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1% Triton-100, pH 7.8). The re-suspended cells were sonicated with ice water bath twice (4 min each, 10 min interval to cool the suspension below 10 ^{o}C before the next run) and the lysate was clarified by centrifugation at 6000 r.p.m. for 20 min at 4 °C. The supernatant was discarded and the pellet was re-suspended with 20 mL lysis buffer, then it was spun down again with 6000 r.p.m. for 20 min at 4°C. The wash was repeated twice. Then the pellet was washed in the same way with wash buffer (50 mM Tris-HCl, 100 mM NaCl, pH 7.8) three times. The final pellet of inclusion body was dissolved in urea buffer (50 mM Tris-HCl, 100 mM NaCl, 8 M urea, pH 7.8) and the solution was clarified by centrifugation at 10000 r.p.m for 30 min at 4°C, then incubated with 3 mL of Ni Sepharose^{TM} 6 Fast Flow from GE Healthcare (Little Chalfont, United Kingdom) for 1 h, and then washed with 50 mL of Ni-NTA wash buffer (50 mM Tris-HCl, 100 mM NaCl, 6 M urea, 20 mM imidazole, pH 7.8). $H3_{AcK}$ mutants was then eluted out with 8 mL of elution buffer (50 mM Tris-HCl, 100 mM NaCl, 250 mM imidazole, 6 M Urea,

pH 7.8) and concentrated by Amicon Ultra-15 Centrifugal Filter Units – 10,000 NMWL from Millipore (Billerica, MA) to 1 mL.. The concentration was determined by BCA protein assay kit from Thermo Fisher Scientific Inc. (Rockford, IL). According to the concentration, $H3_{AcK}$ mutants' expression yield varied from 5-40 mg/L in 2YT. H3 protein was precipitated out by acetone (9 mL, 90%) and re-dissolve in arginine buffer (2 M arginine, 80 mM Tris-HCl, 1 M NaCl, pH 7.8) for further reactions.

Expression of Histone H3 Mutants with 2

The expression of histone H3 mutants with **2** follows the same way as **1** except **2** (1 mM) was supplied into the media instead of **1**.

Expression of Histone H3 Mutants with 3

E. Coli. BL21 cells co-transformed with pETduet-histag-TEV-H3 mutants and pEVOL-mmBuKRS were used for the expression of histone H3 mutants with **3**. The expression of histone H3 mutants with **3** follows the same way as **1** except **3** (1 mM) was supplied into the media instead of **1**.

Expression of Histone H3 Mutants with 4

E. Coli. BL21 cells co-transformed with pETduet-histag-TEV-H3 mutants and pEVOL-mmBuKRS were used for the expression of histone H3 mutants with **4**. The expression of histone H3 mutants with **4** follows the same way as **1** except **4** (1 mM) was supplied into the media instead of **1**.

Expression of Sirtuin 1

pQE-SIRT1 was purchased from Addgene (Cambridge, MA) and transformed into BL21 ($codon^+$) cells. SIRT1 was expressed overnight at 16^oC started with OD at

0.6. The cells was spun down, sonicated in the lysis buffer (50 mM Tris-HCl, 250 mM of NaCl, 20 mM imidazole, pH 7.5) and purified by a Ni Sepharose^{TM} 6 Fast Flow column. The eluent was concentrated, dialyzed against phosphate buffer (50 mM sodium phosphate, pH 7.5), further purified with a superdex 200 10/300 GL from GE Healthcare (Little Chalfont, United Kingdom) and finally concentrated into a stock solution at 1 mg/mL concentration with 50% glycerol and then stored in -80^{o}C. All the other sirtuins are purchased from Sigma-Aldrich (St. Louis, MO).

Deacylation Reactions

H3 mutant sample (30 µL) was mixed with sfGFP (3 µL, 1 mg/mL) to make the reaction stock solution. Then the mix was distributed evenly to six aliquots (5 µL), labelled as control, sir2tm, sirtuin 1, sirtuin 2, sirtuin 6 and sirtuin 7, respectively. All the reaction tubes were supplied with DTT (1 mM), NAD^+ (1 mM) and the corresponding sirtuin (400 nM) except the control. Water was added to the reaction tubes to make the final volume to be 20 µL. The deacylation reactions were performed at 37^{o}C overnight. The reaction mixtures were directly subjected to the protein gel and transferred to the nitrocellulose membrane with standard semi dry western blot protocol. The membrane was coated with 5% fat-free milk (10 mL) for 2 h and then treated with acylation antibody from PTM bio-lab (Chicago, IL) overnight at 4^{o}C(1:2000, 5 mL). The membrane then was washed by PBST buffer (PBS with 0.1% tween-20, 10 mL) on the shaker six times with 10 min intervals. Then the membrane was treated with second antibody (1: 10000, 5 mL) from Jackson ImmunoResearch (West Grove, PA) at room temperature for 1 h. The membrane then was washed by PBST buffer (PBS with 0.1%

tween-20, 10 mL) on the shaker three times with 10 min intervals. And then the result was visualized with Pierce ECL Western Blotting Substrate (Rockford, IL). Images were taken by ChemiDoc XRS+ system from Bio-Rad (Hercules, CA).

Result and Discussion

Previously our group reported that we evolved a mutant of PylRS-tRNA$^{Pyl}_{CUA}$ pair from the wild type that is capable to genetically encode butyryllysine (BuK) and crotonyllysine (CrtK) with an elevated yield comparing with the wild type pair.[24] We also discovered that mmAcKRS1 evolved from *Methanosarcina mazei* PylRS also recognizes propionyllysine (PrK) as a substrate.[118] A His-tag was installed in front of the H3 gene for purification convenience. A TEV cleavage site was place between the His-tag and H3 so that the His-tag can be removed after the refolding of the nucleosome. All the H3 mutants were expressed with a yield ranging from 5 mg/L to 40 mg/L after Ni-NTA resin purification. The expression levels of the mutants with PrK were slightly lower than the ones with BuK and CrtK and mutants at K36 position afforded the lowest expression yield.

SIRT1 and SIRT2 exhibited astonishingly high reactivities towards most of the acylated H3s (**Table 5-8, Figure 33-36**), regardless sites or type of acylations, especially to the ones in the N-terminal free-loop region. To both the sites (K56, K79) in the core region, both SIRT1 and SIRT2 showed significantly decreased reactivities. We have also observed that SIRT2 has a higher reactivity against butyrylated and crotonylated substrates than SIRT1.

SIRT6 and SIRT7 displayed rather low deacylation reactivities. We did observe elevated deacylation level at certain site with specific type of acylation, we cannot find a consistency nonetheless. Considering that all the reactions were performed overnight, we expected an effective deacylase to be able to revert acylated lysines back to lysine. Moreover, there was no apparent deacylations at K79 by SIRT6 or SIRT7 which was consistent with the reactivities decrease of SIRT1 and SIRT2 at the same site.

It has been a question scientists tried to answer for a long time that how sirtuins choose their substrates with specificity. Given that 7 out of 11 H3 lysines are located in the N-terminal free loop region,[53] the difference of protein primary structures seems to be the reason behind it. Garske and Denu used a 5-mer library to find the 'favorite' sequences for SIRT1.[156] They discovered that certain sequences are more prone to deacetylation, with an order higher of reactivity comparing with the rest of the candidates. However, those 'top hits' do not exist in H3. From our data, it shows that SIRT1 and SIRT2 efficiently deacylated all the acylations in the free loop region of H3. We conclude that even if SIRT1 and SIRT2 have a sequence preference on deacylation, it is a rather minor effect of the substrate selectivity with the case of H3. Whenever it comes to the core region, both SIRT1 and SIRT2 demonstrated decreased reactivities. We postulate that the secondary structure of the H3 affects the reactivities of SIRT1 and SIRT2 to the acylation sites. It is very likely that the steric hindrance created by the α-helices in H3 impedes the binding between SIRT1 and SIRT2 to the sites in the core region. More interestingly, SIRT1 and SIRT2 have higher depropionylation and debutyrylation reactivities than those of deacetylation and decrotonylation. It seems that both two

Table 5. Histone H3 mutants with AcK deacylated with different sirtuins.

	Sir2tm	Sirtuin 1	Sirtuin 2	Sirtuin 6	Sirtuin7
K4Ac	W W W	W W W	W W W	W W V	W V V
K9Ac	W W W	W W W	W W W	W V V	W W V
K14Ac	W W W	W W W	W W W	W V V	W V V
K18Ac	W W W	W W W	W W W	V V V	V V V
K23Ac	W W W	W W W	W W W	V V V	V V V
K27Ac	W W W	W W V	W W V	V V V	V V V
K36Ac	W V V	V V V	W V V	V V V	V V V
K56Ac	V V V	V V V	W V V	V V V	V V V
K79Ac	W W V	W V V	W V V	V V V	V V V

W W W: over 90% deacylation. W W V: 40-90% deacylation. W V V: 10-40% deacylation
V V V: less than 10% deacylation.

Table 6. Histone H3 mutants with PrK deacylated with different sirtuins.

	Sir2tm	Sirtuin 1	Sirtuin 2	Sirtuin 6	Sirtuin7
K4Pr	W W V	W W V	W W V	W V V	W W V
K9Pr	W W W	W W W	W W W	W V V	W V V
K14Pr	W W W	W W W	W W W	W W V	W W V
K18Pr	W W W	W W W	W W W	W W V	W V V
K23Pr	W W W	W V V	W W W	V V V	V V V
K27Pr	W W W	W W W	W W W	V V V	V V V
K36Pr	W W V	W W W	W W W	V V V	W W V
K56Pr	W W V	W V V	W W V	V V V	W V V
K79Pr	W W V	W W V	W W V	V V V	W V V

W W W**:** over 90% deacylation. W W V**:** 40-90% deacylation. W V V**:** 10-40% deacylation
V V V**:** less than 10% deacylation.

Table 7. Histone H3 mutants with BuK deacylated with different sirtuins.

	Sir2tm	Sirtuin 1	Sirtuin 2	Sirtuin 6	Sirtuin7
K4Bu	W W W	W W V	W W W	W V V	W W V
K9Bu	W W V	W W V	W W W	W W V	W V V
K14Bu	W W W	W W V	W W V	W V V	W W V
K18Bu	W W W	W W V	W W W	W W V	W V V
K23Bu	W W W	W V V	W V V	V V V	V V V
K27Bu	W W W	W W V	W W W	W W V	W W V
K36Bu	W W W	W W V	W W W	W W V	W W V
K56Bu					
K79Bu	W W V	W W V	W W V	V V V	V V V

W W W: over 90% deacylation. W W V: 40-90% deacylation. W V V: 10-40% deacylation
V V V: less than 10% deacylation.

Table 8. Histone H3 mutants with CrK deacylated with different sirtuins.

	Sir2tm	Sirtuin 1	Sirtuin 2	Sirtuin 6	Sirtuin7
K4Cr	W W W	W W V	W W W	W V V	W V V
K9Cr	W W W	W W V	W W V	W V V	V V V
K14Cr	W W V	V V V	W W V	W V V	W V V
K18Cr	W W W	W W W	W W W	W W V	W W V
K23Cr	W W W	W W V	W W V	V V V	V V V
K27Cr	V V V	W V V	W W W	W V V	W V V
K36Cr	W W V	W W V	W W V	W W V	W W V
K56Cr	W W V	W W V	W W V	W W V	W W V
K79Cr	W W V	W V V	W W V	V V V	V V V

W W W**:** over 90% deacylation. W W V**:** 40-90% deacylation. W V V**:** 10-40% deacylation
V V V**:** less than 10% deacylation.

enzymes prefer longer chain acylated lysines as the substrate. We postulate that crotonylation is hard to be removed because of the rigid structure of the crotonyl group.

Comparing with SIRT1 and SIRT2, the deacylation reactivities of SIRT6 and SIRT7 are significantly lower. Given the fact that the deacylation reactions were performed overnight, we determined that SIRT6 and SIRT7 have very poor deacylation reactivity, if there is any. Even though they have mid-level reactivity towards some specific sites with certain acylations, SIRT1 and SIRT2 are still much better deacylases to those sites. Therefore, with the presence of SIRT1 and SIRT2 in nucleus, SIRT6 and SIRT7 are very unlikely served as H3 deacylases for short-chain acylated lysines although they share a conserved catalytic domain as the rest of the sirtuins. Frye's study strengthens our hypothesis.[157] It was found that the SIRT6 and SIRT7 sequences are much more similar to each another than they are to Sir2 from yeast. Consequently, SIRT6 and SIRT7 possibly have evolved to serve other functions we have not discovered.

So what about SIRT1 and SIRT2? Numerous efforts have been made to solve out their specificities. Nevertheless, here we present that they have similar reactivity towards all the sites in the free loop region. Basing on our preliminary study, SIRT1 and SIRT2 resulted in a similar level of deacylation after one hour. Accordingly, we conclude that SIRT1 and SIRT2 do not have specificity towards different sites or acylation types in H3. Given that global hyperacetylation has been observed with SIRT1 knockout strains,[158] we further propose that SIRT1 and SIRT2 behave as 'global defenders' rather than 'specialists' in nucleus and cytoplasm, respectively. Whenever hyperacylation

happens, SIRT1 and SIRT2 are recruited to take redundant acylations off to ensure regular cellular activities. When cells are exposed to stress, a number of stress-response proteins will be expressed, which may be related to the widely-known elevated level of histone acetylation under stress due to the fact that histone acetylation correlates to the activation of gene transcription. It has been also well studied that SIRT1 gets upregulated under stress.[159-162] It is reasonable to postulate that cells need high SIRT1 level to make the transition from normal status to under-stress status since cells need to precisely control the expression of stress-response proteins and avoid the expression of unrelated proteins.

Acetylation needs acetyl-CoA as the substrate, which is massively produced by glycolysis and fatty acid metabolism. Within the cells or living organisms undergo intensive glycolysis or fatty acid metabolism, it is possible that the proteins inside those cells would get excessive acetylation due to the high concentration of the substrate. Recent study by Ishikawa discovered that the mRNA level of SIRT1 is upregulated in the cats fed with high fat diet.[163] We believe the high concentration of SIRT1 counters the vast production of acetyl-CoA by fatty acid metabolism. Wagner and Payne already demonstrated that the chemical conditions of the mitochondrial matrix are sufficient to cause non-enzymatic lysine acetylation and succinylation.[164] This discovery also supports our idea that SIRT1 serves as ‘global defender’ towards over-acetylation in cells.

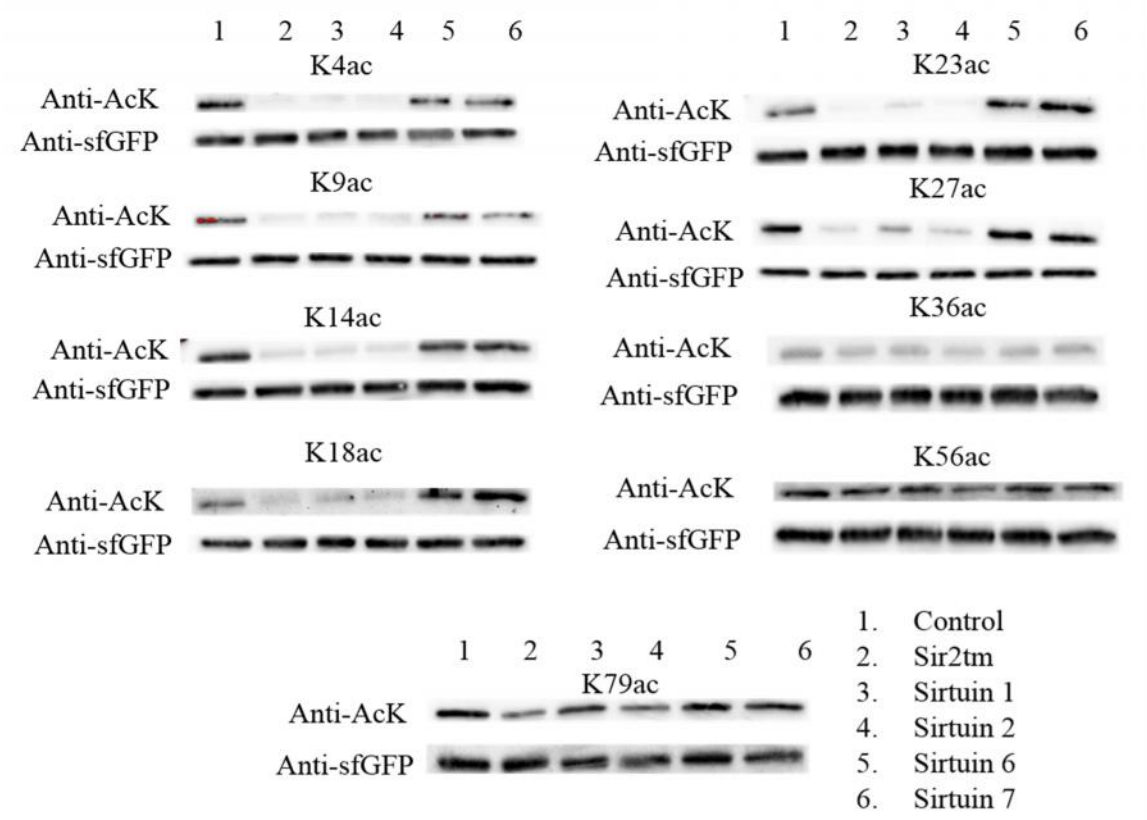


Figure 33. H3 deacetylation.

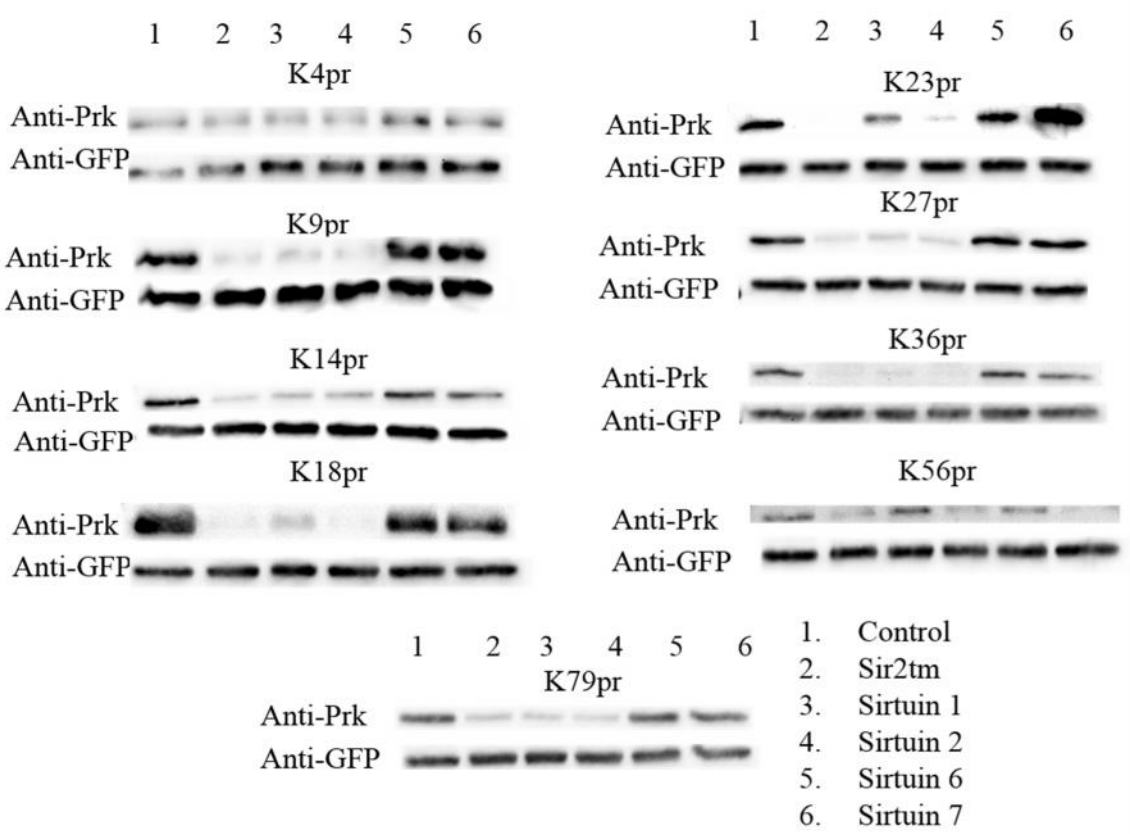


Figure 34. H3 depropionylation.

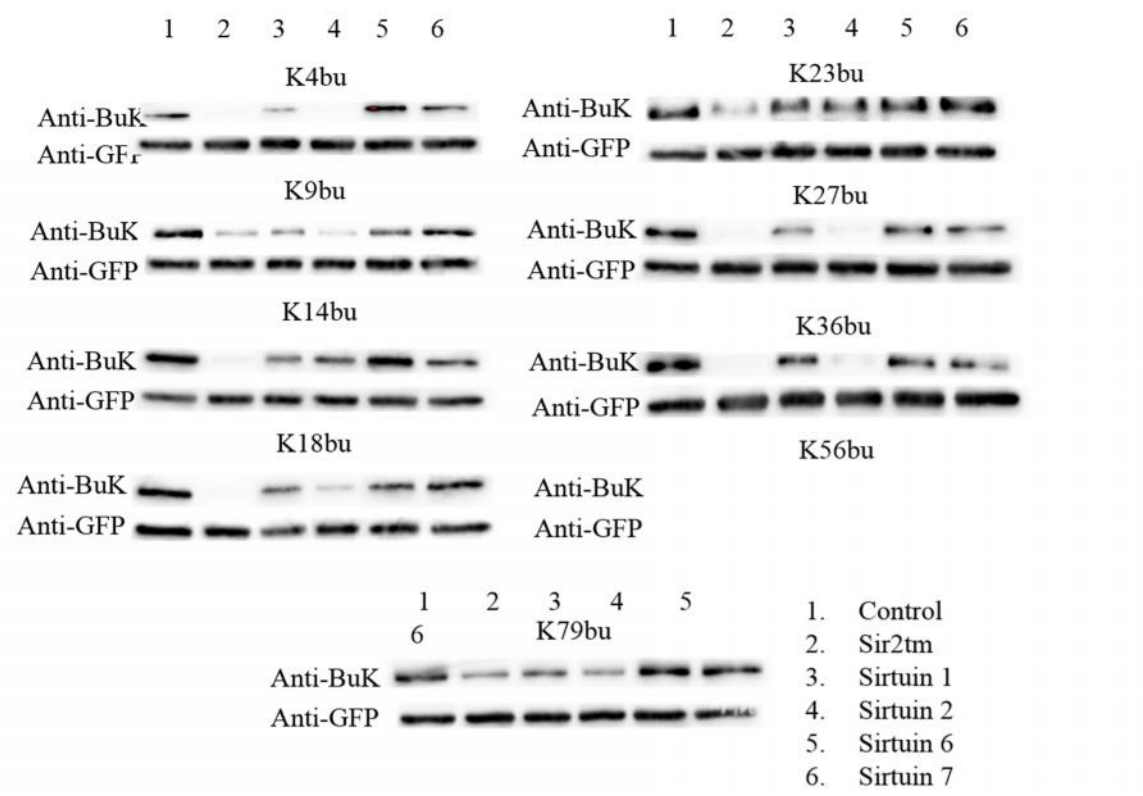


Figure 35. H3 debutyrylation.

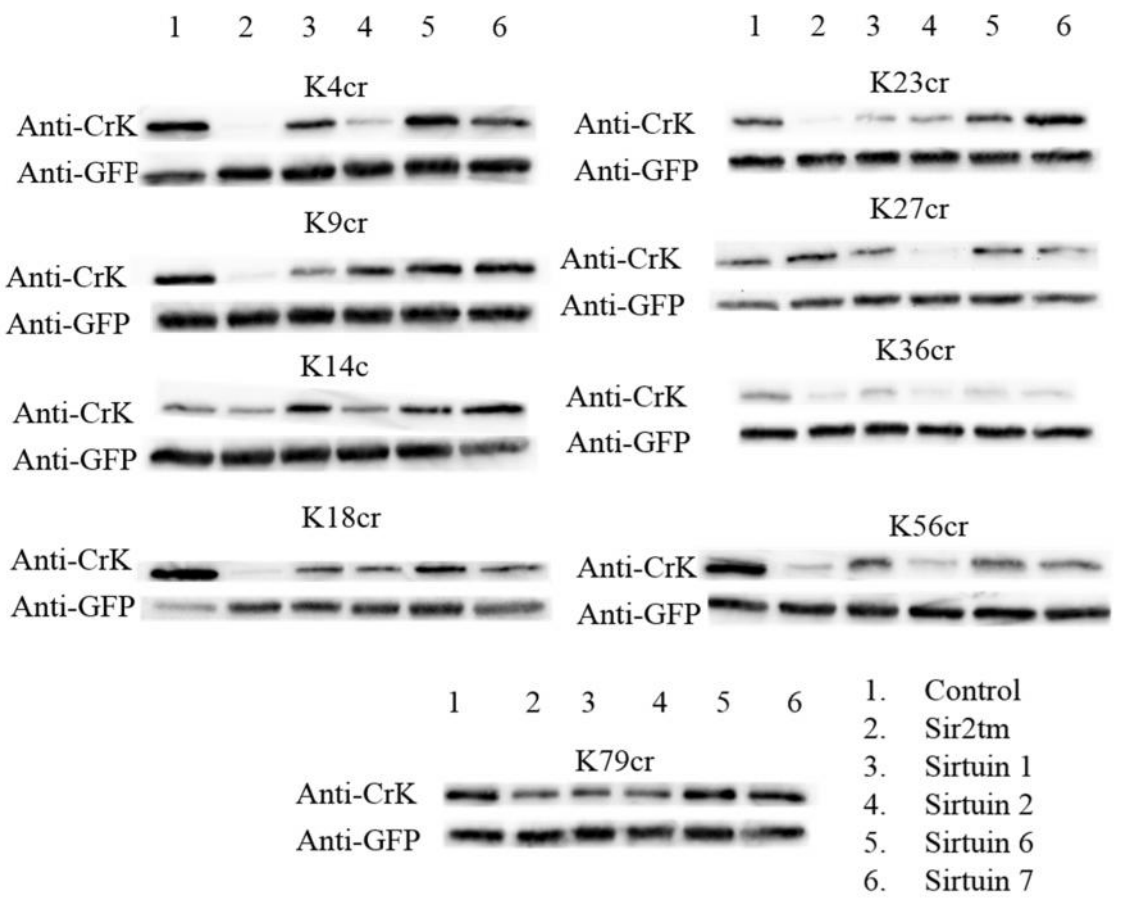


Figure 36. H3 decrotonylation.

Conclusion

Sirtuins are type III histone deacetylases which are closely related to cancer, inflammation and aging. Conventional studies were focused on investigating the specificity of individual sirtuin towards different sites and acylations. However, according to our study on recombinant H3 with different acylations, we discovered that SIRT1 and SIRT2 serve as ‘global protectors’ instead of ‘specialists’. That is to say, they show unexpected reactivities to short-chain deacylations regardless the acylation sites on H3. Meanwhile, SIRT6 and SIRT7 show negligible reactivities towards all acylations. Therefore, our study redefined the function of sirtuins in the nucleus which may potentially help us to understand the real role all sirtuins play in cells.

CHAPTER VII

CONCLUSION: REMARKS AND FUTURE OUTLOOK

In this study, we demonstrate the genetic incorporation of a variety of NCAAs into proteins via wild-type or evolved PylRS-$tRNA^{Pyl}_{CUA}$ pairs. In particular, Kme1 was genetically encoded into proteins. With the following reductive amination, the genetic incorporation of Kme2 can be achieved with ease as well. Due to the success of expressing histone H3 bearing four short-chain acylations, histone proteins with site-specific mono-methylation and di-methylation can be obtained, which are useful to study the interaction between site-specific methylations and the recruitment of transcriptional factors. Moreover, if we manage to construct the nucleosome with site-specific methylated H3, it can be subjected to histone methyltransferases or histone acetyltransferases with corresponding substrates to observe whether one specific methylation triggers cross-talk to further modify all other histone proteins.

In the case of acylations, indeed we showed the substrate preference of sirtuins towards different modified H3, yet the conclusion we drew was based on H3 itself rather than the nucleosome. To understand how sirtuins choose their substrates in cells, nucleosomes with a variety of acylated histone proteins need to be constructed so that we can unveil how sirtuins regulate histone acylations with native substrates. Furthermore, similar to the case the methylation, nucleosomes with site-specific acylations grant an opportunity to investigate cross-talks that acylations initiate, which also help us to further understand histone epigenetics.

Besides short-chain acylations, proteins in cells also undergo long-chain acylations, such as myristoylation. With the genetic incorporation of our ‘Swiss army knife’ into proteins and the further reactions, proteins with myristoylation can be afforded so that the mechanism of its regulation can be investigated.

Although we encountered obstacles to deprotect the alloc-protected hydroxylamine/hydrazine containing lysine mimics, several adjustments can be applied to solve the issue. With an alternative synthetic route, it is very possible that proteins with hydrazine containing lysine mimics can be obtained. Given the fact that a better approach to incorporate N^{ε}-long-chain acylated lysine (the Swiss-knife) has been developed, to synthesize proteins with N^{ε}-long-chain acylated lysine mimics may not sound too appealing. However, proteins with hydrazine are still useful in protein labeling so that it is still worthwhile to synthesize proteins with hydrazine groups.

REFERENCES

(1) Walsh, C. *Posttranslational modification of proteins : expanding nature's inventory*; Roberts and Co. Publishers: Englewood, Colo., 2006.

(2) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. *Angew Chem Int Edit* **2005**, *44*, 7342.

(3) Liu, W. S. R.; Wang, Y. S.; Wan, W. *Mol Biosyst* **2011**, *7*, 38.

(4) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41.

(5) Zhang, Y.; Reinberg, D. *Genes & development* **2001**, *15*, 2343.

(6) Rice, J. C.; Allis, C. D. *Curr Opin Cell Biol* **2001**, *13*, 263.

(7) Shahbazian, M. D.; Grunstein, M. *Annu Rev Biochem* **2007**, *76*, 75.

(8) Jasencakova, Z.; Meister, A.; Walter, J.; Turner, B. M.; Schubert, I. *Plant Cell* **2000**, *12*, 2087.

(9) Grewal, S. I. S.; Moazed, D. *Science* **2003**, *301*, 798.

(10) Woo, D. D. L.; Clarklewis, I.; Chait, B. T.; Kent, S. B. H. *Protein Eng* **1989**, *3*, 29.

(11) Hojo, H.; Kwon, Y.; Kakuta, Y.; Tsuda, S.; Tanaka, I.; Hikichi, K.; Aimoto, S. *B Chem Soc Jpn* **1993**, *66*, 2700.

(12) Inui, T.; Bodi, J.; Kubo, S.; Nishio, H.; Kimura, T.; Kojima, S.; Maruta, H.; Muramatsu, T.; Sakakibara, S. *J Pept Sci* **1996**, *2*, 28.

(13) Zawadzke, L. E.; Berg, J. M. *Proteins* **1993**, *16*, 301.

(14) Akaji, K.; Fujii, N.; Yajima, H.; Hayashi, K.; Mizuta, K.; Aono, M.; Moriga, M. *Chem Pharm Bull* **1985**, *33*, 184.

(15) Miller, M.; Baca, M.; Rao, J. K. M.; Kent, S. B. H. *J Mol Struc-Theochem* **1998**, *423*, 137.

(16) Dawson, P. E.; Muir, T. W.; Clarklewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776.

(17) Danielson, M. A.; Falke, J. J. *Annu Rev Bioph Biom* **1996**, *25*, 163.

(18) Li, C. G.; Wang, G. F.; Wang, Y. Q.; Creager-Allen, R.; Lutz, E. A.; Scronce, H.; Slade, K. M.; Ruf, R. A. S.; Mehl, R. A.; Pielak, G. J. *J Am Chem Soc* **2010**, *132*, 321.

(19) Kitevski-LeBlanc, J. L.; Prosser, R. S. *Prog Nucl Mag Res Sp* **2012**, *62*, 1.

(20) Speers, A. E.; Cravatt, B. F. *Chem Biol* **2004**, *11*, 535.

(21) Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y. *Angew Chem Int Edit* **2012**, *51*, 5852.

(22) Wu, B.; Wang, Z. Y.; Huang, Y.; Liu, W. S. R. *Chembiochem* **2012**, *13*, 1405.

(23) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z. Y.; Russell, D. H.; Liu, W. S. *Bioorg Med Chem Lett* **2010**, *20*, 878.

(24) Lee, Y. J.; Wu, B.; Raymond, J. E.; Zeng, Y.; Fang, X.; Wooley, K. L.; Liu, W. R. *ACS chemical biology* **2013**, *8*, 1664.

(25) Ausubel, F. M. *Current protocols in molecular biology*; Greene Pub. Associates ;J. Wiley, order fulfillment: Brooklyn, N.Y. Media, Pa., 1987.

(26) Erlanson, D. A.; Chytil, M.; Verdine, G. L. *Chem Biol* **1996**, *3*, 981.

(27) Xu, M. Q.; Perler, F. B. *Embo J* **1996**, *15*, 5146.

(28) Chong, S. R.; Mersha, F. B.; Comb, D. G.; Scott, M. E.; Landry, D.; Vence, L. M.; Perler, F. B.; Benner, J.; Kucera, R. B.; Hirvonen, C. A.; Pelletier, J. J.; Paulus, H.; Xu, M. Q. *Gene* **1997**, *192*, 271.

(29) Muir, T. W.; Sondhi, D.; Cole, P. A. *P Natl Acad Sci USA* **1998**, *95*, 6705.

(30) Cotton, G. J.; Ayers, B.; Xu, R.; Muir, T. W. *J Am Chem Soc* **1999**, *121*, 1100.

(31) Noren, C. J.; Anthonycahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182.

(32) Hohsaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J Am Chem Soc* **1999**, *121*, 34.

(33) Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J Am Chem Soc* **1989**, *111*, 8013.

(34) Cornish, V. W.; Mendel, D.; Schultz, P. G. *Angewandte Chemie-International Edition in English* **1995**, *34*, 621.

(35) Martin, G. W.; Harney, J. W.; Berry, M. J. *Rna* **1996**, *2*, 171.

(36) Bock, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F. *Mol Microbiol* **1991**, *5*, 515.

(37) Blight, S. K.; Larue, R. C.; Mahapatra, A.; Longstaff, D. G.; Chang, E.; Zhao, G.; Kang, P. T.; Church-Church, K. B.; Chan, M. K.; Krzycki, J. A. *Nature* **2004**, *431*, 333.

(38) Srinivasan, G.; James, C. M.; Krzycki, J. A. *Science* **2002**, *296*, 1459.

(39) Ibba, M.; Soll, D. *Annu Rev Biochem* **2000**, *69*, 617.

(40) Wang, L.; Schultz, P. G. *Angew Chem Int Edit* **2005**, *44*, 34.

(41) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.

(42) Furter, R. *Protein Sci* **1998**, *7*, 419.

(43) Chin, J. W.; Cropp, T. A.; Chu, S.; Meggers, E.; Schultz, P. G. *Chem Biol* **2003**, *10*, 511.

(44) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G. *Science* **2003**, *301*, 964.

(45) Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G. *P Natl Acad Sci USA* **2003**, *100*, 56.

(46) Deiters, A.; Cropp, T. A.; Summerer, D.; Mukherji, M.; Schultz, P. G. *Bioorg Med Chem Lett* **2004**, *14*, 5743.

(47) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J Am Chem Soc* **2002**, *124*, 9026.

(48) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *P Natl Acad Sci USA* **2002**, *99*, 11020.

(49) Komander, D. *Biochem Soc T* **2009**, *37*, 937.

(50) Zhao, S. M.; Xu, W.; Jiang, W. Q.; Yu, W.; Lin, Y.; Zhang, T. F.; Yao, J.; Zhou, L.; Zeng, Y. X.; Li, H.; Li, Y. X.; Shi, J.; An, W. L.; Hancock, S. M.; He, F. C.; Qin, L. X.; Chin, J.; Yang, P. Y.; Chen, X.; Lei, Q. Y.; Xiong, Y.; Guan, K. L. *Science* **2010**, *327*, 1000.

(51) Martin, C.; Zhang, Y. *Nat Rev Mol Cell Bio* **2005**, *6*, 838.

(52) Wan, W.; Huang, Y.; Wang, Z. Y.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. *Angew Chem Int Edit* **2010**, *49*, 3211.

(53) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. *J Mol Biol* **2002**, *319*, 1097.

(54) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; Lu, Z.; Ye, Z.; Zhu, Q.; Wysocka, J.; Ye, Y.; Khochbin, S.; Ren, B.; Zhao, Y. *Cell* **2011**, *146*, 1016.

(55) Chen, Y.; Sprung, R.; Tang, Y.; Ball, H.; Sangras, B.; Kim, S. C.; Falck, J. R.; Peng, J. M.; Gu, W.; Zhao, Y. M. *Mol Cell Proteomics* **2007**, *6*, 812.

(56) Calalb, M. B.; Polte, T. R.; Hanks, S. K. *Molecular and cellular biology* **1995**, *15*, 954.

(57) Ciesla, J.; Fraczyk, T.; Rode, W. *Acta Biochim Pol* **2011**, *58*, 137.

(58) Deutscher, J.; Saier, M. H. *J Mol Microb Biotech* **2005**, *9*, 125.

(59) Manohar, M.; Mooney, A. M.; North, J. A.; Nakkula, R. J.; Picking, J. W.; Edon, A.; Fishel, R.; Poirier, M. G.; Ottesen, J. J. *Journal of Biological Chemistry* **2009**, *284*, 23312.

(60) Nguyen, A. T.; Zhang, Y. *Genes & development* **2011**, *25*, 1345.

(61) Yuan, J.; Pu, M. T.; Zhang, Z. G.; Lou, Z. K. *Cell Cycle* **2009**, *8*, 1747.

(62) Feng, X.; Zhang, K. L.; Grunstein, M. *Cell* **2005**, *121*, 375.

(63) Latham, J. A.; Dent, S. Y. *Nat Struct Mol Biol* **2007**, *14*, 1017.

(64) Lall, S. *Nat Struct Mol Biol* **2007**, *14*, 1110.

(65) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat Struct Mol Biol* **2007**, *14*, 1025.

(66) Yang, X. J.; Seto, E. *Mol Cell* **2008**, *31*, 449.

(67) Luo, J.; Li, M.; Tang, Y.; Laszkowska, M.; Roeder, R. G.; Gu, W. *Proc Natl Acad Sci U S A* **2004**, *101*, 2259.

(68) Simon, M. D.; Chu, F. X.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. *Cell* **2007**, *128*, 1003.

(69) He, S.; Bauman, D.; Davis, J. S.; Loyola, A.; Nishioka, K.; Gronlund, J. L.; Reinberg, D.; Meng, F. Y.; Kelleher, N.; McCafferty, D. G. *P Natl Acad Sci USA* **2003**, *100*, 12033.

(70) Wang, L.; Schultz, P. G. *Angew Chem Int Ed Engl* **2004**, *44*, 34.

(71) Herring, S.; Ambrogelly, A.; Polycarpo, C. R.; Soll, D. *Nucleic Acids Res* **2007**, *35*, 1270.

(72) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. *J Am Chem Soc* **2009**, *131*, 8720.

(73) 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.; Krzycki, J. A. *Nature* **2004**, *431*, 333.

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

(75) Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K. *Angew Chem Int Ed Engl* **2009**, *48*, 1633.

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

(77) Huang, Y.; Russell, W. K.; Wan, W.; Pai, P. J.; Russell, D. H.; Liu, W. *Mol Biosyst* **2010**, *In press*.

(78) Tatsu, Y.; Shigeri, Y.; Ishida, A.; Kameshita, I.; Fujisawa, H.; Yumoto, N. *Bioorg Med Chem Lett* **1999**, *9*, 1093.

(79) Talaga, P.; Benezra, C.; Stampf, J. L. *Bioorg Chem* **1990**, *18*, 199.

(80) Wiejak, S.; Masiukiewicz, E.; Rzeszotarska, B. *Chem Pharm Bull* **2001**, *49*, 1189.

(81) Dyer, R. G.; Turnbull, K. D. *J Org Chem* **1999**, *64*, 7988.

(82) Yamamoto, H.; Yang, J. T. *Biopolymers* **1974**, *13*, 1093.

(83) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nature biotechnology* **2002**, *20*, 1044.

(84) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W. *Bioorg Med Chem* **2010**, *20*, 878.

(85) Huang, Y.; Russell, W. K.; Wan, W.; Pai, P. J.; Russell, D. H.; Liu, W. *Mol Biosyst* **2010**, *6*, 683.

(86) Xie, J. M.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nature biotechnology* **2004**, *22*, 1297.

(87) Tam, J. P.; Yu, Q. T.; Miao, Z. W. *Biopolymers* **1999**, *51*, 311.

(88) Nguyen, D. P.; Garcia Alai, M. M.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. *J Am Chem Soc* **2009**, *131*, 14194.

(89) Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G. *Angew Chem Int Ed Engl* **2009**, *48*, 4052.

(90) Xie, J.; Liu, W.; Schultz, P. G. *Angew Chem Int Ed Engl* **2007**, *46*, 9239.

(91) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A. *Proc Natl Acad Sci U S A* **2007**, *104*, 11268.

(92) Stryer, L.; Haugland, R. P. *Proc Natl Acad Sci U S A* **1967**, *58*, 719.

(93) Morris, S. J.; Sudhof, T. C.; Haynes, D. H. *Biochim Biophys Acta* **1982**, *693*, 425.

(94) Ha, T.; Enderle, T.; Ogletree, D. F.; Chemla, D. S.; Selvin, P. R.; Weiss, S. *Proc Natl Acad Sci U S A* **1996**, *93*, 6264.

(95) Heim, R.; Tsien, R. Y. *Curr Biol* **1996**, *6*, 178.

(96) Suzuki, Y.; Yasunaga, T.; Ohkura, R.; Wakabayashi, T.; Sutoh, K. *Nature* **1998**, *396*, 380.

(97) Ting, A. Y.; Kain, K. H.; Klemke, R. L.; Tsien, R. Y. *Proc Natl Acad Sci U S A* **2001**, *98*, 15003.

(98) Harpur, A. G.; Wouters, F. S.; Bastiaens, P. I. H. *Nature Biotechnology* **2001**, *19*, 167.

(99) Roy, R.; Hohng, S.; Ha, T. *Nat Methods* **2008**, *5*, 507.

(100) Margittai, M.; Widengren, J.; Schweinberger, E.; Schroder, G. F.; Felekyan, S.; Haustein, E.; Konig, M.; Fasshauer, D.; Grubmuller, H.; Jahn, R.; Seidel, C. A. *Proc Natl Acad Sci U S A* **2003**, *100*, 15516.

(101) Schuler, B.; Pannell, L. K. *Bioconjug Chem* **2002**, *13*, 1039.

(102) Talaga, D. S.; Lau, W. L.; Roder, H.; Tang, J.; Jia, Y.; DeGrado, W. F.; Hochstrasser, R. M. *Proc Natl Acad Sci U S A* **2000**, *97*, 13021.

(103) Jager, M.; Nir, E.; Weiss, S. *Protein Sci* **2006**, *15*, 640.

(104) Jager, M.; Michalet, X.; Weiss, S. *Protein Sci* **2005**, *14*, 2059.

(105) Brustad, E. M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A. *J Am Chem Soc* **2008**, *130*, 17664.

(106) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464*, 441.

(107) Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. *Angew Chem Int Ed Engl* **2010**, *49*, 3211.

(108) Xie, J.; Schultz, P. G. *Methods* **2005**, *36*, 227.

(109) Xie, J.; Schultz, P. G. *Nat Rev Mol Cell Biol* **2006**, *7*, 775.

(110) Wang, Y. S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. *Mol Biosyst* **2010**, *6*, 1557.

(111) Wang, Y.-S.; Wang, Z.; Wan, W.; Lee, Y.-J.; Dodd, L. E.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. *J Am Chem Soc* **2010**, *submitted*.

(112) Wang, J. Y.; Xie, J. M.; Schultz, P. G. *J Am Chem Soc* **2006**, *128*, 8738.

(113) Smith, R. H.; Mehl, A. F.; Shantz, D. L.; Chmurny, G. N.; Michejda, C. J. *J Org Chem* **1988**, *53*, 1467.

(114) Maly, D. J.; Leonetti, F.; Backes, B. J.; Dauber, D. S.; Harris, J. L.; Craik, C. S.; Ellman, J. A. *J Org Chem* **2002**, *67*, 910.

(115) Feng, Y. B.; Burgess, K.; Pledger, D.; Cairns, N.; Linthicum, D. S. *Bioorg Med Chem Lett* **1998**, *8*, 881.

(116) Brouwer, A. J.; Liskamp, R. M. J. *Eur J Org Chem* **2005**, 487.

(117) Berthelot, T.; Talbot, J. C.; Lain, G.; Deleris, G.; Latxague, L. *J Pept Sci* **2005**, *11*, 153.

(118) Umehara, T.; Kim, J.; Lee, S.; Guo, L. T.; Soll, D.; Park, H. S. *Febs Lett* **2012**, *586*, 729.

(119) Wang, J.; Xie, J.; Schultz, P. G. *J Am Chem Soc* **2006**, *128*, 8738.

(120) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew Chem Int Edit* **2001**, *40*, 2004.

(121) Huisgen, R. *Proc Chem Soc* **1961**, 357.

(122) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J Mol Biol* **2010**, *395*, 361.

(123) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J Am Chem Soc* **2002**, *124*, 9026.

(124) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew Chem Int Ed Engl* **2009**, *48*, 9879.

(125) Sivakumar, K.; Xie, F.; Cash, B. M.; Long, S.; Barnhill, H. N.; Wang, Q. *Org Lett* **2004**, *6*, 4603.

(126) Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. *Proc Natl Acad Sci U S A* **2007**, *104*, 16793.

(127) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W. *Bioorg Med Chem Lett* **2010**, *20*, 878.

(128) Umehara, T.; Kim, J.; Lee, S.; Guo, L.-T.; Soll, D.; Park, H.-S. *in preparation* **2012**.

(129) Bernstein, B. E.; Mikkelsen, T. S.; Xie, X. H.; Kamal, M.; Huebert, D. J.; Cuff, J.; Fry, B.; Meissner, A.; Wernig, M.; Plath, K.; Jaenisch, R.; Wagschal, A.; Feil, R.; Schreiber, S. L.; Lander, E. S. *Cell* **2006**, *125*, 315.

(130) Kondo, Y.; Shen, L. L.; Issa, J. P. J. *Molecular and cellular biology* **2003**, *23*, 206.

(131) Jones, P. A.; Baylin, S. B. *Nat Rev Genet* **2002**, *3*, 415.

(132) Sarg, B.; Koutzamani, E.; Helliger, W.; Rundquist, I.; Lindner, H. H. *Journal of Biological Chemistry* **2002**, *277*, 39195.

(133) Greer, E. L.; Shi, Y. *Nat Rev Genet* **2012**, *13*, 343.

(134) Brownell, J. E.; Zhou, J. X.; Ranalli, T.; Kobayashi, R.; Edmondson, D. G.; Roth, S. Y.; Allis, C. D. *Cell* **1996**, *84*, 843.

(135) Grunstein, M. *Nature* **1997**, *389*, 349.

(136) Lachner, M.; Jenuwein, T. *Curr Opin Cell Biol* **2002**, *14*, 286.

(137) Kouzarides, T. *Curr Opin Genet Dev* **2002**, *12*, 198.

(138) Wang, Z. Y. U.; Wang, Y. S.; Pai, P. J.; Russell, W. K.; Russell, D. H.; Liu, W. S. R. *Biochemistry-Us* **2012**, *51*, 5232.

(139) Simon, M. D.; Shokat, K. M. *Method Enzymol* **2012**, *512*, 57.

(140) Wang, Y. S.; Wu, B.; Wang, Z. Y.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Moe, Y. N.; Russell, D. H.; Liu, W. R. *Mol Biosyst* **2010**, *6*, 1575.

(141) Stevenson, F. T.; Bursten, S. L.; Fanton, C.; Locksley, R. M.; Lovett, D. H. *P Natl Acad Sci USA* **1993**, *90*, 7245.

(142) Lin, S. S.; Manchester, J. K.; Gordon, J. I. *Journal of Biological Chemistry* **2003**, *278*, 13390.

(143) Toska, E.; Campbell, H. A.; Shandilya, J.; Goodfellow, S. J.; Shore, P.; Medler, K. F.; Roberts, S. G. E. *Cell Rep* **2012**, *2*, 462.

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

(145) Schwer, B.; Verdin, E. *Cell Metab* **2008**, *7*, 104.

(146) Yamamoto, H.; Schoonjans, K.; Auwerx, J. *Mol Endocrinol* **2007**, *21*, 1745.

(147) Sauve, A. A.; Wolberger, C.; Schramm, V. L.; Boeke, J. D. *Annu Rev Biochem* **2006**, *75*, 435.

(148) Richon, V. M.; Sandhoff, T. W.; Rifkind, R. A.; Marks, P. A. *P Natl Acad Sci USA* **2000**, *97*, 10014.

(149) Sauve, A. A.; Youn, D. Y. *Current opinion in chemical biology* **2012**, *16*, 535.

(150) Vollmuth, F.; Geyer, M. *Angewandte Chemie* **2010**, *49*, 6768.

(151) Vaquero, A.; Scher, M. B.; Lee, D. H.; Sutton, A.; Cheng, H. L.; Alt, F. W.; Serrano, L.; Sternglanz, R.; Reinberg, D. *Genes & development* **2006**, *20*, 1256.

(152) Vaquero, A.; Scher, M.; Lee, D. H.; Erdjument-Bromage, H.; Tempst, P.; Reinberg, D. *Mol Cell* **2004**, *16*, 93.

(153) Barber, M. F.; Michishita-Kioi, E.; Xi, Y. X.; Tasselli, L.; Kioi, M.; Moqtaderi, Z.; Tennen, R. I.; Paredes, S.; Young, N. L.; Chen, K. F.; Struhl, K.; Garcia, B. A.; Gozani, O.; Li, W.; Chua, K. F. *Nature* **2012**, *487*, 114.

(154) Michishita, E.; McCord, R. A.; Berber, E.; Kioi, M.; Padilla-Nash, H.; Damian, M.; Cheung, P.; Kusumoto, R.; Kawahara, T. L. A.; Barrett, J. C.; Chang, H. Y.; Bohr, V. A.; Ried, T.; Gozani, O.; Chua, K. F. *Nature* **2008**, *452*, 492.

(155) Yang, B.; Zwaans, B. M. M.; Eckersdorff, M.; Lombard, D. B. *Cell Cycle* **2009**, *8*, 2662.

(156) Garske, A. L.; Denu, J. M. *Biochemistry-Us* **2006**, *45*, 94.

(157) Frye, R. A. *Biochemical and biophysical research communications* **2000**, *273*, 793.

(158) Chen, Y.; Zhao, W. H.; Yang, J. S.; Cheng, Z. Y.; Luo, H.; Lu, Z. K.; Tan, M. J.; Gu, W.; Zhao, Y. M. *Mol Cell Proteomics* **2012**, *11*, 1048.

(159) Merksamer, P. I.; Liu, Y. F.; He, W. J.; Hirschey, M. D.; Chen, D.; Verdin, E. *Aging-Us* **2013**, *5*, 144.

(160) Gu, X.; Sun, J.; Li, S.; Wu, X.; Li, L. *Neurobiology of aging* **2013**, *34*, 1069.

(161) Cornelius, C.; Salinaro, A. T.; Scuto, M.; Fronte, V.; Cambria, M. T.; Pennisi, M.; Bella, R.; Milone, P.; Graziano, A.; Crupi, R.; Cuzzocrea, S.; Pennisi, G.; Calabrese, V. *Immun Ageing* **2013**, *10*.

(162) Lee, S. I.; Park, K. H.; Kim, S. J.; Kang, Y. G.; Lee, Y. M.; Kim, E. C. *Clinical and experimental immunology* **2012**, *168*, 113.

(163) Ishikawa, S.; Li, G. B.; Takemitsu, H.; Fujiwara, M.; Mori, N.; Yamamoto, I.; Arai, T. *Bmc Vet Res* **2013**, *9*.

(164) Wagner, G. R.; Payne, R. M. *The Journal of biological chemistry* **2013**, *288*, 29036.