GENETIC INCORPORATION OF NONCANONICAL AMINO ACIDS INTO PROTEINS FOR PROTEIN FUNCTION INVESTIGATION

A Dissertation

by

YING HUANG

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

DOCTOR OF PHILOSOPHY

May 2012

Major Subject: Chemistry

Genetic Incorporation of Noncanonical Amino Acids into Proteins for

Protein Function Investigation

GENETIC INCORPORATION OF NONCANONICAL AMINO ACIDS INTO PROTEINS FOR PROTEIN FUNCTION INVESTIGATION

A Dissertation

by

YING HUANG

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Wenshe Liu
Committee Members,	Paul A. Lindahl
	Coran M. Watanabe
	Jiong Yang
Head of Department,	David H. Russell

May 2012

Major Subject: Chemistry

ABSTRACT

Genetic Incorporation of Noncanonical Amino Acids into

Proteins for Protein Function Investigation. (May 2012)

Ying Huang, B.A., Peking University

Chair of Advisory Committee: Dr. Wenshe Liu

With the objective to functionalize proteins for the understanding of their biological roles and developing protein-based biosensors, I have been developing methods to synthesize proteins with defined modifications and applying them to study protein functional roles and generate proteins with new properties. These methods rely on the read-through of an in-frame stop codon in mRNA by a nonsense suppressor tRNA specifically acylated with a noncanoncial amino acid (NAA) by a unique aminoacyl-tRNA synthetase and the genetic incorporation of this NAA at the stop codon site. NAAs either provide chemical handles for site-specific manipulation or mimic the posttranslational modifications, which are critical for understanding cellular regulations and signal transduction.

The pyrrolysine synthetase (PylRS) has been wildly used to incorporate NAAs into proteins in *E. coli*. Taking advantage of PylRS, I have developed method to genetically incorporate ketone-containing *N*-*ε*-acetyl-L-lysine analog, 2-amino-8-oxononanoic acid (KetoK), into proteins for their site-specific modifications and used it to mimic the protein lysine acetylation process.

I have also modified the ribosome in order to improve the amber suppression efficiency and therefore to achieve incorporation of multiple copies of NAA into one protein. By overexpressing a truncated ribosomal protein, L11C, I have demonstrated 5-fold increase of amber suppression level in *E. coli*, leading to higher expression levels for proteins incorporated with NAAs. I have also demonstrated this method can be applied successfully to incorporate at least 3 NAAs into one protein in *E. coli*.

With the success of incorporating multiple NAAs into one protein, I have further introduced two distinct NAAs into one protein simultaneously. This is done by using a wild type or evolved PylRS-pylT$_{UUA}$ pair and an evolved *M. jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)-tRNA$_{CUA}$ pair. By suppressing both UAG and UAA stop codons in one mRNA, a protein incorporated with two NAAs is synthesized with a decent yield.

There is of great interest to incorporate new NAAs into proteins, which is done by library selection. By introducing both positive and negative selective markers into one plasmid, I have developed a one-plasmid selection method. In this method, the positive and negative selections are accomplished by in a single type of cells hosting a single selection plasmid.

To My Mother

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

1. INTRODUCTION: PYLRS, A MEDIOCRE ENZYME OR A GIFT FROM NATURE

1.1 Introduction to protein biosynthesis

In a cell, protein translation is the process that a nucleotide sequence in an mRNA is translated to be a corresponding amino acid sequence in a protein (Figure 1). During this process, a nucleotide triplet in an mRNA, which is called a codon, is used to code one amino acid. In total, there are 64 triplet codons. Shown as the standard genetic code for most organisms in Figure 2, 61 codons code 20 canonical amino acids and 3 codons code for protein translation termination. To undergo protein translation, ribosome mediates the correct recognition of a codon in mRNA by a tRNA that is acylated with an amino acid by an aminoacyl-tRNA synthetase (aaRS). Acylation of a tRNA by an aaRS is a two-step process, the first of which is to activate amino acid by ATP to form aminoacyl-AMP and the second is to transfer the aminoacyl group from aminoacyl-AMP to the 3' end of the tRNA.[1]

1.2 Fidelity of the canonical amino acid incorporation

Accurate acylation of tRNAs catalyzed by aaRSs is essential to keep high fidelity of protein translation in a cell. Typically, an aaRS consists of the ATP binding site, the amino acid activation site, and the tRNA binding domain. Some aaRSs have additional editing domains to hydrolyze the misacylated tRNAs. Based on where ATP is bound in

---

This dissertation follows the style of *Journal of the American Chemical Society.*

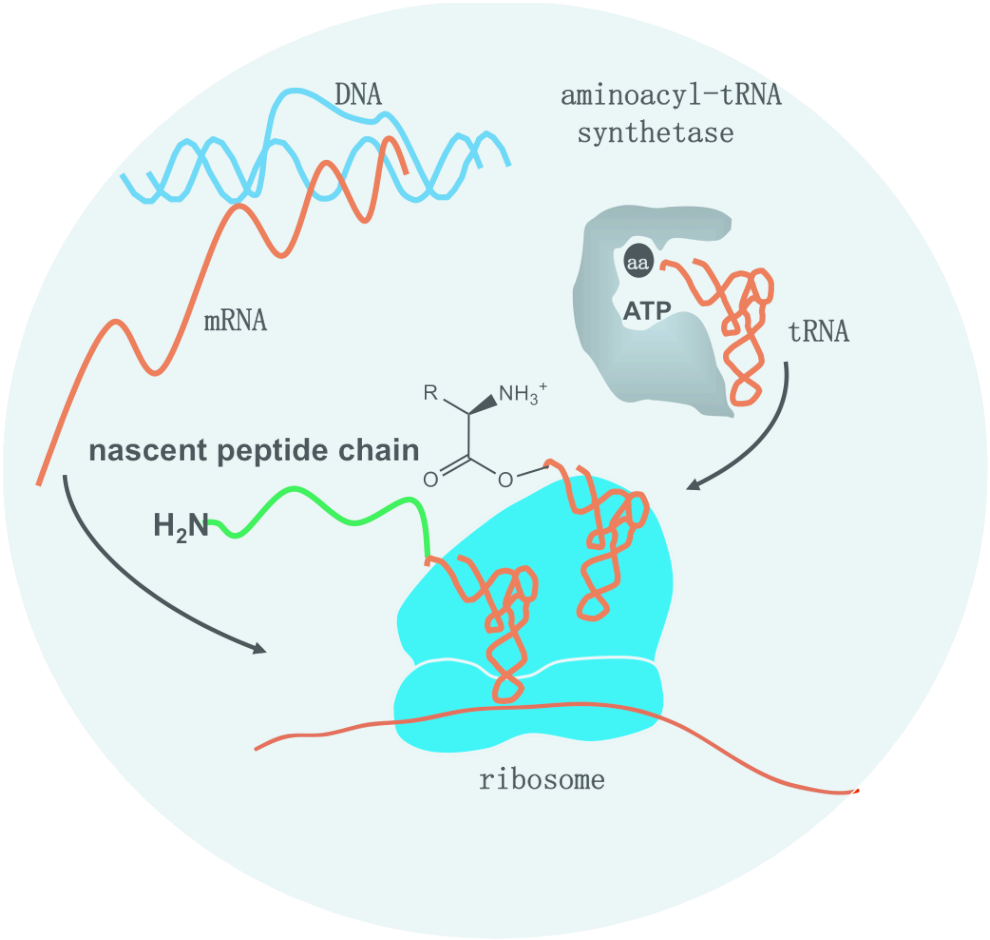


Figure 1. Protein biosynthesis machinery.

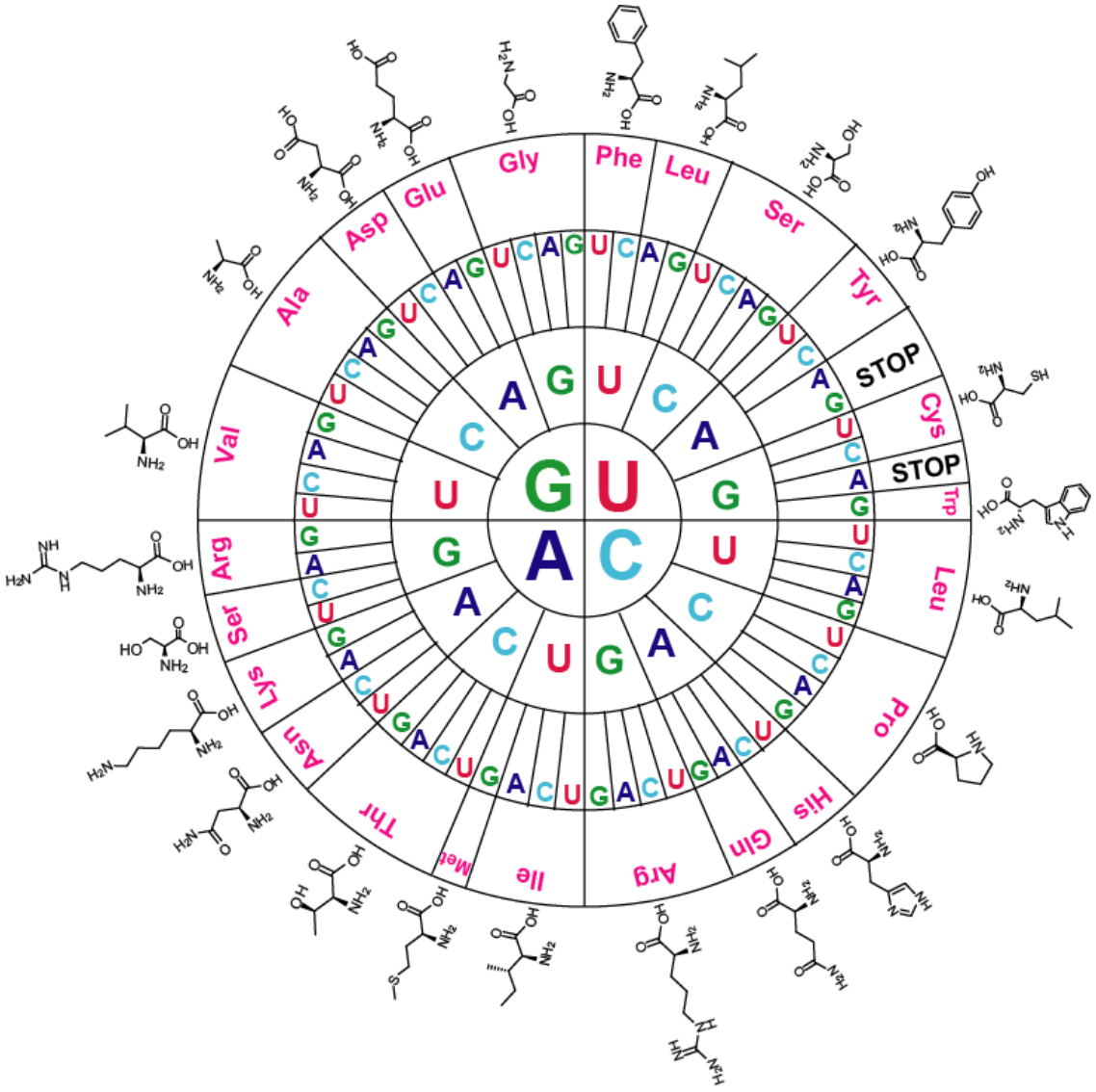


Figure 2. Standard genetic code.

the active site, and which ribose hydroxyl group of A76 in a tRNA is acylated with an amino acid, aaRSs are classified into two structurally unrelated groups, namely class I and class II.[2, 3] Two classes of aaRSs have different binding patterns when they associate tRNAs. A class I aaRS approaches its tRNA acceptor end from the tRNA minor groove. On the contrary, a class II aaRS approaches its tRNA acceptor end from the tRNA major groove.[4] For class I aaRSs, the activated amino acids are attached to the 2' ribose hydroxyl group of A76. For class II aaRSs, with the exception of PheRS, the activated amino acids are attached at the 3' ribose hydroxyl group of A76. Among 20 regular aaRSs, those for methionine, valine, isoleucine, cysteine, glutamate, leucine, lysine, arginine, tryptophan and tyrosine belong to class I and the rest belong to class II.[5]

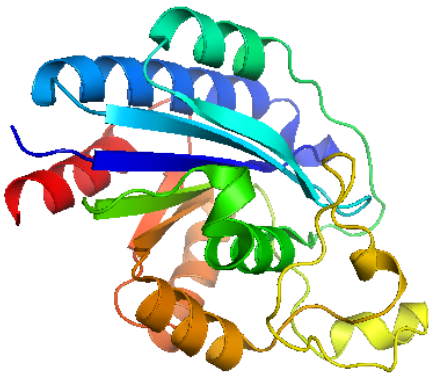


Figure 3. Structure of Rossmann fold. PDB: 2IUE.

1.2.1 Class I aaRSs

Class I aaRSs have the basic Rossmann fold, a three-layer α/β/α structure, with an inner core of five parallel β strands (Figure 3). In addition, two activation sites HXGH and KMSKS are conserved in all class I aaRSs.[2] The tetrapeptide HXGH (X denotes a hydrophobic amino acid) is at the end of one element sequence that consists of 11 residues; pentapeptide KMSKS is located near to the C terminus of the nucleotide-binding domain.[6] In general, due to the deep open pocket, the amino acids activated by class I aaRSs are usually larger than those for class II.[3]

The 10 class I aaRSs are further divided into three subclasses, subclass Ia, subclass Ib, and subclass Ic (Table 1).[7, 8] Subclass Ia contains CysRS, ArgRS, IleRS, LeuRS, ValRS, and MetRS. Subclass Ib contains GlnRS, GluRS, and LysRS. Subclass Ic includes TyrRS and TrpRS.

Table 1. Subclasses of AaRSs

	Class I	Class II
a	Cys, Arg, Ile, Leu, Val, Met	Ser, Pro, His, Thr
b	Gln, Glu, Lys	Asp, Asn, Lys
c	Tyr, Trp	Ala, Gly, Phe

1.2.2 Class II aaRSs

Class II aaRSs contain a very special fold structure, with mixed α helices and β strands, and a core with antiparallel β strands flanked by α helices. Class II aaRSs share three short conserved motifs where ATP is bound and the amino acid is activated.[2] Compared to class I aaRSs, which are most monomeric, most of class II are dimeric or multimeric.[1]

The 10 class II aaRSs are further grouped into three subclasses, subclass IIa, subclass IIb, and subclass IIc (Table 1).[7, 8] Subclass IIa, responsible for acylation of small amino acids and small polar amino acids, consists of SerRS, ProRS, HisRS, and ThrRS. Subclass IIb aaRSs that acylate tRNAs with large polar and charged amino acids, include AspRS, AsnRS, and LysRS. Subclass IIc aaRSs comprise AlaRS, GlyRS, and PheRS.

1.2.3 LysRS: a violator

The classification of 20 aaRSs was thought to be generic in all species, until the discovery of LysRS from *Methanococcus maripaludis* in 1997.[9] With the presence of characteristic Rossmann fold that are found in all class I aaRSs and absence of motives 1, 2, and 3 that are found in a typical class II LysRS,[2] *M. maripaludis* LysRS was classified as a class I aaRS (Figure 4). To date, only LysRSs have been found to exist as both class I and class II aaRSs. The class I LysRS is found in most archaea and some bacterial strains. In some archaea strains, both class I and class II LysRSs exist. Microbial genome sequencing discovered over 30 class I LysRSs.[10]

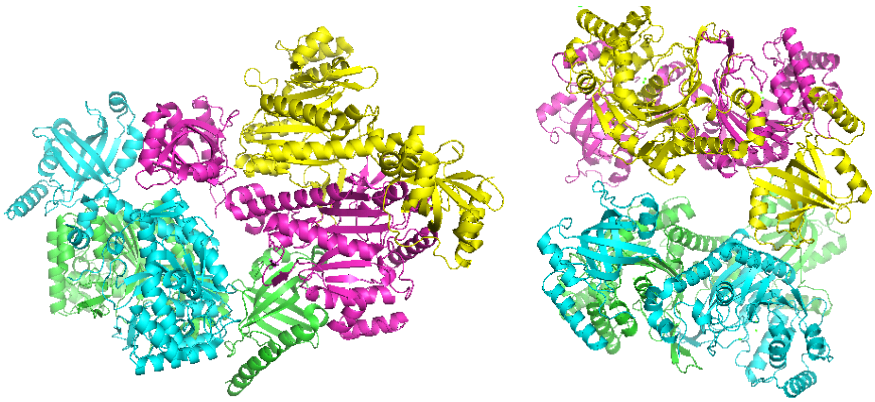


Figure 4. Structures for class I and II LysRSs. PDB (LysRS1): 3E9I, PDB (LysRS2): 3BJU.

1.2.4 Indirect biosynthesis of Cys-tRNA^{Cys}

In most organisms, Cys-tRNA^{Cys} is made by CysRS using cysteine.[11] However, genomic screening showed that CysRS is absent in at least three archaea, *Methanococcus jannaschii* and *Methanobacterium thermoautotrophicum* and *Methanopyrus kandleri*. Since cysteine has a critical role as a nucleophilic catalyst, a metal ligand and an electron-carrying thiol and cysteine exists in most of proteins in these organisms, it is of great interest to see how cysteine is incorporated into their proteins. In 2002, Dr. Dieter Söll group at Yale University proposed that ProRS might be involved in the process of cysteine incorporation, based on the observation that ProRS was able to form Cys-tRNA^{Pro} using cysteine as the substrate.[12] Nevertheless, contradicting data were reported later. By using acid urea gel electrophoresis analysis, it was shown that ProRS isolated from *M jannaschii* was unable to acylate mature tRNA^{Cys} purified from *M. jannaschii*.[13] The discovery of a novel enzyme MJ1660 from the

bioinformatic analysis solved the puzzle.[14] MJ1660, found in *M. jannaschii,* was able to catalyze the ligation between a phosphoserine (Sep) (Figure 5) and $tRNA^{Cys}$. MJ1660, later named as phosphoseryl-tRNA synthetase (SepRS), might play an equivalent role as CysRS.[15] Different from a classical class I CysRS, SepRS is a typical class II aaRS.[15] The biosynthesis of Cys-$tRNA^{Cys}$ is started from the synthesis of Sep-$tRNA^{Cys}$ and completed by another novel enzyme, MJ1678. MJ1678, a Sep-tRNA:Cys-tRNA synthetase (SepCysS), further converts Sep-$tRNA^{Cys}$ to Cys-$tRNA^{Cys}$, using pyridoxal 5'-phosphate (PLP) as a cofactor.[15] Therefore, a new route for the biosynthesis of Cys-$tRNA^{Cys}$ was demonstrated (Figure 5). In addition, this pathway serves as the only pathway in *Methanococcus maripaludis* to synthesize Cys-$tRNA^{Cys}$.[15] The SepRS is a unique synthetase that involves in the biosynthesis of a NAA, Sep.

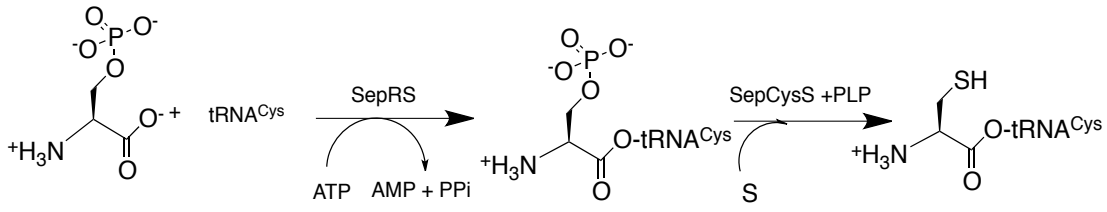


Figure 5. Indirect biosynthetic route of Cys-$tRNA^{Cys}$.

1.2.5 Indirect biosynthesis of Gln-tRNA^{Gln}

GlnRS was first found to be absent in the *Bacillus subtilis* in the late 1960's.[16] After extensive studies, GlnRS turns out to be rare, only found in eukaryotes and a few bacteria.[1, 17] Since all archaea and most of bacteria are lack of GlnRS, Gln-tRNA^{Gln} is synthesized by an indirect pathway (Figure 6).[18, 19] This alternative pathway requires two kinds of enzymes. One kind of enzymes is non-discriminating GluRS (ND-GluRS) that acylates both tRNA^{Glu} and tRNA^{Gln} with glutamate or misacylating GluRS (GluRS2) that only acylates tRNA^{Gln} with glutamate,[20, 21] the other one is glutamine dependent Glu-tRNA^{Gln} amidotransferase (Glu-Adt) or glutamine dependent Asp-tRNA^{Asn}/Glu-tRNA^{Gln} amidotransferase (Asp/Glu-Adt).[22] Glu-Adt and Asp/Glu-Adt use glutamine as the primary nitrogen source to convert Glu-tRNA^{Gln} to Gln-tRNA^{Gln}. The universal absence of GlnRS and presence of Glu-Adt in all known archaea genomes indicates it might be the only Gln-tRNA^{Gln} synthesis pathway.

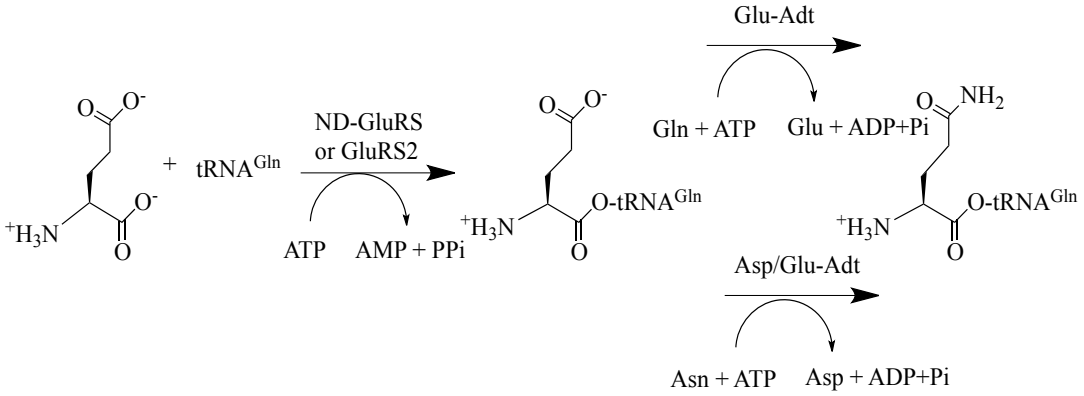


Figure 6. Indirect Gln-tRNA^{Gln} synthesis pathway.

There are two kinds of GluRSs, ND-GluRS and discriminating GluRS (D-GluRS).[10] Species with functional GlnRS such as eukaryotes contain only D-GluRS that generates Glu-tRNA^{Glu}. In all archaea and most bacteria, only ND-GluRS exists. ND-GluRS has a relaxed binding specificity toward tRNAs and binds both tRNA^{Glu} and tRNA^{Gln}.[23, 24] ND-GluRS has a larger anticodon binding site than D-GluRS that accommodates both C36 of tRNA^{Glu} and G36 of tRNA^{Gln}.[25] ND-GlnRS mediates the misacylation of tRNA^{Gln} with glutamate to form Glu-tRNA^{Gln} that undergoes a transamidation reaction to form Gln-tRNA^{Gln}. This transamidation reaction is an ATP dependent process. ATP is used to activate the carboxyl group to form an acyl phosphate. An ammonia molecule provided by hydrolysis of glutamine or asparagine then undergoes an SN2 reaction with the acyl phosphate intermediate to form the final product Gln-tRNA^{Gln} (Figure 7). In archaea, these reactions are carried out by Glu-Adt, a heterodimer composed of the GatD and GatE subunits, using glutamine as the nitrogen source.[26] On the contrary, in bacteria, organelles of some eukaryotes, and some archaea, these reactions are catalyzed by the Asp/Glu-Adt, a heterotrimer composed of the GatC, GatA, and GatB.[26, 27] In some special organisms, both Glu-Adt and Asp/Glu-Adt exist.[26]

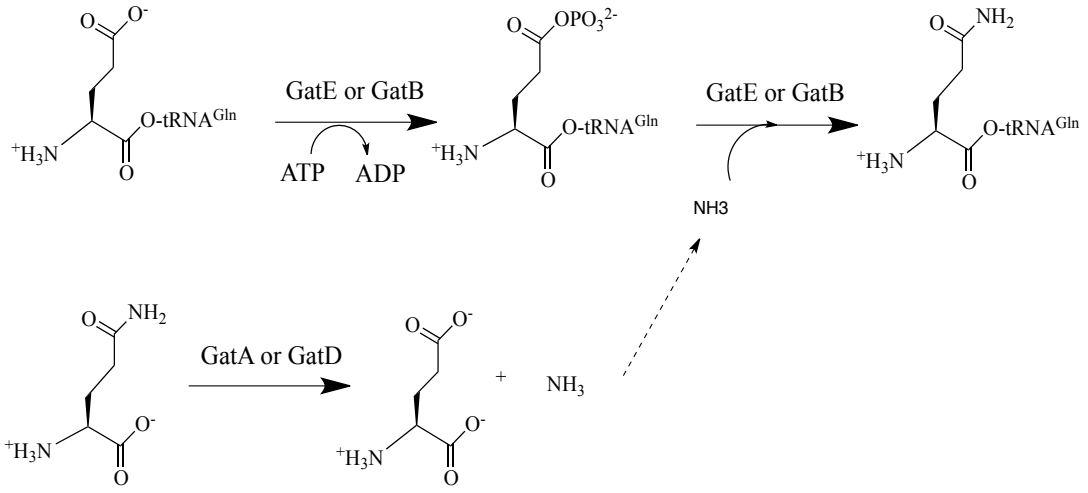


Figure 7. Transamidation reaction.

1.2.6 Indirect biosynthesis of Asn-tRNA^{Asn}

Similar as the indirect biosynthesis of Gln-tRNA^{Gln}, in some organisms that lack AsnRS, Asp-tRNA^{Asn} is generated by misacylation of tRNA^{Asn} with aspartate followed by the conversion of Asp-tRNA^{Asn} to Asn-tRNA^{Asn}.[28, 29] The misacylation reaction is catalyzed by a non-discriminating AspRS (ND-AspRS). The transamidation reaction is accomplished by Asp/Glu-Adt (Figure 8). Although analogous to each other, there are some unique features in the biosynthesis of Asn-tRNA^{Asn}. Unlike the universal absence of GlnRS in archaea, not all archaea are lack of AsnRS. Therefore, there is only one indirect pathway to generate Gln-tRNA^{Gln}, while both direct and indirect pathways to produce Asn-tRNA^{Asn} exist in archaea.[30] In addition, there is only one enzyme that catalyzes the transamidation reaction of Asp-tRNA^{Asn} to Asn-tRNA^{Asn}, as Asp-tRNA^{Asn} is not a substrate for Glu-Adt.[31, 32]

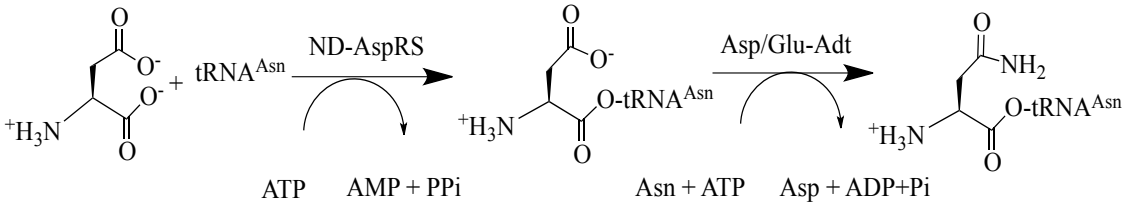


Figure 8. Indirect synthetic pathway of Asn-tRNA^{Asn}.

1.3 The genetic incorporation of NAAs in nature

Proteins are the most important players within a cell by carrying out reactions in most of cellular processes such as DNA replication, transcription, translation, metabolism, signal transduction, etc.[33] The diversity of protein functions arises from three aspects of their structures: primary amino acid sequence, post-translational modification, and protein folding.

Generally speaking, the genetic code of all known organisms specifies the same 20 canonical amino acids. NAAs are amino acids with additional functions groups beyond what 20 canonical amino acids could provide. It is clear that many proteins require additional chemical groups to carry out their native functions. Nature satisfies this requirement by one process that is called posttranslational modifications. Many of functional groups are installed into proteins through covalent posttranslational modifications of amino acid side chains, including acetylation, methylation, phosphorylation, sulfation, glycosylation, lipidation, etc. (Figure 9).[34] These

modifications, to a great extent, grant proteins with expanded opportunities for catalysis, mediating signal transduction, and alteration of cellular locations.

There are more than 140 kinds of amino acids that have been found in natural proteins, most of which are due to the posttranslation modification process. However, two other NAAs selenocysteine and pyrrolysine are incorporated into protein cotranslationally (Figure 10). Selenocysteine, the 21st amino acid, was first found in a subunit of *Clostridium stricklandii* glycine reductase in 1976.[35] Pyrrolysine, the 22nd amino acid, was discovered in methylamine methyltransferase in *Methanosarcina barkeri* in 2002.[36, 37]

1.3.1 Biosynthesis of selenocysteine

Selenium, a trace element required for mammals, serves as pivotal catalytic roles in many enzymes. Selenium carries out its functions in proteins as a form of selenocysteine.[35] Compared with cysteine that has a pK_a close to 8.5, selenocysteine with a pK_a of 5.2 exists as a selenolate anion form at physiological pH.[38] In addition, the selenol group of selenocysteine is more redox sensitive and higher nucleophilic than the thiol group of cysteine. Indeed, most known selenoproteins are oxidoreductases with an essential selenocysteine active site residue. Mutation of selenocysteine in the active sites of these enzymes to cysteine causes more than 100-fold decrease in their catalytic turnover.[38]

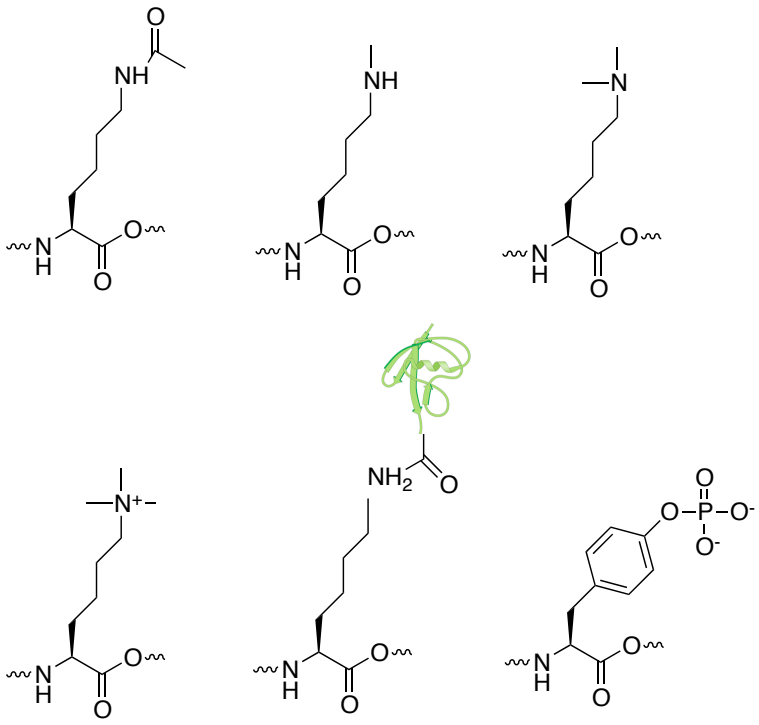
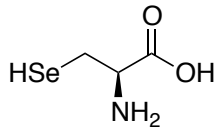
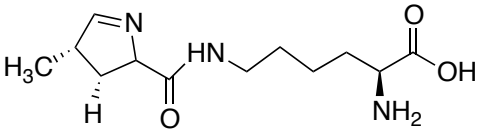


Figure 9. Representative post-translation modifications.



selenocysteine, 21st AA, Coded by TGA



pyrrolysine, 22nd AA, Coded by TAG

Figure 10. Structures of selenocysteine and pyrrolysine.

After its initial identification in a subunit of *Clostridium stricklandii* glycine reductase in 1976,[35] it took ten years to further demonstrated that selenocysteine is directed incorporated into proteins by an in-frame UGA opal stop codon.[39, 40] Because it is genetically encoded, selenocysteine has been named as the 21st amino acid. It is now well accepted that selenocysteine has been used in all three domains of life, but not in all species. Although there are more than 30 selenocysteine-containing proteins in mammalian cells, selenocysteine is not genetically encoded from fungi and higher plants.[41, 42] To date, the biosynthesis pathway of selenocysteine-containing proteins has been elucidated that selenocysteine is incorporated into protein via selenocysteinyl-tRNA^{Sec} (Sec-tRNA^{Sec}).

1.3.1.1 The biosynthesis of selenocysteine-containing proteins in bacteria

The biosynthesis of Sec-tRNA^{Sec} in bacteria follows an indirect pathway, similar as the indirect pathway of Gln-tRNA^{Gln}. There are two steps involved, the acylation of tRNA^{Sec} (SelC) by SerRS to generate Ser-tRNA^{Sec} and the following conversion from Ser-tRNA^{Sec} to Sec-tRNA^{Sec} (Figure 11).[43] The conversion from Ser-tRNA^{Sec} to Sec-tRNA^{Sec} is catalyzed by the enzyme selenocysteine synthase (SelA), a PLP-dependent enzyme, using selenophosphate, which is made by selenophosphate synthase (SelD) from selenide and phosphate, as the selenium donor.[44] Later, the synthesized Sec-tRNA^{Sec} is incorporated into a protein at an opal UAG codon with the help of a Sec-tRNA^{Sec} specific elongation factor (SelB) since the endogenous elongation could not recognize Sec-tRNA^{Sec}.[45] The site-specificity is control by a special hairpin loop element structure that is called selenocysteine insertion sequence (SECIS) in the encoding

mRNA. In bacteria, the SECIS hairpin is located immediately downstream from the in-frame UGA opal stop codon.[46]

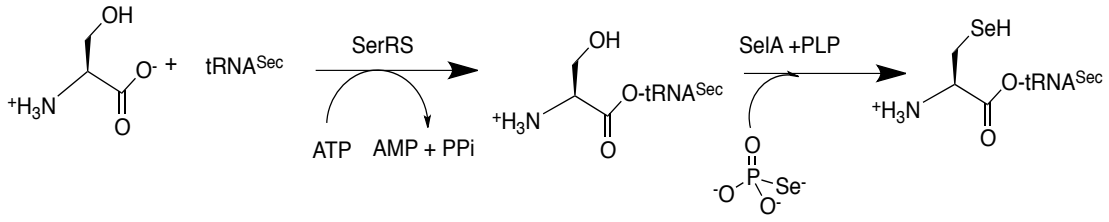


Figure 11. Indirect Sec-tRNA^{Sec} synthetic pathway in bacteria.

1.3.1.2 The biosynthesis of selenocysteine-containing protein in eukaryotes and archaea

In mammals, Sec-tRNA^{Sec} biosynthesis is via a pathway different from that used in bacteria. tRNA^{Sec} is acylated with serine by SerRS and then phosphorylated by an *O*-phosphorylseryl-tRNA^{Sec} kinase (PstK) to form Sep-tRNA^{Sec}. Sep-tRNA^{Sec} is converted to Sec-tRNA^{Sec} by SepSecS, a PLP dependent enzyme (Figure 12).[47, 48] Different from bacteria, the SECIS motif is no longer located right after the stop codon. Instead, the SECIS is located in the 3' noncoding region of mRNA.[49] In addition, a special elongation factor (EF_{Sec}) and a unique SECIS-binding protein (SBP2) need to form a complex to mediate the binding of SECIS with Sec-tRNA^{Sec} for the correct recognition of an opal codon in mRNA.[50]

The incorporation of selenocysteine into proteins in archaea is the combination of those in bacteria and eukaryotes. The biosynthesis of Sec-tRNA^{Sec} is as same as the

pathway in mammals,[48] and the SECIS motif is far away from opal stop codon. However, only SelB is required to associate Sec-tRNA^{Sec} with SECIS for the correct recognition of an mRNA opal codon.[51]

1.3.2 Biosynthesis of pyrrolysine

In 1998, Krzycki et al. reported that there is a traditional UAG amber stop codon in the open reading frame of monomethylamine methyltransferase (MtmB) in *Methanosarcina barkeri*.[36] The crystal structure of MtmB clearly indicated the presence of a lysine analog at this amber codon site that was later confirmed to be pyrrolysine.[37, 52] MtmB is a critical enzyme in *Methanosarcinaceae*.[37] *Methanosarcinaceae* are a group of methanogenic archaea that are able to reduce a wide variety of compounds to methane. *Methanosarcinaceae* are special for that they produce methane from methylamines (mono-, di-, and trimethylamines). This process requires specific methyltransferases, such as monomethylamine methyltransferase (mtmB), dimethylamine methyltransferase (mtbB), or trimethylamine methyltransferase (mttB).[53] The genes of MttB, MtbB, and MtmB contain a single in-frame amber codon.[54, 55] Based on the fact that pyrrolysine was observed in the crystal structure to bind ammonia,[52] the mechanism of methyltransferation is proposed in Figure 13.[56] MttC, MtbC, and MtmC are cognate corrinoid proteins, which involve in the conversion of the methyl group to carbon dioxide.[57, 58]

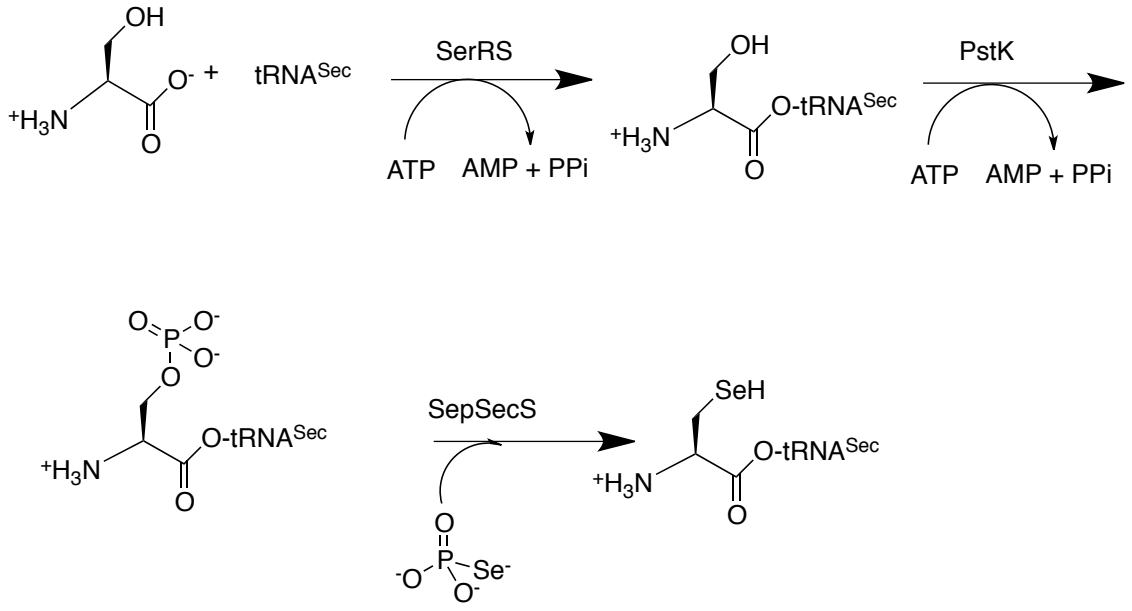


Figure 12. Indirect Sec-tRNA^{Sec} synthetic pathway in eukaryotes.

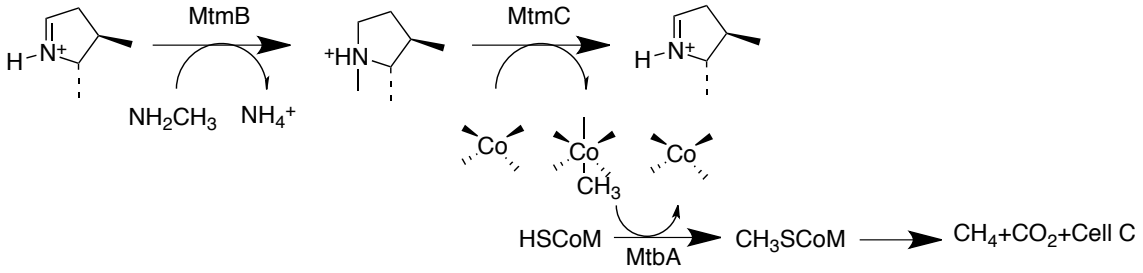


Figure 13. Monomethylamine metabolism in *Methanosarcinaceae*.

1.3.3 The cotranslational incorporation of pyrrolysine

Efforts were made to elucidate how pyrrolysine was cotranslationally inserted into proteins after its discovery. Bioinformatic analysis of the *Methanosarcina barkeri* genome revealed a pyl gene cluster that contains pylT, pylS, pylB, pylC, and pylD (Figure 14).[59] pylT codes a special tRNA, $tRNA^{Pyl}$ that has a CUA anticodon and pylS codes a putative class II aminoacyl-tRNA synthetase.[37] PylB, pylC and pylD have putative functions in the biosynthesis of pyrrolysine. To date, there are six bacterial and six archaeal species found with similar pyl gene clusters.

After pylT was identified, two possible routes for the biosynthesis of pyrrolysine-containing proteins were proposed. PylS might acylate pylT with pyrrolysine for its direct incorporation at an amber codon or lysine is first linked to pylT either by pylS or another aaRS, followed by further modification to generate pyrrolysyl (Pyl)-$tRNA^{Pyl}$. Originally, it was uncovered that both class I and class II LysRSs in *M. barkeri* were required to form lysyl-$tRNA^{Pyl}$, indicating the indirect synthesis route to Pyl-$tRNA^{Pyl}$, like the 21st amino acid selenocysteine.[60] It is worth noting that pylS itself could not ligate lysine to pylT. However, later in 2004,[61] Krzycki group demonstrated that pyrrolysine is attached as a free molecule to pylT by pylS *in vitro*. This is the first example of direct aminoacylation of a tRNA with a NAA, earlier than SepRS. The idea was further confirmed by the fact that the pair of pylS and pylT was moved to *E. coli* for the genetic incorporation of NAAs at amber mutation sites. Unlike sec-$tRNA^{Sec}$, Pyl-$tRNA^{Pyl}$ could be recognized by endogenous elongation factor, EF-Tu.[62] The mechanism of the genetic incorporation of pyrrolysine is shown in Figure 15.[59]

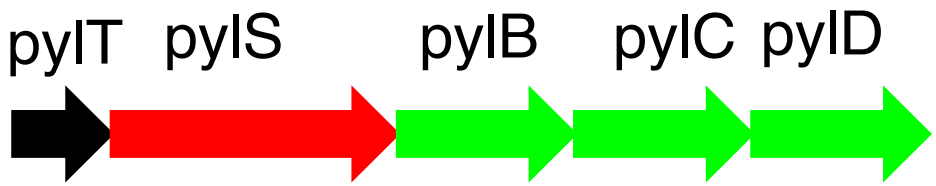


Figure 14. Pyl gene cluster.

1.3.4 The introduction to pylS

PylRS coded by pylS is an archaeal class II aaRS.[61] Since PylRS has been found in different organisms, the crystal structures of several PylRS variants in archaea, such as *Methanosarcina mazei* (*Mm*PylRS),[63] and bacteria *D. hafniense* (*Dh*PylRS) has been solved (Figure16).[64-66] *Mm*PylRS has two major domains, the structurally unresolved N-terminal domain, and the catalytic C-terminal domain.[63] However, *Dh*PylRS only contains the catalytic domain. The N-terminal domain equivalent is coded by a different gene that is called pylSn.[66] With classic β-sheet core and three special motifs, PylRS is a class II aaRS. As a matter of fact, PylRS belongs to subclass IIc, together with AlaRS, SepRS, and PheRS. Subclass IIc aaRSs all have a similar core domain. In addition, all of them could potentially form quaternary complexes.[63] Actually, phylogenetic tree suggests that they may evolve from a same ancestor aaRS.

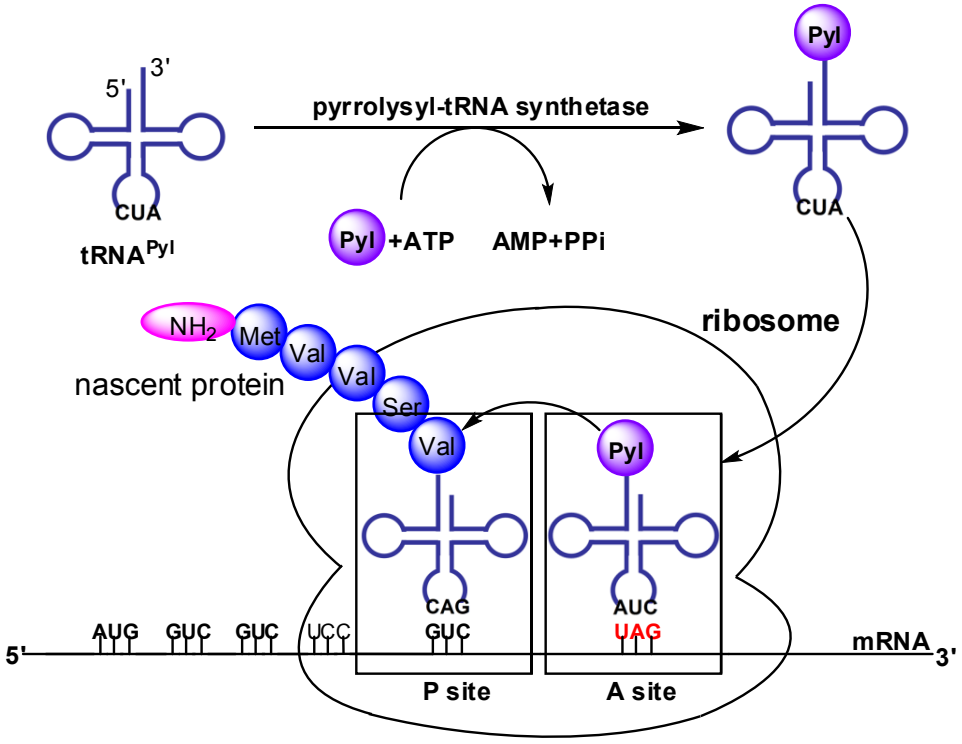


Figure 15. Incorporation of pyrrolysine.

Another question about aaRS is whether a similar mRNA secondary structure like SECIS is required for the incorporation of pyrrolysine, just like selenocysteine. There are some data showed that the cis-element PYLIS enhanced the suppression level of pyrrolysine incorporation in *E. coli*.[67] Since there is no particular motif for PylRS to discriminate TAG as an in-frame sense codon or stop codon, how could pyrrolysine-containing species function normally? Indeed, pyrrolysine-containing species have shown a low occurrence frequency of this stop codon.[29]

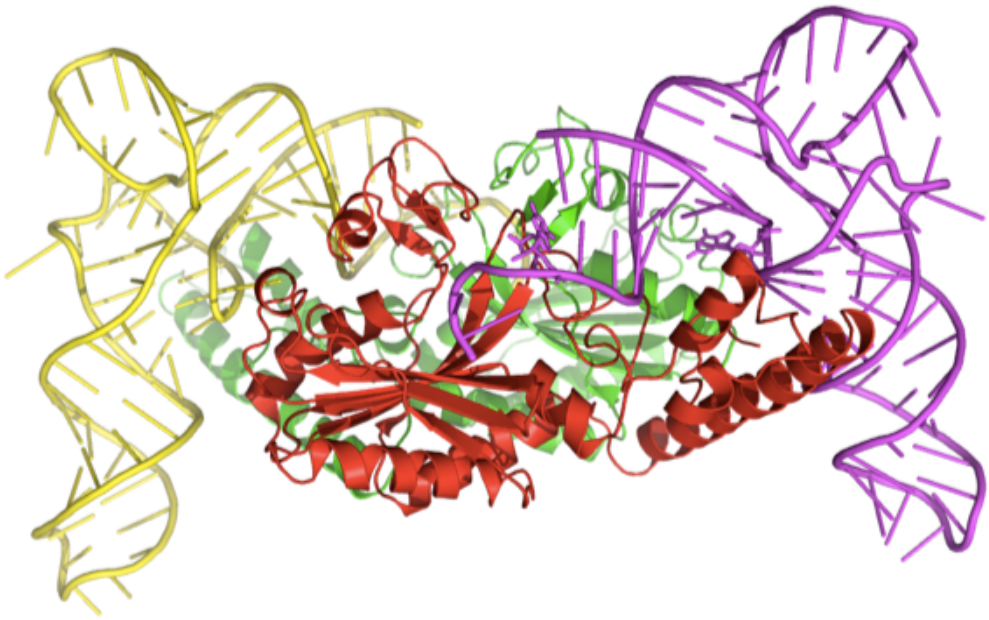


Figure 16. Crystal structure of *Dh*PylRS with pylT.

1.3.5 The introduction to pylT

The DNA sequences of pylT differ a lot. PylT variants from archaeal share 87% sequence identity, while pylT variants from bacteria only share 45% sequence identity. In spite of very diversified DNA sequences, all pylT homologs have same unusual secondary structures (Figure 17), including a three-base variable loop (not 4 bases), small D-loop, long anticodon stem (six nucleotide pairs instead five), and a single base between the D-stem and acceptor stem (not 2 bases).[59, 64]

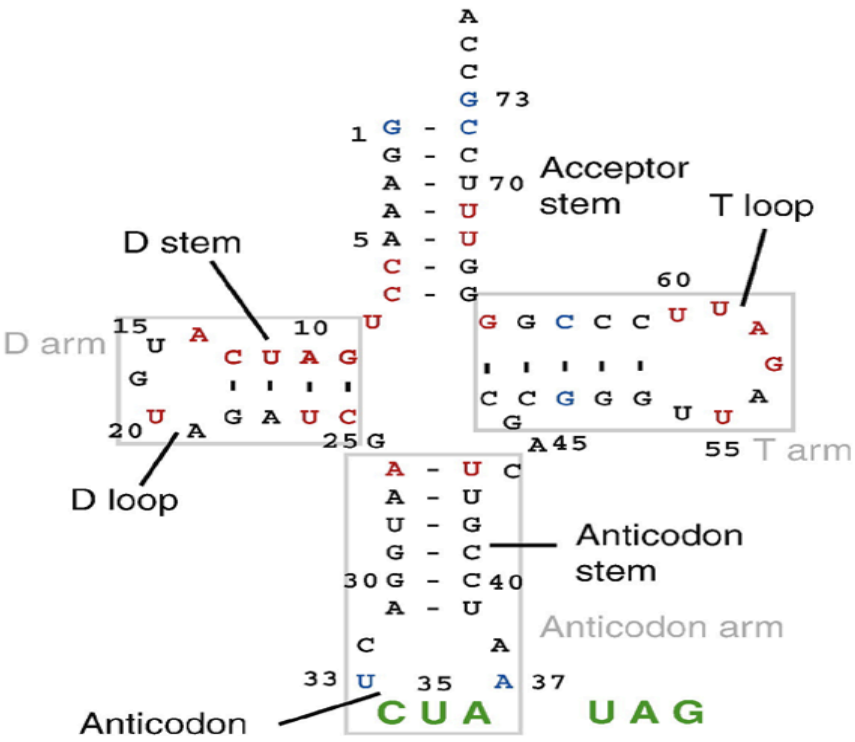


Figure 17. Structure of *Mm*pylT.

1.3.6 The biosynthesis of pyrrolysine

Like all other 20 amino acids, pyrrolysine should first be synthesized and then be incorporated into proteins. Thus, there leaves another question about the biosynthesis of pyrrolysine. It has been demonstrated that transferring the whole pyl gene cassette into *E. coli* led to the suppression of amber mutations. Besides pylT and pylS, the same pylBCD arrangement has been maintained in all pyrrolysine-containing species.[59] The remaining pylBCD genes in the pyl gene cluster are expected to involve in the enzymatic pathway of biosynthesis of pyrrolysine.[68] Protein sequence alignment provided clues for the functions of pylBCD. PylB has typical structure of SAM family. PylC belongs to the ligase family. pylD has the feature of dehydrogenases.[69]

At the beginning, one lysine and other metabolic products, such as ornithine[61] were proposed to be the precursors of the pathway. Recently, Krzycki and his colleagues have demonstrated only lysine is the precursor. The route for the conversion of two lysines to pyrrolysine has been elucidated using stable isotope labeling and mass spectrometry (Figure 18).[70] The SAM-dependent protein pylB conducts the conversion of a lysine to 3-methylornithine, followed by a ligation reaction with another lysine by pylC. After the oxidation reaction carried out by pylD, pyrrolysine is formed by the spontaneous elimination of H_2O.

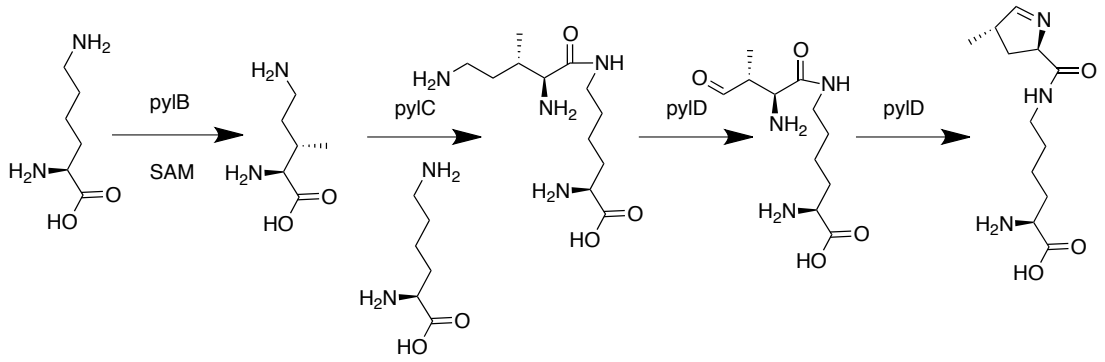


Figure 18. Biosynthetic route for pyrrolysine.

1.4 PylRS: a mediocre enzyme

With increasing understanding of pyrrolysine, the more surprising features of pyrrolysine have been uncovered. These unique features make the PylRS-pylT pair gain more and more attentions.

1.4.1 The orthogonality of the PylRS-pylT pair from archaea in both bacteria and eukaryotes

The orthogonality mentioned here means that the tRNA/codon/aaRS set, termed the orthogonal set, must not crosstalk with endogenous tRNA/codon/aaRS sets. The orthogonal tRNA should not be recognized by any endogenous aaRS, and it should decode the orthogonal codon, which is not assigned to any canonical amino acid. The orthogonal aaRS should not charge any endogenous tRNA but the orthogonal tRNA. In addition, it should charge the orthogonal tRNA with a NAA only. The requirement of orthogonality is essential for introducing a new tRNA/codon/aaRS set into any organism. Basically, there are three domains of life, archaea, bacteria and eukaryotes.

From the evolution history, bacteria first separated from the other two, and then the remaining life has divided into archaea and eukaryotes. Thus, the tRNA/codon/aaRS sets in the bacteria are possibly orthogonal to the pairs in the archaea and eukaryotes. The tRNA/codon/aaRS sets in the archaea and eukaryotes may crosstalk to each other. It is very hard to find one aaRS-tRNA pair that is both orthogonal in bacteria and eukaryotes, two systems we use a lot for the genetic code expansion. Therefore, the orthogonality of the PylRS-tRNA pair in both bacteria and eukaryotes makes it an optimal choice because the pair can be easily evolved in bacteria and then transfer to other domains of life. This special orthogonality of the PylRS-pylT pair basically in three domains of life might be due to the fact that PylRS appeared before the division of life.[63]

1.4.2 The deep hydrophobic binding pocket of PylRS

Due to the large size of pyrrolysine, the amino acid active site of PylRS has a very deep hydrophobic binding pocket (Leu305, Tyr306, Leu309, Cys348, and Tyr384) to sandwich the pyrrole group of pyrrolysine.[63] The phenolic hydroxyl group of Tyr384 forms a hydrogen bond with nitrogen of the pyrrole ring. However, there is no hydrogen bond with ε-amino group (Figure 19).

Given that there is only one hydrogen bond involving in the binding of pyrrolysine and no other natural amino acids could be taken by PylRS, the specific interaction between PylRS and pyrrolysine is mediated by van der Waals or hydrophobic interactions between the binding pocket and the pyrrole group. In comparison to the size of the pyrrole group, the binding pocket is exceptionally large, which clearly indicates the binding site may fit other large NAAs that contain hydrophobic side chains. The lack

of an editing domain will allow misacylated $tRNA^{Pyl}$ to leave PylRS and deliver the linked NAA to ribosome for its incorporation at an amber mutation site.

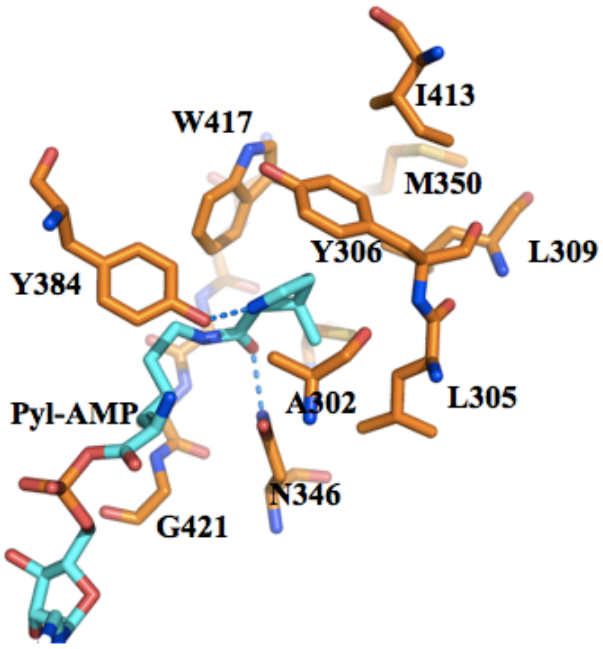


Figure 19. Binding pocket of *Mm*PylRS with Pyl-AMP.

An editing domain is very important to maintain translation fidelity. The charging of the tRNA with its cognate amino acid is critical for the accurate translation. An aaRS needs to specifically acylate its tRNA with its cognate amino acid, but against structurally similar amino acids. The existence of an editing domain solves this problem. Taking IleRS for example, valine differs from isoleucine only by the lack of a single methyl group. Surprisingly, IleRS, with an editing domain, transfer 40000 or so

isoleucine molecules to $tRNA^{Ile}$ for every valine it can transfer (Figure 20).[71, 72] However, there is no editing domain in the PylRS, making the use of PylRS to charge pylT with NAAs possible.

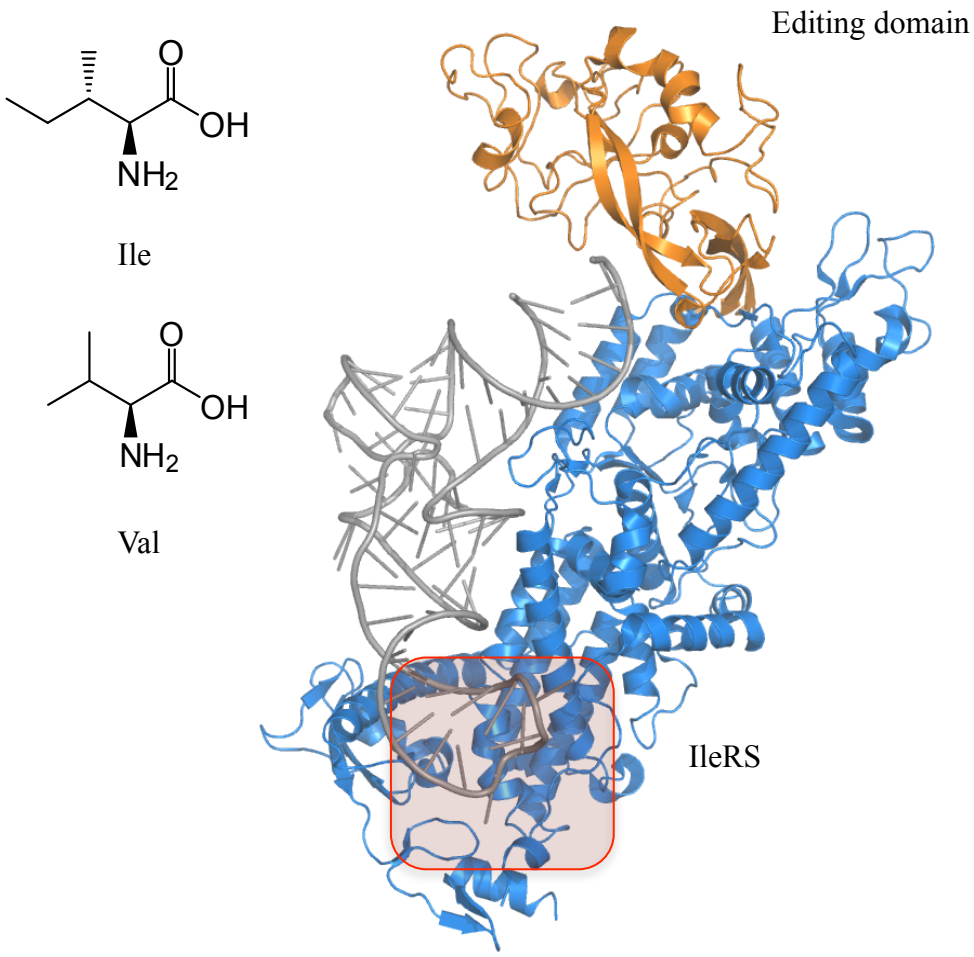


Figure 20. Editing domain of IleRS.

1.4.3 The weak interaction between α-amino group of pyrrolysine and PylRS

According to the PylRS crystal structure, there are two significant residues. One is the Tyr384 mentioned above, and the other one is Asn346. Asn346 is important for binding of pyrrolysine to the active site at two aspects. Asn346 forms hydrogen bonds to both the side chain amide oxygen of pyrrolysine and a water molecule that also forms a hydrogen bond with the α-amino group of pyrrolysine (Figure 21).[63] PylRS is the only known aaRS that does not have any direct interaction with the α-amino group of its cognate amino acid. The α-amino group of pyrrolysine is not involved in direct interactions with residues of PylRS. This makes it possible for PylRS to recognize α-hydroxyl acid derivative. The possible reason for PylRS not to specifically recognize pyrrolysine could be due to the fact that there is no similar hydroxyl acid existing in cells and the specific recognition of the α-amino group may increase the binding of the ligand to an extent that other natural amino acid could be recognized by the enzyme.

1.4.4 The lack of stringent interactions between the pylT anticodon and PylRS

As an aaRS, PylRS interacts with both pyrrolysine and pylT. Dieter Söll group demonstrated the high conservation of the interaction between pylT and PylRS.[64] There is 54% sequence identity for pylT binding regions of PylRS from different origins, compared with only 39% sequence identity for the whole protein.[64] The binding of pylT by PylRS relies on a combination of direct interactions with certain nucleotides of pylT and recognition of the overall tertiary structure of pylT.[64] Detailed studies of the interactions uncovered the pylT major identity elements, including the discriminator base G73 and the first base pair of the acceptor stem.[73] The mutation at D-loop, T-loop

and variable stem also showed the decreasing binding of pylT to PylRS,[73] while no direct interaction between these regions of pylT with PylRS. These data further confirmed the idea that the recognition of pylT is a combination of the specific binding of identity elements and the association of overall structure of pylT to PylRS.

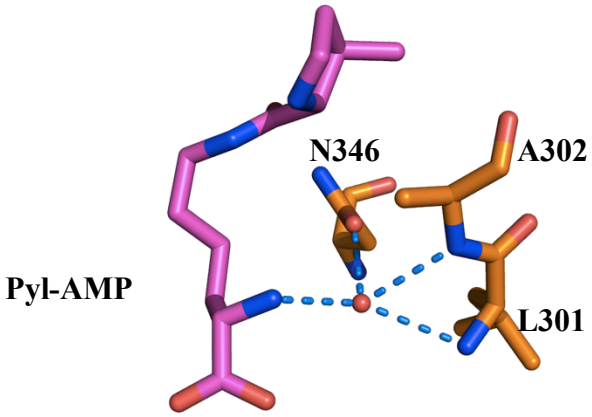


Figure 21. Water-mediated binding of α-amino group.

It's worthy of note that there is no specific interaction between pylT anticodon and PylRS.[73] The mutations of the anticodon from TAG to other codons have almost no effect on the binding of pylT to PylRS. PylRS may represent a derivative of a primitive state of aaRS arose before the last universal common ancestral state that is also the

ancestor of several other aaRSs. The lack of the interactions with pylT anticodon might make the primitive aaRS easily evolve to other aaRSs, such as PheRS.

1.5 PylRS: a gift from nature

With few exceptions, the genetic code of all known organisms specifies the same 20 canonical amino acid building blocks. However, it is clear that many proteins require additional chemical groups beyond what the 20 building blocks can provide to carry out their native functions.[34] Many of these groups are installed into proteins through covalent posttranslational modifications of amino acid side chains, including acetylation, methylation, phosphorylation, sulfation, etc. These modifications, to a great extent, expand the amino acid inventory of proteins by assimilating NAAs and grant proteins expanded opportunities for catalysis, mediating signal transduction, integration of information at many metabolic intersections, and alteration of cellular locations.

A major challenge in studying posttranslationally modified proteins is that they exist typically as a mixture of different forms. This makes it difficult to purify uniquely modified forms from their natural sources. For this reason, a variety of approaches have been developed to synthesize proteins with defined posttranslational modifications that can be easily isolated. One of these approaches uses the nucleophilicity of cysteine to incorporate NAAs into proteins.[74] Native chemical ligation and its extension, expression protein ligation, have also been used to install NAAs into proteins.[75] These two approaches provide great opportunities to study functional roles of posttranslational modifications. However, they both are *in vitro* semisynthetic techniques and cannot been applied *in vivo*. An alternative NAA installation approach that can be carried out *in vivo*

is the genetic incorporation of NAAs directly into proteins during translation in living cells.

1.5.1 Introduction of genetic incorporation of NAAs

The method of genetic incorporation of NAAs in live cells has been developed by Schultz et al.[76-82] This method relies on the read-through of an in-frame amber (UAG) stop codon in mRNA by an amber suppressor tRNA ($tRNA_{CUA}$) specifically acylated with a NAA by an evolved aaRS. An orthogonal aaRS-$tRNA_{CUA}$ pair can be developed to specifically charge its cognate $tRNA_{CUA}$ with a NAA. When expressed in cells, this aaRS-$tRNA_{CUA}$ pair enables the NAA to be site-specifically incorporated into a protein.

Several aaRSs have been applied for genetic incorporation of NAAs, such as TyrRS from *E. coli*[83] and *Methanococcus jannaschii*,[84] TrpRS from *E. coli*,[84] PheRS from *S. cerevisiae*,[85] LysRS,[84] and LeuRS from *E. coli*.[84] Among all the available aaRSs, TyrRS from *M. jannaschii* is the most used one for the genetic code expansion, for more than 50 NAAs have been incorporated into proteins by evolved *Mj*TyrRS.[86]

1.5.2 The unique orthogonality of PylRS

The introduction of aaRS-$tRNA_{CUA}$ into a new system has one prerequisite, the orthogonality. As mentioned above, the aaRS-$tRNA_{CUA}$ sets in the bacteria are possibly orthogonal to the pairs in the archaea and eukaryotes. The aaRS-$tRNA_{CUA}$ sets in the archaea and eukaryotes may crosstalk to each other. Thus, the aaRSs from archaea have been introduced into *E. coli*, while not eukaryotes. On the contrary, the aaRSs from *E. coli* are used to expand the genetic codon in eukaryotes. The incompatibility of using one aaRS-$tRNA_{CUA}$ pair in different organisms such as in both *E. coli* and eukaryotes

has greatly hindered the genetic code expansion for the NAA incorporation in these organisms.

For above reasons, the discovery that PylRS-pylT pair is orthogonal to both *E. coli* and eukaryotes, makes it a very attractive genetic code expanding tool. As a matter of fact, ever since Yokoyama group first put the pyrrolysine analogs into mammalian cells in 2008,[87] many NAAs have been incorporated into proteins in eukaryotes by evolving the PylRS-pylT pair in *E. coli* and then transferred to eukaryotic cells such as yeast and mammalian cells.[86, 88, 89]

1.5.3 The deep hydrophobic binding pocket

The evolution of aaRS mutants to specifically recognize one NAA begins with the generation of a library in the active site. The residues around the active site have been randomLy mutated to generate the library to take other amino acids. However, the variety of the amino acids it could take largely depends on the original size of the active site. Taking TyrRSs for example, the mutants mainly take phenylalanine and tyrosine derivatives, and some other NAAs with short side chains.

The deep pocket of PylRS offers a great opportunity to introduce the NAAs with long side chains. Indeed, many lysine analogs have been incorporated into proteins by PylRS or its mutants (Figure 22). In addition, phenylalanine and tyrosine analogs have also been reported to be taken by PylRS.[90] Therefore, the large pocket provides the diversity of NAAs.

1.5.4 The lack of strong interactions between PylRS and pyrrolysine α-amino group

PylRS is the only known aaRS that interacts with the α-amino group of the substrate indirectly through a water molecule. The lack of strong interactions grants a chance of incorporating non-α-amino substrates. It's of our knowledge that the backbone of proteins is consisted of the amide bond. The possibility of introducing non-α-amino acids into proteins is expected to expand the backbone diversity of proteins. Yokoyama et al. reported the successful incorporation of one non-α-amino acid: α-hydroxyl acid (Figure 23).

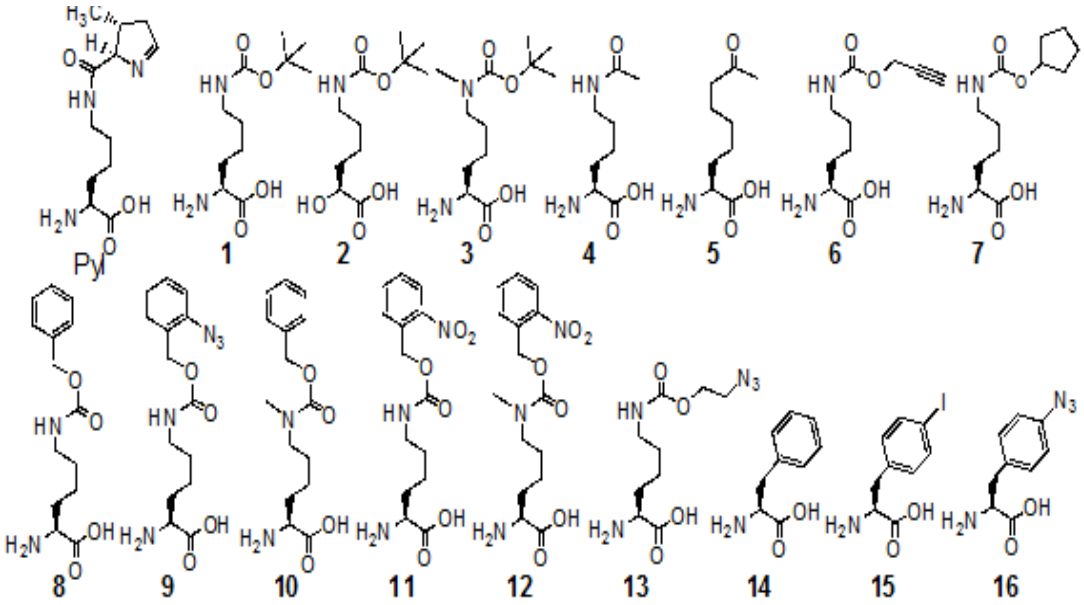


Figure 22. Representative NAAs taken by wild type and evolved PylRSs.

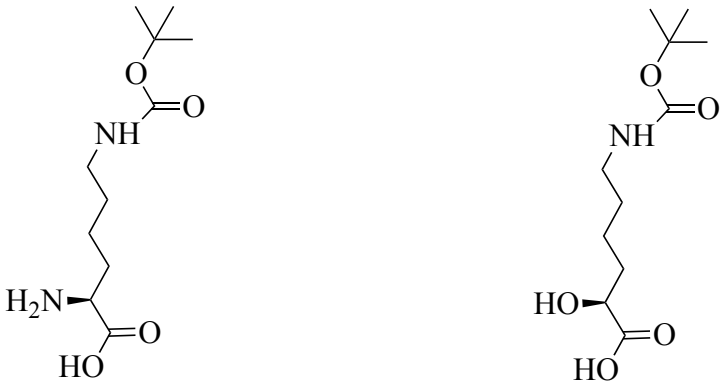


Figure 23. Non-α-amino acid incorporated into proteins.

1.5.5 The lack of specific interactions between PylRS and pylT anticodon

The amber stop codon is the primarily used codon for codon expansion *in vivo* due to the specific interactions between most aaRSs used for genetic code expansion and their cognate tRNAs. It's not the case for PylRS. Without the concern of specific interactions with PylRS, the anticodon can theoretically be mutated to any codon, which greatly promotes the codon expansion. The introduction of other codons is extremely critical for simultaneous translational incorporation of two different NAAs into a single protein. In addition, different codons could be tested for optimized suppression in different systems.

2. GENETICALLY ENCODING KETOK INTO ONE PROTEIN IN *ESCHERICHIA COLI**

2.1 Introduction

The keto group is one of the most versatile functional groups in organic chemistry due to its high reactivity with hydrazide, and hydroxylamino-bearing compounds under physiological conditions.[91] This unique feature has been widely used, such as biopolymer modification (Figure 24).[92] Due to the usually absence from proteins, a variety of approaches have been developed to achieve the of the keto or aldehyde group into proteins. For instance, an N-terminal serine could be transferred to aldehyde group in the presence of periodates.[93] The keto group has also been obtained by enzymatic modifications of specific tags that are genetically linked to proteins and utilized to site-specifically label proteins on cell surface.[94-96] Although both chemical and enzymatic methodologies provide ways for installation of keto group, the site of installation has been limited to the two termini of protein. Given the above fact, site-specific incorporation of NAAs by an orthogonal aaRS/tRNA pair is an alternative powerful approach to selectively install keto group at any desired site of a protein.

After the genetic incorporation of *p*-acetyl-phenylalanine (*p*-Ac-Phe) into proteins in *E. coli*,[97] several more phenylalanine derivatives bearing ketone group have

---

*Reprinted with permission from "Genetic Incorporation of an Aliphatic Keto-Containing Amino Acid into Proteins for Their Site-Specific Modifications" by Huang, Y., Wan, W., Russell, W. K., Pai, P. J., Wang, Z., Russell, D. H., Liu, W., 2010. *Bioorg. Med. Chem. Lett.*, 20, 878-880, Copyright [2010] by Elsevier.

been reported to be successfully incorporated into proteins.[98, 99] However, besides the potential protein misfolding problem caused by the introduction of bulky aromatic ring, the aromatic ring decreases the activity of ketone group, for no significant reaction could occur under physiological conditions. Thus, it's desirable to incorporate an aliphatic keto-containing amino acid into protein to gain high reactivity.

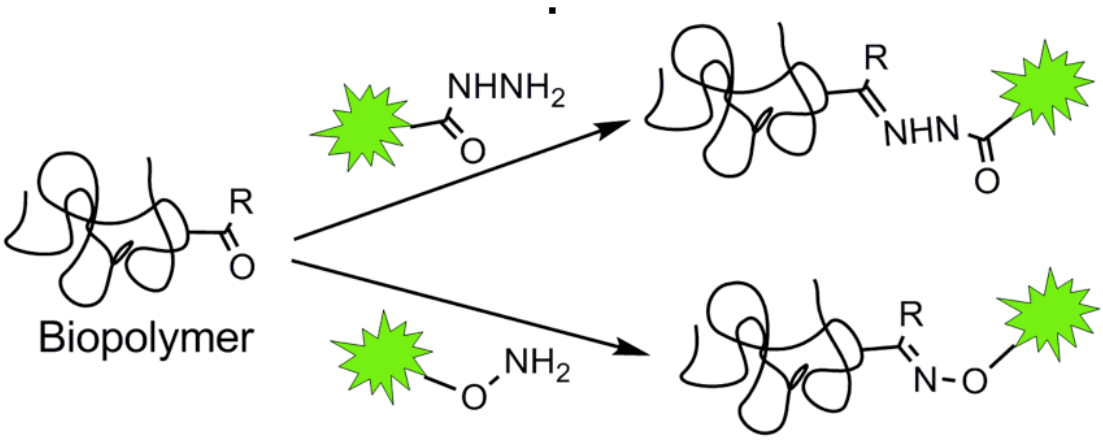


Figure 24. Reaction of the keto group in biopolymers.

Recently, Chin group demonstrated that the evolved *Mb*PylRS (AcKRS)/$tRNA_{CUA}$ pair can be directly used to incorporate acetyllysine (AcK) into proteins.[100] Since the *N*-*ε*-amide NH group of AcK is not involved in any apparent hydrogen bonding interactions with vicinal residues of AcKRS when bound to the active site,[63] we reasoned that changing the amino acid substrate from AcK to KetoK should not significantly affect the substrate's binding potential to AcKRS. Herein, I proposed to

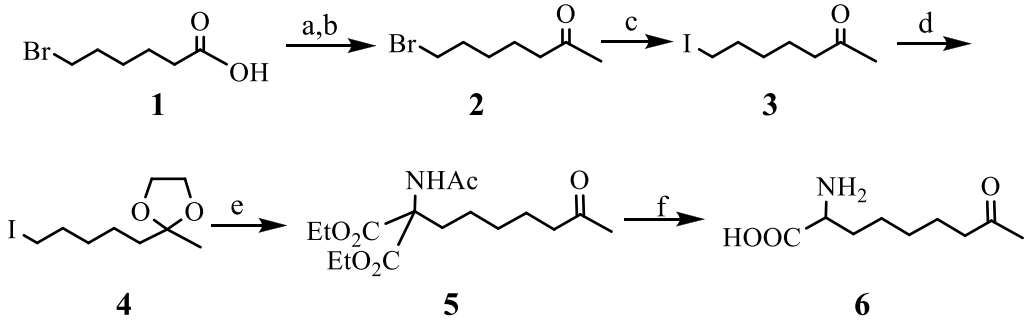
use the AcKRS/pylT pair to genetically incorporate KetoK into proteins in *E. coli.* Our interest of KetoK is due to the chemical activity of aliphatic keto group, and it is an unhydrolysable mimic of AcK.

2.2 Experiments and results

2.2.1 The synthesis of 2-amino-8-oxononanoic acid (KetoK)

Starting from 6-bromohexanoic acid, racemic KetoK was conveniently synthesized in gram quantities over a 6-step sequence in 50% overall yield (Scheme 1).

Scheme 1. Synthesis of KetoK



Conditions: (a) EDC, DMAP (cat.), *N*,*O*-dimethylhydroxylamine, CH_2Cl_2; (b) CH_3MgBr, THF, 83% for two steps; (c) NaI, acetone, reflux, 95%; (d) $HOCH_2CH_2OH$, *p*-TsOH, toluene, 89%; (e) NaH, diethylacetamidomalonate, DMF, 82%; (f) 12 *N* HCl, reflux, quant.

2.2.1.1 7-Bromoheptan-2-one[101]

A modified literature procedure[101] was employed. A mixture of **1** (2.73 g, 14.0 mmol), 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (3.22 g, 16.8 mmol), 4-dimethylaminopyridine (0.17 g, 1.4 mmol), and *N,O*-dimethylhydroxylamine (1.5 g, 15.4 mmol) in dichloromethane (75 mL) was stirred at room temperature for 16 h. The mixture was diluted in dichloromethane (100 mL), washed with water (30 mL), hydrochloric acid (0.5 *N*, 30 mL), sodium hydroxide (0.5 *N*, 30 mL) and brine (15 mL), dried (Na_2SO_4), and concentrated under reduced pressure to give the crude 6-bromo-*N*-methoxy-*N*-methylhexanamide (3.39 g, quant.) as yellow oil. The material was directly used without further purification.

To a stirred solution of the above amide (3.39 g, ~14.0 mmol) in anhydrous THF (40 mL) cooled in an ice bath under argon protection was added a solution of methylmagnesium bromide (3.0 M in diethyl ether, 7.4 mL, 22.2 mmol) dropwise over 20 min. After stirring for 3 h, hydrochloric acid (1.0 *N*, 21.9 mL) was cautiously added dropwise. The mixture was diluted in ethyl acetate (50 mL), washed with water (20 mL) and brine (20 mL), dried (Na_2SO_4), and concentrated under reduced pressure to give crude **2** (2.35 g, 83%) as yellow oil, which was directly used in the next step. ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.41 (2 H, t, J = 6.9 Hz), 2.46 (2 H, t, J = 7.2 Hz), 2.15 (3 H, s), 1.88 (2 H, appar. quintet, J = 6.9 Hz), 1.65-1.54 (2 H, m), 1.49-1.37 (2 H, m) (Figure 25).

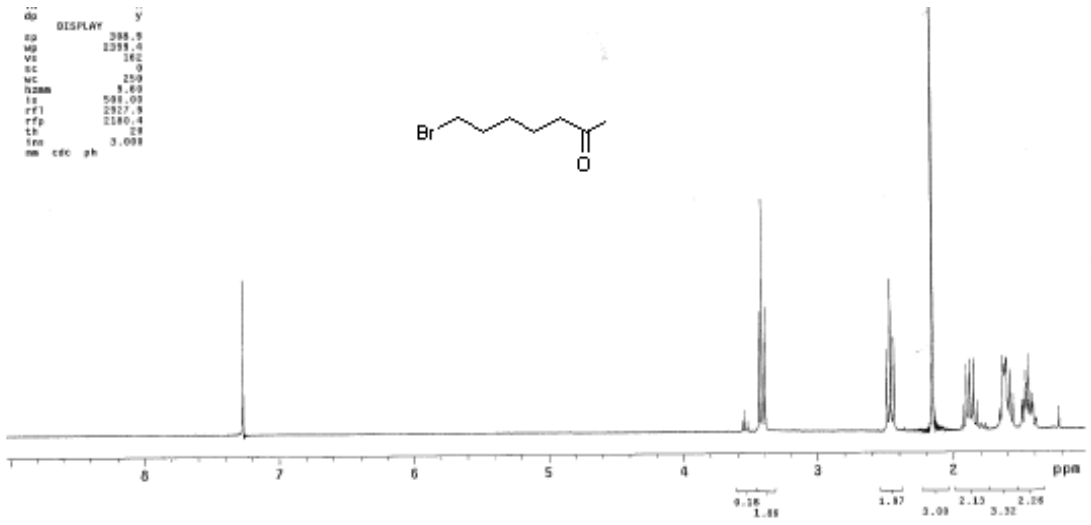


Figure 25. ^{1}H NMR spectrum for compound **2**.

2.2.1.2 7-Iodoheptan-2-one[101]

A literature procedure[101] was followed. A mixture of **2** (2.35 g, 12.1 mmol) and sodium iodide (5.47 g, 36.4 mmol) in acetone (40 mL) was refluxed for 5 h. The solvent was then evaporated under reduced pressure and ether (50 mL) was added. The suspension was filtered and washed with ether (20 mL). The combined filtrate was washed with water (15 mL), sodium thiosulfate (1.0 M, 10 mL) and brine (10 mL), dried (Na_2SO_4), and concentrated under reduced pressure to give **3** (2.90 g, 95%) as yellow oil, which was directly used in the next step. ^{1}H NMR (300 MHz, $CDCl_3$) δ 3.19 (2 H, t, J = 6.9 Hz), 2.46 (2 H, t, J = 7.2 Hz), 2.15 (3 H, s), 1.86 (2 H, appar. quintet, J = 6.9 Hz), 1.60 (2 H, appar. quintet, J =7.5), 1.45-1.32 (2 H, m) (Figure 26).

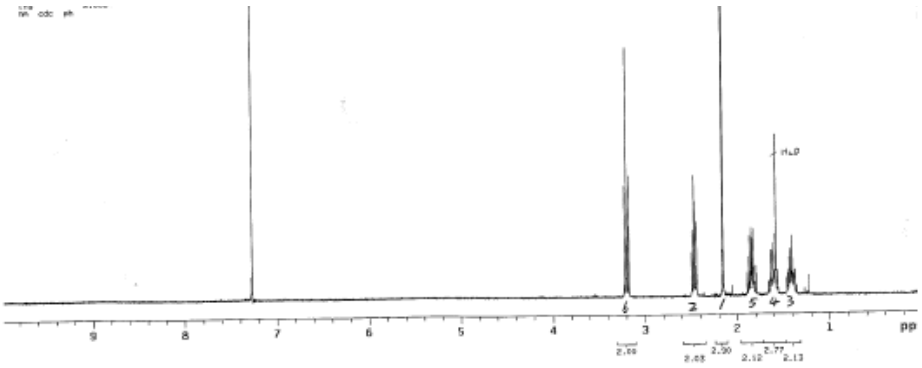


Figure 26. ^{1}H NMR spectrum for compound **3**.

2.2.1.3 2-(5-Iodopentyl)-2-methyl-1,3-dioxolane[101]

A literature procedure[101] was followed. A mixture of **3** (3.33 g, 13.2 mmol), ethylene glycol (4.09 g, 66 mmol), and *p*-toluenesulfonic acid (0.75 g, 3. 96 mmol) in toluene (60 mL) was heated at reflux for 48 h, during which time the generated water was removed by a Dean-Stark apparatus. The reaction was monitored by TLC. Most of the solvent was removed through the Dean-Stark apparatus, and the residue was redissolved of ethyl acetate (60 mL), washed with sodium hydroxide (0.5 *N*, 20 mL) and brine (15 mL), dried (Na_2SO_4), concentrated, and chromatographed (ethyl acetate/hexanes, 1:9) to give **4** (3.35 g, 89%) as yellow oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ 4.00-3.83 (4 H, m), 3.20 (2 H, t, J = 7.2 Hz), 2.46 (2 H, t, J = 7.5 Hz), 2.15 (3 H, s), 1.90-1.79 (2 H, m), 1.70-1.54 (2 H, m), 1.50-1.36 (2 H, m) (Figure 27).

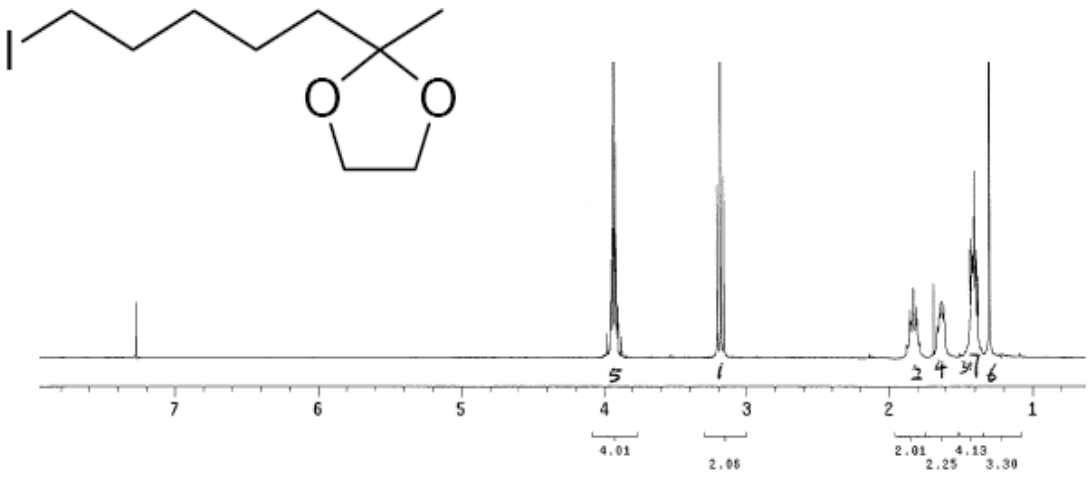


Figure 27. ^{1}H NMR spectrum for compound **4**.

2.2.1.4 Diethyl 2-acetamido-2-(5-(2-methyl-1,3-dioxolan-2-yl)pentyl) malonate

To a suspension of diethylacetamidomalonate (0.70 g, 3.60 mmol) in anhydrous DMF (2 mL) was added sodium hydride (60% dispersion in mineral oil, 0.15 g, 3.84 mmol), and the mixture was stirred at room temperature for 10 min. Compound **4** (1.00 g, 3.52 mmol) in DMF (2 mL) was then added dropwise over 5 min, and the resulting yellow solution was stirred at room temperature overnight. Water (1 mL) was added, and the mixture was diluted in ethyl ether (50 mL), washed with a mixture of brine (9 mL) and hydrochloric acid (1 *N*, 1 mL), dried (Na_2SO_4), concentrated under reduced pressure, and chromatographed (ethyl acetate/hexanes, 1:3 to 1:1) to give 5 (0.98 g, 82%) as yellow oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ 4.631 (4 H, t, J = 7.2 Hz), 4.36-4.26 (4 H, m), 2.70 (2 H, appar. t, J = 8.7 Hz), 2.42 (3 H, s), 2.10-1.96 (2 H, m), 1.80-1.65 (4 H, m), 1.68-1.60 (9 H, m), 1.56-1.42 (2 H, m) (Figure 28).

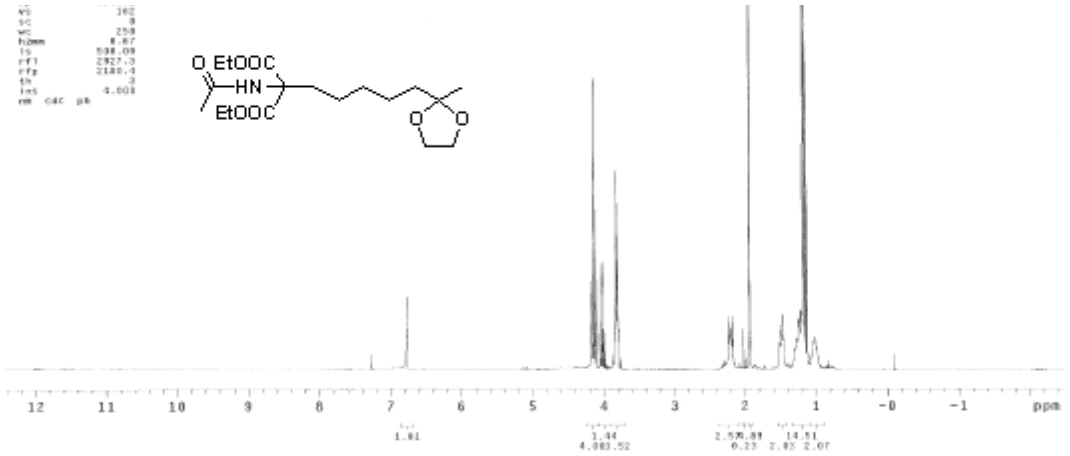


Figure 28. ^{1}H NMR spectrum for compound **5**.

2.2.1.5 2-Amino-8-oxononanoic acid hydrochloride

A suspension of **5** (3.63 g) in concentrated hydrochloric acid (12 *N*, 20 mL) was heated at refluxed overnight. The solvent was evaporated under a high vacuum generated by an oil pump, and the residue was dissolved in minimal amount of water and loaded onto an ion-exchange column made from Dowex 50WX4-400 cation-exchange resin (wet volume ~10 mL) which had been prewashed with hydrochloric acid (6 *N*, 30 mL) and then copious amount of water until neutral. The column was desalted with excessive water washing (200 mL) and then eluted with hydrochloric acid (0.5 *N* to 3 *N*) to give 6 (2.17 g, quant.) as white solid. ^{1}H NMR (300 MHz, D_2O) δ 3.87 (1 H, t, J = 6.6 Hz), 2.39 (2 H, t, J = 7.2 Hz), 2.02 (3 H, s), 1.81-1.65 (2 H, m), 1.38 (2 H, appar. quintet, J=7.5), 1.28-1.12 (4 H, m).

2.2.2 Construction of plasmids

A plasmid pAcKRS-pylT-GFP1Amber bearing the genes encoding AcKRS, pylT and GFP_{UV} with an amber mutation at position 149 and a His tag at the C-terminus was constructed from the pETduet-1 vector (Stratagene Inc.). In this plasmid, AcKRS and GFP^{UV} are both under the control of IPTG-inducible T7 promoter and pylT is flanked by the *lpp* promoter and the *rrnC* terminator (Figure 29).

2.2.2.1 DNA and protein sequences

pylT:

5’-ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca-3’

***lpp* promoter:**

5’-cccatcaaaaaaatattctcaacataaaaaactttgtgtaatacttgtaacgct-3’

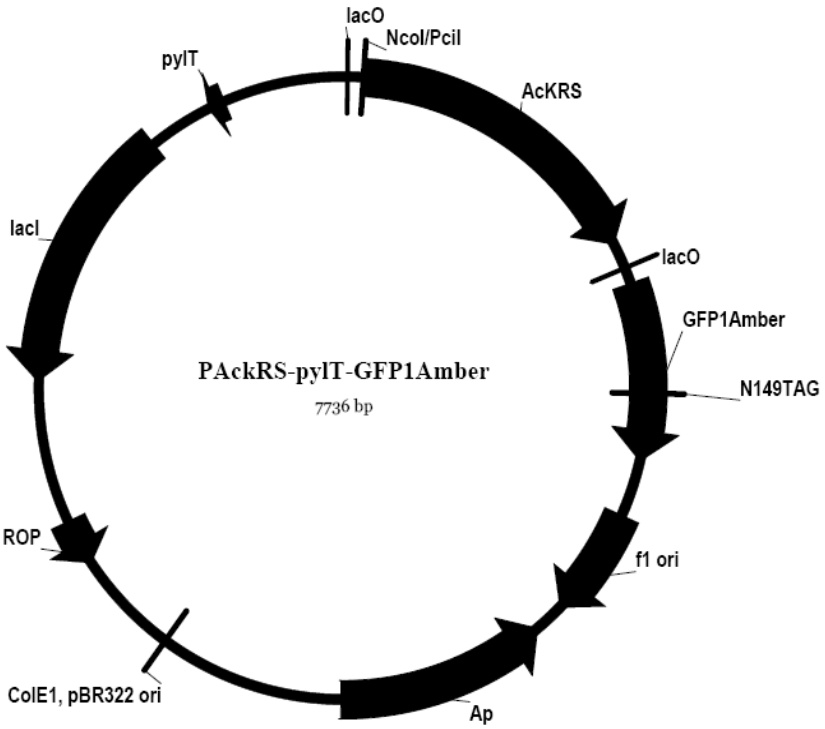


Figure 29. Plasmid map for pAcKRS-pylT-GFP1Amber.

***rrnC* terminator:**

5’- atccttagcgaaagctaaggatttttttta-3’

AcKRS:

5’-atgtcagataaaaaaccattagatgttttaatatctgcgaccgggctctggatgtccaggactggcacgctccacaaaatcaa
gcaccatgaggtctcaagaagtaaaatatacattgaaatggcgtgtggagaccatcttgttgtgaataattccaggagttgtagaa
cagccagagcattcagacatcataagtacagaaaaacctgcaaacgatgtagggtttcgggcgaggatatcaataattttctcac
aagatcaaccgaaagcaaaaacagtgtgaaagttagggtagtttctgctccaaaggtcaaaaaagctatgccgaaatcagtttc
aagggctccgaagcctctggaaaattctgtttctgcaaaggcatcaacgaacacatccagatctgtaccttcgcctgcaaaatca
actccaaattcgtctgttcccgcatcggctcctgctccttcacttacaagaagccagcttgatagggttgaggctctcttaagtcca
gaggataaaatttctctgaatatggcaaagcctttcagggaacttgagcctgaacttgtgacaagaagaaaaaacgattttcagc
ggctctataccaatgatagagaagactacctcggtaaactcgaacgtgatattacgaaatttttcgtagaccggggttttctggag
ataaagtctcctatccttattccggcggaatacgtggagagaatgggtattaataatgatactgaactttcaaaacagatcttccgg
gtggataaaaatctctgcttgaggccaatggttgccccgactattttcaactatgcgcgaaaactcgataggattttaccaggccc
aataaaaatttttcgaagtcggaccttgttaccggaaagagtctgacggcaaagagcacctggaagaatttactatggtgaacttct
ttcagatgggttcgggatgtactcgggaaaatcttgaagctctcatcaaagagtttctggactatctggaaatcgacttcgaaatcg
taggagattcctgtatggtctatggggatactcttgatataatgcacggggacctggagctttcttcggcagtcgtcgggccagttt
ctcttgatagagaatggggtattgacaaaccatggataggtgcaggttttggtcttgaacgcttgctcaaggttatgcacggcttta
aaaacattaagagggcatcaaggtccgaatcttactataatgggatttcaaccaatctgtaa-3’

GFP1Amber (amber mutation at position 149):

5’-atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaatttttctgt
cagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggc
caacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgcca

tgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaag

gtgataccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaact

ataactcacactaggtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaag

atggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgt

cgacacaatctgcccttttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattaca

catggcatggatgaactctacaaagagctccatcaccatcaccatcactaa-3’

AcKRS:

MSDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRS

CRTARAFRHHKYRKTCKRCRVSGEDINNFLTRSTESKNSVKVRVVSAPKVKKA

MPKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNSSVPASAPAPSLTRSQLDR

VEALLSPEDKISLNMAKPFRELEPELVTRRKNDFQRLYTNDREDYLGKLERDITK

FFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRPMVAPTIFNY

ARKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFFQMGSGCTRENLEALI

KEFLDYLEIDFEIVGDSCMVYGDTLDIMHGDLELSSAVVGPVSLDREWGIDKPW

IGAGFGLERLLKVMHGFKNIKRASRSESYYNGISTNL

GFP1Amber:

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP

VPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYK

TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHK*VYITADKQKNGI

KANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKR

DHMVLLEFVTAAGITHGMDELYKELHHHHHH

2.2.2.2 Construction of pBK-AcKRS-pylT

*Mb*PylRS was amplified from *M. barberi* genomic DNA purchased from ATCC by polymerase chain reaction (PCR) using two primers, 5'-GAGGAATCCCATATGGATAAAAAACCATTAG-3' and 5'-CGTTTGAAACTGCAGTTACAGATTGGTTG-3'. AcKRS (L266V, L270I, Y271F, L274A, C313F and D76G)[100] was subsequently synthesized by overlap extension PCR using *Mb*PylRS as the template and eight oligodeoxynucleotide primers (5'-GAGGAATCCCATATGGATAAAAAACCATTAG-3', 5'-AAATTATTGATATCCTCGCCCGAAACCCTACATCGTTTGC-3', 5'-GATGTAGGGTTTCGGGCGAGGATATCAATAATTTTC-3', 5'-GTTGAAAATAGTCGGGGCAACCATTGGCCTCAAGCAG-3', 5'-GCCCCGACTATTTTCAACTATGCGCGAAAACTCGATAGG-3', 5'-CCGAACCCATCTGAAAGAAGTTCACCATAG-3', 5'-CTATGGTGAACTTCTTTCAGATGGGTTCGG-3', and 5'-CGTTTGAAACTGCAGTTACAGATTGGTTG-3'). Two restriction sites *Nde1* at 5' head and *Pst1* at 3' tail were designed into the synthesized AcKRS, which was subsequently digested by *NdeI* and *PstI* restriction enzymes and cloned into the same two sites in a pBK plasmid[76] to afford pBK-AcKRS. In pBK-AcKRS, AcKRS is under the control of a constitutive glnS promoter. The gene of pylT flanked by the *lpp* promoter at 5' end and the *rrnC* terminator at 3' end was constructed by using overlap extension PCR of six oligodeoxynucleotides (5'-CCCGGGATCCCCCATCAAAAAAATATTCTCAACAT-3', 5'-

TTACAAGTATTACACAAAGTTTTTTATGTTGAGAATATTTTTTTG-3', 5'-ACTTTGTGTAATACTTGTAACGCTGAATCCGGAAACCTGATCATGTAGAT-3', 5'-CTAACCCGGCTGAACGGATTTAGAGTCCATTCGATCTACATGATCAGGTTT-3', 5'-TCAGCCGGGTTAGATTCCCGGGGTTTCCGCCACTGCCCATCCTTAGCGAA-3', and 5'-GAACCCAGATCTTAAAAAAAATCCTTAGCTTTCGCTAAGGATG-3'). Two restriction sites, *BamH1* at 5' end and *Bgl11* at 3' end, were introduced in the synthesized DNA, which was subsequently digested by these two enzymes and cloned into the *BamHI* site in pBK-AcKRS to afford pBK-AcKRS-pylT.

2.2.2.3 Construction of pAcKRS-pylT-GFP1Amber

Plasmid pAcKRS-pylT-GFP1Amber encodes the AcKRS/pylT pair and GFP1Amber with one amber mutation at position 149. Both of AcKRS and GFP1Amber were under the control of the T7 promoter. AcKRS was amplified from pBK-AcKRS-pylT by two oligodeoxynucleotides (5'-GATATAACATGTCAGATAAAAAACCATTAGATG-3', and 5'-GTCGACCTGCAGTTACAGATTGGTTGAAATCCC-3'). The amplified DNA was digested by *Pci1* end and *Pst1* restriction enzymes and cloned into *Nco1* and *Pst1* sites of the vector of pETduet-1, which was purchased from Stratagene Inc. to afford pAcKRS. GFP1Amber was amplified from pleiG-N149[89] (a gift from Dr. Peter G. Schultz) by two oligodeoxynucleotides (5'-GAAGGAGATATACATATGAGTAAAGGAGAAG-3', and 5'-GACTCGAGGGTACCTTAGTGATGGTGATGGTGATG-3'), digested by *NdeI* and *KpnI*, and then cloned into *NdeI* and *KpnI* restriction sites of pAcKRS to afford

pAcKRS-GFP1Amber. The pylT gene with the *lpp* promoter and the *rrnC* terminator was amplified from pBK-AcKRS-pylT by two oligodeoxynucleotides (5'-GCTAGATCTGGAAACCTGATGTAGATC-3', and 5'-GATACTAGTTGGCGGAAACCCCGGG-3'), digested by *SphI,* and then cloned into the *SphI* site of pAcKRS-GFP1Amber to afford pAcKRS-pylT-GFP1Amber.

2.2.3 Protein expression procedure

To express GFP_{UV}, *E. coli* Bl21 cells were transformed with pAcKRS-pylT-GFP1Amber grown in LB medium that contained 100 μg/mL ampicillin and 25 μg/mL kanomycin and induced with the addition of 500 μM IPTG when OD_{600} was 0.6. After induction, 2 mM KetoK was subsequently added. Cells were then let grown overnight or 10 h at 37 degree. Cells were harvested by centrifugation (4500 r.p.m., 20 min, 4 degree) and resuspended in 20 mL of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM DTT, 10% glycerol, 0.1% Triton X-100, 5 mM imidazole, and 1 μg/mL lysozyme (pH 7.4)). The resuspended cells were sonicated and the lysate was clarified by centrifugation (10200 r.p.m., 60 min, 4 degree). The supernatant was decanted and loaded to Ni-NTA superspeed agarose (Qiagen Inc.) column on FPLC. The column was washed by 5 × bed volume of buffer A that contained 50 mM HEPES, 300 mM NaCl, 5 mM imidazole (pH 7.5) and then eluted by running a gradient that changed from buffer A to buffer B in 10 × bed volume. Buffer B contained 50 mM HEPES, 300 mM NaCl, 250 mM imidazole (pH 7.5). Proteins were concentrated by Amicon (Millipore, NMWL10 KDa) and analyzed by 12% SDS-PAGE (Figure 30).

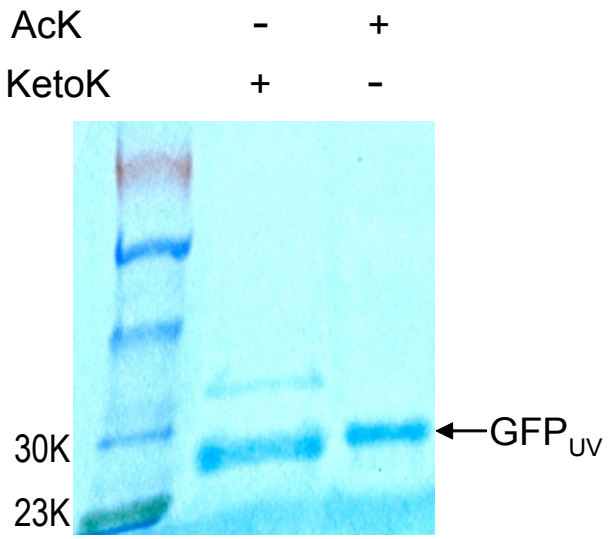


Figure 30. SDS-PAGE analysis of expression of GFP_{UV} with amber mutation.

2.2.4 Protein Characterization

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 flow rate of 10 µL/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 for B was as follows: starting at 2% 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 +4900 V. BioAnalyst software (Applied Biosystems) was used for spectral deconvolution. 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 a S/N threshold of 20.

GFP_{UV} was dissolved in 25 mM of ammonia bicarbonate and denatured at 90 degree for 15 min. A prepared trypsin solution in 0.01% TFA (pH 3) was added to the denatured protein solution (w/w=1:50) and incubated at 37 degree overnight.

Peptides resulting from tryptic digests were mixed (1:1 v/v) with the matrix (5 mg/mL α-cyano-4-hydroxycinnamic acid, 50% (v/v) acetonitrile, 10 mM ammonium dihydrogen phosphate, and 1% trifluoroacetic acid) and 1 µL of the resulting mixture was spotted onto a stainless steel target plate. Mass spectra and tandem MS spectra were collected using an Applied Biosystems 4800 Tof/Tof (Framingham, MA). Collision

induced dissociation tandem MS spectra were acquired using air at the medium pressure setting and at 2 kV of collision energy. Tandem MS data was manually interpreted using the Data Explorer™ software package (Applied Biosystems, Framingham, MA).

The transformation of pAcKRS-pylT-GFP1Abmer into BL21 cells and subsequent growth in LB medium supplemented with 2 mM KetoK and 500 μM IPTG afforded full-length GFP_{UV} incorporated with KetoK (GFP-KetoK) with a yield of 0.5 mg/L, which was comparable to the expression yield of full-length GFP_{UV} incorporated with AcK (GFP-AcK) (0.8 mg/L) when the supplemented amino acid was changed to 5 mM AcK. The deconvoluted electrospray ionization mass spectrometry (ESI-MS) spectrum of GFP-KetoK revealed two intense peaks corresponding to the full-length GFP-KetoK with and without N-terminal methionine, respectively (detected: 27,898.9 Da, 27,767.8 Da; calculated: 27,896 Da, 27,765 Da) (Figure 31A). The site-specific incorporation of KetoK at position 149 was further validated by the tandem mass spectral analysis of the tryptic KetoK containing fragment of LEYNYNSHK*VYITADK (K* denotes KetoK). The K*-containing ions (y8 to y12 and b9 to b15) all had the expected mass (Figure 31B).

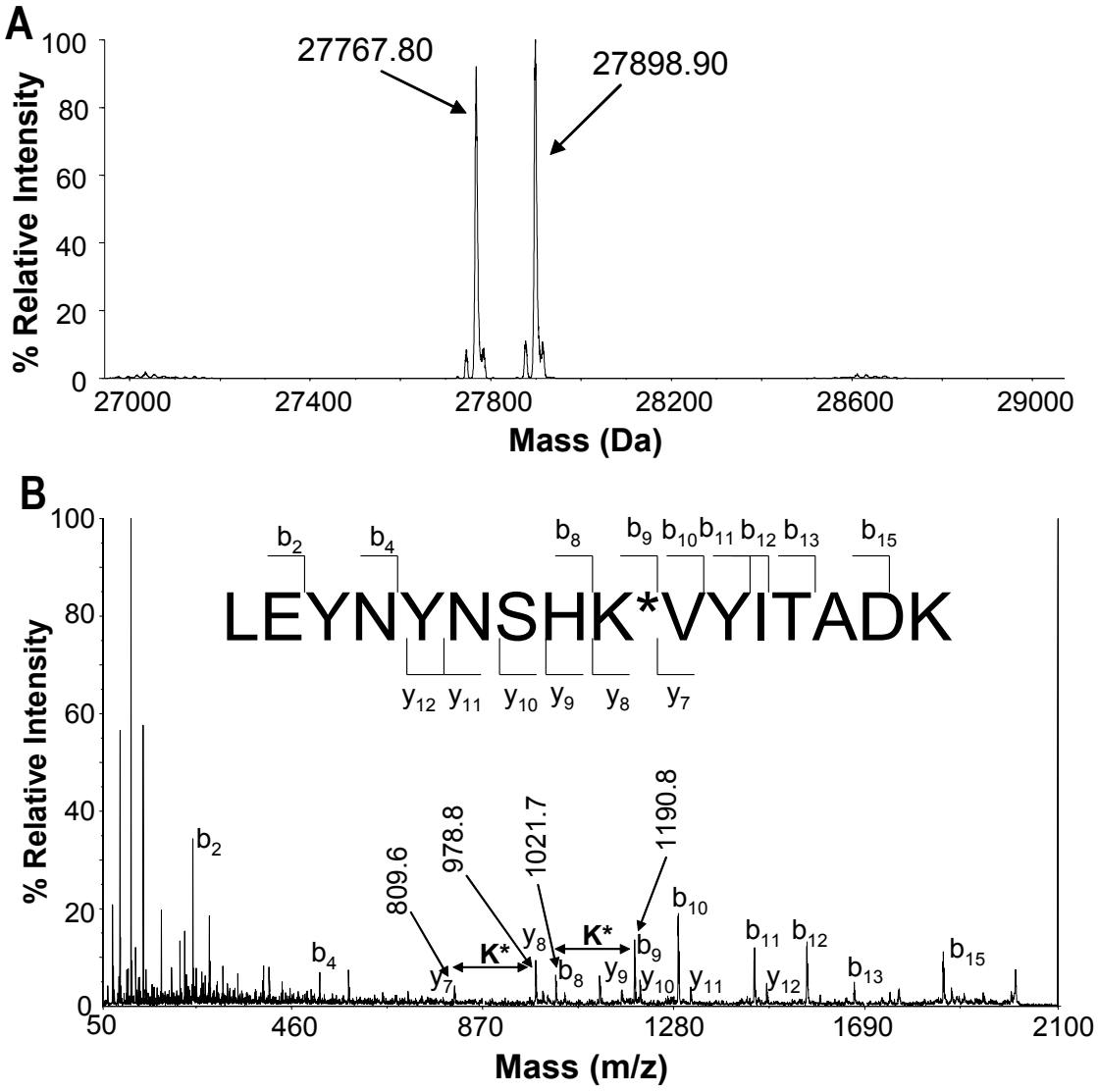


Figure 31. Mass spectra. (**A**) ESI-MS spectrum of GFP-KetoK, and (**B**) tandem MS spectrum of LEYNYNSHK*VYITADK.

2.2.5 Protein labeling

2.2.5.1 Protein labeled with a fluorescent dye

To test the efficiency of using the genetically incorporated KetoK to specifically label proteins with fluorescent dyes under a mild condition, the purified GFP-KetoK and GFP-AcK were both reacted with 1 equiv. Texas Red hydrazide in PBS buffer at 37 degree, pH 6.3 overnight and then analyzed on a SDS-PAGE gel. The gel was visualized by silver staining and fluorescent imaging (Figure 32). Whereas both GFP-KetoK and GFP-AcK showed a strong protein band after silver staining, only labeled GFP-KetoK had an intense fluorescent band. This indicates that labeling only occurred on GFP-KetoK and the reaction between the keto group and a hydrazide is specific.

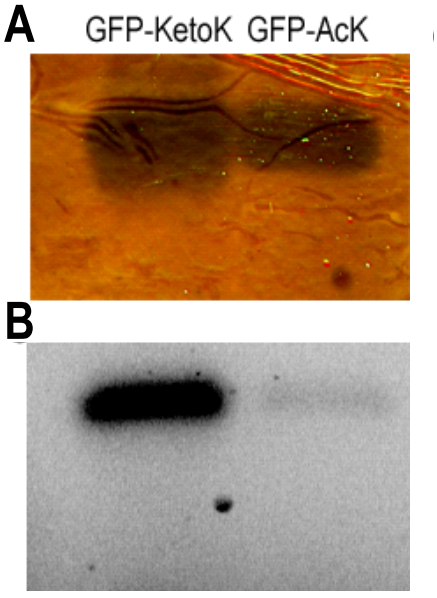


Figure 32. (**A**) Silver staining. (**B**) Fluorescent imaging of GFP-KetoK and GFP-AcK after their reaction with Texas Red hydrazine dye.

2.2.5.2 Protein labeled with biotin derivative

To further characterize the labeling efficiency, the purified GFP-KetoK was reacted with 5 equiv. of biotin alkoxyamine in PBS buffer at 37 degree, pH 6.3 overnight and the modified protein was analyzed by ESI-MS (Figure 33B). Two intense peaks (28,213.6 Da, 28,081.0 Da) were shown in the deconvoluted ESI-MS spectrum and they matched the biotin labeled full-length GFP-KetoK with and without N-terminal methionine (calculated mass: 28,209 Da, 28,078 Da), respectively. Given the fact that no original full-length GFP-KetoK was detected, the labeling reaction was highly efficient. As a control, the same labeling reaction was carried out on wild type GFP_{UV}. The deconvoluted ESI-MS spectra of wild type GFP_{UV} before and after the labeling reaction showed no difference (Figure 34), demonstrating the reaction specificity between KetoK and alkoxyamine probes.

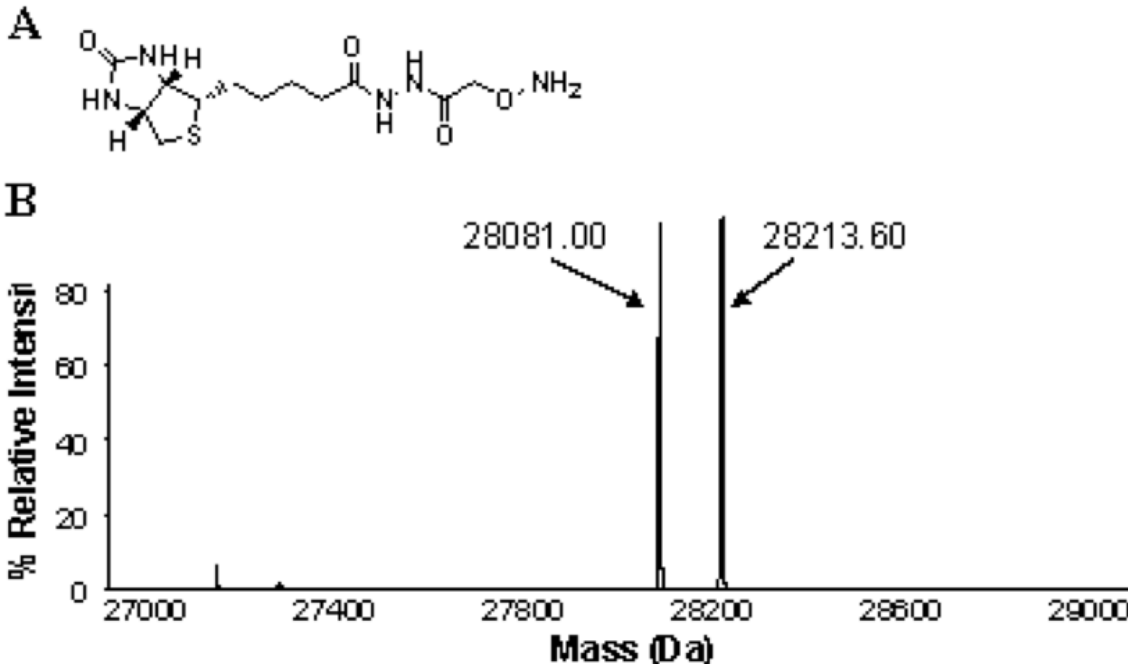


Figure 33. (**A**) Structure of biotin alkoxyamine. (**B**) ESI-MS spectrum of GFP-KetoK after its reaction with biotin alkoxyamine.

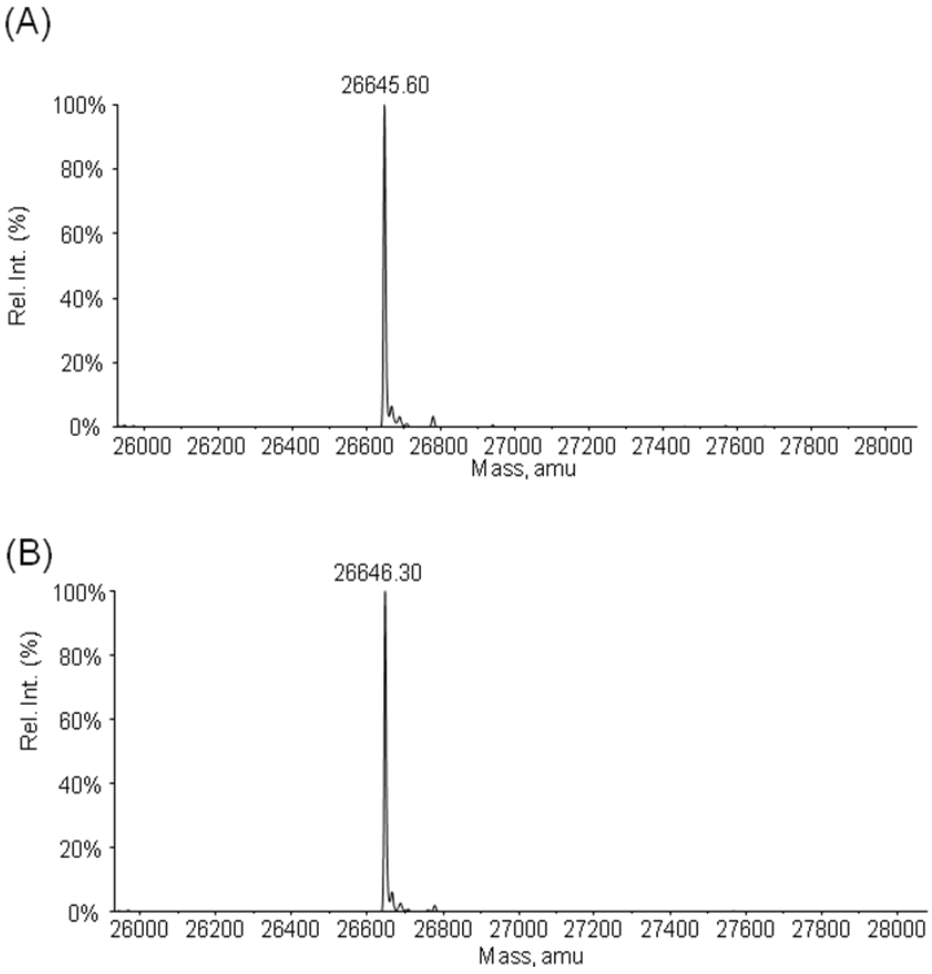


Figure 34. ESI-TOF-MS spectra of Wild type $GFP_{UV.}$ Before (**A**) and after (**B**) labeling reaction with biotin hydroxylamine. No reaction was detected by MS analysis.

2.3 Summary

In summary, we have achieved the genetic incorporation of KetoK into proteins in *E. coli* and applied it to label proteins with different probes. The labeling features high specificity and high efficiency under a mild reaction condition. One potential application of our method is to site-specifically introduce various biochemical and biophysical probes into proteins, including biotin tag, fluorescent labels, NMR probes, EPR probes, etc. Moreover, KetoK is an unhydrolysable analogue of AcK that represents one of the most important posttranslationally modified amino acids in eukaryotic cells. Since protein acetylation is a reversible process catalyzed by enzymes like histone acetyltransferases and histone deacetyltransferases, direct incorporation of KetoK into proteins *in vivo* at sites with naturally occurring lysine acetylation will permanently install an AcK mimic and may provide an efficient way to decipher the regulation roles of acetylation in histones, p53 and other transcription regulatory proteins.

3. GENETIC INCORPORATION OF MULTIPLE NAAs INTO ONE PROTEIN IN *ESCHERICHIA COLI**

3.1 Introduction

With few exceptions, the genetic codes of all known organisms specify the same 20 canonical amino acid building blocks.[102] However, it is clear that many proteins require additional chemical groups beyond what the 20 building blocks can provide to carry out their native functions.[34] Many of these groups are installed into proteins through covalent posttranslational modifications, such as acetylation, methylation, phosphorylation, sulfation, etc. These modifications, to a great extent, expand the amino acid inventory of proteins by assimilating NAAs and grant proteins expanded opportunities for catalysis, mediating signal transduction, integration of information at many metabolic intersections, and alteration of cellular locations. The diversity of these covalent modifications greatly exceeds the number of proteins predicted by DNA coding capacities, leading to proteomes in higher organisms 2-3 orders of magnitude more complex than the encoding genomes and therefore adding a great complexity to the functional investigation of a lot of proteins.[34]

A major challenge in studying posttranslationally modified proteins is that they exist typically as a mixture of different forms, making it difficult to purify uniquely

---

modified forms from their natural sources. For this reason, a variety of approaches have been developed to synthesize proteins with defined posttranslational modifications that can be easily isolated. One of these approaches uses the nucleophilicity of cysteine to install *N*-ε-methylated lysine analogs into proteins. Native chemical ligation and its extension, expression protein ligation, have also been used to install NAAs into proteins.[74] These two approaches provide great opportunities to study functional roles of posttranslational modifications. However, they both are *in vitro* semisynthetic techniques. In addition, NAAs installed by cysteine replacement are not exactly the same as what exist in nature; incorporating NAAs into internal sites of proteins using native chemical ligation is also problematic. An alternative NAA installation approach that can be carried out *in vivo* is the genetic incorporation of NAAs directly into proteins during translation in live cells. This approach relies on the read-through of an in-frame amber (UAG) stop codon in mRNA by an amber suppressor tRNA specifically acylated with a NAA by an evolved aaRSs.[103] One essential advantage of this approach is that a NAA can be installed at any site of a protein regardless of the protein size. Using this approach, two NAAs, *O*-sulfo-L-tyrosine and AcK which occur naturally in posttranslationally modified proteins have been incorporated into proteins in *E. coli*.[100, 104] Although the genetic NAA incorporation approach has opened a new avenue for functional studies of posttranslationally modified proteins both *in vitro* and *in vivo*, it has its intrinsic limitation. The method relies on the suppression of an amber stop codon and has to compete with the release factor-1 (RF1)-mediated translation termination which terminates protein translation at the amber codon (Figure 35).[105] In general, the

suppression efficiency for a single amber codon is limited to 10-20%.[106] As the number of amber codons in a gene increases, the expression yield of the full-length protein decreases multiplicatively. This creates a huge difficulty in the expression of proteins

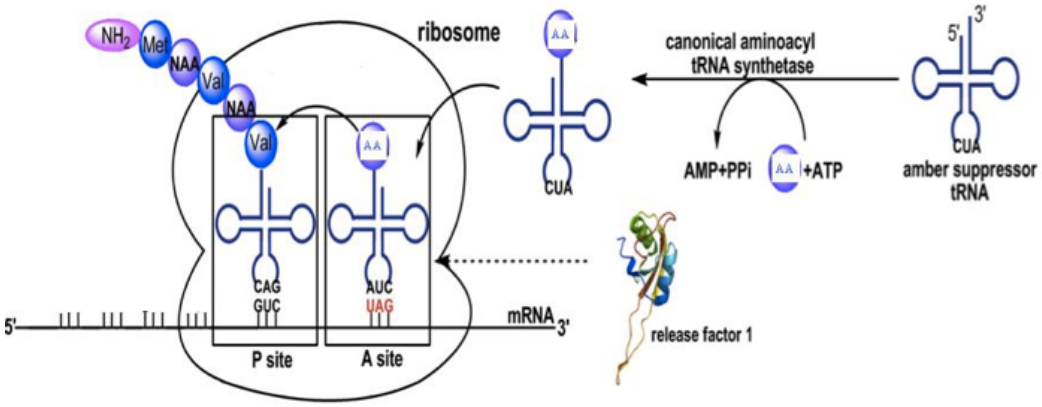


Figure 35. Scheme for RF-1 termination.

with multiple posttranslational modifications, which are abundant in nature. For example, seven acetylation sites have been discovered in histone 3.[107]

Different acetylation patterns at these sites regulate cellular processes ranging from transcription activation to DNA repair. A modified genetic NAA incorporation approach that allows synthesis of proteins with multiple modifications is critical for their functional investigation and is in dire need. Chin et al. recently developed an elegant system that used an evolved orthogonal ribosome (ribo-X) to achieve the incorporation of two NAAs into one protein.[106] As demonstrated here, we have developed a simpler method for the incorporation of multiple NAAs into one protein in *E. coli* simply by overexpressing the C-terminal domain of the large ribosomal subunit protein L11 and applied it to synthesize green fluorescent protein (GFP_{UV}) with three AcKs.

The underlying reason of a low NAA incorporation efficiency is due to RF-1 mediated release of premature peptide from ribosome at a designated amber codon. A decrease in the RF1-mediated translation termination rate will certainly enhance the amber suppression efficiency and make it possible to incorporate multiple NAAs into one protein. To achieve this goal, we decided to modify the large ribosomal subunit protein L11 in *E. coli*. L11 is a highly conserved small ribosomal protein that associates 23S rRNA and contains two domains, N- and C-terminal domains linked together by a small peptide hinge (Figure 36).[108] It has been suggested from biochemical studies that L11 plays an important role in the RF1-mediated peptide release.[109, 110] Knockout of the N-terminal domain of L11 (L11N) from *E. coli* led to a slightly slower cell growth but with a much higher amber suppression rate. However, *E. coli* lacking the entire L11

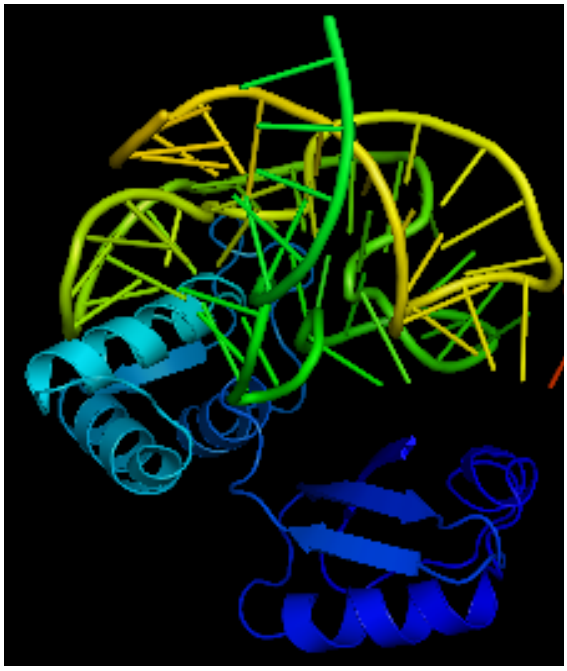


Figure 36. Crystal structure of the L11-RNA complex. PBD: 1MMS.

protein was thermosensitive and showed a drastically decreased cell growth rate.[109] These indicate that the C-terminal domain of L11 (L11C) alone efficiently binds 23S rRNA and the resulted ribosome maintains almost regular protein translation efficiency but with a high amber suppression rate. The *in vitro* analysis also proved 4-6 folds reduction in the RF1-mediated termination efficiency when ribosome did not contain L11N.[110] Based on these observations, we hypothesized that overexpression of L11C inside *E. coli* should efficiently replace L11 in ribosome, consequently decrease the RF1-meidated translation termination at amber codon, and increase amber suppression efficiency. This enhanced amber suppression level may allow the incorporation of multiple NAAs into one protein (Figure 37).

We chose to demonstrate our hypothesis by the incorporation of multiple AcKs into one protein in *E. coli*. Mutant AcKRSs that are specific for AcK have been evolved from *Methanosarcina barkeri* PylRS. Together with *M. barkeri* pylT, the cognate amber suppressor tRNA of PylRS, these mutant variants have been used to incorporate AcK into proteins in *E. coli*.[100]

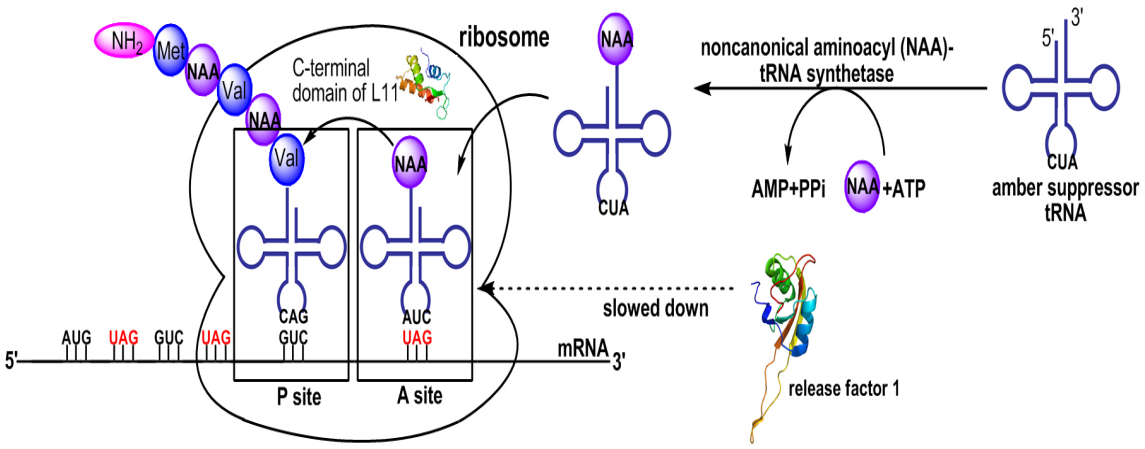


Figure 37. Scheme for incorporation of multiple NAAs into one protein.

3.2 Experiments and results

3.2.1 Construction of plasmids

I chose to demonstrate our hypothesis by the incorporation of multiple AcKs into one protein in *E. coli*. Thus, pAcKRS-pylT-GFP1Amber was used. The gene for overexpression L11C had been cloned into pET30a to generate pET-L11C. One and two additional amber codons had been introduced into pAcKRS plasmid separately to afford pAcKRS-pylT-GFP2Amber and pAcKRS-pylT-GFP3Amber. These three pAcKRS plasmids together with pET-L11C had been used to express GFP_{UV} with one and two and three UAAs (Figure 38).

The construction of pAcKRS-pylT-GFP1Amber, pET-L11C, pAcKRS-pylT-GFP2Amber, and pAcKRS-pylT-GFP3Amber all followed standard cloning and QuikChange site-directed mutagenesis procedures using Platinum Pfx (Invitrogen) and PfuTurbo (Stratagene) DNA polymerases. All the plasmid structures have been confirmed by sequencing. All oligonucleotide primers were purchased from Integrated DNA Technologies, Inc.

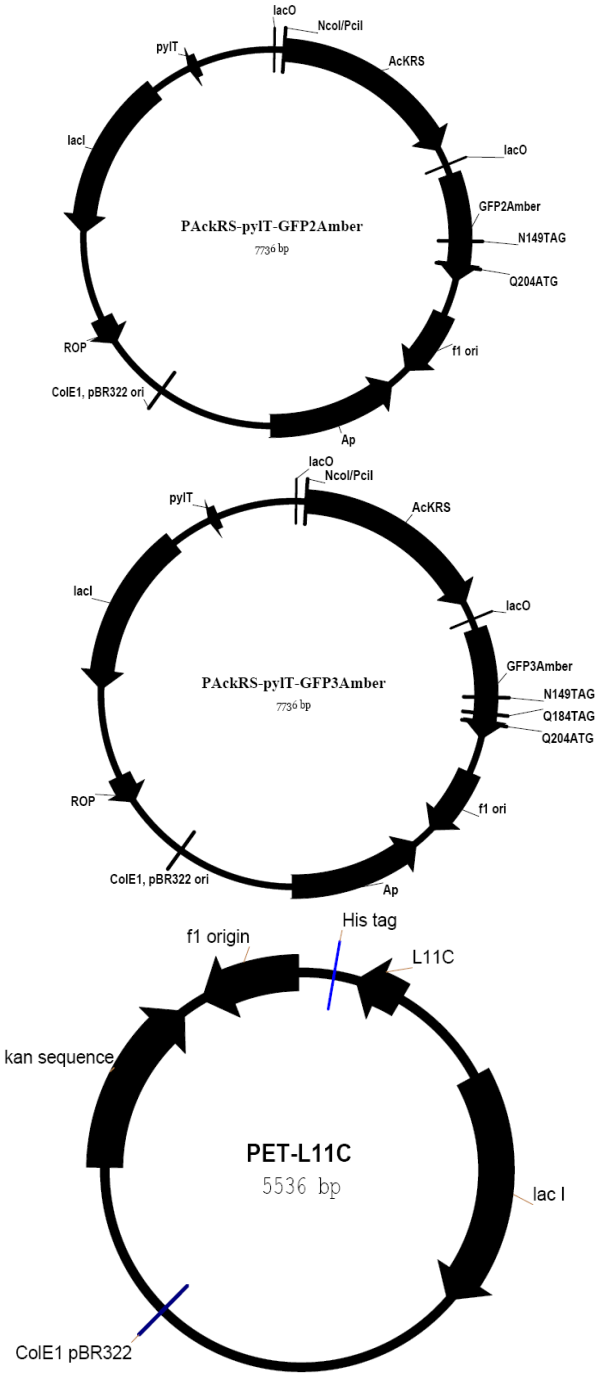


Figure 38. Plasmid maps.

3.2.1.1 DNA and protein sequences

GFP2Amber:

5'-atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctg
tcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatgg
ccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgcc
atgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaa
ggtgataccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaac
tataactcacactaggtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaa
gatggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacct
gtcgacatagtctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggatta
cacatggcatggatgaactctacaaagagctccatcaccatcaccatcactaa-3'

GFP3Amber:

5'-atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctg
tcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatgg
ccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgcc
atgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaa
ggtgataccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagtacaac
tataactcacactaggtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaa
gatggatccgttcaactagcagaccattatcaatagaatactccaattggcgatggccctgtccttttaccagacaaccattacctg
tcgacatagtctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattaca
catggcatggatgaactctacaaagagctccatcaccatcaccatcactaa-3'

GFP2Amber':

5’-atggcatagagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaa

ttttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttc

catggccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaaga

gtgccatgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagt

ttgaaggtgataccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagt

acaactataactcacactaggtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaaca

ttgaagatggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccatt

acctgtcgacatagtctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctggg

attacacatggcatggatgaactctacaaagagctccatcaccatcaccatcactaa-3’

L11C:

5’-atgaccaagaccccgccggcagcagttctgctgaaaaaagcggctggtatcaagtctggttccggtaagccgaacaaag

acaaagtgggtaaaatttcccgcgctcagctgcaggaaatcgcgcagaccaaagctgccgacatgactggtgccgacattga

agcgatgactcgctccatcgaaggtactgcacgttccatgggcctggtagtggaggactaa-3’

GFP2Amber:

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP

VPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYK

TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHK*VYITADKQKNGI

KANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTK*SALSKDPNEK

RDHMVLLEFVTAAGITHGMDELYKELHHHHHH

GFP3Amber:

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP
VPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYK
TRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHK*VYITADKQKNGI
KANFKIRHNIEDGSVQLADHYQK*NTPIGDGPVLLPDNHYLSTK*SALSKDPNEK
RDHMVLLEFVTAAGITHGMDELYKELHHHHHH

GFP2Amber':

MAK*SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTG
KLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHK*VYITADKQ
KNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDP
NEKRDHMVLLEFVTAAGITHGMDELYKELHHHHHH

L11C:

MTKTPPAAVLLKKAAGIKSGSGKPNKDKVGKISRAQLQEIAQTKAADMTGADI
EAMTRSIEGTARSMGLVVED

3.2.1.2 Construction of pAcKRS-pylT-GFP2Amber

Plasmid pAcKRS-pylT-GFP2Amber was derived from pAcKRS-pylT-GFP1Amber with an addition amber mutation at position 204. The mutagenesis was carried out by the standard QuikChange site-directed mutagenesis procedure using PfuTurbo DNA polymerase and two oligonucleotides, 5'-CTTTCGAAAGGGCAGACTATGTCGACAGGTAATG-3' and 5'-CATTACCTGTCGACATAGTCTGCCCTTTCGAAAG-3'.

3.2.1.3 Construction of pAcKRS-pylT-GFP3Amber

Plasmid pAcKRS-pylT-GFP3Amber was derived from pAcKRS-pylT-GFP2Amber with an additional amber mutation at position 184. The mutagenesis was carried out by the standard QuikChange site-directed mutagenesis procedure using PfuTurbo DNA polymerase and two oligonucleotides, 5'-ATCGCCAATTGGAGTATTCTATTGATAATGGTCTGC-3' and 5'-GCAGACCATTATCAATAGAATACTCCAATTGGCGAT-3'.

3.2.1.4 Construction of pAcKRS-pylT-GFP2Amber'

Plasmid pAcKRS-pylT-GFP2Amber' was derived from pAcKRS-pylT-GFP1Amber with an addition amber mutation at position 2. The mutagenesis was carried out by the standard QuikChange site-directed mutagenesis procedure using PfuTurbo DNA polymerase and two oligonucleotides, 5'-GATATACATATGGCATAGAGTAAAGGAGAAGAA-3' and 5'-CATTACCTGTCGACATAGTCTGCCCTTTCGAAAG-3'.

3.2.1.5 Construction of pET-L11C

Plasmid pET-L11C contains the gene of L11C whose transcription is under the control of T7 promoter. L11C gene was amplified from the genomic DNA of *E. coli* by standard PCR using two primers, 5'-GGAGATATACATATGACCAAGACCCCGCCGGCA-3' and 5'-GTCGTCGGTACCTTAGTCCTCCACTACCAG-3'. The amplified DNA was digested by *NdeI* and *KpnI* restriction enzymes and cloned into *NdeI* and *KpnI* sites in pET30a (Stratagene Inc.) to afford pET-L11C.

3.2.2 Protein expression procedure

To express different kinds of GFP_{UV} variants, *E. coli* Bl21 cells was transformed with pAcKRS-pylT-GFP1Amber, pAcKRS-pylT-GFP2Amber, or pAcKRS-pylT-GFP3Amber together with or without pET-L11C. Cells transformed with one plasmid were grown in LB media that contained 100 μg/mL ampicillin and induced with the addition of 500 μM IPTG when OD_{600} reached 0.6. 5 mM AcK and 5 mM nicotinamide were subsequently added into the media 0.5 h after induction. Nicotinamide was inhibitor for the deacetylation of AcK, which was common inside of *E. coli*. Cells were then let grown for 10 h at 37 degree. The protein expression in cells transformed with two plasmids followed exactly same procedures except the addition of 25 μg/mL kanamycin into the media to force cells to maintain pET30-L11C. The GFP_{UV} expression in cells transformed with either one or two plasmids at the absence of AcK also followed the same procedures. GFP_{UV} were purified as previous descried.

3.2.3 The study for L11C to increase the protein expression level for single incoproation of NAA

The plasmid pAcKRS-pylT-GFP1Amber that bear the genes encoding GFP_{UV} with an amber mutation at position 149 was used to test our hypothesis. The transformation of BL21 cells with this plasmid and subsequent growing cells in LB medium supplemented with 5 mM AcK and 500 μM IPTG afforded full-length GFP_{UV} incorporated with AcK at position 149 (GFP-AcK) with a yield of ~0.8 mg/L. Trace amount of full-length GFP_{UV} was expressed when AcK was absent in LB media (Figure 39).

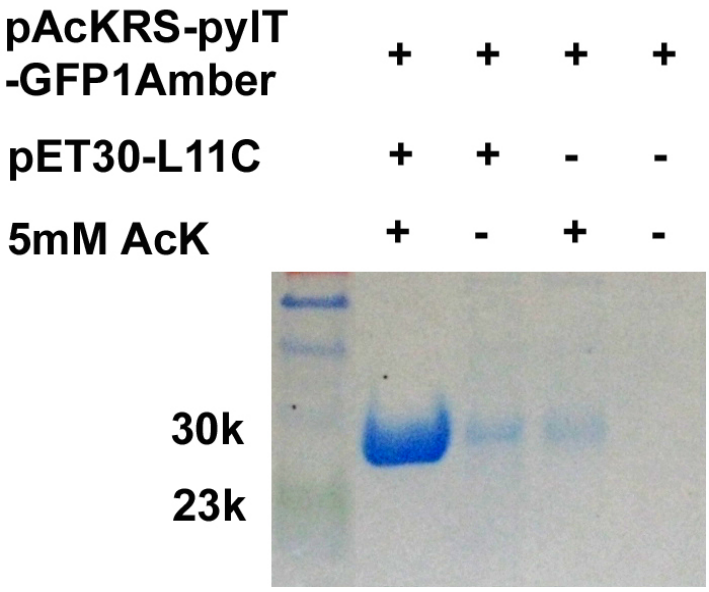


Figure 39. SDS-PAGE analysis of GFP-AcK expressed in the absence and presence of L11C.

We then tested whether overexpression of L11C could improve amber suppression rate. Transformation of BL21 cells with both pAcKRS-pylT-GFP1Amber and pET-L11C and subsequent growth in LB medium with the addition of 5 mM AcK and 500 μM IPTG led to GFP-AcK production with a yield of ~3.5 mg/L which was four times higher than that from cells transformed only with pAcKRS-pylT-GFP1Amber. The electrospray ionization mass spectrometry (ESI-MS) of full-length GFP_{UV} purified from cells transformed with two plasmids revealed two intense peaks corresponding to GFP-AcK with and without the N-terminal methionine, respectively (Table 2 and Figure 40).

The tandem mass spectrum of the tryptic AcK-containing fragment of LEYNYNSHK*VYITADK (K* denotes AcK) validated the incorporation of AcK at position 149. The presence of K* containing ions (y8 to y15 and b9 to b15) all had the expected mass (Figure 41).

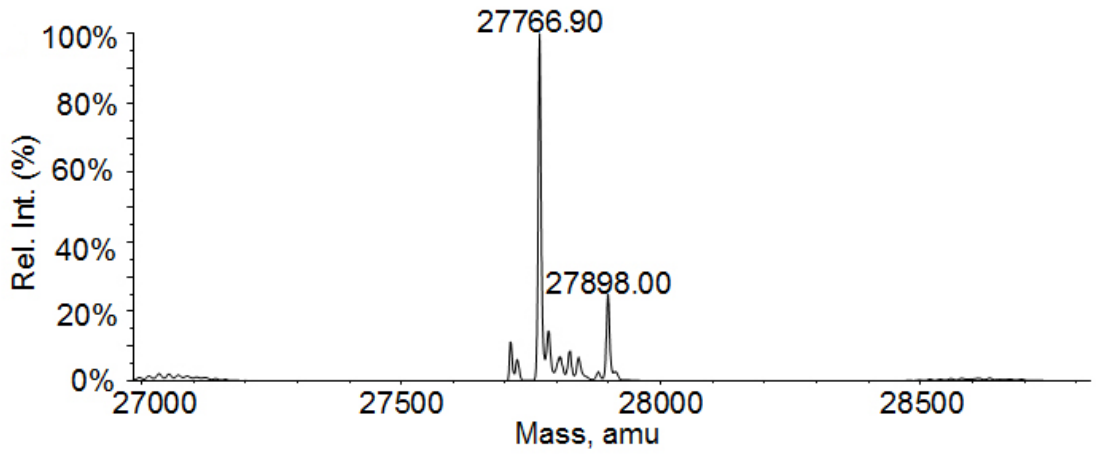


Figure 40. ESI spectrum of GFP-AcK.

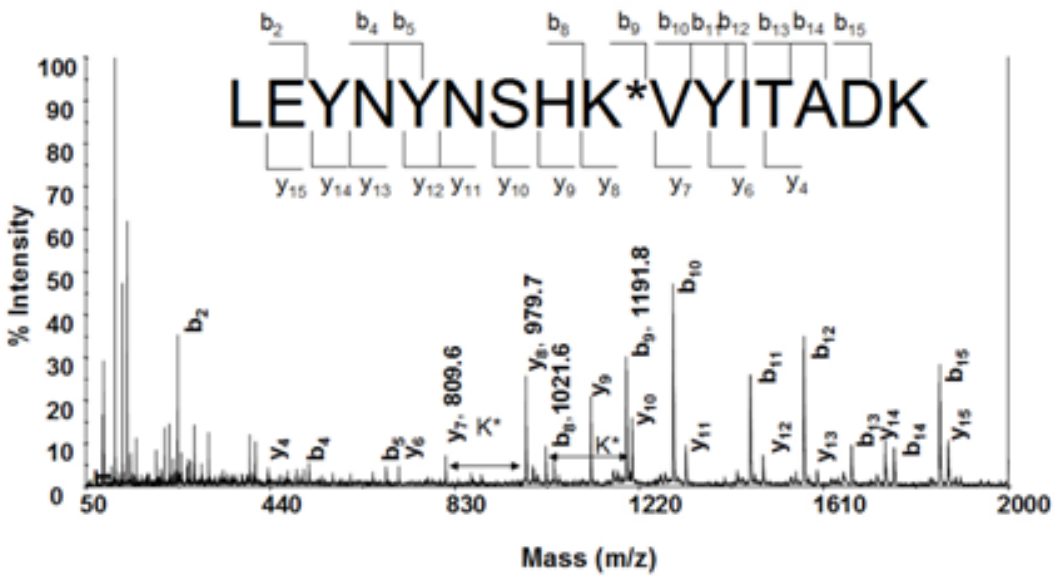


Figure 41. Tandem MS spectrum of LEYNYNSHK*VYITADK from GFP-AcK. Similar spectra were obtained from GFP-2AcK and GFP-3AcK. K* is at position 149.

Table 2. GFP_{UV} Expression Yields and MS Characterization

Protein[a]	L11C coexpression	Concentration (mg/L)	Calculated Mass (Da)	Actual Mass (Da)
GFP	-	56	26645[b]	26646
GFP-AcK	+	3.5	27766[b] 27897[c]	27767 27898
	-	0.8		
GFP-2AcK	+	0.2	27808[b] 27939[c]	27809 27940
	-	N.D.		
GFP-3AcK	+	0.1	27850[b] 27981[c]	27850 27981
	-	N.D.		

[a] Proteins were expressed in the presence of 5 mM AcK and 500 μM IPTG. [b] Molecular mass of full-length protein without N-terminal methionine. [c] Molecular mass of full-length protein with N-terminal methionine.

To test whether L11C overexpression improves the basal level of amber suppression, the BL21 cells transformed with both pAcKRS-pylT-GFP1Amber and pET-L11C were grown in LB medium supplemented only with 500 μM IPTG. In the absence of AcK, full-length GFP_{UV} was expressed with a yield of 0.2 mg/L, indicating a significant basal amber suppression improvement. Since the evolved AcKRS still showed low activity toward natural amino acids in LB medium,[100] this basal amber suppression improvement may represent the improved ability of AcKRS to incorporate natural amino acids into proteins. ESI-MS of the purified full-length GFP_{UV} expressed in this condition revealed five peaks (Figure 42), of which one (27766 Da) matched the mass of GFP-AcK without N-terminal methionine but the other four (27710 Da, 27725 Da, 27782 Da, and 27842 Da) were not identified in ESI-MS of GFP-AcK purified from the same cells growing in the presence of AcK.

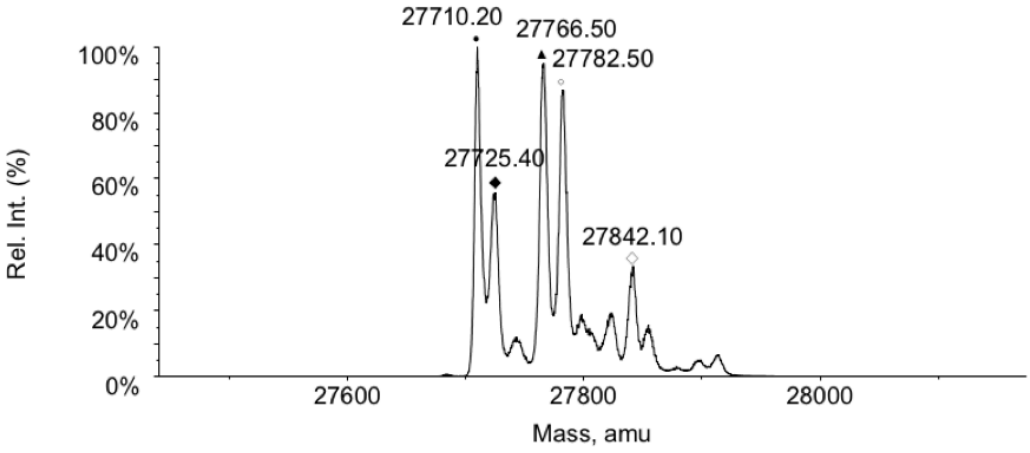


Figure 42. ESI mass spectrum for GFP-AcK misincorporation.

These indicate the presence of AcK efficiently inhibits the incorporation of other amino acids, assuring the high fidelity of AcK incorporation. These experiments demonstrated that the overexpression of L11C indeed enhances amber suppression efficiency and maintains high NAA incorporation fidelity as well.

3.2.4 The study for L11C to increase the protein expression level for multiple incoproation of NAAs into one protein

We first tested the feasibility to use the L11C overexpression to incorporate two AcKs into one protein in *E. coli*. An additional amber mutation was introduced into position 204 of the GFP_{UV} gene in pAcKRS-pylT-GFP1Amber to afford pAcKRS-pylT-GFP2Amber. Together with or without pET-L11C, the modified plasmid was used to transform BL21 cells. Subsequent cell growth in LB medium supplemented with 5 mM AcK and 500 μM IPTG led to the production of GFP_{UV} incorporated with AcKs at positions 149 and 204 (GFP-2AcK) in cells that were transformed with two plasmids (Figure 43). The expression yield was ~0.2 mg/L. In contrast, GFP-2AcK was not detected in cells that were transformed with only pAcKRS-pylT-GFP2Amber.

The ESI-MS analysis of the purified GFP-2AcK from cells transformed with two plasmids confirmed the expected incorporation (Figure 44).

The incorporation of AcKs at positions 149 (same as GFP-AcK, not showing) and 204 was also independently confirmed by tandem mass spectral analysis of tryptic and endoproteinase Asp-N-digested AcK-containing fragments (Figures 45).

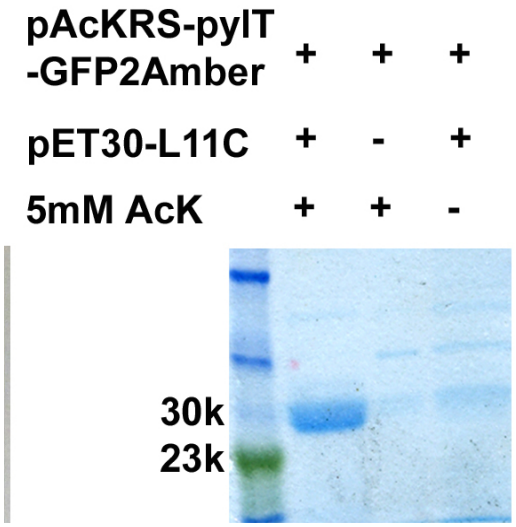


Figure 43. SDS-PAGE analysis of GFP-2AcK expressed in the absence and presence of L11C.

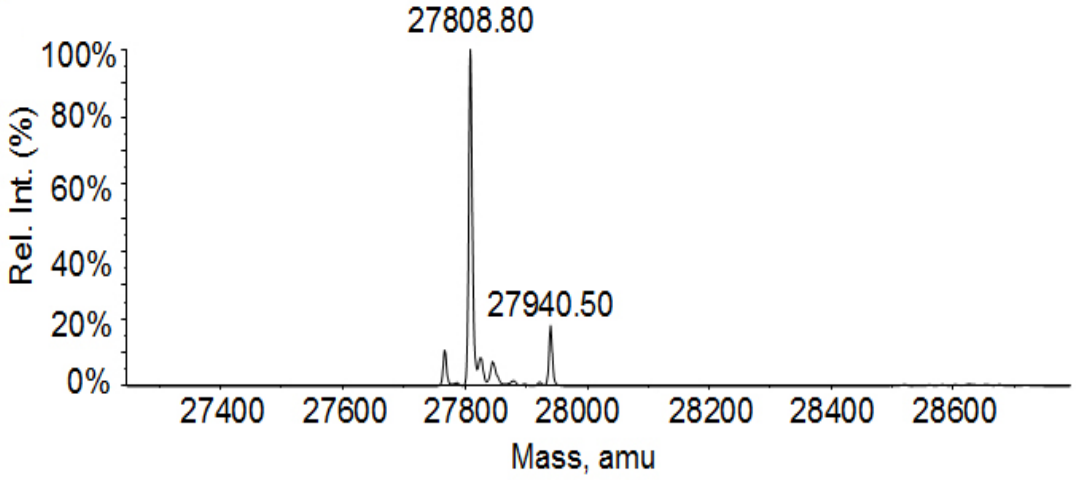


Figure 44. ESI spectrum of GFP-2AcK.

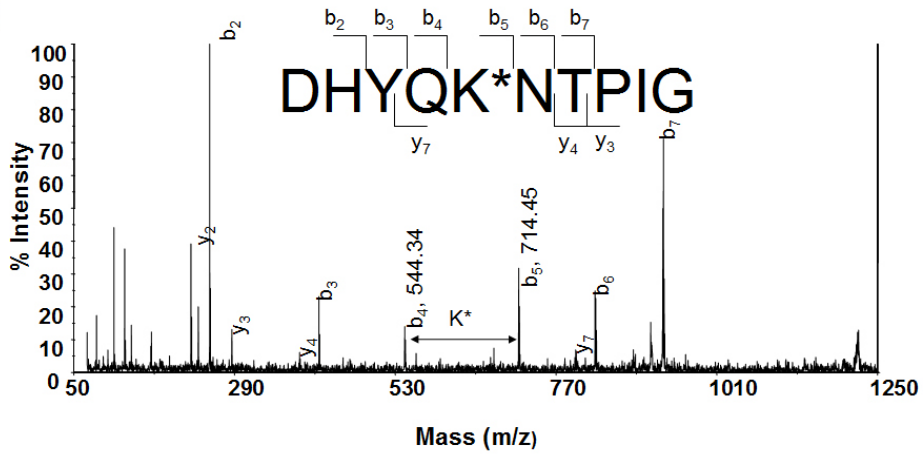


Figure 45. Tandem MS spectrum of DNHYLSTK*SALSK from GFP-2AcK. Almost identical spectra were obtained from GFP-3AcK. K* is at position 204.

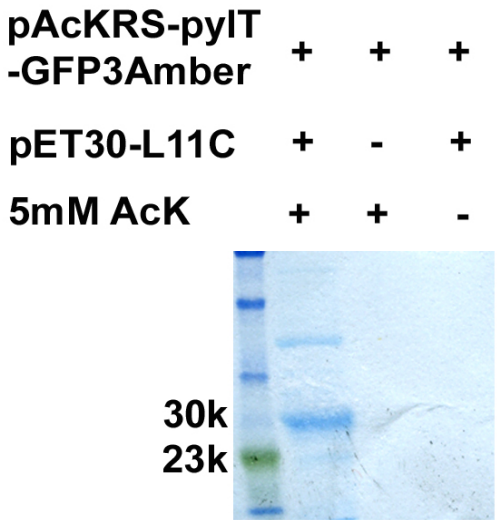


Figure 46. SDS-PAGE analysis of GFP-3AcK expressed in the absence and presence of L11C.

To test the limit of our strategy, we carried on to incorporate three AcKs into GFP_{UV} in *E. coli*. One more amber mutation was installed at position 184 of the GFP_{UV} gene in the plasmid pAcKRS-pylT-GFP2Amber to afford pAcKRS-pylT-GFP3Amber. Together with pET-L11C, this plasmid was used to express GFP_{UV} incorporated with three AcKs (GFP-3AcK) in *E. coli*, following the exactly same procedures used in GFP-2AcK expression. GFP-3AcK was expressed in cells transformed with both pAcKRS-pylT-GFP3Amber and pET-L11C with a final yield of ~0.1 mg/L. No GFP-3AcK was detected in cells transformed only with pAcKRS-pylT-GFP3Amber (Figure 46).

ESI-MS of purified full-length GFP-3AcK and tandem MS analysis of proteolytic AcK-containing peptide fragments confirmed that three AcKs were indeed incorporated into positions 149, 184, and 204 (Figure 47 and Figure 48).

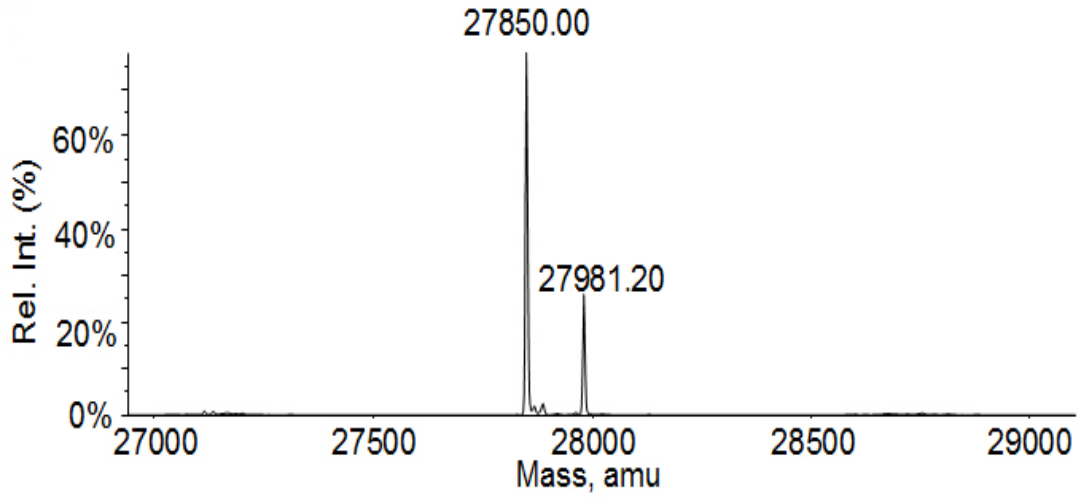


Figure 47. ESI spectrum of GFP-3AcK.

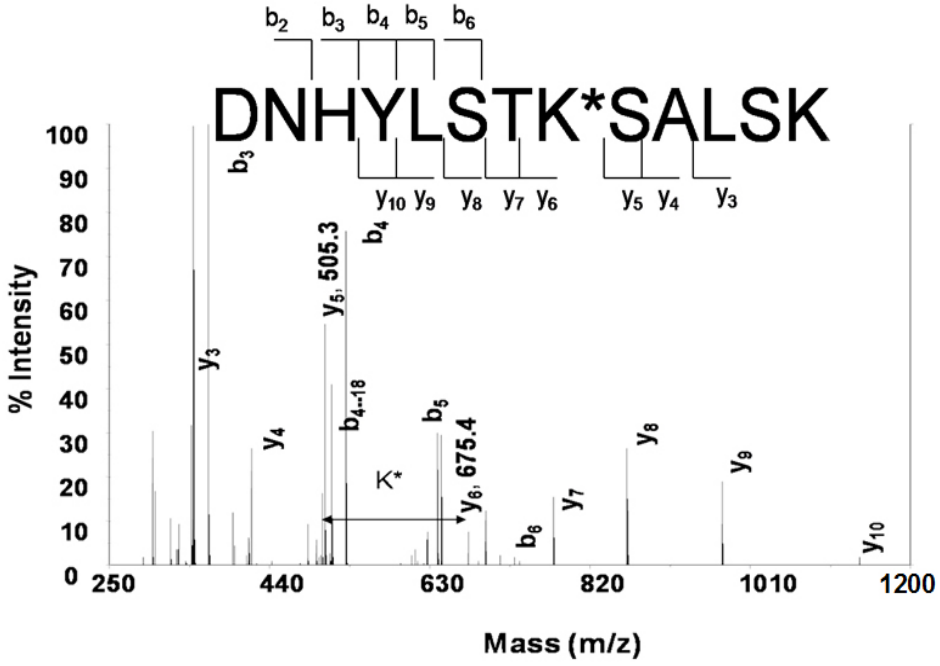


Figure 48. Tandem MS spectrum of DHYQK*NTPIG from GFP-3AcK. K* is at position 184.

Although the enhancement of amber suppression rate by overexpressing L11C allowed the expression of GFP-2AcK and GFP-3AcK in *E. coli*, the drastic decline of the expression level from GFP-AcK to GFP-2AcK was of a big concern. Based on our previous experiences, we think this might be due to protein folding difficulty when AcK is introduced at position 204 because we frequently observed that the choice of the position to introduce a NAA into a protein usually affects the protein expression yield to a great extent. To approve the drastic expression level decrease is due to the choice of the position, we carried out to express GFP_{UV} incorporated with two AcKs at positions 2 and 149 (GFP-2AcK'). An additional amber mutation was introduced into position 2 of the GFP_{UV} gene in pAcKRS-pylT-GFP1Amber to afford pAcKRS-pylT-GFP2Amber' in which an alanine insertion was also introduced into the GFP_{UV} gene between the start codon ATG and the amber codon at position 2 to relieve possible difficulty of protein translation when ribosome immediately meets a stop codon right after a start codon. Together with or without pET-L11C, the modified plasmid was used to transform BL21 cells. Subsequent cell growth in LB medium supplemented with 5 mM AcK and 500 μM IPTG led to the production of GFP-2AcK' with an expression yield of ~1.4 mg/L that is seven times higher than the GFP-2AcK expression under the same condition (Figure 49).

Cells transformed with only pAcKRS-pylT-GFP2Amber' yielded non-detectable GFP-2AcK'. This demonstrates that the low protein expression level of GFP-2AcK is a consequence of choosing position 204 to introduce the second AcK that may affect protein folding.

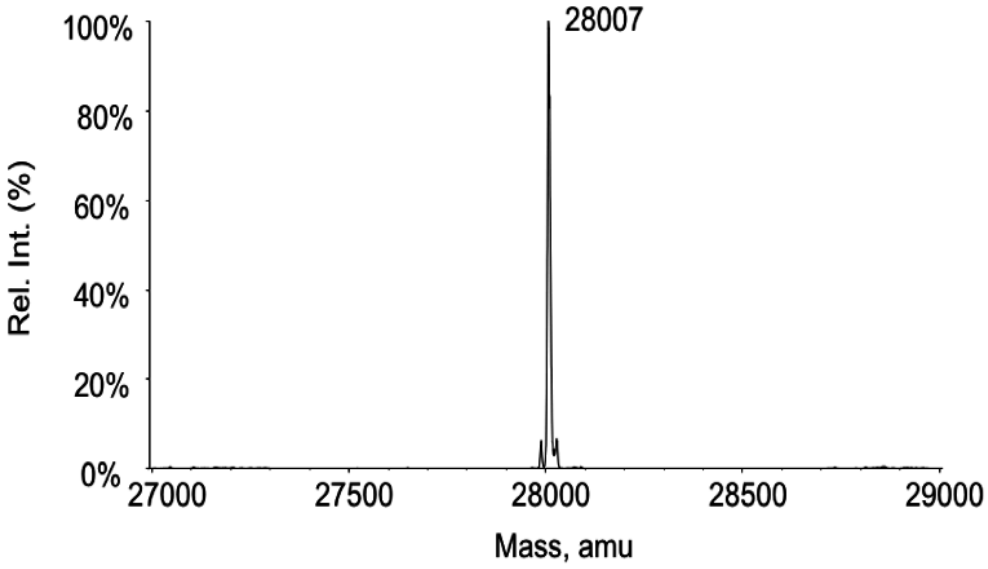


Figure 49. ESI spectrum of GFP-2AcK′.

3.2.5 Unpublished results

To further test the generality of our approach, another protein, dihydrofolate reductase (DHFR) was used as an example. DHFR was widely used for demonstration due to its high expression yield. The genes codes for DHFR, DHFR1Amber, DHFR2Amber replaced the position of GFP in the pAcKRS-pylT-GFP1Amber to obtain pAcKRS-pylT-DHFR, pAcKRS-pylT-DHFR1Amber and pAcKRS-pylT-DHFR2Amber. *E. coli* Bl21 cells transformed with pAcKRS-pylT-DHFR1Amber, or pAcKRS-pylT-DHFR2Amber together with or without pET-L11C were let grow to express the desired proteins.

3.2.5.1 DNA and protein sequences

DHFR:

5’-atgatcagtctgattgcggcgttagcggtagatcgcgttatcggcatggaaaacgccatgccgtggaacctgcctgccgat
ctcgcctggtttaaacgcaacaccttaaataaacccgtgattatgggccgccatacctgggaatcaatcggtcgtccgttgccag
gacgcaaaaatattatcctcagcagtcaaccgggtacggacgatcgcgtaacgtgggtgaagtcggtggatgaagccatcgc
ggcgtgtggtgacgtaccagaaatcatggtgattggcggcggtcgcgtttatgaacagttcttgccaaaagcgcaaaaactgta
tctgacgcatatcgacgcagaagtggaaggcgacacccatttcccggattacgagccggatgactgggaatcggtattcagcg
aattccacgatgctgatgcgcagaactctcacagctattgctttgagattctggagcggcgggagctccatcaccatcaccatca
ctaa-3’

DHFR1Amber:

5’-atgatcagtctgattgcggcgttagcggtagatcgcgttatcggcatggaaaacgccatgccgtggaacctgcctgccgat
ctctagtggtttaaacgcaacaccttaaataaacccgtgattatgggccgccatacctgggaatcaatcggtcgtccgttgccag
gacgcaaaaatattatcctcagcagtcaaccgggtacggacgatcgcgtaacgtgggtgaagtcggtggatgaagccatcgc

ggcgtgtggtgacgtaccagaaatcatggtgattggcggcggtcgcgtttatgaacagttcttgccaaaagcgcaaaaactgta

tctgacgcatatcgacgcagaagtggaaggcgacacccatttcccggattacgagccggatgactgggaatcggtattcagcg

aattccacgatgctgatgcgcagaactctcacagctattgctttgagattctggagcggcgggagctccatcaccatcaccatca

ctaa-3’

DHFR2Amber:

5’-atgatcagtctgattgcggcgttagcggtagatcgcgttatcggcatggaaaacgccatgccgtggaacctgcctgccgat

ctctagtggtttaaacgcaacaccttaaataaacccgtgattatgggccgccatacctgggaatcaatcggtcgtccgttgccag

gacgcaaaaatattatcctcagcagtcaaccgggtacggacgatcgcgtaacgtgggtgaagtcggtggatgaagccatcgc

ggcgtgtggtgacgtaccagaaatcatggtgattggcggcggtcgcgtttatgaacagttcttgccaaaagcgcaaaaactgta

tctgacgcatatcgacgcagaagtggaaggcgacacccatttcccggattacgagccggatgactgggaatcggtattcagcg

aattccacgatgctgatgcgtagaactctcacagctattgctttgagattctggagcggcgggagctccatcaccatcaccatca

ctaa-3’

DHFR:

MISLIAALAVDRVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWESIGRP

LPGRKNIILSSQPGTDDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKA

QKLYLTHIDAEVEGDTHFPDYEPDDWESVFSEFHDADAQNSHSYCFEILERREL

HHHHHH

DHFR1Amber:

MISLIAALAVDRVIGMENAMPWNLP*DLAWFKRNTLNKPVIMGRHTWESIGRPL

PGRKNIILSSQPGTDDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQ

KLYLTHIDAEVEGDTHFPDYEPDDWESVFSEFHDADAQNSHSYCFEILERRELH

HHHHH

DHFR2Amber:

MISLIAALAVDRVIGMENAMPWNLP*DLAWFKRNTLNKPVIMGRHTWESIGRPL
PGRKNIILSSQPGTDDRVTWVKSVDEAIAACGDVPEIMVIGGGRVYEQFLPKAQ
KLYLTHIDAEVEGDTHFPDYEPDDWESVFSEFHDADA*NSHSYCFEILERRELHH
HHHH

3.2.5.2 Construction of pAcKRS-pylT-DHFR

Plasmid pAcKRS-pylT-DHFR was derived from pAcKRS-pylT-GFP1Amber with the replacement of GFP1Amber with DHFR. DHFR was amplified from *E. coli* Top10 by two oligodeoxynucleotides, (5'-GAGATATACATATGATCAGTCTGATTGCGGCG-3', and 5'-GTGATGGAGCTCCCGCCGCTCCAGAATCTC-3'), digested by *NdeI* and *SacI*, and then cloned into *NdeI* and *SacI* restriction sites of pAcKRS-pylT-GFP1Amber to afford pAcKRS-pylT-DHFR.

3.2.5.3 Construction of pAcKRS-pylT-DHFR1Amber

Plasmid pAcKRS-pylT-DHFR1amber was derived from pAcKRS-pylT-DHFR with an additional amber mutation at position 26. The mutagenesis was carried out by the standard QuikChange site-directed mutagenesis procedure using PfuTurbo DNA polymerase and two oligonucleotides, 5'-CTGCCTGCCGATCTCTAGTGGTTTAAACGCAAC-3', and 5'-GTTGCGTTTAAACCACTAGAGATCGGCAGGCAG-3'.

3.2.5.4 Construction of pAcKRS-pylT-DHFR2Amber

Plasmid pAcKRS-pylT-DHFR2amber was derived from pAcKRS-pylT-DHFR1Amber with an additional amber mutation at position 146. The mutagenesis was carried out by the standard QuikChange site-directed mutagenesis procedure using PfuTurbo DNA polymerase and two oligonucleotides, 5'-CACGATGCTGATGCGTAGAACTCTCACAGCTAT-3' and 5'-ATAGCTGTGAGAGTTCTACGCATCAGCATCGTG-3'.

3.2.5.5 DHFR mutants expression and characterization

The expression and purification of wtDHFR, DHFR-AcK, and DHFR-2AcK were followed the methods described above. DHFR-AcK and DHFR-2AcK were expressed in the absence or presence of L11C. In the presence of L11C, both DHFR-2AcK and DHFR-2AcK was obtained. As expected, in the absence of L11C, smaller amount of DHFR-AcK was expressed, and no detectable amount of DHFR-2AcK was expressed (Figure 50).

The electrospray ionization mass spectrometry (ESI-MS) of full-length DHFR, DFHR1Amber, and DHFR2Amber purified from cells transformed with two plasmids confirmed the desired incorporation (Figure 51).

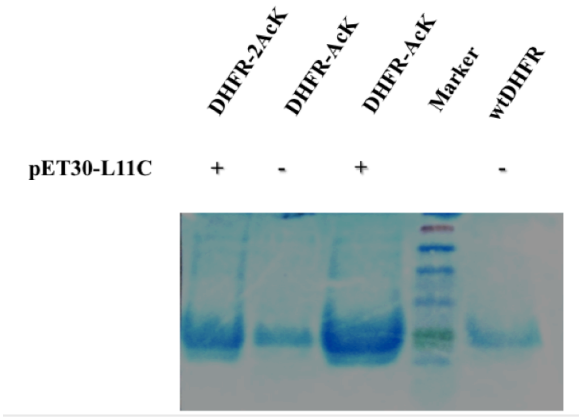


Figure 50. SDS-PAGE analysis of DHFR mutants expression. WtDHFR was loaded 1/5 of amount of other proteins. No DHFR-2AcK was expressed in the absence of L11C.

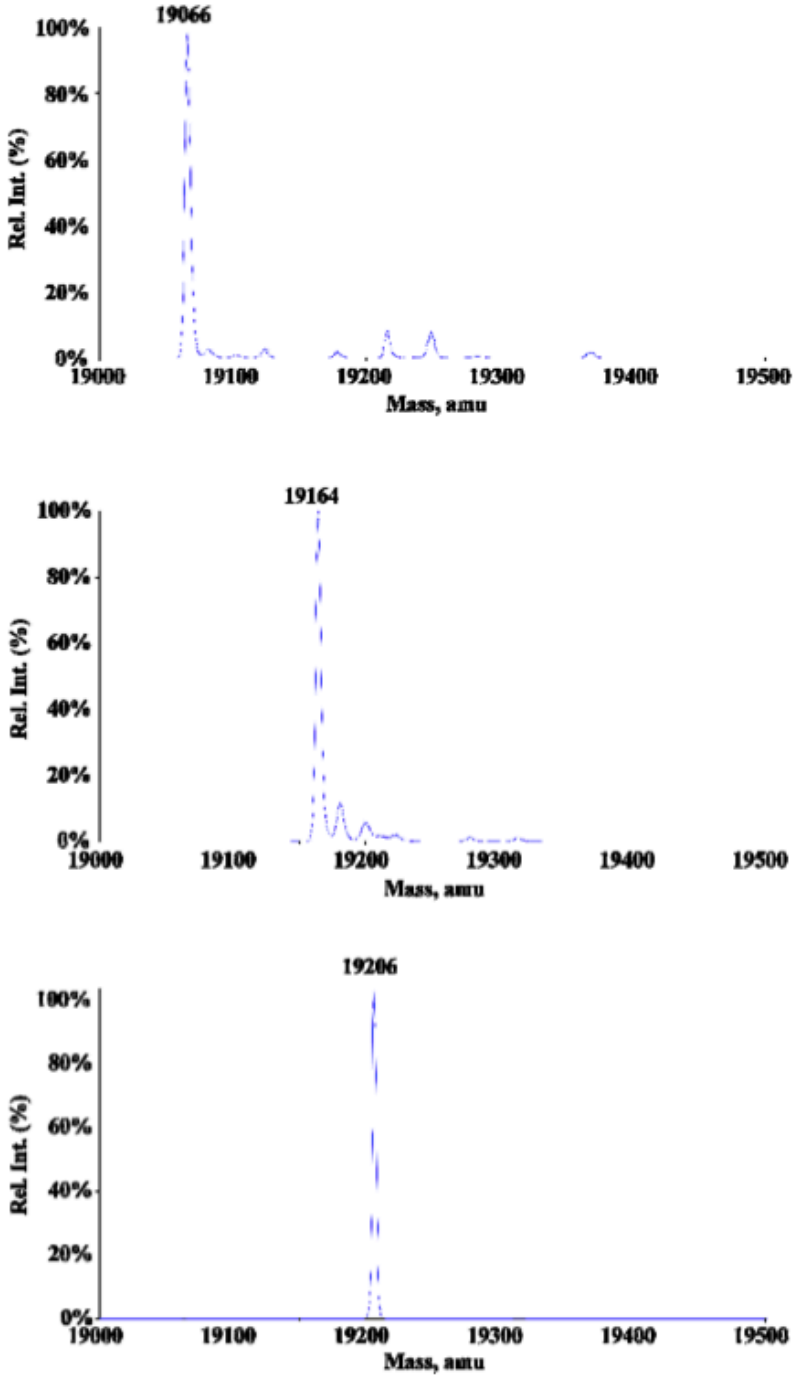


Figure 51. ESI spectra for DHFR, DHFR-AcK, and DHFR-2AcK.

3.3 Summary

In summary, we have devised a convenient approach for genetic incorporation of multiple NAAs into one protein in *E. coli* and used it successfully to incorporate three AcKs into GFP_{UV}. To our knowledge, this is the first report that demonstrated the synthesis of a protein with three acetylations in *E. coli.* Given the fact that lysine acetylation represents one of the most important posttranslational modifications that is crucial in modulating chromatin-based transcriptional control and shaping inheritable epigenetic programs, this reported method will find broad applications in deciphering the regulatory roles of acetylations in histones, p53 and other transcription regulatory proteins.[111, 112] Besides the incorporation of multiple AcKs into a protein, the developed method can be virtually used to incorporate multiple copies of any genetic encoded NAA into single proteins. The amber suppression improvement achieved by the reported approach can also be directly used to increase expression yield of NAA-incorporated proteins in *E. coli* in general. Since the genetic NAA incorporation has been applied for constructing therapeutic proteins, the application of this approach in therapeutic protein production is also anticipated.[113] Another potential application of the reported approach is to use it to achieve the incorporation of two distinct NAAs. We are currently pursuing encoding two distinct NAAs in one protein by amber and ochre (UAA) codons, respectively. Since the translation termination at the ochre codon is also partly mediated by RF1, the overexpression of L11C should enhance both amber and ochre suppression rates and allow the expression of proteins incorporated with two distinct NAAs with reasonable yields. Moreover, our method and the ribo-X system developed by Chin et al.

are independent.[106] The integration of two methods could further enhance the amber suppression rate and may allow the use of amber codon as a regular sense codon.

4. A FACILE SYSTEM FOR GENETIC INCORPORATION OF TWO DIFFERENT NONCANONICAL AMINO ACIDS INTO ONE PROTEIN IN *ESCHERICHIA COLI**

4.1 Introduction

Since the first report in 1998,[76] genetic incorporation of NAAs into proteins at amber UAG codons in living cells using orthogonal aaRS-amber suppressor tRNA ($tRNA_{CUA}$) pairs has flourished.[114] Evolved *Mj*TyrRS-$_{Mj}tRNA^{Tyr}_{CUA}$ pairs together with the naturally occurring wild-type or evolved PylRS- pylT pairs have enabled the incorporation of more than 70 NAAs into proteins in *E. coli*.[87, 100, 103, 115-119] Although the incorporation of these NAAs equipped with unique chemical properties and reactivities into proteins has dramatically increased our ability to manipulate protein structure and function, there still exists one major limitation for the technique. Namely, the technique in general only allows the incorporation of a single NAA into a single protein because the amber codon is the only one available for the incorporation of NAAs and the nonsense suppression rate in living cells is low. We have previously demonstrated the incorporation of up to three AcKs at amber mutation sites of GFP_{UV} in *E. coli* under an enhanced amber suppression condition.[120]

---

*Repreinted with permission from "A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in *Escherichia Coli*" by Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P. J., Russell, D. H., Liu, W. R., 2010. *Angew. Chem. Int. Ed. Engl.*, 49, 3211-3214, Copyright [2010] WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

The crystal structure of the PylRS-pylT complex revealed no direct interaction between PylRS and the anticodon region of pylT.[64] We suspected that mutation of the anticodon of pylT might not affect its interaction with PylRS and the mutated PylRS-pylT pair could be used to suppress a blank codon such as opal UGA codon, ochre UAA codon or even a four-base codon UAGA (Figure 52). Therefore, I aimed to utilize one additional blank codon suppressed by PylRS mutants to couple with amber codon suppressed by *Mj*TyrRS mutants to facilitate the incorporation of two different NAAs into one protein. Herein, we show that the PylRS-pylT pair can be mutated to suppress the ochre UAA codon and the mutated pair can be coupled with an evolved *Mj*TyrRS-$Mj\text{tRNA}^{\text{Tyr}}_{\text{CUA}}$ pair to efficiently incorporate two different NAAs at two defined sites of a single protein in *E. coli* by both amber and ochre suppressions.

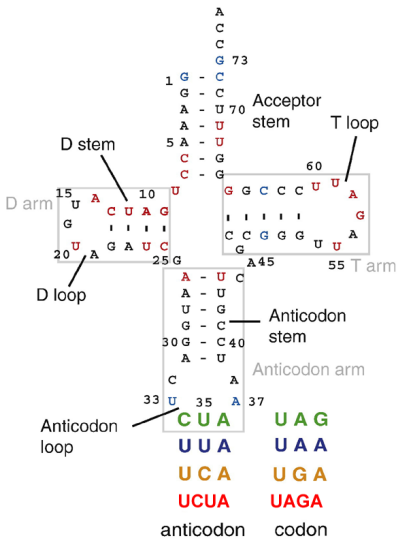


Figure 52. PylT structure and anticodon design.

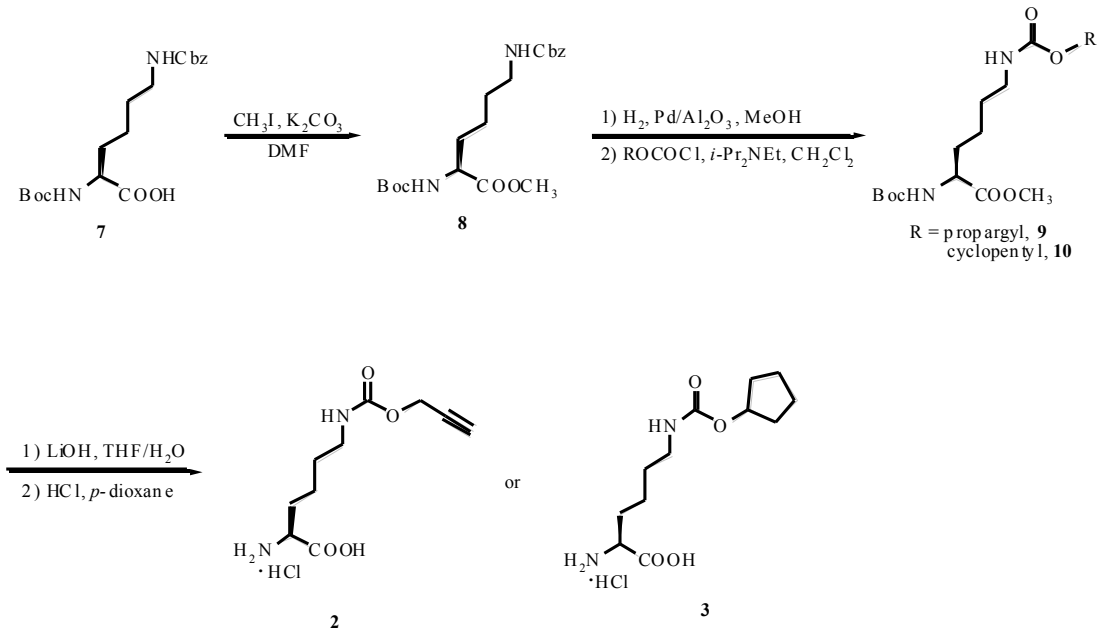
4.2 Experiments and results

4.2.1 Synthesis of **2**, **3**, and **6**

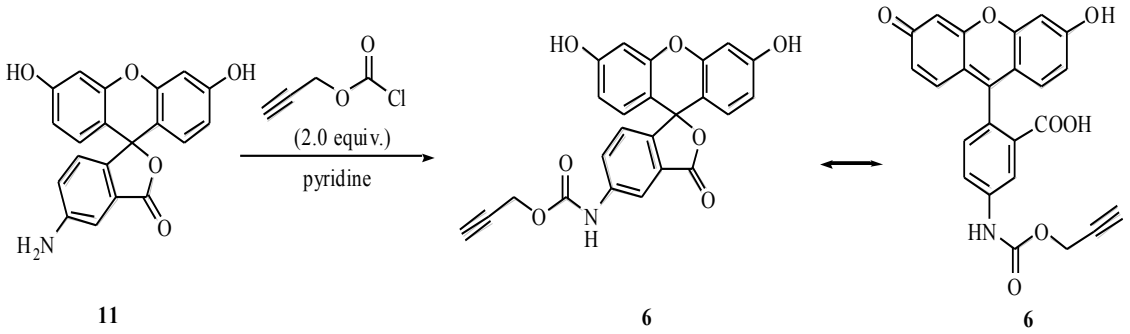
4.2.1.1 Synthetic schemes

Compounds **2** and **3** were synthesized in a divergent route (Scheme 2). Compound **6** was synthesized in one step following the literature protocol for similar compounds (Scheme 3).[121]

Scheme 2. Synthetic Route of 2 and 3



Scheme 3. Synthetic Route of 6



11 **6** **6**

4.2.1.2 Boc-Lys(Z)-OMe (**8**)[122]

To a suspension of Boc-Lys(Z)-OH (**7**, 40.3 g, 0.106 mol) and potassium carbonate (27.6 g, 0.200 mol) in DMF (200 mL) was added iodomethane (9.90 mL, 0.159 mol), and the mixture was stirred at room temperature for 30 h. The mixture was filtered, and the filter cake was washed with ethyl acetate (50 mL), dissolved in water (100 mL), and extracted with ethyl acetate (100 mL × 2). All the ethyl acetate solutions were combined with the filtrate, and the solution was evaporated under vacuum until most of the DMF has been removed. The residue was redissolved in ether (250 mL), washed with water (100 mL) and brine (50 mL), dried (Na_2SO_4), and evaporated to afford 8 (41.8 g, quant.) as yellow oil. The material was pure enough for the next-step reaction without further treatment.

A fraction of pure **8** was obtained by flash chromatography (EtOAc/hexanes, 1:5) for characterization: R_f = 0.35 (EtOAc/hexanes, 1:2); $[\alpha]_D^{22}$ +6.1(*c* 2.90, $CHCl_3$) $[\alpha]_D^{20}$ +7.8 (*c* 3.93, $CHCl_3$)); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.36-7.35 (m, 4 H), 7.33-7.30 (m,

1 H), 5.12 (d, 1 H, J = 8.5 Hz), 5.09 (s, 2 H), 4.89 (bs, 1 H), 4.30-4.26 (m, 1 H), 3.73 (s, 3 H), 3.19 (dt, 2 H, J = 6.7, 6.7 Hz), 1.80-1.77 (m, 1 H), 1.68-1.60 (m, 1 H), 1.55-1.49 (m, 2 H), 1.43 (s, 9 H), 1.39-1.33 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.5, 156.6, 155.6, 136.7, 128.7, 128.32, 128.27, 80.1, 66.8, 53.3, 52.5, 40.8, 32.5, 29.5, 28.5, 22.5.

In order to check the optical purity of **8**, a small amount of racemic **8** was obtained in the same way starting from racemic **7**. The racemic sample was dissolved in isopropanol (10 mg/mL), filtered over 0.2 μm PTFE membrane filter (VWR), and injected (4 μL) onto the Shimadzu LC system equipped with a Chiralpak IB column. The sample was isocratically eluted with 10% isopropanol in hexanes and the peaks were monitored at 210 nm. The (*S*)-enantiomer was eluted at 12.72 min and the (*R*)-enantiomer was eluted at 11.59 min. A sample of **8** (10 mg/mL, 2 μL) was subsequently analyzed under the same conditions, and its e.e. was determined to be 100%.

4.2.1.3 *N*-α-Boc-*N*-ε-propargyloxycarbonyl-L-lysine methyl ester (**9**)

A solution of **8** (3.61 g, 9.15 mmol) in methanol (100 mL) was hydrogenated under a H_2 balloon in the presence of palladium on alumina (10 wt.% Pd, 0.61 g, 0.57 mmol) at room temperature for 4 h, and TLC analysis showed a complete conversion. The mixture was then filtered over a pad of Celite and evaporated to give the crude amine as grey oil. The material should be immediately used without purification for the next-step reaction since both prolonged storage at room temperature and flash chromatography would facilitate lactam formation.

To a solution of the above amine (~9.15 mmol) in anhydrous dichloromethane (90 mL) cooled in an ice bath was added *N*,*N*-diisopropylethylamine (2.80 mL, 16.07

mmol) dropwise, followed by a solution of propargyl chloroformate (1.34 mL, 13.18 mmol) in dichloromethane (10 mL) dropwise over 20 min. The mixture was then stirred at room temperature for 12 h, and it was washed with sodium hydroxide solution (0.5 *N*, 20 mL) and brine (20 mL × 2), dried (Na_2SO_4), evaporated, and flash chromatographed (EtOAc/hexanes, 1:3) to give **9** (2.47 g, 79% for two steps) as colorless oil. R_f = 0.28 (EtOAc/hexanes, 1:2); $[\alpha]_D^{20}$ +3.9 (*c* 4.96, CH_2Cl_2); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.16 (d, 1 H, *J* = 8.5 Hz), 5.12 (t, 1 H, *J* = 5.8 Hz), 4.62 (d, 2 H, *J* = 2.5 Hz), 4.26-4.21 (m, 1 H), 3.14 (dt, 2 H, *J* = 6.5, 6.5 Hz), 2.45 (t, 1 H, *J* = 2.2 Hz), 1.80-1.73 (m, 1 H), 1.64-1.57 (m, 1 H), 1.53-1.44 (m, 2 H), 1.40 (s, 9 H), 1.36-1.30 (m, 2 H); ^{13}C NMR ($CDCl_3$, 125 MHz) δ 173.4, 155.7, 155.6, 80.0, 78.4, 74.7, 53.2, 52.4, 40.7, 32.3, 29.3, 28.4, 22.4; HRMS (ESI) calcd for $C_{16}H_{27}N_2O_6$ ($[M + H]^+$) 343.1869, found 343.1871.

4.2.1.4 *N*-*ε*-Propargyloxycarbonyl-L-lysine hydrochloride (**2**)[119]

To a solution of **9** (2.46 g, 7.18 mmol) in THF (30 mL) was added lithium hydroxide solution (0.5 M, 30.0 mL, 15.0 mmol), and the mixture was stirred at room temperature for 2 h. The mixture was diluted in water (20 mL) and extracted with ether (30 mL × 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 (60 mL × 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 colorless oil, which was used without further purification.

The above crude acid (~7.18 mmol) was dissolved in 1,4-dioxane (20 mL), and hydrogen chloride in 1,4-dioxane (4.0 M, 17.0 mL, 68.0 mmol) was added. The resulting white suspension was stirred at room temperature for 20 h, filtered, washed with dichloromethane, and dried to give **2** (1.44 g, 76% for two steps) as white solid. $[\alpha]_D^{22}$ +15.7 (*c* 1.26, 3 *N* HCl); ^{1}H NMR (D_2O, 500 MHz) δ 4.67 (s, 2 H), 4.04 (t, 1 H, *J* = 6.5 Hz), 3.16 (t, 2 H, *J* = 7.0 Hz), 2.90 (s, 1 H), 2.04-1.89 (m, 2 H), 1.57 (quintet, 2 H, *J* = 7.1 Hz), 1.53-1.38 (m, 2 H); ^{13}C NMR (D_2O, 125 MHz) δ 173.0, 158.3, 79.2, 76.2, 53.6, 53.3, 40.5, 30.0, 28.9, 22.0; HRMS (ESI) calcd for $C_{10}H_{17}N_2O_4$ ($[M + H]^+$) 229.1188, found 229.1193. The characterization data matched well with that of the corresponding trifluoroacetic acid salt of **2**.[119]

4.2.1.5 *N*-α-Boc-*N*-ε-cyclopentyloxycarbonyl-L-lysine methyl ester (**10**)

Compound **8** (2.23 g, 5.65 mmol) was converted into the corresponding amine by hydrogenolysis, which was then treated with cyclopentyl chloroformate (1.01 g, 6.79 mmol) according to the procedure for **9** to give **10** (1.12 g, 53% for two steps) as white solid. R_f = 0.35 (EtOAc/hexanes, 1:2); $[\alpha]_D^{22}$ +3.8 (*c* 1.20, CH_2Cl_2); ^{1}H NMR ($CDCl_3$, 500 MHz) δ 5.13 (d, 1 H, *J* = 7.5 Hz), 5.06 (bs, 1 H), 4.71 (bs, 1 H), 4.28-4.24 (m, 1 H), 3.72 (s, 3 H), 3.14 (dt, 2 H, *J* = 6.3, 6.3 Hz), 1.82-1.76 (m, 3 H), 1.66-1.64 (m, 5 H), 1.55-1.44 (m, 4 H), 1.42 (s, 9 H), 1.39-1.29 (m, 2 H); ^{13}C NMR (CDC_{l3}, 125 MHz) δ 173.5, 156.8, 155.6, 80.0, 53.3, 52.4, 40.5, 32.94, 32.93, 32.5, 29.6, 28.5, 23.8, 22.6; HRMS (ESI) calcd for $C_{18}H_{33}N_2O_6$ ($[M + H]^+$) 373.2339, found 373.2343.

4.2.1.6 *N*-*ε*-Cyclopentyloxycarbonyl-L-lysine hydrochloride (**3**)

According to the same procedure for **2**, **10** (1.06 g, 2.84 mmol) afforded **3** (0.71 g, 85% for two steps) as a white solid. $[\alpha]_D^{22}$+15.0 (*c* 1.14, 3 *N* HCl)[123] $[\alpha]_D^{24}$ +13.0 (*c* 2.0, 80% acetic acid) for the corresponding free amino acid); ^{1}H NMR (D_2O, 500 MHz) δ 4.94 (m, 1 H), 3.67 (t, 1 H, *J* = 6.2 Hz), 3.06 (t, 2 H, *J* = 7.0 Hz), 1,87-1.75 (m, 4 H), 1.62 (m, 4 H), 1.54-1.51 (m, 2 H), 1.48 (quintet, 2 H, *J* = 7.0 Hz), 1.39-1.29 (m, 2 H); ^{13}C NMR (D_2O, pH = 10, 125 MHz) δ 183.8, 159.5, 79.4, 56.5, 40.7, 34.7, 32.9, 29.5, 23.8, 22.8; HRMS (ESI) calcd for $C_{12}H_{23}N_2O_4$ ($[M + H]^+$) 259.1658, found 259.1650.

4.2.1.7 5-(Propargyloxycarbonylamino) fluorescein (**6**)[121]

To a solution of fluoresceinamine isomer I (**11**, 0.102 g, 0.29 mmol) (Aldrich) in anhydrous pyridine (1.0 mL) cooled in an ice bath was added propargyl chloroformate (58 μL, 0.59 mmol) dropwise, and the mixture was stirred at room temperature for 30 h. Ethyl acetate (50 mL) was then added, and the solution was washed with water (10 mL), saturated sodium bicarbonate (10 mL), hydrochloric acid (1 *N*, 10 mL) and brine (10 mL), dried (Na_2SO_4), evaporated, and flash chromatographed (EtOAc/hexanes, 1:1) to give yellow oil, which was dissolved in a minimal amount of ethyl acetate (~0.5 mL) and precipitated with excessive hexanes (~10 mL) to afford **6** (66 mg, 52%) as bright orange solid. R_f = 0.55 (EtOAc/hexanes, 2:1); ^{1}H NMR (CD_3O_D, 500 MHz) δ 8.17 (s, 1 H), 7.76 (dd, 1 H, *J* = 8.5, 2.0 Hz), 7.13 (d, 1 H, *J* = 9.0 Hz), 6.67 (d, 2 H, *J* = 2.5 Hz), 6.65 (d, 2 H, *J* = 9.0 Hz), 6.54 (dd, 2 H, *J* = 8.5, 2.5 Hz), 4.82 (d, 2 H, *J* = 2.5 Hz), 2.97 (t, 1 H, *J* = 2.5 Hz); ^{13}C NMR (CD_3OD, 75 MHz) δ 171.6, 162.3, 155.0, 154.6, 142.3,

130.5, 126.7, 126.2, 114.9, 114.1, 111.9, 103.6, 79.2, 76.4, 53.6; HRMS (ESI) calcd for $C_{24}H_{16}NO_7$ ($[M + H]^+$) 430.0927, found 430.0931.

4.2.2 Construction of plasmids

All the plasmid structures have been confirmed by DNA sequencing. All oligonucleotide primers were purchased from Integrated DNA Technologies, Inc. pAcKRS-pylT-GFP1Amber was used as template. In order to test the feasibility of all stop codons and one four base codon TAGA, pylT and mutations were introduced to both anticodon of pylT and gene of GFP1Amber, corresponding to afford pAcKRS-pylT-GFP1Ochre, pAcKRS-pylT-GFP1Opal, and pAcKRS-pylT-GFP1four base. In addition to check the suppression rate of two codons, another mutation was introduced to pAcKRS-pylT-GFP1Ochre to afford pAcKRS-pylT-GFP1TAG149TAA. In this plasmid, the tRNA was $pylT_{UUA}$, which recognized Ochre codon. Two stop codons were introduced into GFP, TAG at third position following first Met and second Ala, TAA at 149^{th} position. pEVOL, carrying two copy of *Mj*TyrRS mutants, was used to suppress TAG Amber codon (Figure 53).

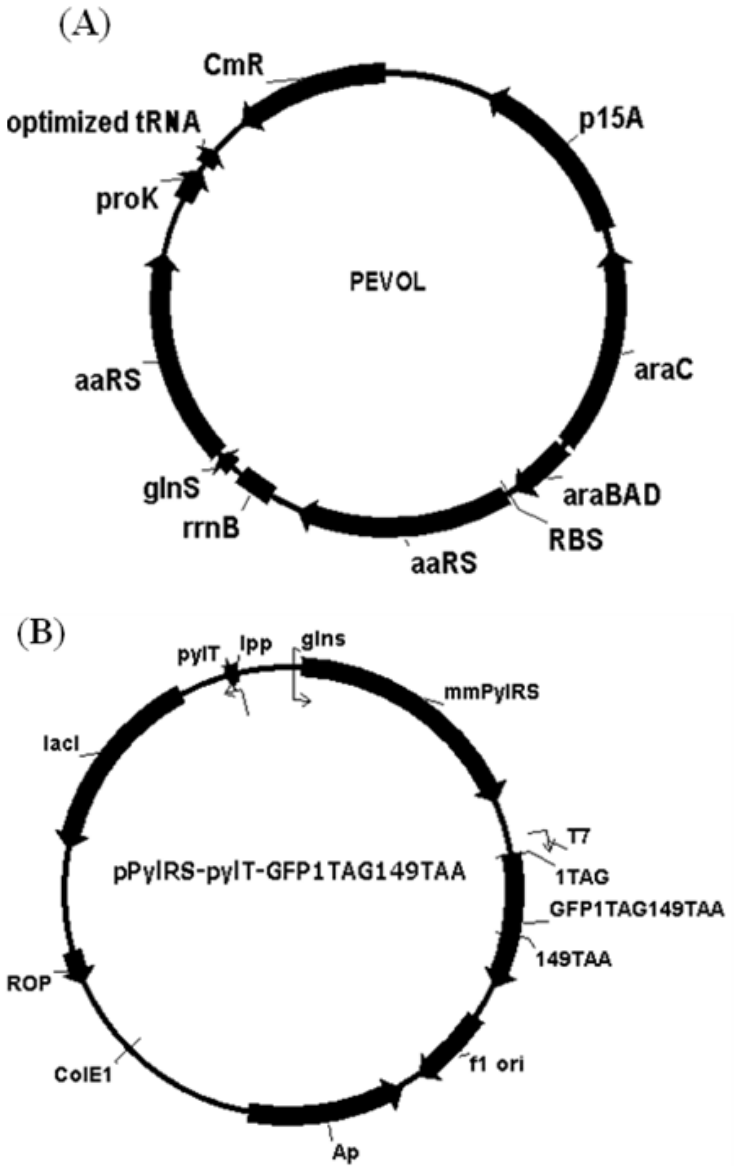


Figure 53. Plasmid maps of (**A**) pEVOL and (**B**) pPylRS-pylT-GFP1TAG149TAA.

4.2.2.1 DNA and protein sequences

pylT$_{CUA}$:

5’-ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca-3’

pylT$_{UUA}$:

5’-ggaaacctgatcatgtagatcgaatggacttaaatccgttcagccgggttagattcccggggtttccgcca-3’

pylT$_{UCA}$:

5’-ggaaacctgatcatgtagatcgaatggacttcaaatccgttcagccgggttagattcccggggtttccgcca-3’

pylT$_{UCUA}$:

5’-ggaaacctgatcatgtagatcgaatggacttctaaatccgttcagccgggttagattcccggggtttccgcca-3’

GFP1TAG149TAA:

5’-atggcatagagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaa

ttttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttc

catggccaacacttgtcactactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaaga

gtgccatgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagt

ttgaaggtgataccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgagt

acaactataactcacactaagtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaaca

ttgaagatggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtcctttaccagacaaccatt

acctgtcgacacaatctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctggg

attacacatggcatggatgaactctacaaagagctccatcaccatcaccatcactga-3’

GFP1TAG149TAA (X and X* are the incorporated noncanonical amino acids):

MAXSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGK

LPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGN

YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHX*VYITADKQK

NGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNE

KRDHMVLLEFVTAAGITHGMDELYKELHHHHHH

4.2.2.2 Construction of pAcKRS-pylT-GFP1Ochre, pAcKRS-pylT-GFP1Opal, and pAcKRS-pylT-GFP1four base

pAcKRS-pylT-GFP1Opal contains genes encoding AcKRS, $pylT_{UCA}$ and GFP_{UV} with an opal mutation at position 149. This plasmid was derived from pAcKRS-pylT-GFP1Amber.[124] The standard QuikChange site-directed mutagenesis was used to mutate the anticodon of pylT in pAcKRS-pylT-GFP1Amber to UCA first. The resulting plasmid then underwent the second QuikChange site-directed mutagenesis to mutate TAG at position 149 of GFP_{UV} to TGA.

Similarly, two sequential QuikChange site-directed mutagenesis experiments were carried out to mutate the anticodon of pylT and the amber mutation of GFP_{UV} in pAcKRS-pylT-GFP1Amber to form pAcKRS-pylT-GFP1Ochre and pAcKRS-pylT-GFP1four base.

4.2.2.3 Construction of pAcKRS-pylT-GFP1TAG149TAA

The plasmid pAcKRS-pylT-GFP1TAG149TAA was derived from pAcKRS-pylT-GFP1Ochre that contains $pylT_{UUA}$ and GFP_{UV} with an ochre mutation at position 149 (GFP1Ochre). GFP1TAG149TAA was first generated by PCR amplification of

GFP1Ochre with primers that add one amber mutation at the first codon, two additional codons (ATGGCA) in front of the first codon, and TGA stop codon at the end of the gene. This gene was then cloned into the pAcKRS-pylT-GFP1Ochre between *NdeI* and *KpnI* sites to afford pAcKRS-pylT-GFP1TAG149TAA. In the final plasmid, GFP1Ochre was replaced by GFP1TAG149TAA.

4.2.2.4 Construction of pPylRS-pylT-GFP1TAG149TAA

The plasmid pPylRS-pylT-GFP1TAG149TAA was derived from pAcKRS-pylT-GFP1AG149TAA. The PylRS gene flanked by the constitutive glutamine promoter at the 5' end and the glutamine terminator at the 3' end was PCR amplified from pBK-PylRS. Two restriction sites, *ClaI* at the 5' end and *Hind111* at the 3' end, were introduced in the synthesized DNA, which was subsequently digested by these two enzymes and cloned into pAcKRS-pylT-GFP1AG149TAA to replace AcKRS.

4.2.2.5 Construction of pEVOL-sTyrRS

The plasmid pEVOL-sTyrRS was derived from pEVOL-AzFRS.[125] sTyrRS was PCR amplified from pBK-sTyrRS.[104] It was then cloned into *Nde1*and *Pst1* sites of pEVOL-AzFRS to replace the first AzFRS copy. After confirmation by DNA sequencing, the second sTyrRS copy was cloned into *Bgl11* and *Pst1* sites to replace the second copy of AzFRS to afford pEVOL-sTyrRS.

4.2.3 Orthogonality test

We attempted to combine the PylRS-pylT pair and the *Mj*TyrRS- *Mj*tRNA$^{Tyr}_{CUA}$ pair to incorporate two different NAAs into a single protein because of their proven efficiency in NAA incorporation.[87, 100, 117, 118, 126-129] In order to use both pairs in a single

E. coli cell, we first demonstrated that they were orthogonal to each other. When coexpressed in *E. coli*, either one of the aaRSs could not efficiently charge $tRNA_{CUA}$ from the other pair with its cognate amino acid (BocK for PylRS[87] and L-tyrosine for *Mj*TyrRS) to suppress an amber mutation at position 112 of a chloramphenicol acetyltransferase gene to give detectable chloramphenicol resistance.

To test the interaction between PylRS and $Mj\text{tRNA}^{\text{Tyr}}_{\text{CUA}}$, two plasmids, pBK-pylRS and pREP, were used to transform *E. coli* Top10 cells. pREP contains genes encoding $Mj\text{tRNA}^{\text{Tyr}}_{\text{CUA}}$ and a chloramphenicol acetyltransferase with an amber mutation at position 112. The transformed cells were grown on LB plates containing 25 μg/mL kanamycin and 12 μg/mL tetracycline. Five single colonies from the kanamycin/tetracycline plate were transferred onto a LB plate containing 25 μg/mL kanamycin, 12 μg/mL tetracycline, 34 μg/mL chloramphenicol, and 1 mM BocK. None of them were viable. A similar experiment was carried out to test the interaction between *Mj*TyrRS and pylT. pBK-JYRS that contains wild type *Mj*TyrRS and pREP in which $Mj\text{tRNA}^{\text{Tyr}}_{\text{CUA}}$ was replaced with pylT were used to transform *E. coli* Top10 cells. The growth of the transformed cells on a LB plate containing 25 μg/mL kanamycin, 12 μg/mL tetracycline and 34 μg/mL chloramphenicol did not lead to any viable clones (Figure 54).

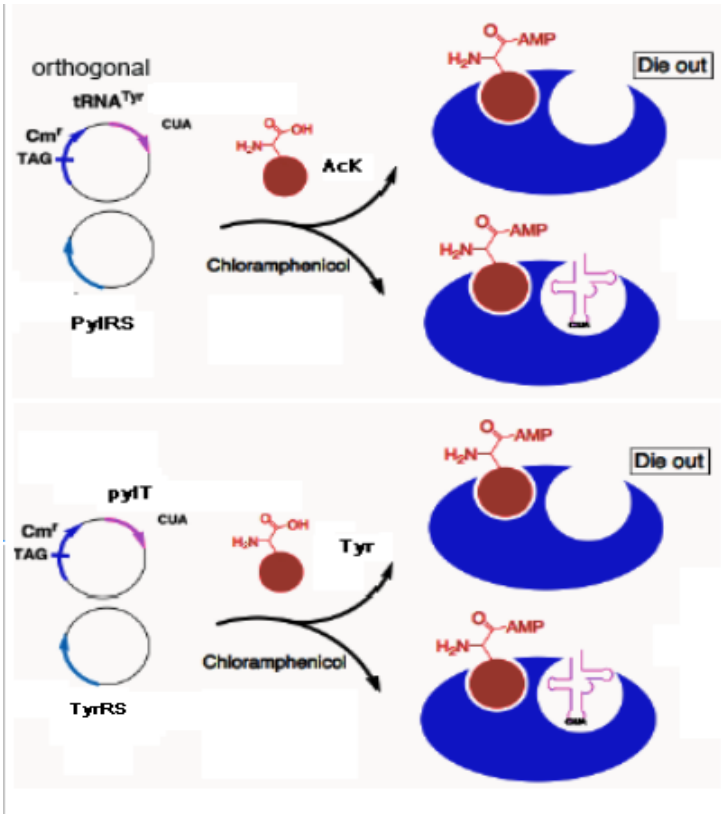


Figure 54. Orthogonality test.

4.2.4 Codon suppression test

We then tested whether we can use mutated pylT to suppress other blank codons. Although naturally encoded by the UAG codon, it has been reported that pyrrolysine is not hardwired for cotranslational insertion at UAG codon positions.[73] The crystal structure of the PylRS-pylT complex also revealed no direct interaction between PylRS and the anticodon region of pylT.[64] We suspected that mutation of the anticodon of pylT might not affect its interaction with PylRS and the mutated PylRS-pylT pair could be used to suppress a blank codon such as opal UGA codon, ochre UAA codon or even a four-base codon UAGA. A pAcKRS-pylT-GFP1Amber plasmid from our previous work[120, 130] was employed. This plasmid contains genes coding a mutant PylRS (AcKRS) specific for AcK, pylT and GFP_{UV}. The GFP_{UV} gene has one amber mutation at position 149. Growing cells transformed with this plasmid in LB medium supplemented with 5 mM AcK led to full-length GFP_{UV} expression that was easily detected by fluorescent emission of GFP_{UV} when excited (Figure 55). When the anticodon of pylT was mutated to the complimentary one for opal, ochre or a four-base UAGA codon and the corresponding codon was introduced to position 149 of GFP_{UV} in the pAcKRS-pylT-GFP1Amber plasmid, cells transformed with the plasmid and grown in LB medium supplemented with 5 mM AcK also exhibited detectable GFP_{UV} expression for all three mutated pylTs (Figure 55). In comparison to wild-type pylT, $pylT_{UUA}$ that suppresses ochre codon gave significantly high suppression levels. This indicates that the PylRS-$pylT_{UUA}$ pair can be used to efficiently incorporate NAAs into proteins at its corresponding suppressed codons and it might also be feasible to couple

the PylRS-pylT$_{UUA}$ pair together with an evolved *Mj*TyrRS- *Mj*tRNA$^{Tyr}_{CUA}$ pair to incorporate two different NAAs into a single protein in *E. coli* by both amber and ochre suppressions.

E. coli BL21 cells transformed with pAcKRS-pylT-GFP1Amber were grown in 5 mL of LB medium at 37 degree overnight, and the culture was subsequently inoculated into 50 mL of LB supplemented with 100 μg/mL ampicillin. The expression of GFP$_{UV}$ was then induced with the addition of 500 μg/mL IPTG and 1 mM AcK when the OD$_{600}$ reached 0.6. After induction, the culture was let grown at 37 degree for 6 h. Cells were then harvested by centrifugation at 4500 r.p.m. for 20 min at 4 degree and resuspended in 20 mL of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM DTT, 10% glycerol, 0.1% Triton X-100, 5 mM imidazole, and 1 μg/mL lysozyme, pH 7.4). The resuspended cells were sonicated in ice water bath three times (4 min each, 5 min interval to cool the suspension below 10 degree before the next run) and the lysate was clarified by centrifugation at 10200 r.p.m. for 60 min at 4 degree. More lysis buffer was added to the supernatant to make a final volume of 50 mL. The expression of GFP$_{UV}$ from cells transformed with pAcKRS-pylT-GFP1Amber, pAcKRS-pylT-GFP1Opal, pAcKRS-pylT-GFP1Ochre and pAcKRS-pylT-GFP1four base were carried out following exactly the same procedures. Fluorescence spectroscopic studies of the final clarified GFP$_{UV}$ solution were performed on a Cary Eclipse fluorometer. The slit width was 5 nm for both excitation and emission. The fluorescence of each sample was excited at 397 nm and then measured at 505 nm.

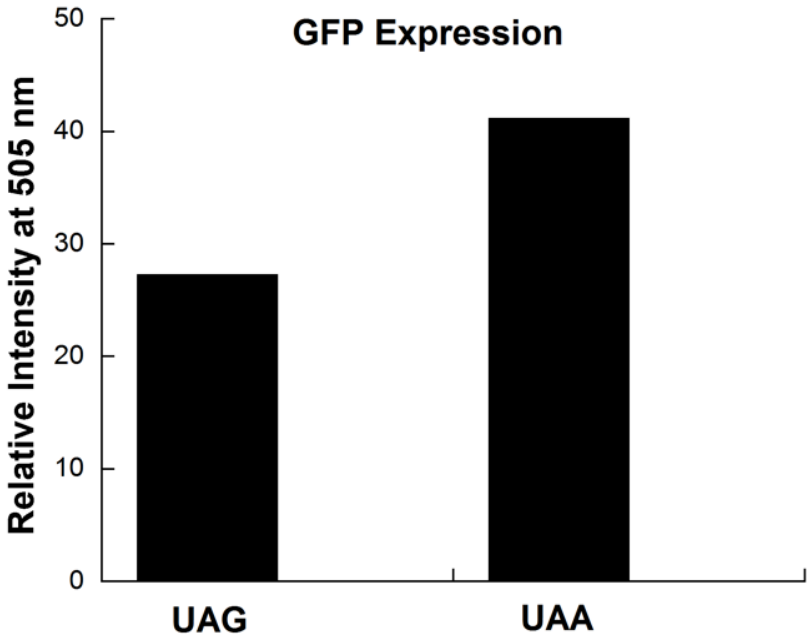


Figure 55. Suppression levels of UAG and UAA mutations at position 149 of GFP_{UV} by their corresponding mutant pylT suppressors.

4.2.5 Expression of wild-type GFP_{UV}

A pREP plasmid that contains wild type GFP_{UV} gene under control of the T7 promoter was transformed into BL21 cells. The transformed cells were grown in 5 mL of LB medium supplemented with 12 μg/mL tetracycline at 37 °C overnight, and the culture was inoculated into 200 mL of LB medium containing 12 μg/mL tetracycline and let grown at 37 degree for 8 h. Cells were then harvested by centrifugation at 4500 r.p.m. for 20 min at 4 degree and resuspended in 20 mL of lysis buffer (50 mM HEPES, 500 mM NaCl, 10 mM DTT, 10% glycerol, 0.1% Triton X-100, 5 mM imidazole, and 1 μg/mL lysozyme, pH 7.4). The resuspended cells were sonicated in ice water bath three times (4 min each, 5 min interval to cool the suspension below 10 degree before the next run) and the lysate was clarified by centrifugation at 10200 r.p.m. for 60 min at 4 degree. The supernatant that contained high concentration of GFP was then fractionated by the addition of 70% ammonium sulfate to effect precipitation. The precipitate was then redissolved in 100 mM sodium phosphate buffer (pH 7.6) to make a 4 mg/mL solution, which was directly used for protein labeling assay.

4.2.6 Expression and purification of GFP_{UV}(**1**+**4**), GFP_{UV}(**2**+**4**), GFP_{UV}(**3**+**4**) from cells transformed with pPylRS-pylT-GFP1TAG149TAA and pEVOL-AzFRS

In order to demonstrate the utility of the PylRS-$pylT_{UUA}$ pair together with an evolved *Mj*TyrRS- *Mj*$tRNA^{Tyr}_{CUA}$ pair to incorporate two different NAAs into a single protein in *E. coli* by both amber and ochre suppressions, two plasmids, pEVOL-AzFRS and pPylRS-pylT-GFP1TAG149TAA, were used to transform *E. coli* BL21 cells. The pEVOL-AzFRS plasmid contains genes encoding an optimized *Mj*$tRNA^{Tyr}_{CUA}$ and two

copies of an evolved *Mj*TyrRS (AzFRS) specific for *p*-azido-L-phenylalanine (**4** in Figure 56). This plasmid provides an enhanced amber suppression in *E. coli*.[131] The pPylRS-pylT-GFP1TAG149TAA plasmid contains genes encoding wild-type *M. mazei* PylRS, $pylT_{UUA}$, and GFP_{UV}. The GFP_{UV} gene has an amber mutation at position 1, an ochre mutation at 149, an N-terminal Met-Ala leader dipeptide in front of the amber mutation and an opal stop codon at the C-terminal end.

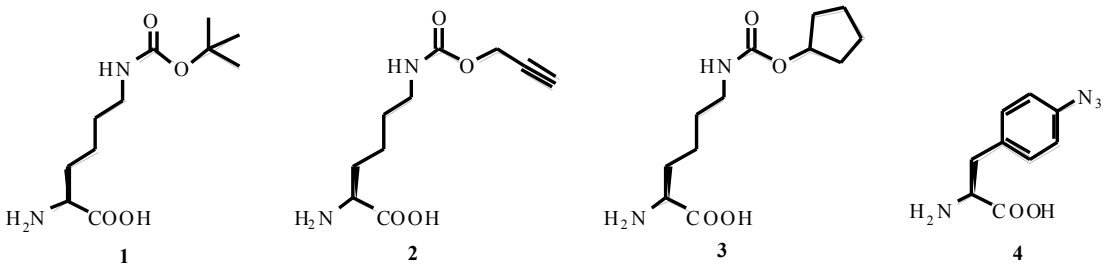


Figure 56. Chemical structure for compound **1**, **2**, **3** and **4**.

E. coli BL21 cells transformed with pPylRS-pylT-GFP1TAG149TAA together with pEVOL-AzFRS were grown in 2YT medium (50 mL) at 37 degree overnight. The culture was inoculated into 2YT medium (150 mL) containing 100 μg/mL ampicillin and 25 μg/mL kanamycin and then induced with the addition of 500 μg/mL IPTG and 0.02% arabinose when the OD_{600} reached 0.6. Compound **4** with either one of **1**, **2**, and **3** were added so that there was 1 mM each of the two NAAs. The culture was then incubated at 37 degree for 8 h, and cells were harvested and lysed as described above. The

supernatants were loaded onto a GE Healthcare AKTApurifier UPC 10 FPLC system equipped with a Ni-NTA agarose (Qiagen Inc.) column. The column was washed with 5 × bed volumes of buffer A (50 mM HEPES, 300 mM NaCl and 5 mM imidazole, pH 7.5) and then eluted by running a gradient from 100% buffer A to 100% buffer B (50 mM HEPES, 300 mM NaCl and 50 mM imidazole, pH 7.5) in 10 × bed volumes. The eluted proteins were concentrated by Amicon Ultra-4 Centrifugal Filter Units (Millipore, NMWL 10 KDa) and analyzed by 12% SDS-PAGE. The protein concentrations were determined by BCA assay kit, which is purchased from Pierce Inc.

Growing the transformed cells in 2YT medium supplemented with 1 mM BocK (**1** in Figure 57A) and 1 mM **4** afforded full-length GFP_{UV} with a yield of 11 mg/L (Figure 57B, lane 1). No cellular toxicity due to strong amber and ochre suppressions was observed. Exclusion of either NAA from the medium led to no detectable full-length GFP_{UV} expression (Figure 57B, lanes 4 & 7). The results indicate that the suppressions of amber and opal mutations are dependent on the presence of their corresponding NAAs. The ESI-MS of the purified full-length GFP_{UV} incorporated with **4** at position 1 and **1** at position 149 (GFP_{UV}(**1**+**4**)) confirmed the expected incorporations (Figure 58). The detected mass (28085 Da) agrees within 70 parts per million with the calculated mass (28083 Da) of full-length GFP_{UV}(**1**+**4**) without N-terminal methionine. The cleavage of N-terminal methionine from expressed GFP_{UV} in *E. coli* has been observed in related studies.[89, 120] A mass peak (28059 Da) that is 26 Da smaller than the major peak is probably due to the decomposition of the azide group in **4** to form the

corresponding amine during ESI-MS analysis, which has been observed previously.[132, 133]

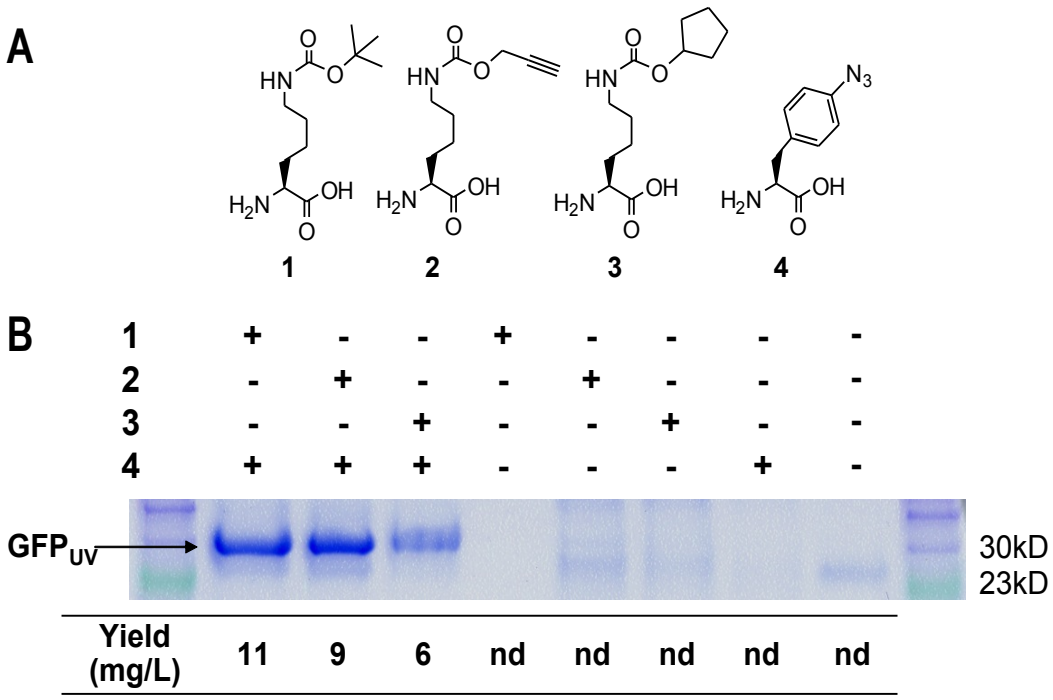


Figure 57. (**A**) Structures of four NAAs. (**B**) The expression level of full-length GFP_{UV} with amber mutation at position 1 and ochre mutation at position 149 from cells transformed with pEVOL-AzFRS and pPylRS-pylT-GFP1TAG149TAA at different conditions. All NAAs supplemented into media were at 1 mM.

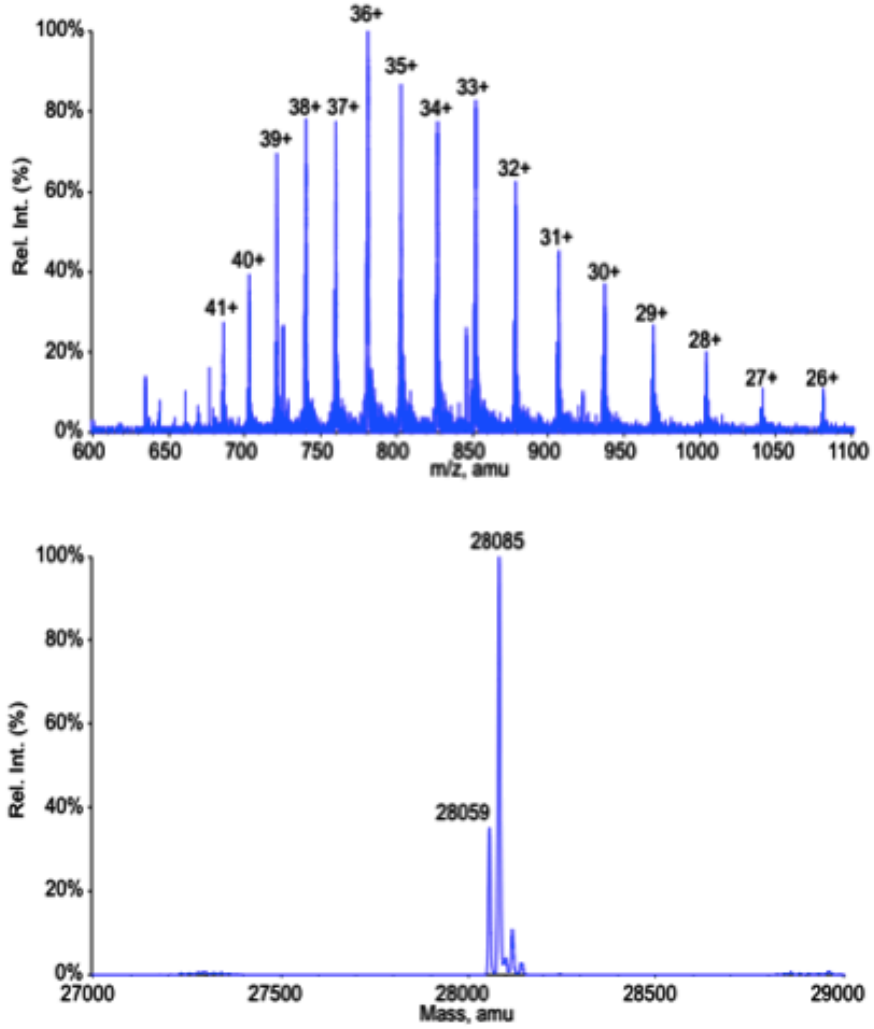


Figure 58. ESI spectrum and deconvoluted MS spectrum of GFP_{UV}(**1**+**4**).

Since wild-type PylRS also charges pylT with *N*-*ε*-propargyloxycarbonyl-L-lysine[87, 134] (**2** in Figure 57A) and *N*-*ε*-cyclopentyloxycarbonyl-L-lysine[87, 134] (**3** in Figure 56A), the incorporation of either of these two NAAs together with **4** into GFP_{UV} was also tested in cells transformed with pEVOL-AzFRS and pPylRS-pylT-GFP1TAG149TAA. Growing cells in 2YT medium supplemented with 1 mM **2** (or **3**) and 1 mM **4** afforded full-length GFP_{UV} in good yields (Figure 57B). No full-length GFP_{UV} expression was detected when only one NAA was present in the medium. ESI-MS analysis of the purified proteins confirmed the expected incorporations (GFP_{UV} incorporated with **4** and **2** (GFP_{UV}(**2**+**4**)): 28065 Da (calculated), 28067 Da (detected) (Figure 59). GFP_{UV} incorporated with **4** and **3** (GFP_{UV}(**3**+**4**)): 28095 Da (calculated), 28096 Da (detected)) (Figure 60).

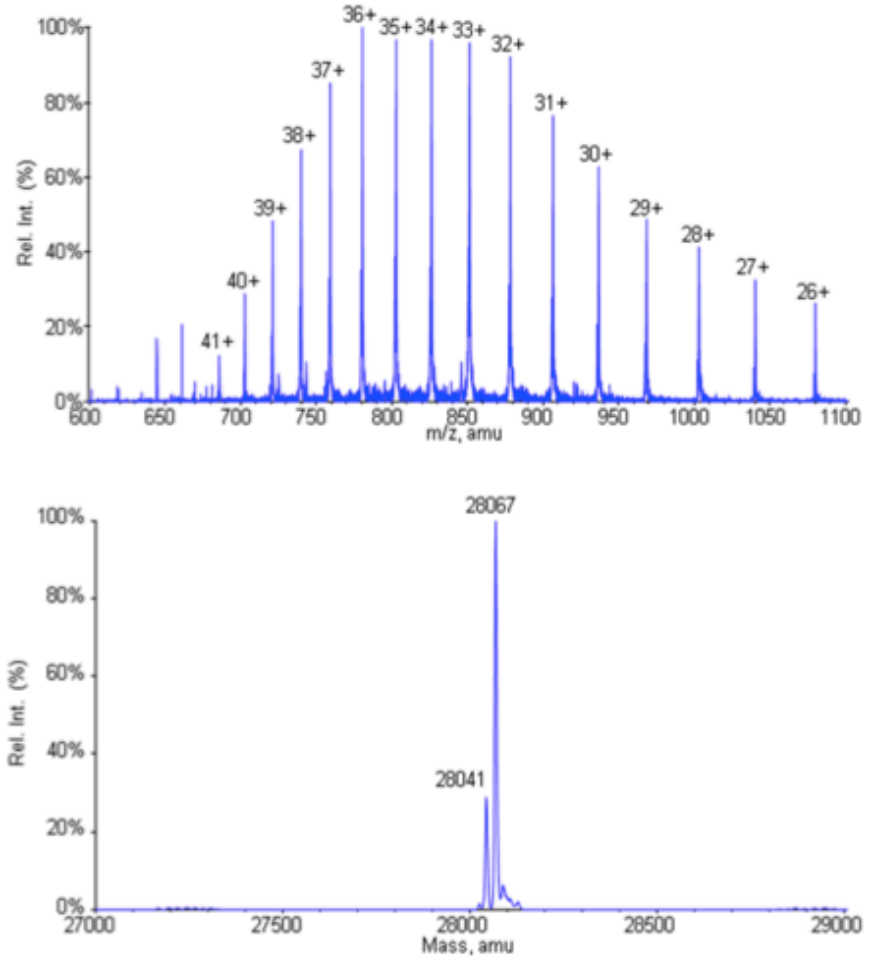


Figure 59. ESI spectrum and deconvoluted MS spectrum of GFP_{UV}(**2**+**4**).

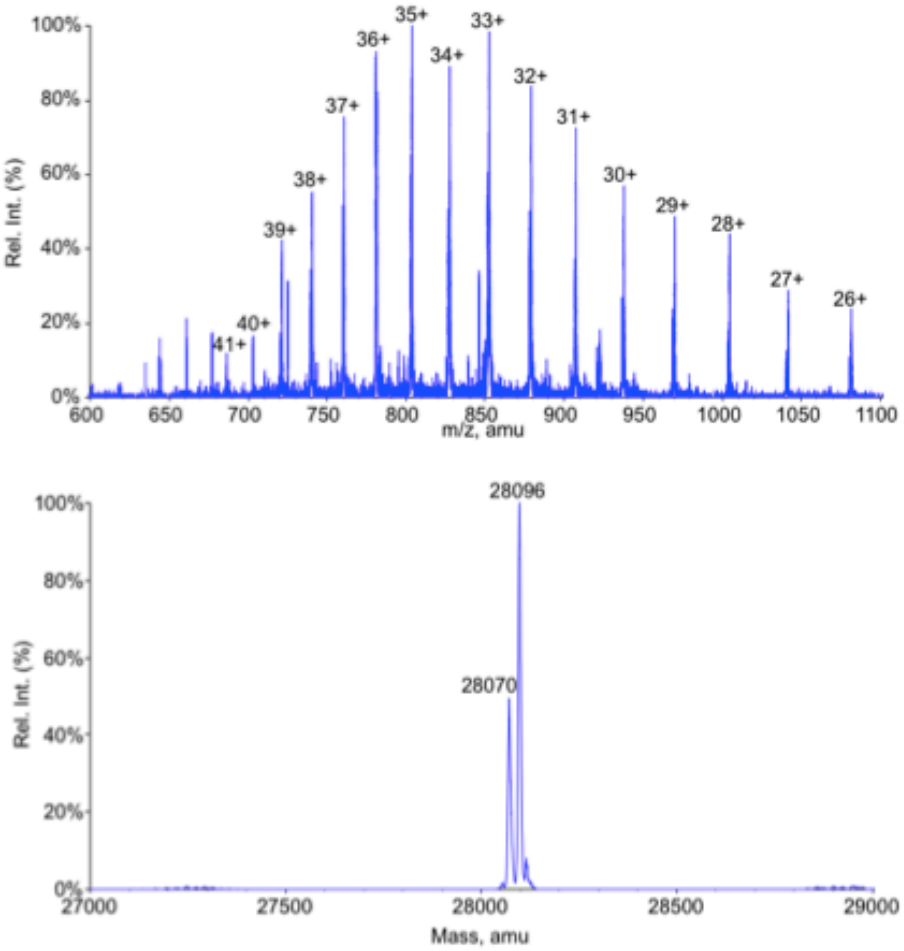


Figure 60. ESI spectrum and deconvoluted MS spectrum of GFP_{UV}(**3**+**4**).

4.2.7 Protein Labeling

Since GFP_{UV}(**2**+**4**) contains both an alkyne group and an azide group, we tested the feasibility of separately labeling this protein with different fluorescent dyes by performing click reactions on these two functional groups.[119, 135] The reaction of GFP_{UV}(**2**+**4**) with 3-azido-7-hydroxycoumarin[119, 135] in the presence of a Cu(I) catalyst[119, 135] led to a labeled GFP_{UV} that emitted strong blue fluorescence under long wavelength UV light (365 nm) after the protein was denatured and analyzed in a SDS-PAGE gel (Figure 61**,** lane 2). The same labeling reaction between wild-type GFP_{UV} ($wtGFP_{UV}$) did not give any detectable blue fluorescence when excited. Since both proteins were denatured prior to SDS-PAGE analysis, the endogenous fluorophore of GFP_{UV} was quenched and did not interfere with the analysis. The similar click chemistry reaction was also carried out between GFP_{UV}(**2**+**4**) and a propargyl-conjugated fluorescein. The labeled protein emitted strong green fluorescence when excited at 365 nm (Figure 61, lane 4). The control reaction on $wtGFP_{UV}$ again yielded no detectable fluorescently labeled protein (Figure 61, lane 3). These experiments demonstrate that both side-chain functional groups of **2** and **4** are active after their incorporation into proteins and can be used separately to effect site-specific protein modifications. Since a large excess of **5** or **6** relative to the protein were used during labeling experiments, the self-coupling reaction between the azide and alkyne groups from two GFP_{UV}(**2**+**4**) molecules was prevented. No GFP_{UV}(**2**+**4**) dimer was observed after the reactions.

GFP_{UV}(**2**+**4**) in 100 mM sodium phosphate buffer (pH 7.6) was concentrated to 4 mg/mL. To an aqueous sodium ascorbate solution (100 mM, 6 μL, 0.6 μmol) was added

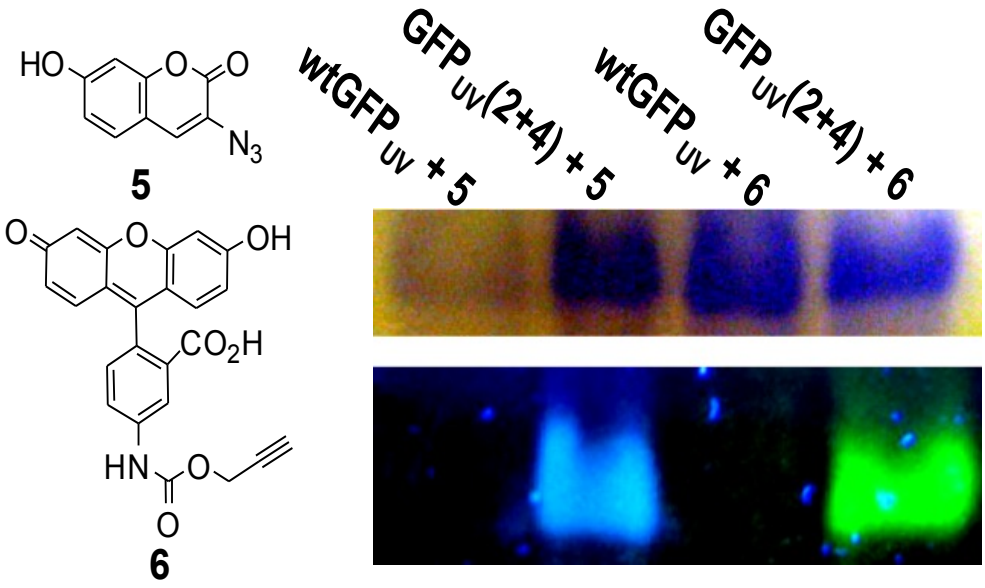


Figure 61. Labeling wtGFP$_{UV}$ and GFP$_{UV}$(**2**+**4**) with **5** and **6**. The top panel shows Coomassie blue stained proteins in a SDS-PAGE gel. The bottom panel shows fluorescent imaging of the same gel under UV light (365 nm). The image shows real colors captured by a regular camera. The faint top left protein band was due to the partial precipitation of wtGFP$_{UV}$ during the labeling reaction.

aqueous sulfonated bathophenanthroline sodium salt solution (100 mM, 8 μL, 0.8 μmol) followed by copper sulfate (100 mM, 5 μL, 0.5 μmol), and the mixture was incubated at room temperature for 5 min. Compound **5** (10 mM in CH_2Cl_2, 20 μL, 0.2 μmol) or **6** (10 mM in CH_2Cl_2, 20 μL, 0.2 μmol) was deposited in a 0.5 mL centrifuge tube and the solvent was evaporated by gentle air blowing. GFP_{UV}(**2**+**4**) (4 mg/mL, 10 μL, 1.4 nmol) was then added, followed by the addition of the above catalyst (2 μL each). Mixture was reacted at **4** was reacted at 4 degree for 2 h to give clear solution which was then directly loaded onto 12% SDS-PAGE for analysis.

4.2.8 Expression and purification of GFP_{UV}(**4**+**12**) and GFP_{UV}(**2**+**13**)

To generalize the method, we tested the feasibility of using other evolved PylRS-pylT$_{UUA}$ and *Mj*TyrRS-$_{Mj}$tRNA$^{Tyr}_{CUA}$ pairs to genetically incorporate their cognate NAAs into one protein in *E. coli*. We replaced the PylRS gene in the pPylRS-pylT-GFP1TAG149TAA plasmid with the AcKRS gene[100] and then transformed *E. coli* BL21 cells with the modified plasmid together with pEVOL-AzFRS. Growing the transformed cells in 2YT medium supplemented with 1 mM **4** and 5 mM AcK led to full-length GFP_{UV} expression with a yield of 6.4 mg/L (Figure 62). Similarly we also replaced AzFRS in the pEVOL-AzFRS with an evolved *Mj*TyrRS (sTyrRS)[104] that is specific for *O*-sulfo-L-tyrosine (sTyr) and used the resulting plasmid and pPylRS-pylT-GFP1TAG149TAA to transform *E. coli* BL21 cells. Growing cells in 2YT medium supplemented with 1 mM sTyr and 1 mM **2** led to full-length GFP_{UV} expression with a yield of 0.4 mg/L (Figure 62). The low GFP_{UV} expression yield in this case is likely due to the low efficiency of the evolved sTyrRS.[104] In both cases, no full-length GFP_{UV} was

expressed when only one NAA was provided in the media. These results indicate that our method can be generally applied to combine any two evolved PylRS-pylT$_{UUA}$ and *Mj*TyrRS-*Mj*tRNA$^{Tyr}_{CUA}$ pairs to genetically incorporate their cognate NAAs into a single protein in *E. coli*.

GFP$_{UV}$ incorporated with **4** at position 1 and **12** at position 149 (GFP$_{UV}$(**4**+**12**); **12**: AcK) and GFP$_{UV}$ incorporated with **13** at position 1 and **2** at position 149 (GFP$_{UV}$(**2**+**13**); **13**: sTyr) were expressed by using two other sets of plasmids. GFP$_{UV}$(**4**+**12**) was expressed in *E. coli* BL21 cells transformed with pAcKRS-pylT-GFP1TAG149TAA together with PEVOL-AzFRS. The transformed cells were grown in 2YT medium (50 mL) at 37 degree overnight, and the culture was subsequently inoculated into 150 mL of 2YT containing 100 μg/mL ampicillin and 25 μg/mL kanamycin. The GFP$_{UV}$ expression was induced with the addition of 500 μg/mL IPTG, 0.02% arabinose, 1 mM **4**, and 5 mM **12** when the OD$_{600}$ reached 0.6. The culture was then let grown at 37 degree for 8 h. Cells were harvested and the proteins were purified using the same procedures as described above. GFP$_{UV}$(**2**+**13**) was similarly expressed in *E. coli* BL21 cells transformed with pPylRS-pylT-GFP1TAG149TAA together with PEVOL-sTyrRS in the presence of 1 mM **2** and 2 mM **13**.

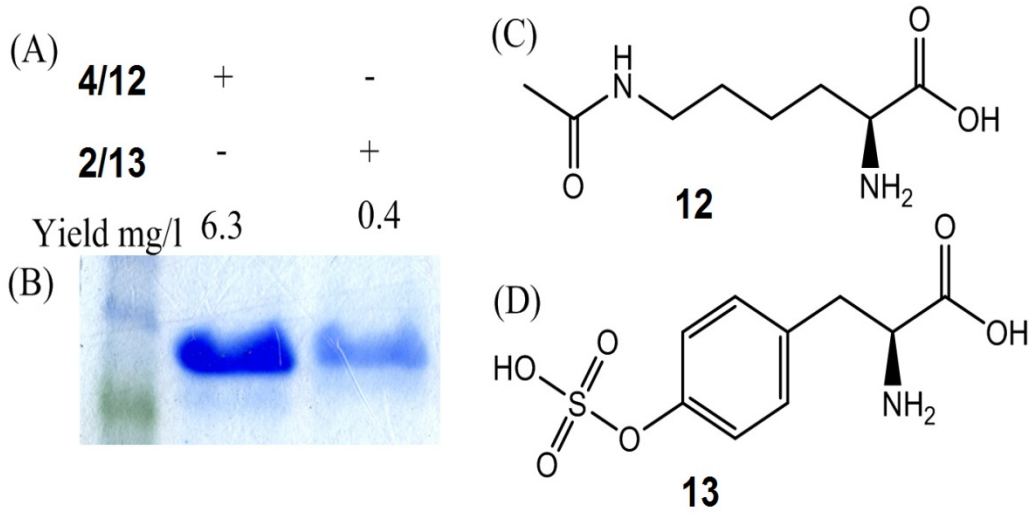


Figure 62. (**A**) Expression yield for GFP_{UV}(**4**+**12**) and GFP_{UV}(**2**+**13**). (**B**) SDS-PAGE gel, (**C**) structure of AcK (**12**) and (**D**) structure of sTyr (**13**). Staining was effected by Coomassie blue.

4.2.9 Unpublished data

4.2.9.1 DNA and protein sequences of sfGFP

DNA sequence:

5’-atggttcatcatcatcatcatagcaaaggtgaagaactgtttaccggcgttgtgccgattctggtggaactggatggtga
tgtgaatggccataaatttagcgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgaccctgaaatttatttgcacc
accggtaaactgccggttccgtggccgaccctggtgaccaccctgacctatggcgttcagtgctttagccgctatccggatcata
tgaaacgccatgatttctttaaaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcaccta
taaaacccgtgcggaagttaaatttgaaggcgataccctggtgaaccgcattgaactgaaaggtattgattttaaagaagatggc
aacattctgggtcataaactggaatataatttcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagc
gaactttaaaatccgtcacaacgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatgg
cccggtgctgctgccggataatcattatctgagcacccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggt
gctgctggaatttgttaccgccgcgggcattacccacggtatggatgaactgtataaaggcagc-3’

Protein sequence:

MVHHHHHHSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKF
ICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISF
KDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITA
DKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLS
KDPNEKRDHMVLLEFVTAAGITHGMDELYKGS

4.2.9.2 Plasmids introduction

pBAD-sfGFP134TAG: Plasmid pBAD-sfGFP134TAG that has a sequence-optimized superfolder GFP (sfGFP) gene with an amber mutation at N134 and a 6 × His

tag at its C terminus was a kind gift from Dr. Ryan Mehl of Franklin & Marshall College.

pETtrio-pylT(UUA)-PylRS-MCS: This plasmid was derived from pPylRS-pylT-GFP1TAG149TAA and contains a $tRNA^{Pyl}_{UUA}$ gene (a C34U form of $tRNA^{Pyl}_{CUA}$) under control of the *lpp* promoter and the *rrnC* terminator, the wild type *Methanosarcina mazei* PylRS gene under control of the *glnS* promoter and terminator, and multiple cloning sites under control of the T7 promoter and terminator.

Three plasmids that vary at anticodon of $tRNA^{Pyl}$ and have different nonsense mutations at N134 of the sfGFP gene were derived from pETtrio-pylT(UUA)-PylRS-sfGFP134TAG. Constructions of these plasmids were carried out using a site-directed mutagenesis protocol that was based on Phusion DNA polymerase. In brief, two oligonucleotide primers, one of which covers the mutation site and contains mutagenized nucleotides were used to amplify the whole plasmid of pETtrio-pylT(UUA)-PylRS-sfGFP134TAG to give a blunt-end PCR product. This PCR product was phosphorylated by T4 PNK and then ligated to itself using T4 DNA ligase. Primers pylT-F with a sequence of 5'-GTC CAT TCG ATC TAC ATG ATC AGG TT-3' and pylT-TAG-R with a sequence of 5'-TCT AAA TCC GTT CAG CCG GGT TAG-3' were used to make pETtrio-pylT(CUA)-PylRS-sfGFP134TAG that has a $tRNA^{Pyl}_{CUA}$ gene and a sfGFP gene with an amber mutation at N134. Primers sfGFPN134-F with a sequence of 5'-GGC AAC ATT CTG CAT AAA CTG GA-3'and sfGFPN134TGA-R with a sequence of 5'-TTA TTC TTT AAA ATC AAT ACC TTT CAG TTC AAT GC-3' were used to

make pETtrio-pylT(UUA)-PylRS-sfGFP134TAA that has a $tRNA^{Pyl}_{UUA}$ gene and a sfGFP gene with an ochre mutation at N134. Two set of primers: (1) pylT-F and pylT-TGA-R with a sequence of 5'-TTC AAA TCC GTT CAG CCG GGT TAG-3' and (2) sfGFPN134-F and sfGFPN134TGA-R with a sequence of 5'-TCA TTC TTT AAA ATC AAT ACC TTT CAG TTC AAT GC-3' were used to run two consecutive site-directed mutagenesis reactions to obtain plasmid pETtrio-pylT(UCA)-PylRS-sfGFP134UGA that contains a $tRNA^{Pyl}_{UCA}$ gene and a sfGFP gene with a opal mutation at N134. Two sets of primers (1) pylT-F and pylT-AGGA-R with a sequence of 5'-TTC CTA ATC CGT TCA GCC GGG TTA G-3' and (2) sfGFPN134-F and sfGFPN134AGGA-R with a sequence of 5'-TCC TTT CTT TAA AAT CAA TAC CTT TCA GTT CAA TGC-3' were also used to run two consecutive site-directed mutagenesis reactions on pETtrio-pylT(UUA)-PylRS-sfGFP134TAG to generate plasmid pETtrio-pylT(UCCU)-sfGFP134AGGA that contains a $tRNA^{Pyl}_{UCCU}$ gene with a quadruple UCCU anticodon and a sfGFP gene with a quadruple AGGA mutation at N134.

4.2.9.3 Amber, opal, and ochre, AGGA suppression analysis of the PylRS-pylT pairs

To demontrate amber suppression efficiency of the PylRS-$tRNA^{Pyl}_{CUA}$ pair, *E. coli* BL21 cells transformed with pETtrio-pylT(CUA)-PylRS-sfGFP134TAG were used to express sfGFP in the absence or presence of 5 mM BocK, a substrate of PylRS. Without BocK in the growth medium, only basal level expression of full-length sfGFP was observed. On the contrary, the addition of BocK promoted the overexpression of full-length sfGFP (Figure 63). This clearly indicates that the PylRS-pylT pair is an efficient

amber suppressing aaRS-tRNA pair and its mediated suppression is dependent on the presence of a NAA.

It has been previously demonstrated that $tRNA^{Pyl}$ is not hardwired for recognizing an amber codon. Mutagenizing the anticodon of $tRNA^{Pyl}$ from CUA to other triplet anticodons does not significantly alter the acylation potential of $tRNA^{Pyl}$ by PylRS. As expected, mutating C34 in $tRNA^{Pyl}_{CUA}$ to U34 in $tRNA^{Pyl}_{UUA}$ did not significantly change the orthogonality of the tRNA.

Although expected to be orthogonal, $tRNA^{Pyl}_{UCA}$ that is a C34U/U35C mutant form of $tRNA^{Pyl}_{CUA}$ is apparently not orthogonal in *E. coli*. Cells transformed with pETtrio-pylT(UCA)-PylRS-sfGFP134TGA exhibited a very high expression level of full-length sfGFP both in the absence and in the presence of BocK in the media (Figure 63).

The AGGA suppression is also very interesting. $tRNA^{Pyl}_{UCCU}$ mutant form $tRNA^{Pyl}_{CUA}$ is apparently not orthogonal in *E. coli*. Cells transformed with pETtrio-pylT(UCCU)-PylRS-sfGFP134AGGA exhibited a very high expression level of full-length sfGFP both in the absence and in the presence of BocK in the media (Figure 64).

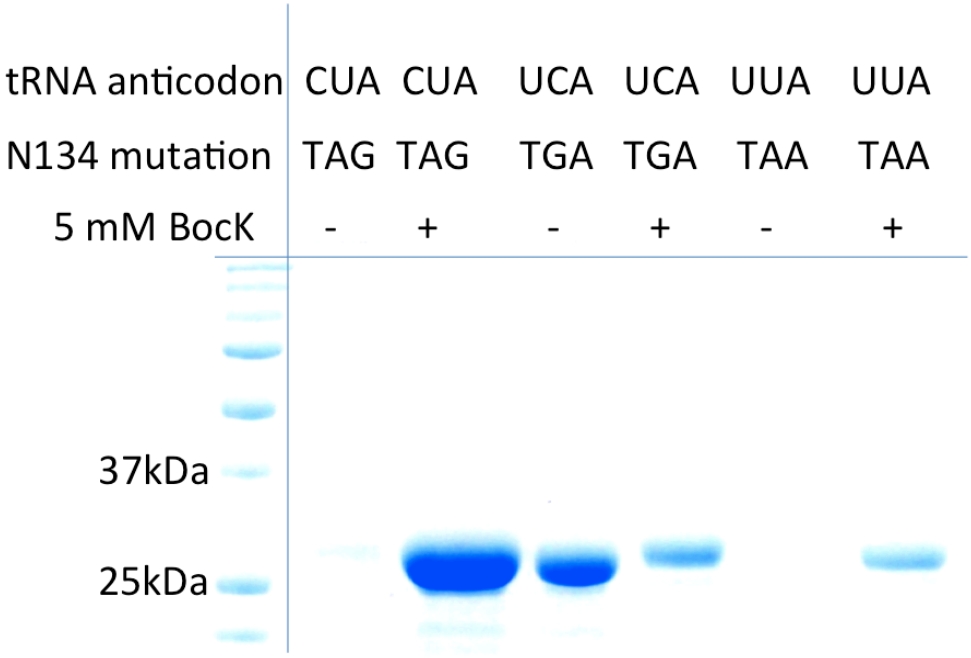


Figure 63. Suppression of amber, opal, and ochre mutations. Suppression of amber, opal, and ochre mutations at N134 of sfGFP by their corresponding suppressing PylRS-$tRNA^{Pyl}$ pairs in the absence and presence of BocK. Proteins shown in the gel represent their really expression levels. The first two lanes show sfGFP expression levels in cells transformed with pETtrio-PylT(CUA)-PylRS-sfGFP134TAG; the second two lanes show sfGFP expression levels in cells transformed with pETtrio-PylT(UCA)-PylRS-sfGFP134TGA; the last two lane show sfGFP expression levels in cells transformed with pETtrio-PylT(UUA)-PylRS-sfGFP134TAA.

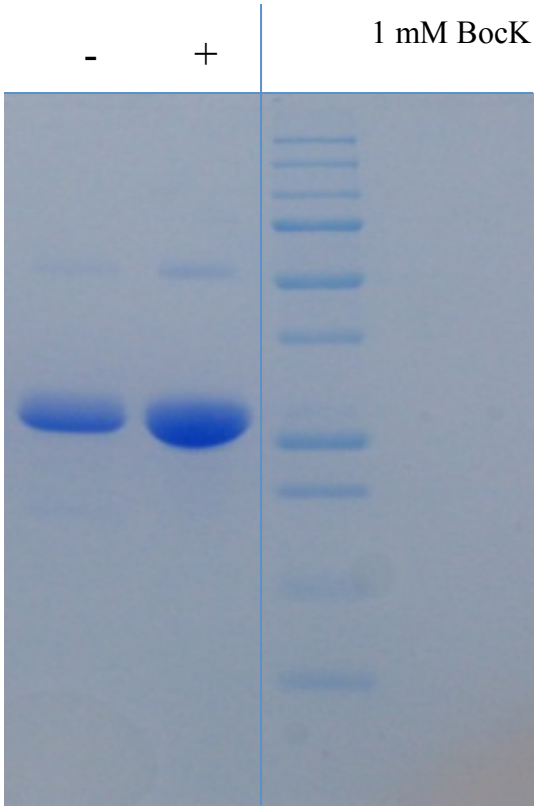


Figure 64. $PylT_{UCCU}$ is not orthogonal in *E. coli*. Expression of sfGFP in cells transformed with pETtrio-pylT(UCCU)-sfGFP134AGGA and grown in the presence or absence of 5 mM BocK.

4.3 Summary

In summary, we have developed a facile system for genetic incorporation of two different NAAs at two defined sites of a single protein in *E. coli* with moderate to high protein production yields. This technique will greatly expand the scope of potential applications for the genetic NAA incorporation approach, and it can be applied to install a FRET pair to a protein for conformation and dynamics studies, synthesize proteins with two different post-translational modifications for functional analysis, or generate phage-displayed peptide libraries with the expanded diversity of the displayed peptides.

5. A STRAIGHTFORWARD ONE PLASMID SELECTION SYSTEM FOR EVOLUTION OF AMINOACYL-TRNA SYNTHETASES

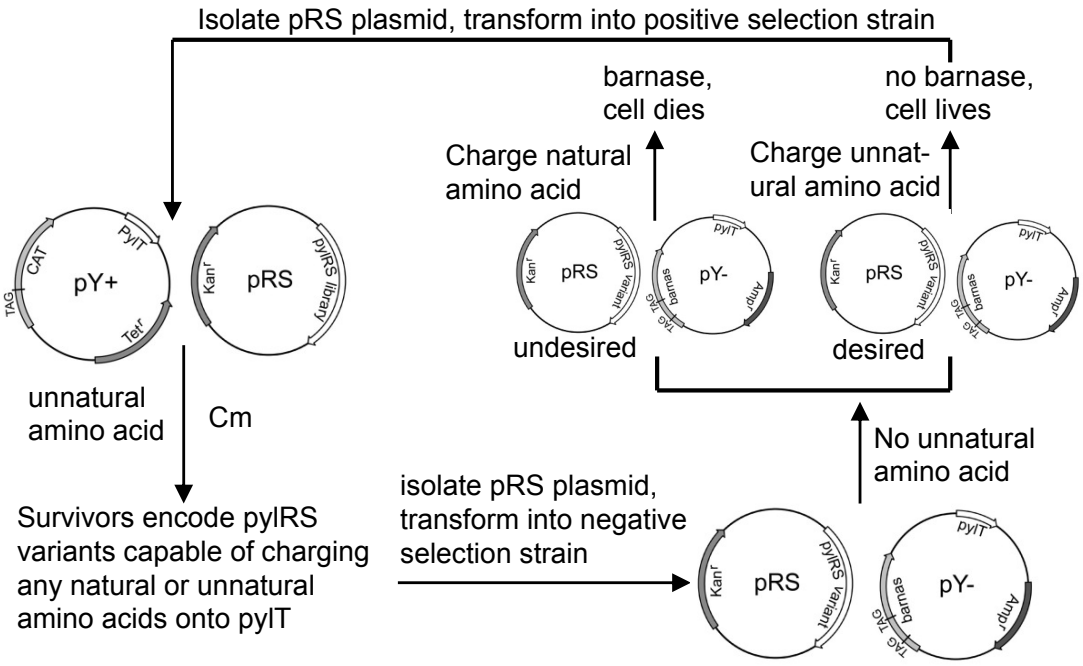
5.1 Introduction

Given the importance of proteins in biology, it's desirable to be able to manipulate proteins for understanding of their structures and functions, such as protein folding, protein-protein interactions, protein localization, and protein-involved biological processes, etc. The ability to genetically encode NAAs into proteins would enable site-specific modification of proteins *in vivo* and *in vitro*, thus providing novel tool for understanding of protein functions. One of the most popular methods is the genetic incorporation of NAAs directly into proteins during translation in living cells, which was developed by Schultz et al.[76-82] To date, several orthogonal aaRS/tRNA pairs, such as evolved *Mj*TyrRS-$Mj\text{tRNA}^{\text{Tyr}}_{\text{CUA}}$ and PylRS-pylT pairs, have been introduced to live cells to facilitate NAA incorporation. More than 70 NAAs have been successfully incorporated into proteins in *E. coli*.[87, 100, 103, 115-119, 136]

To evolve an orthogonal aaRS/tRNA pair in *E. coli*, the most critical step is the redesign and evolution of the amino acid binding pocket of the existed orthogonal aaRS. A library of aaRS variants carrying random mutations is subjected to several rounds of positive and negative selections. Nowadays, several positive and negative selections were developed, such as most commonly used antibiotic resistance genes[126], greens fluorescence protein[137], barnase[76], and acetyltransferase/uracil phosphoribosyltranserease (*cat-uprt*) [138].

In our previous work, we have been able to develop an evolved pylRS/pylT pair, which could specifically take a photocaged *N*-*ε*-methyl-L-lysine. To select PylRS variants specific for different NAAs in *E. coli*, we have adopted an existing positive and negative selection scheme developed by Schultz et al.[103, 126] As shown in Scheme 4, the positive selection is based on resistance to chloramphenicol, which was conferred by the suppression of an amber mutation at a permissive site, Q107, in the chloramphenicol acetyltransferase (CAT) gene. PylRS variants can either acylate pylT with the NAA or any natural amino acid will grow on the plate with antibiotics chloramphenicol. Negative selection utilizes the toxic barnase gene with amber mutations at two permissive sites, Q2 and D44, and is carried out in the absence of the NAA. Growth in the presence of arabinose leads to the death of cells in which PylRS variants can acylate pylT with any natural amino acid. Only PylRS variants that could acylate pylT with the NAA but not with the endogenous amino acids could survive both selections. Additional cycles of positive and negative selections can be carried out to increase the specificity of PylRS variants toward individual NAA (Scheme 4).

Scheme 4. Two-Plasmid Selection System



Effective as it is, the current selection system has several disadvantages. First, the surviving pool of aaRS plasmids must be isolated from selective plasmid after each step, and be transformed to the new host cells with high competence. These reoccurring plasmid isolation and transformation procedures require significant laboratory labor work and resources as well as time investment. Due to the lethalness of the negative selection marker barnase, the flexibility of the system is restricted. One plasmid selection system using chloramphenicol acetyltransferase/uracil phosphoribosyltranserease (*cat-uprt*) fusion gene was developed by Schutlz et al. Herein, we have developed a series of dual positive/negative selection vectors using different positive and negative selection markers.

5.2 Plasmid construction: positive selection marker

To build a series of dual plasmid, I started with construct the two-plasmid selection system. The positive selection marker pY+ (Figure 65) contains an amber suppressor tRNA (pylT) and the positive selection marker chloramphenicol acetyltransferase (CAT) with an amber stop codon (D112TAG). In the presence of chloramphenical (cm) and provided NAA, the positive surviving aaRS variants are capable of incorporating either the NAA or endogenous amino acids.

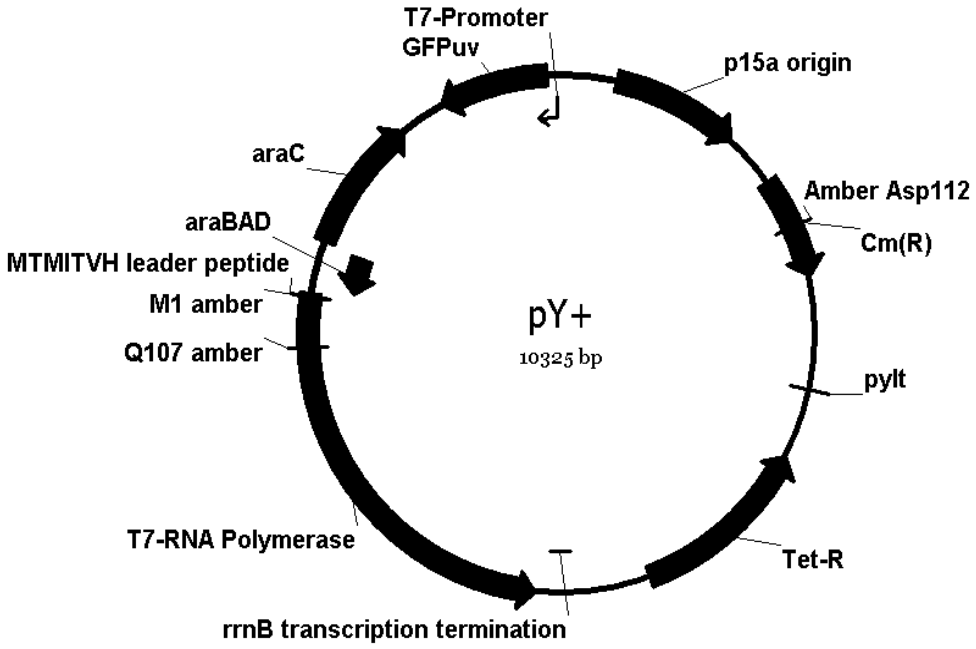


Figure 65. Plasmid structure of pY+.

5.2.1 Plasmid construction of pY+

Plasmid pY+ is used to carry out positive selection. It was derived from pRep,[115] the gene of $tRNA_{CUA}$ for *Mj*TyrRS was replaced by pylT, $tRNA_{CUA}$ for *Mm*PylRS. 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,[139] using two oligodeoxynucleotide primers (5'-CCCATCAAAAAAATATTCTCAACATAAAAAACTTTG-3' and 5'-CGGGACAGGCTGACAACCCGAGGATCT-3'). pylT-*lpp* franking gene with *EcoR1* at the 5' end was amplified by two oligodeoxyuncleotide primers (5'-CATCCGGAATTCCGTATGGCAATGAAAG-3' and 5'-CCCATCAAAAAAATATTCTCAACATAAAAAACTTTG-3'), using Prep as template. pylT-rrnc franking gene with *Eagl* at the 3' end was amplified by two oligodeoxynucleotide primers (5'-GGTTGTCAGCCTGTCCCGGCTTGGCACTGGCCGTCG-3' and 5'-GCATGGCGGCCGACGCGCTGGGC-3'). Then the gene of pylT that was annealed with these two fragments was digested by *EcoR1* and *Eag1* restriction enzymes, and then cloned into the predigested pRep to afford pY+.

5.3 Plasmid construction: negative selection marker

The negative selection plasmid, pY-, bears the pylT gene and the barnase gene with two amber mutations (Q2TAG and D44TAG). The barnase gene is under control of a pBad promoter (Figure 66). The toxicity of negative selection marker eliminates the aaRS variants that recognize endogenous amino acids. In the presence of negative

selection marker and absence of NAA, the aaRS variants taking endogenous amino acids won't survive.

5.3.1 Plasmid construction of pY-

Plasmid pY- is used to do positive selection. It was derived from pNeg,[137] the gene of $tRNA_{CUA}$ for *Mj*TyrRS was replaced by pylT, $tRNA_{CUA}$ for *Mm*PylRS. 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, using two oligodeoxynucleotide primers (5'-CCCATCAAAAAAATATTCTCAACATAAAAAACTTTG-3' and 5'-CGGGACAGGCTGACAACCCGAGGATCT-3'). pylT-rrnc franking gene with *Nde1* at the 5' end was amplified by two oligodeoxynucleotide primers (5'-GATATACATATGGCATAGGTTATCAACACG-3' and 5'-GTTGTCAGCCTGTCCCGTCGTTTTACAACGTC-3'), using Prep as template. pylT-*lpp* franking gene with *Pst1* at the 3' end was amplified by two oligodeoxynucleotide primers (5'-CCCATCAAAAAAATATTCTCAACATAAAAAACTTTG-3' and 5'-CGATGCCTGCAGCAATGGCAACAACGTTG-3'). Then the gene of *pylT* that was annealed with these two fragments was digested by *Nde1* and *Pst1* restriction enzymes, and cloned into the predigested pNeg to afford pY-.

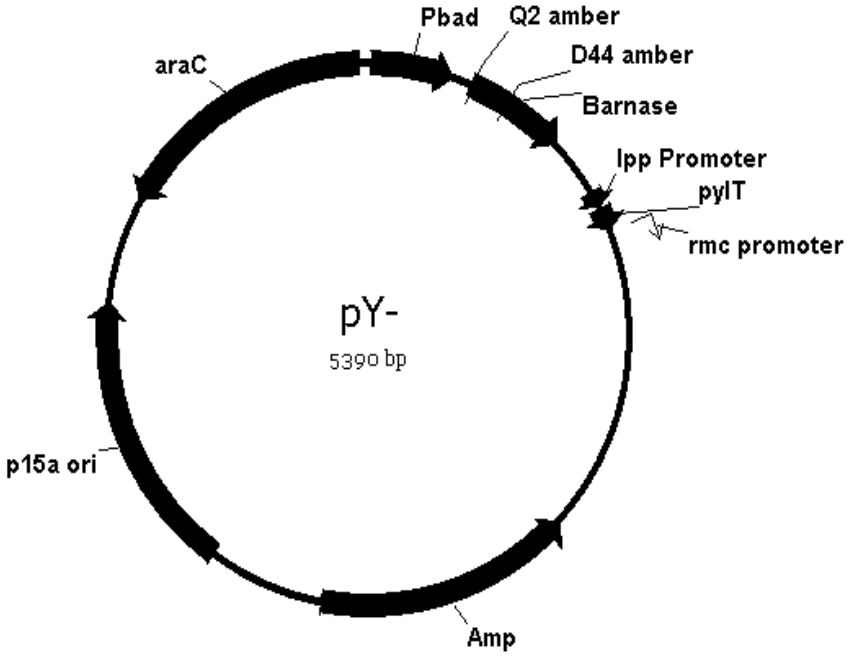


Figure 66. Plasmid structure of pY-.

5.3.2 DNA and protein sequences of barnaseQ2TAGD44TAG

DNA sequence:

5’-atggcataggttatcaacacgtttgacggggttgcggattatcttcagacatatcataagctacctgataattacattacaaa
atcagaagcacaagccctcggctgggtggcatcaaaagggaaccttgcataggtcgctccggggaaaagcatcggcgga
gacatcttctcaaacagggaaggcaaactcccgggcaaaagcggacgaacatggcgtgaagcggatattaactatacatc
aggcttcagaaattcagaccggattctttactcaagcgactggctgatttacaaaacaacggaccattatcagacctttacaaa
aatcagataa-3’

Protein sequence:

MA*VINTFDGVADYLQTYHKLPDNYITKSEAQALGWVASKGNLA*VAPGKSI
GGDIFSNREGKLPGKSGRTWREADINYTSGFRNSDRILYSSDWLIYKTTDHYQ
TFTKIR

5.4 Pduel **1**: pNeg-barnaseQ2TAGQ3TAGD44TAG-Cm(R)

In order to simplify the library selection process, we tried to develop a single plasmid that contains both positive and negative selection markers. We first took advantage of the current positive and positive plasmids, by integrating the positive and negative selection markers into pY- plasmid. The first generation of single selection plasmid, Pduel **1**, bears the pylT gene and the barnase gene with three amber mutations (Q2TAG, Q3TAG and D44TAG) and chloramphenicol acetyltransferase (CAT) with an amber stop codon (D112TAG) (Figure 67). The reason to put one more amber stop codon into barnase is because the toxicity is still too high for effective negative selection with two amber codons from previous study. For the positive selection, the amber stop codon in CAT will be suppressed in the presence of NAAs or

misincorporated natural amino acids. On the contrary, for the negative selection, the barnase gene, under the control of pBad, will be expressed and lead to cell death in the presence of misincorporated natural amino acids and 0.2% of arabinose.

5.4.1 Plasmid construction of Pduel **1**

Plasmid Pduel 1 was derived from pY-. BarnaseQ2TAGD44TAG was replaced by barnaseQ2TAGQ3TAGD44TAG. BarnaseQ2TAGQ3TAGD44TAG was amplified from pY- by two oligodeoxynucleotides (5'-GGAGATATACATATGGCATAGTAGATCAACACGTTTGAC-3', and 5'-AATGGTGCATGCTTATCTGATTTTTGTAAAG-3'). Two restriction sites *Nde1* at 5' head and *Sph1* at 3' tail were designed into the synthesized DNA that was then cloned into pY- digested with *Nde1* and *Sph1* to make pNeg-barnaseQ2TAGQ3TAGD44TAG. Cm(R)D112TAG with its promoter and terminator was amplified from pY+ by two oligodeoxynucleotides (5'-GCGGCCAAGCTTTGATGTCCGGCGGTGC-3', and 5'-GCGCCAAGCTTACTTATTTTTGATCGTT-3'), and cloned into predigested pNeg-barnaseQ2TAGQ3TAGD44TAG by *Hind111* at both 5'head and 3' tail to afford Pduel **1**.

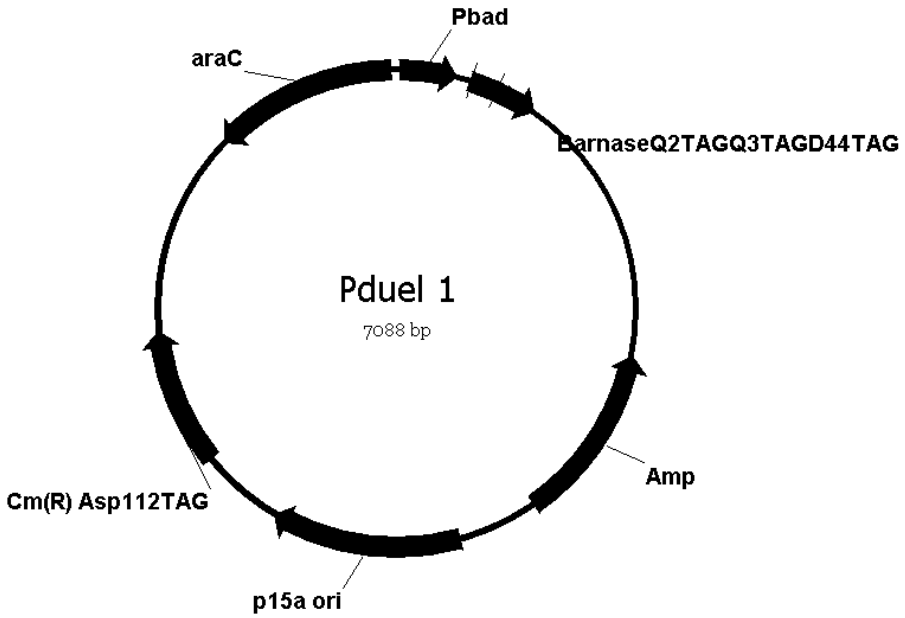


Figure 67. Plasmid structure of Pduel **1**.

5.4.2 The activity test of Pduel **1**

To demonstrate the system applicability, Pduel **1** was used together with the pBK-*Mm*PylRS containing wild-type PylRS gene to transform *E. coli* Top10 strain. For positive selection, the Top 10 cells were let grown on the GMML plates, with 1% of glucose, and 0.3 mM leucine. The minimal media GMML was used due to the presence of large amount of natural amino acids, which is too rich for positive selection. 100 μg/mL of ampicillin and 25 μg/mL of kanamycin were provided to force cells containing Pduel **1** and pBK-*Mm*PylRS. 34 μg/mL or 68 μg/mL of chloramphenicol were used with and without 1 mM BocK, which could be applied to suppress ambor codon by *Mm*PylRS. Without chloramphenicol, cells were healthy.

However, in the presence of chloramphenicol and 1 mM BocK, cells were all dead. It indicated either the positive selective marker, Cm(R) inside Pduel **1**, didn't function well or the negative selective marker, barnase, was too toxic for cells. For the negative selection, the Top 10 cells were let grown on LB plates, with 100 μg/mL of ampicillin and 25 μg/mL of kanamycin to keep cells maintaining Pduel **1** and pBK-*Mm*PylRS. 5% of glucose or 0.2% of arabinose was provided, with or without of 1 mM BocK. For the Pduel **1**, glucose and arabinose were used to inhibit or promote the expression of barnase. Without BocK, cells were healthy. With BocK, no matter whether glucose or arabinose was present, cells were all dead (Table 3). It indicated the barnase, even with three amber codons, was still too toxic. It also interfered with the positive selection, leading to the failure of positive selection undetermined.

Table 3. Activity Test of Pduel 1

Positive (GMML) 100 μg/mL amp 25 μg/mL kan 0.3 mM Leucine	1% Glucose	+	+	+
	34 or 68 μg/mL cm	-	+	+
	1 mM BocK	-	-	+
		Grow	**Die**	**Die**
Negative (LB) 100 μg/mL amp 25 μg/mL kan	5% Glucose	-	+	-
	0.2% arabinose	-	-	+
	1 mM BocK	-	+	+
		Grow	**Die**	**Die**

5.5 Pduel **2**: pNeg-barnaseQ2TAGQ3TAGK27AD44TAG-Cm(R)

Due to the high toxicity of barnaseQ2TAGQ3TAGD44TAG, another modification was made to barnase. The mutation K27A to barnase was reported previously to great reduce its toxicity.[140] BarnaseQ2TAGQ3TAGK27AD44TAG was further made to replace barnaseQ2TAGQ3TAGD44TAG in Pduel **1** to afford pDuel **2** (Figure 68).

5.5.1 Plasmid construction of Pduel **2**

Plasmid Pduel **2** was derived from Pduel **1**. Using Pduel **1** as template, barnase Q2TAGQ3TAGK27AD44TAG was generated by site-directed mutagenesis using four oligodeoxynucleotides (5'-GGAGATATACATATGGCATAGTAGATCAACACGTTTGAC-3', 5'-GATAATTACATTACAGCATCAGAAGCACAAGCC-3', 5'-GGCTTGTGCTTCTGATGCTGTATGTAATTATC-3', and 5'-AATGGTGCATGCTTATCTGATTTTTGTAAAG-3'), and then cloned into digested Pduel **1** by two restriction sites *Nde1* at 5' head and *Sph1* at 3' tail to make Pduel **2**.

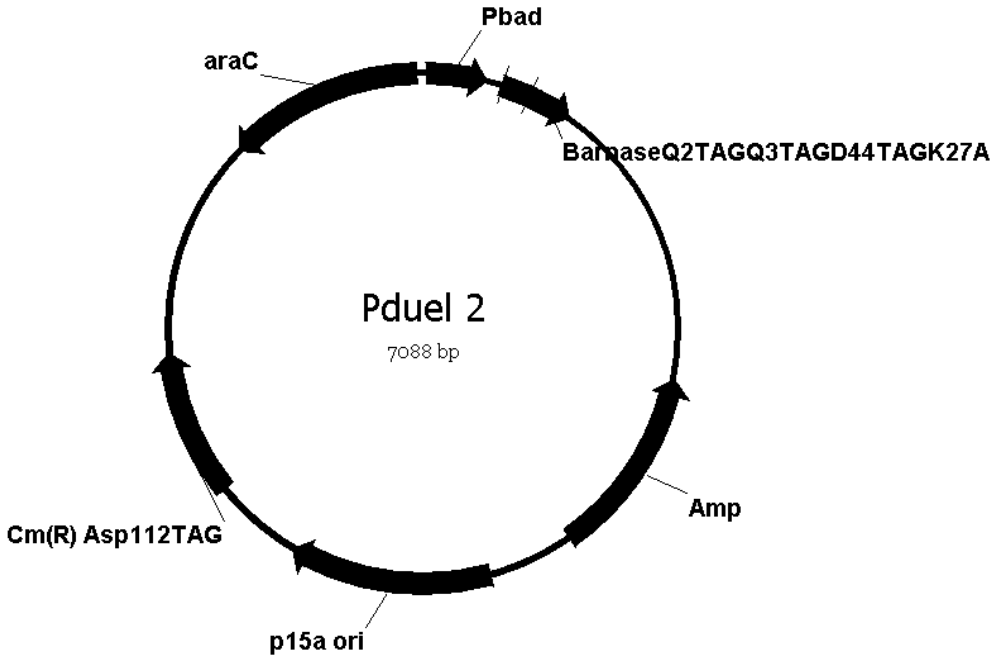


Figure 68. Plasmid structure of Pduel **2**.

5.5.2 The activity test of Pduel **2**

Same as Pduel **1**, Pduel **2** and pBK-*Mm*PylRS were used to cotransform *E. coli* Top10 strain. For positive selection, the Top 10 cells were let grown on the LB plates, with 1% of glucose, 100 μg/mL of ampicillin, and 25 μg/mL of kanamycin. Only 34 μg of chloramphenicol were used with and without 1 mM BocK. As for the result, without chloramphenicol, cells were healthy. In the presence of chloramphenicol and 1 mM BocK, only a few cells survived. For the negative selection, the Top 10 cells were let grown on LB plates. 5% of glucose or 0.2% of arabinose was provided, with or without of 1 mM BocK. With or without BocK, cells were survived. With 1 mM BocK and 0.2% arabinose, cells were grown but unhealthy (Table 4). It indicated the mutation of K29A to barnase, greatly decreased the toxicity of it. The toxicity of barnase, with three amber codons and K29A mutation, was too weak to kill cells.

Table 4. Activity Test of Pduel 2

Positive (LB) 100 μg/mL amp 25 μg/mL kan	1% Glucose	+	+	+
	34 μg/mL cm	-	+	+
	1 mM BocK	-	-	+
		Grow	**Die**	**Die**
Negative (LB) 100 μg/mL amp 25 μg/mL kan	5% Glucose	-	+	-
	0.2% arabinose	-	-	+
	1 mM BocK	-	+	+
		Grow	**Grow**	**Grow**

5.6 Pduel **3**: pNeg-barnaseQ2TAGK27AD44TAG-Cm(R)

Due to the low toxicity of barnaseQ2TAGQ3TAGK27AD44TAG, the Q3TAG was mutated back to Q3. Therefore, a new plasmid Pduel **3** was constructed to better control the toxicity of barnase (Figure 69).

5.6.1 Plasmid construction of Pduel **3**

Plasmid Pduel **3** was derived from Pduel **2**. Using Pduel **2** as template, barnaseQ2TAGK29AD44TAG was amplified by two oligodeoxynucleotides (5'-GATATACATATGGCATAGGTTATCAACACG-3', and 5'-AATGGTGCATGCTTATCTGATTTTTGTAAAG-3'), and then cloned into digested Pduel **2** by two restriction sites *Nde1* at 5' head and *Sph1* at 3' tail to make Pduel **3**.

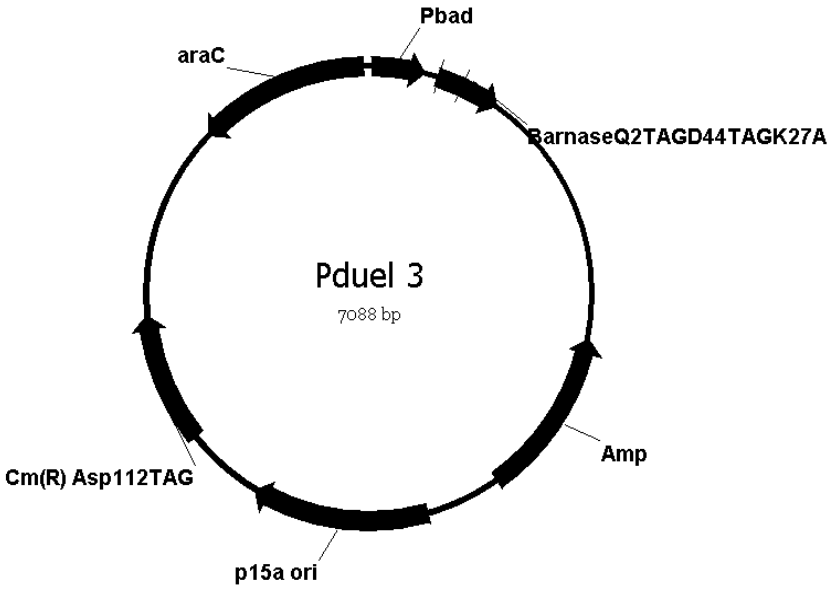


Figure 69. Plasmid structure of Pduel **3**.

5.6.2 The activity test of Pduel **3**

Pduel **3** and pBK-*Mm*PylRS were used to cotransform *E. coli* Top10 strain. For positive selection, the Top 10 cells were let grown on the LB plates, with 1% of glucose, 100 μg/mL of ampicillin, and 25 μg/mL of kanamycin. Only 34 μg of chloramphenicol were used with and without 1 mM BocK. As for the result, without chloramphenicol, cells were healthy. In the presence of chloramphenicol and 1 mM BocK, only a few cells survived. For the negative, the Top 10 cells were let grown on LB plates. 5% of glucose or 0.2% of arabinose was provided, with or without of 1 mM BocK. Without BocK, cells were healthy. With 1 mM BocK and glucose, cells were healthy. With 1 mM BocK and 0.2% of arabinose, cells were all dead (Table 5). It indicated the toxicity of barnase, with two amber codons and K29A mutation, was under good control. It also indicates the inefficiency of positive selection due to the positive selective marker itself.

Table 5. Activity Test of Pduel 3

Positive (LB) 100 μg/mL amp 25 μg/mL kan	1% Glucose	+	+	+
	34 μg/mL cm	-	+	+
	1 mM BocK	-	-	+
		Grow	**Die**	**Die**
Negative (LB) 100 μg/mL amp 25 μg/mL kan	5% Glucose	-	+	-
	0.2% arabinose	-	-	+
	1 mM BocK	-	+	+
		Grow	**Grow**	**Die**

5.7 Pduel **4**: pRep-barnaseQ2TAGK27AD44TAG

While the toxicity of barnase in Pduel **3** was under good control, the activity of Cm(R)D112TAG did not function appropriately. Since the Cm(R)D112TAG in pY+ worked, we decided to use pY+ as new scaffold to generate Pduel **4** plasmid by replacing the gene of GFP in pY+ with barnaseQ2TAGK27AD44TAG (Figure 70). In the plasmid Pduel **4**, for the positive selection, the amber stop codon in CAT will be suppressed in the presence of NAAs or misincorporated natural amino acids. On the contrary, for the negative selection, the barnase gene, under control of T7 promoter, will be produced and kill cells if the T7 polymerase is expressed. In the Pduel **4**, T7 polymerase, with two amber codons, is under the control of pBad promoter. T7 polymerase will be expressed, and induce the expression of barnase in the presence of misincorporated natural amino acids and 0.2% of arabinose.

5.7.1 Plasmid construction of Pduel **4**

Plasmid Pduel **4** was derived from pY+. The gene of GFP was replaced by gene of barnaseQ2TAGK29AD44TAG. Using Pduel **3** as template, barnase Q2TAGK29AD44TAG was amplified by two oligodeoxynucleotides (5'-TTTCCCTCTAGAGAAATAATTTTGTTTAACTTTAAGAAGG-3', and 5'-TTTGTAGAGCTCTTATCTGATTTTTGTAAAGGTCTG-3'), and then cloned into digested pY+ at two restriction sites *Nde1* and *Sac1* to make Pduel **4**.

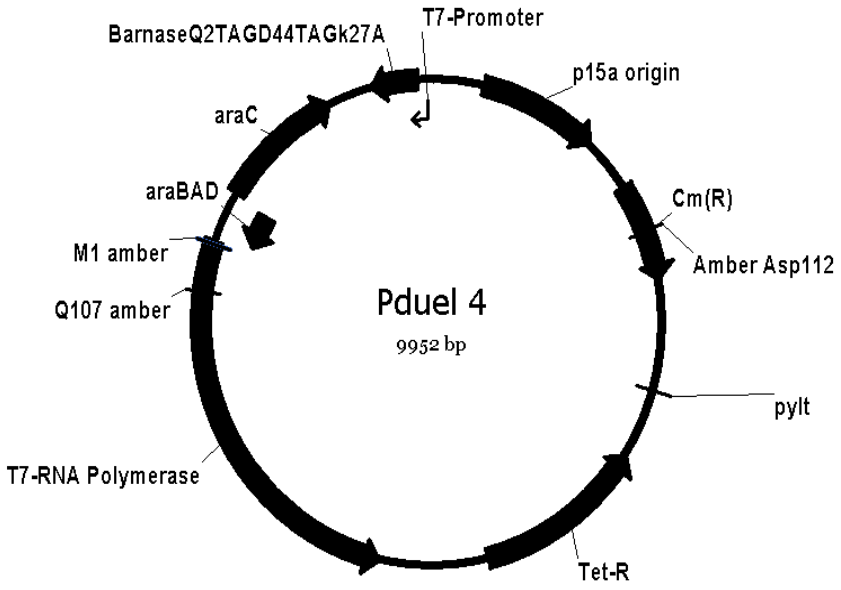


Figure 70. Plasmid structure of Pduel **4**.

5.7.2 The activity test of Pduel **4**

Pduel **4** and pBK-*Mm*PylRS were used to cotransform *E. coli* Top10 strain. The Top 10 cells were let grown on the LB plates, with 5% of glucose, 12.5 μg/mL of tetracycline, and 25 μg/mL of kanamycin. Only 34 μg of chloramphenicol were used with and without BocK. As for the result, without chloramphenicol and BocK, cells were healthy. In the presence of chloramphenicol and 0.1, 0.25, 0.5 and 1 mM of BocK, no cell was grown. With 0.1 mM BocK, many cells were grown. With 0.25 mM of BocK, a few cells were grown. With 0.5 and 1 mM of BocK, no cell was grown (Table 6). It indicated 0.25 mM of BocK was not enough to produce barnase to kill all cells in the presence of glucose. 0.1 mM of BocK would give lower toxicity of

barnase, which mean the activity of barnase inside Pduel **4** was adjustable. However, if chloramphenicol was present, no cell could survive. Therefore, we faced the dilemma. With high concentration of BocK, the toxicity of barnase would lead to cell death. With low concentration of BocK, Cm(R) could not suppress the chloramphenicol, which would also cause cell death. This was no way out for efficient positive selection at this stage.

Table 6. Activity Test of Pduel 4

34 μg/mL cm	-	+	-	+	-	+	-	+	-
5% Glucose	+	+	+	+	-	-	-	-	-
0.1 mM BocK	+	+	-	-	-	-	-	-	-
0.25 mM BocK	-	-	+	+	-	-	-	-	-
0.5 mM BocK	-	-	-	-	+	+	-	-	-
1 mM BocK	-	-	-	-	-	-	+	+	-
	Die	**Die**	**Die**	**Die**	**Die**	**Die**	**Die**	**Die**	**Grow**

5.8 Pduel **5**: pRep-barnaseQ2TAGQ2TAGK27AD44TAG

While the toxicity of barnase in Pduel **4** was still too lethal to cells, the mutation Q3TAG was added back to barnase to generate Pduel **5** (Figure 71).

5.8.1 Plasmid construction of Pduel **5**

Plasmid Pduel **5** was derived from Pduel **4**. The gene of barnaseQ2TAGK29AD44TAG was replaced by gene of barnaseQ2TAGQ3TAGK29AD44TAG. Using Pduel **4** as template, barnase Q2TAGQ3TAGK29AD44TAG was amplified by two oligodeoxynucleotides (5'-GGAGATATACATATGGCATAGTAGATCAACACGTTTGAC-3', and 5'-TTTGTAGAGCTCTTATCTGATTTTTGTAAAGGTCTG-3'), and then cloned into digested pY+ at two restriction sites *Nde1* and *Sac1* to make Pduel **5**.

5.8.2 The activity test of Pduel **5**

The activity test of Pduel **5** was confusing and therefore was not provided.

5.9 Pduel **6**: pRep-barnaseQ2TAGD44TAG

Due to the fact that the toxicity of barnase under control of T7 promoter was hard to adjust, the T7 polymerase with two amber codons was replaced by the gene of barnaseQ2TAGD44TAG (Figure 72). In the plasmid Pduel **6**, for the negative selection, the barnase gene, under the control of pBad promoter, would be expressed in the presence of 0.2% of arabinose and when its amber mutations are suppressed.

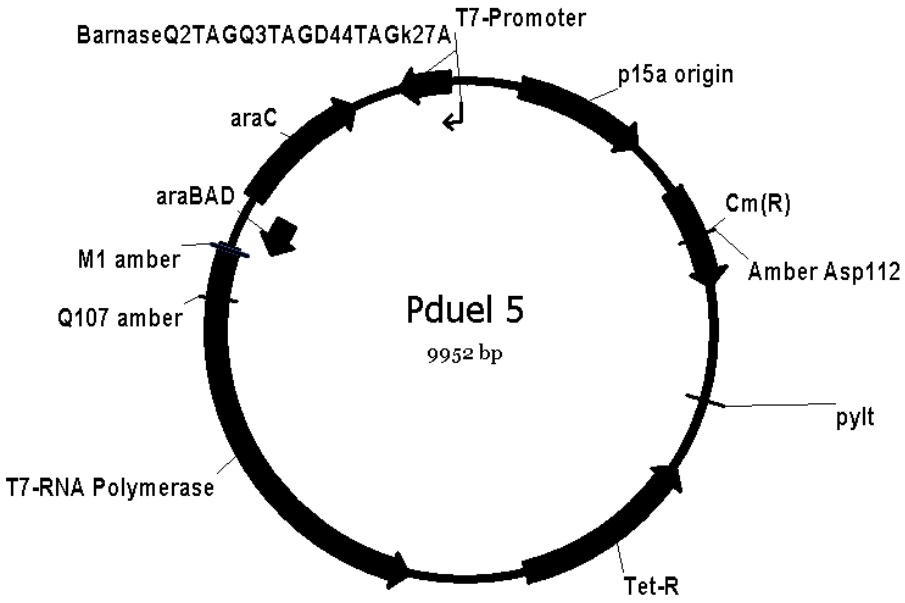


Figure 71. Plasmid structure of Pduel **5**.

5.9.1 Plasmid construction of Pduel **6**

Plasmid Pduel **6** was derived from pY+. The gene of T7 polymerase was replaced by gene of barnaseQ2TAGD44TAG. Using pY- as template, barnase Q2TAGD44TAG was amplified by two oligodeoxynucleotides (5'-GGCCGGTGCACATGGCATAGGTTATC-3', and 5'-CGGCCGACGTCTTATCTGATTTTTGTA-3'), and then cloned into digested pY+ at two restriction sites *Apal1* and *Aat11* to make Pduel **6**.

5.9.2 The activity test of Pduel **6**

Pduel **6** and pBK-mmpylRS were cotransformed into *E. coli* Top10 strain. For positive selection, the Top 10 cells were let grown on the LB plates, 12.5 μg/ml of tetracycline and 25 μg/ml of kanamycin. 68 μg/ml of chloramphenicol were used together with and without 1 mM BocK. In the presence of chloramphenicol, with 1 mM BocK, cells all survived; without BocK, cells all died. However, for the negative condition, with 0.2% of arabinose and 1 mM BocK, the suppression was not obvious was added (Table 7).

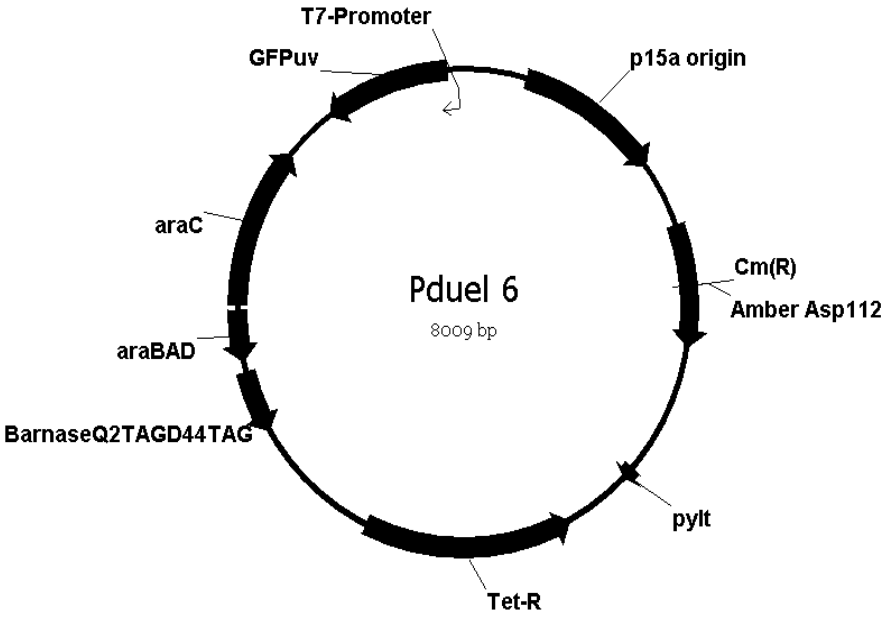


Figure 72. Plasmid structure of Pduel **6**.

Table 7. Activity Test of Pduel 6

Positive (LB) 12.5 μg/mL tet 25 μg/mL kan	68 μg/mL cm	-	+	+
	1 mM BocK	-	-	+
		Grow	**Die**	**Grow**
Negative (LB) 12.5 μg/mL tet 25 μg/mL kan	0.2% arabinose	-	-	+
	1 mM BocK	-	+	+
		Grow	**Grow**	**Inhibition**

5.10 Pduel **7**: pRep-SacB

Although we tried very hard, the toxicity of barnase in Pduel was hard to adjust. Therefore, we had to choose other negative selective marker, namely SacB. The *sacB* gene encodes the enzyme levansucrase, which catalyzes the formation of high molecular weight fructose polymers. The accumulation of these polymers in the periplasm interferes with metabolism, and causes cell death. Thus, the *sacB* gene is lethal to a gram-negative cell growing on a medium containing sucrose. Plasmid Pduel **7** was obtained by replacing the gene of GFP in pY+ with the gene of *sacB* (Figure 73). In the plasmid Pduel **6**, for the negative selection, the *sacB* gene, under control of T7 promoter, will be produced and cause cell death in the presence of sucrose. T7 polymerase, with two amber codons, is under the control of pBad promoter. T7 polymerase will be expressed, and induce the expression of SacB in the presence of 0.2% of arabinose and when its amber mutations are suppressed.

5.10.1 DNA and protein sequences of SacB

DNA sequence:

5’-atgaacatcaaaaagtttgcaaaacaagcaacagtattaacctttactaccgcactgctggcaggaggcgcaactcaag
cgtttgcgaaagaaacgaaccaaaagccatataaggaaacatacggcatttcccatattacacgccatgatatgctgcaaatc
cctgaacagcaaaaaaatgaaaaatatcaagttcctgagttcgattcgtccacaattaaaaatatctcttctgcaaaaggcctg
gacgtttgggacagctggccattacaaaacgctgacggcactgtcgcaaactatcacggctaccacatcgtctttgcattagc
cggagatcctaaaaatgcggatgacacatcgatttacatgttctatcaaaaagtcggcgaaacttctattgacagctggaaaa
acgctggccgcgtctttaaagacagcgacaaattcgatgcaaatgattctatcctaaaagaccaaacacaagaatggtcagg
ttcagccacatttacatctgacggaaaaatccgtttattctacactgatttctccggtaaacattacggcaaacaaacactgaca

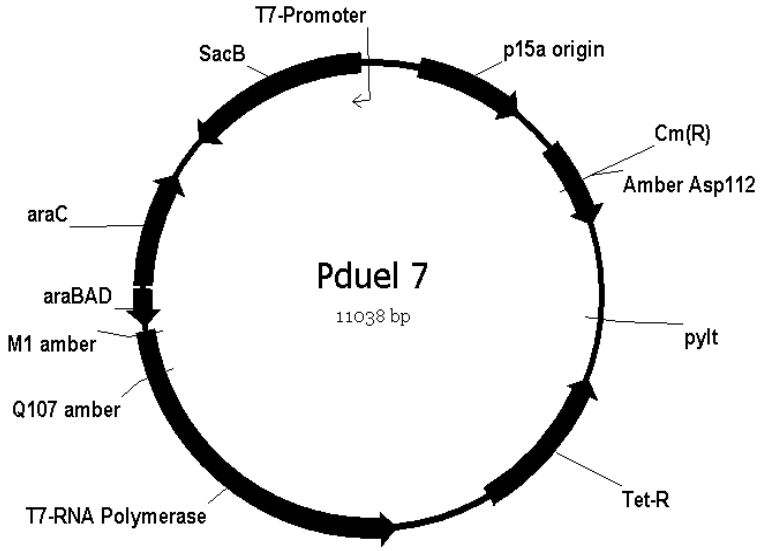


Figure 73. Plasmid structure of Pduel **7**.

actgcacaagttaacgtatcagcatcagacagctctttgaacatcaacggtgtagaggattataaatcaatctttgacggtgac
ggaaaaacgtatcaaaatgtacagcagttcatcgatgaaggcaactacagctcaggcgacaaccatacgctgagagatcct
cactacgtagaagataaaggccacaaatacttagtatttgaagcaaacactggaactgaagatggctaccaaggcgaagaa
tctttatttaacaaagcatactatggcaaaagcacatcattcttccgtcaagaaagtcaaaaacttctgcaaagcgataaaaaac
gcacggctgagttagcaaacggcgctctcggtatgattgagctaaacgatgattacacactgaaaaaagtgatgaaaccgct
gattgcatctaacacagtaacagatgaaattgaacgcgcgaacgtctttaaaatgaacggcaaatggtacctgttcactgact
cccgcggatcaaaaatgacgattgacggcattacgtctaacgatatttacatgcttggttatgtttctaattctttaactggcccat
acaagccgctgaacaaaactggccttgtgttaaaaatggatcttgatcctaacgatgtaacctttacttactcacacttcgctgta
cctcaagcgaaaggaaacaatgtcgtgattacaagctatatgacaaacagaggattctacgcagacaaacaatcaacgtttg
cgccaagcttcctgctgaacatcaaaggcaagaaaacatctgttgtcaaagacagcatccttgaacaaggacaattaacagtt
aacaaataa-3’

Protein sequence:

MNIKKFAKQATVLTFTTALLAGGATQAFAKETNQKPYKETYGISHITRHDML
QIPEQQKNEKYQVPEFDSSTIKNISSAKGLDVWDSWPLQNADGTVANYHGYH
IVFALAGDPKNADDTSIYMFYQKVGETSIDSWKNAGRVFKDSDKFDANDSILK
DQTQEWSGSATFTSDGKIRLFYTDFSGKHYGKQTLTTAQVNVSASDSSLNING
VEDYKSIFDGDGKTYQNVQQFIDEGNYSSGDNHTLRDPHYVEDKGHKYLVFE
ANTGTEDGYQGEESLFNKAYYGKSTSFFRQESQKLLQSDKKRTAELANGALG
MIELNDDYTLKKVMKPLIASNTVTDEIERANVFKMNGKWYLFTDSRGSKMTI
DGITSNDIYMLGYVSNSLTGPYKPLNKTGLVLKMDLDPNDVTFTYSHFAVPQ
AKGNNVVITSYMTNRGFYADKQSTFAPSFLLNIKGKKTSVVKDSILEQGQLTV
NK

5.10.2 Plasmid construction of Pduel **7**

Plasmid Pduel **7** was derived from pY+. The gene of GFP was replaced by gene of *sacB*. Using PYJ1712-SacB (Addgene) as template, *sacB* was amplified by two oligodeoxynucleotides (5'-CCCCTCTAGATTTAAGAAGGAGATATACCATGAACATCAAAAAG-3', and 5'-CCGGCGAGCTCTTATTTGTTAACTGTTAA-3'), and then cloned into digested pY+ at two restriction sites *Xba1* and *Sac1* to make Pduel **7**.

5.10.3 The activity test of Pduel **7**

Pduel **7** was used together with the pBK-*Mm*PylRS to transform *E. coli* Top10 strain. For positive selection, the Top 10 cells were let grown on the LB plates, 12.5 μg/mL of tetracycline and 25 μg/mL of kanamycin. 34 μg/mL or 68 μg/mL of chloramphenicol were used with and without 1 mM BocK. Without chloramphenicol, cells were healthy. In the presence of chloramphenicol, with 1 mM BocK, cells all survived; without BocK, cells all died. It indicated the positive selective marker function appropriately. For the negative, in the presence of sucrose, cells all died with and without BocK. 1% and 5% of sucrose both were enough to lead to cell death. 5% of glucose did not help (Table 8).

Table 8. Activity Test of Pduel 7

Positive (LB) 12.5 μg/mL tet 25 μg/mL kan	34 or 68 μg/mL cm	-	+	+		
	1 mM BocK	-	-	+		
		Grow	**Die**	**Grow**		
Negative (LB) 12.5 μg/mL tet 25 μg/mL kan	1% or 5% sucrose	-	+	+	+	+
	5% Glucose	-	-	-	+	+
	1 mM BocK	-	-	+	-	+
		Grow	**Die**	**Die**	**Die**	**Die**

5.11 Pduel **8**: pRep-SacBK144TAG

Since the control of T7 promoter is not very tight, we introduced one amber codon to SacB, to decrease the expression level of SacB (Figure 74).

5.11.1 Plasmid construction of Pduel **8**

Plasmid Pduel **8** was derived from Pduel **7**. The gene of SacB was replaced by gene of SacBK144TAG. Using Pduel **7** as template, SacBK144TAG was generated by site-directed mutagenesis using four oligodeoxynucleotides (5'-CCCCTCTAGATTTAAGAAGGAGATATACCATGAACATCAAAAAG-3', 5'-GAATTTGTCGCTGTCCTAAAAGACGCGGCCAGC-3', 5'-GCTGGCCGCGTCTTTTAGGACAGCGACAAATTC-3', and 5'-CCGGCGAGCTCTTATTTGTTAACTGTTAA-3'), and then cloned into digested pY+ at two restriction sites *Xba1* and *Sac1* to make Pduel **8**.

5.11.2 The activity test of Pduel **8**

Pduel **8** was used together with the pBK-*Mm*PylRS to transform *E. coli* Top10 strain. For positive selection, the Top 10 cells were let grown on the LB plates, 12.5 μg/mL of tetracycline and 25 μg/mL of kanamycin. 34 μg/mL or 68 μg/mL of chloramphenicol were used with and without 1 mM BocK. Without chloramphenicol, cells were healthy. In the presence of chloramphenicol, with 1 mM BocK, cells all survived; without BocK, cells all died. It indicated the positive selective marker function appropriately. For the negative, in the presence of sucrose, cells were all grown with and without BocK (Table 9). It indicated the mutation made SacB lose all activity.

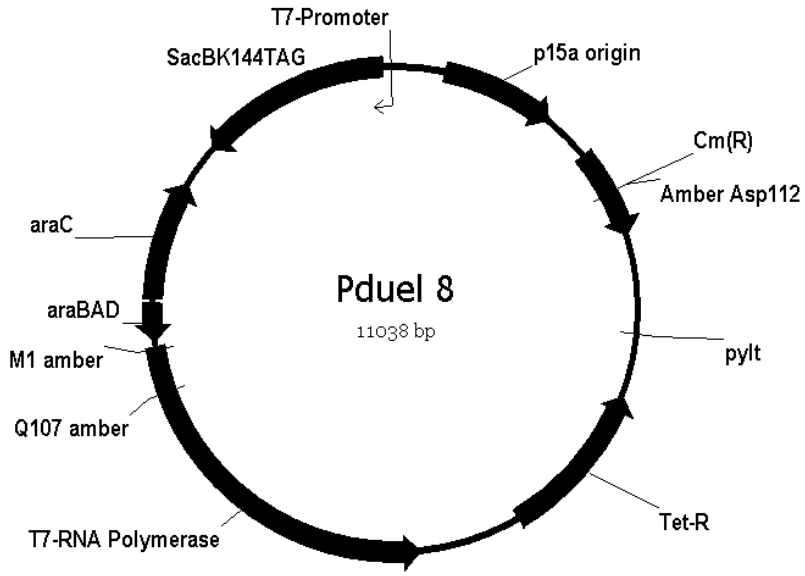


Figure 74. Plasmid structure of Pduel **8**.

Table 9. Activity Test of Pduel 8

Positive (LB) 12.5 μg/mL tet 25 μg/mL kan	34 or 68 μg/mL cm	-	+	+
	1 mM BocK	-	-	+
		Grow	**Die**	**Grow**
Negative (LB) 12.5 μg/mL tet 25 μg/mL kan	5% sucrose	-	+	+
	1 mM BocK	-	-	+
		Grow	**Grow**	**Grow**

5.12 Pduel **9**: pRep-upp

Since it's very tricky to tune the activity of SacB, we moved to another negative selective marker, uracil phosphoribosyltransferase (uprt). Uprt coded by the gene of *upp*, converts 5-fluorouracil (5-FU) to 5-fluoro-dUMP, which inhibits thymidylate synthase, causing cell death. The new plasmid, Pduel **9** was obtained by replacing the gene of GFP in pY+ with the gene of *upp* (Figure 75). In the plasmid Pduel **9**, for the negative selection, the *upp* gene, under control of T7 promoter, similar as Pduel **7**, will be expressed by T7 polymerase, in the presence of 5-FU.

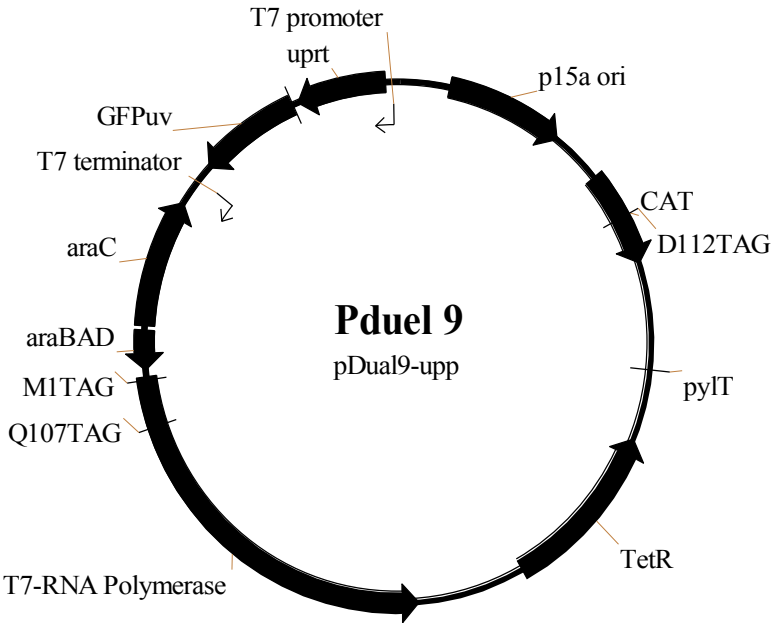


Figure 75. Plasmid structure of Pduel **9**.

5.12.1 DNA and protein sequences of upp

DNA sequence:

5’-atgaagatcgtggaagtcaaacacccactcgtcaaacacaagctgggactgatgcgtgagcaagatatcagcaccaa
gcgctttcgcgaactcgcttccgaagtgggtagcctgctgacttacgaagcgaccgccgacctcgaaacggaaaaagtaac
tatcgaaggctggaacggcccggtagaaatcgaccagatcaaaggtaagaaaattaccgttgtgccaattctgcgtgcggg
tcttggtatgatggacggtgtgctggaaaacgttccgagcgcgcgcatcagcgttgtcggtatgtaccgtaatgaagaaacg
ctggagccggtaccgtacttccagaaactggtttctaacatcgatgagcgtatggcgctgatcgttgacccaatgctggcaac
cggtggttccgttatcgcgaccatcgacctgctgaaaaaagcgggctgcagcagcatcaaagttctggtgctggtagctgc
gccagaaggtatcgctgcgctggaaaaagcgcacccggacgtcgaactgtataccgcatcgattgatcagggactgaacg
agcacggatacattattccgggcctcggcgatgccggtgacaaaatctttggtacgaaataa-3’

Protein sequence:

MKIVEVKHPLVKHKLGLMREQDISTKRFRELASEVGSLLTYEATADLETEKVT
IEGWNGPVEIDQIKGKKITVVPILRAGLGMMDGVLENVPSARISVVGMYRNEE
TLEPVPYFQKLVSNIDERMALIVDPMLATGGSVIATIDLLKKAGCSSIKVLVLV
AAPEGIAALEKAHPDVELYTASIDQGLNEHGYIIPGLGDAGDKIFGTK

5.12.2 Plasmid construction of Pduel **9**

Plasmid Pduel 9 was derived from pY+. The gene of *upp* was cloned into pY+, followed by GFP. The *upp* gene was amplified from the genomic DNA of *E. coli* Top10 strain by two oligodeoxynucleotides (5’-TCCCTCTAGAGTGGCTGCCCCTCAAAGGAG-3’, and 5’-TATTTCTAGATTATTTCGTACCAAAGATTTTGTC-3’), and then cloned into digested pY+ at two restriction sites *Xba1* at both ends to make Pduel **9**.

5.12.3 The activity test of Pduel **9**

The Top 10 cells transformed with Pduel **9** and pBK-*Mm*PylRS were let grown on the LB plates, with 12.5 μg/mL of tetracycline and 25 μg/mL of kanamycin. The positive selective marker functioned very well. However, for negative conditions, with 0.2% of arabinose, 1 mg/mL 5-FU, 1 mM Bock, the suppression was not obvious. Even when we changed to more sensitive GH371 *E. coli* whose genomic copy of *upp* was disrupted, the inhibition dynamic range was still restricted (Table 10).

Table 10. Activity Test of Pduel 9

Positive (LB) 12.5 μg/mL tet 25 μg/mL kan	68 μg/mL cm	-	+	+
	1 mM BocK	-	-	+
		Grow	**Die**	**Grow**
Negative (LB) 12.5 μg/mL tet 25 μg/mL kan	1mg/mL 5-FU	-	+	+
	1 mM BocK	-	-	+
		Grow	**Grow**	**Inhibition**

5.13 Pduel **10**: pRep-ccdB2m

After using uprt, we tried one last negative selection marker, ccdB. CcdB, one part of ccd system of the F plasmid in *E. coli*, is responsible for the plasmid's high stability by postsegregational killing of plasmid-free cells. The toxicity of the ccdB was demonstrated to reduce colony formation by more than five orders of magnitude and showed a great potential as an ideal negative selection marker. Thus, Pduel **10** was designed by replacing *upp* with ccdB2m in the Pduel **9** (Figure 76). CcdB2m, D44TAG and A13TAG, instead of ccdB, was chosen because of the high toxicity of ccdB. For the negative selection, the ccdB2*m* gene, under control of T7 promoter, will be produced and cause cell death in the presence of T7 polymerase.

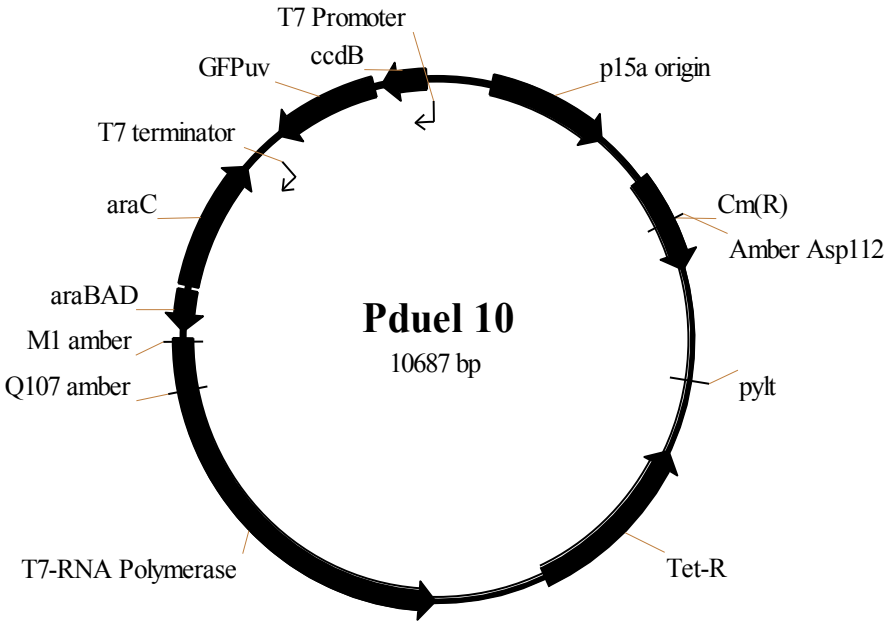


Figure 76. Plasmid structure of Pduel **10**.

5.13.1 DNA and protein sequences of ccdB

DNA sequence:

5’-atgcagtttaaggtttacacctataaaagagagagccgttatcgtctgtttgtggatgtacagagtgatattattgacacgc
ccgggcgacggatggtgatccccctggccagtgcacgtctgctgtcagataaagtctcccgtgaactttacccggtggtgca
tatcggggatgaaagctggcgcatgatgaccaccgatatggccagtgtgccggtctccgttatcggggaagaagtggctga
tctcagccaccgcgaaaatgacatcaaaaacgccattaacctgatgttctggggaatataa-3’

Protein sequence:

MQFKVYTYKRESRYRLFVDVQSDIIDTPGRRMVIPLASARLLSDKVSRELYPV
VHIGDESWRMMTTDMASVPVSVIGEEVADLSHRENDIKNAINLMFWGI

5.13.2 Plasmid construction of Pduel **10**

Plasmid Pduel **10** was derived from pY+. The ccdB gene was amplified from pAraCB2 plasmid, which is a gift from Dr. Söll at Yale University by two oligodeoxynucleotides (TGTAGAGCTCTTATATTCCCCAGAACATCAGG, and TGTATCTAGATTATATTCCCCAGAACATCAGGTTAATGGCGTT), and then cloned into digested Pduel **9** by two restriction sites *Xba1* at both ends to make Pduel **10**.

5.13.3 The activity test of Pduel **10**

To demonstrate the applicability of the system, Top10/pRep-ccdB2m, pBK-*Mm*PylRS and Top10/pRep-ccdB1m, pBK-*Mm*PylRS cells were prepared. Cell survival is highly NAA dependent (Figure 77). At positive condition, when 68 μg/mL of chloramphenicol were used, cells grew normally in media supplied with 1 mM BocK, but cell growth is suppressed in media without BocK. At negative conditions,

cell growth is suppressed in media supplied with BocK, but cells grew normally in media without BocK. Cells with pDual9-ccdB1m showed the same result under the negative conditions. However, cell growth is slightly suppressed at positive condition in the presence of BocK (Table 11).

Table 11. Activity Test of Pduel 10

Positive (LB) 12.5 μg/mL tet 25 μg/mL kan		68 μg/mL cm	-	+	+
		1 mM BocK	-	-	+
	ccdB1m		**Grow**	**Die**	**Inhibition**
	ccdB2m		**Grow**	**Die**	**Grow**
Negative (LB) 12.5 μg/mL tet 25 μg/mL kan		0.2% arabinose	-	-	+
		1 mM BocK	-	+	+
	ccdB1m		**Grow**	**Grow**	**Die**
	ccdB2m		**Grow**	**Grow**	**Die**

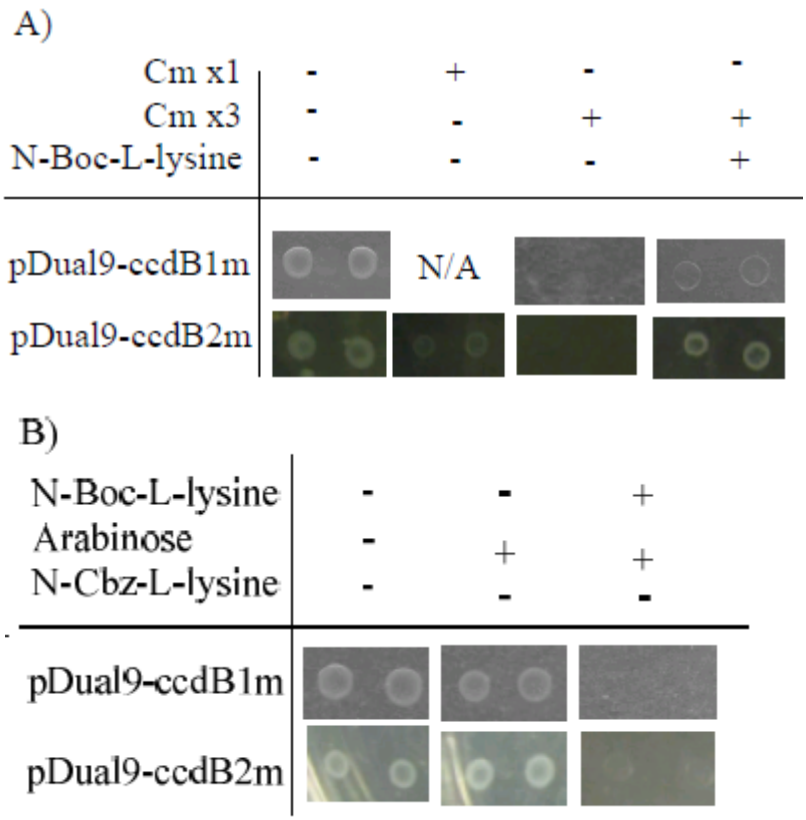


Figure 77. Demonstration of Pduel **10**. The phenotype of Top10/pRep-ccdB2m, pBK-*Mm*PylRS, and Top10/pRep-ccdB1m, pBK-*Mm*PylRS at positive A) negative conditions B).

5.14 Summary

In summary, we have developed the one-plasmid selective system, which could carry out both the positive and negative library selection. This system could be applied to any library selection based on NAAs, and greatly simplified the selection process. Nowadays, more and more NAAs have been incorporated into proteins, and utilized to study proteins. Our one-plasmid selection system will speed up the process of incorporation of NAAs, and therefore the understanding of protein functions. In the meantime, by generating a serial of one-plasmid system plasmids, we have thoroughly studied the positive and negative markers, and it's helpful for us to optimize the selection process.

6. CONCLUSION

Genetic incorporation of NAAs into proteins is a powerful approach to extend the diversity of amino acids. This method relies on the read-through of an in-frame amber (UAG) stop codon in mRNA by an amber suppressor tRNA ($tRNA_{CUA}$) specifically acylated with a NAA by an evolved aaRS. This system is equivalent to the naturally occurring pyrrolysine incorporation machinery discovered first in methanogenic archaea and a Gram positive bacterium *Desulfitobacterium hafniense*.[59, 61, 69, 141-143] In these organisms, Pyl is co-translationally inserted into proteins by an in-frame amber codon. Similarly to the Pyl incorporation machinery, an orthogonal aaRS-$tRNA_{CUA}$ pair can be developed in which the aaRS is evolved to specifically charge its cognate $tRNA_{CUA}$ with a NAA. When expressed in cells, this aaRS-$tRNA_{CUA}$ pair enables a NAA to be site-specifically incorporated into a protein at the amber codon with high fidelity and efficiency. Using this approach, a variety of NAAs have been incorporated into proteins in bacteria, yeast and mammalian cells.[84, 98, 116, 144-146]

PylRS/pylT, with unique features, demonstrates itself as the perfect tool to expand the genetic code. PylRS/pylT pair, which is orthogonal to the endogenous aaRS/tRNA pairs in both bacteria and eukaryotes, can be easily evolved in *E. coli*, and later transferred to eukaryotes. The large hydrophobic pocket in pylRS leads to notable substrate flexibility. Indeed, pylRS has been used to incorporate a variety of NAAs, from lysine analogs to tyrosine analogs (Figure 78). The lack of interaction between α-amino group and pylRS allows the incorporation of non-α-amino acid, which expands the

category of protein backbone. In addition, crystal structure revealed the missing direct interactions between the anticodon of pylT and pylRS. Taking advantage of all these unique features of pylRS/pylT pair, I have been able to carry out the following studies.

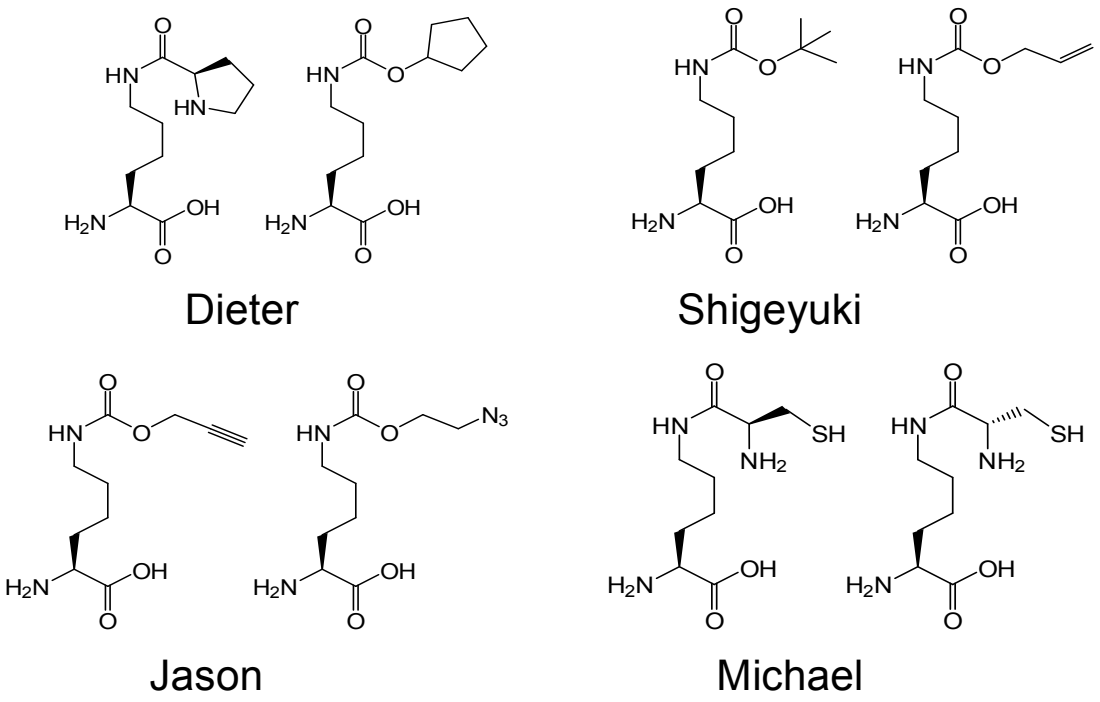


Figure 78. Reprehensive NAAs taken by wtPylRS or its mutants.

6.1 Genetic incorporation of KetoK into one protein

I demonstrated genetic incorporation of KetoK, an unhydrolysable analog of AcK for the installation of a permanent acetylation analog to proteins. Since KetoK has a keto group, it was also applied to achieve site-specific protein modification. KetoK synthesis followed the previous route with a 58% overall yield. The genetic encoding of KetoK by amber codon in *E. coli* was subsequently tested by transforming cells with pAcKRS-pylT-GFP1Amber and growing in media containing 2 mM KetoK. Since KetoK contains a keto group, we also tested the efficiency of using this group to specifically modify a protein incorporated with KetoK with hydrazine- and alkoxyamine-bearing compounds. To demonstrate the feasibility of using KetoK to perform protein fluorescent labeling, the purified GFP-KetoK was reacted with Texas Red hydrazine and biotin alkoxyamine at pH 6.3 for overnight. Both labeling gave high labelling specificity and efficiency (Figure 79).

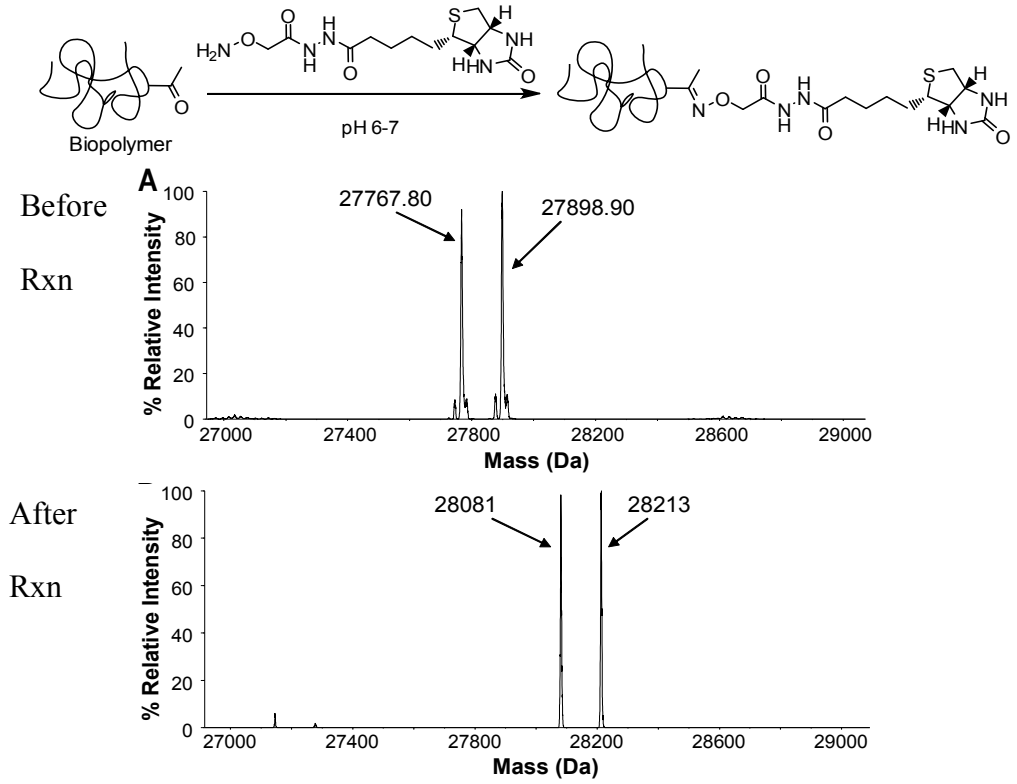


Figure 79. Biotin alkoxyamine reaction with ketoK in GFP_{UV}.

6.2 Genetic incorporation of multiple NAAs into one protein

I developed a novel strategy to incorporate multiple NAAs into one protein. The method is based on the decrease of the translation termination rate mediated by RF1. It has been suggested that the large ribosomal subunit protein L11 plays an important role in peptide release by RF1. By replacing L11 in ribosome with overexpressed L11C to enhance amber suppression, GFP_{UV} with up to three AcKs has been expressed in *E. coli*. We first confirmed that overexpressing L11C enhances amber suppression. The results of protein expression in *E. coli* BL21 cells transformed with different plasmids and grown in different conditions are presented in Figure 80. The expression of GFP1Amber in the presence of overexpressed L11C was four times higher than that obtained in the absence of L11C. We then demonstrated that overexpressing L11C allows synthesizing proteins incorporated with multiple AcKs. GFP2Amber has amber mutations at positions 149 and 204. GFP3Amber has one additional amber mutation at position 184.

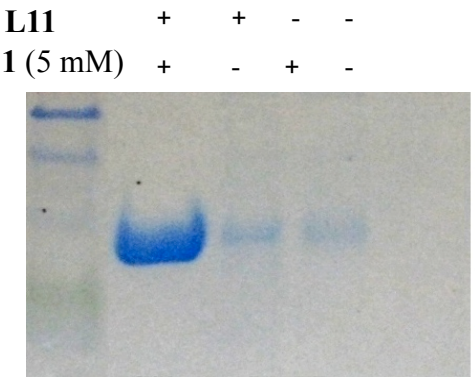


Figure 80. GFP_{UV} expressed w/ and w/o L11C.

6.3 Genetic incorporation of two distinctive NAAs into a single protein

I have developed a method to incorporate two distinctive NAAs into one protein in *E. coli*, which can be used to synthesize proteins with two different concomitant modifications. The method takes advantage of two aaRS-$tRNA_{CUA}$ pairs evolved from *E. coli* (Figure 81). We attempted to combine the PylRS-pylT pair and the *Mj*TyrRS-tRNA pair to incorporate two different NAAs into a single protein because of their proven efficiency in NAA incorporation.[87, 100, 117, 118, 126-129] In order to use both pairs in a single *E. coli* cell, we first demonstrated that they are orthogonal to each other. In order to demonstrate the utility of the PylRS-$pylT_{UUA}$ pair together with an evolved *Mj*TyrRS-$tRNA_{CUA}$ pair to incorporate two different NAAs into a single protein in *E. coli* by both amber and ochre suppressions, two plasmids, pEVOL-AzFRS and pPylRS-pylT-GFP1TAG149TAA, were used to transform *E. coli* BL21 cells. Full-length GFP_{UV} were expressed when two corresponding NAAs were supplemented in media. Since GFP_{UV}(**2**+**4**) contains both an alkyne group and an azide group, we tested the feasibility of separately labeling this protein with different fluorescent dyes by performing click reactions on these two functional groups. The labeling gave very high specify.

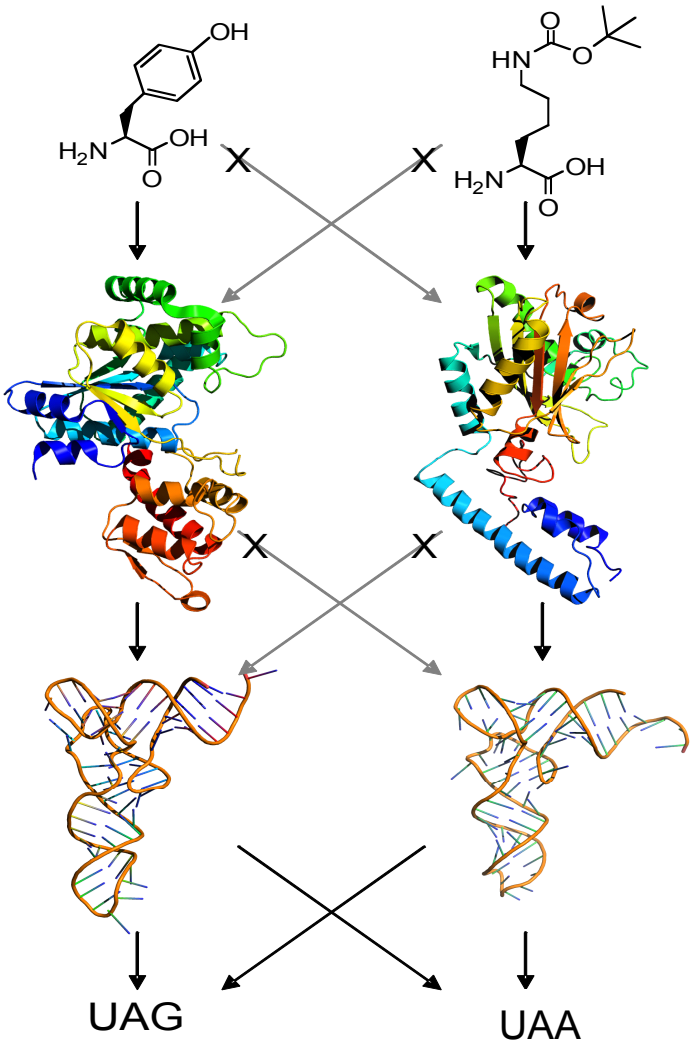


Figure 81. Strategy for two NAAs incorporation.

6.4 A straightforward one plasmid selection system for evolution of aaRSs

I have developed an easy, straightforward selection method for the evolution of aaRSs in *E. coli*. This method relies on one plasmid, which is responsible for dual positive/negative selection. This plasmid utilizes an amber stop codon containing chloramphenicol acetyltransferase as positive selection marker, and two amber stop codon containing ccdB as negative selection marker. I demonstrated the utility of the system by identifying a variant of the *Mm*pylRS from a library of 10^9 variants that selectively incorporates *N*-*ε*-carbobenzyloxy-lysine in response to an amber stop codon.

REFERENCES

(1) Ibba, M.; Soll, D. *Annu. Rev. Biochem.* **2000**, *69*, 617-650.

(2) Eriani, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D. *Nature* **1990**, *347*, 203-206.

(3) Delarue, M.; Moras, D. *Bioessays.* **1993**, *15*, 675-687.

(4) Moras, D. *Trends Biochem. Sci.* **1992**, *17*, 159-164.

(5) Ibba, M.; Curnow, A. W.; Soll, D. *Trends Biochem. Sci.* **1997**, *22*, 39-42.

(6) Rould, M. A.; Perona, J. J.; Soll, D.; Steitz, T. A. *Science* **1989**, *246*, 1135-1142.

(7) Cusack, S. *Nat. Struct. Biol.* **1995**, *2*, 824-831.

(8) O'Donoghue, P.; Luthey-Schulten, Z. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 550-573.

(9) Ibba, M.; Morgan, S.; Curnow, A. W.; Pridmore, D. R.; Vothknecht, U. C.; Gardner, W.; Lin, W.; Woese, C. R.; Soll, D. *Science* **1997**, *278*, 1119-1122.

(10) Praetorius-Ibba, M.; Ibba, M. *Mol. Microbiol.* **2003**, *48*, 631-637.

(11) Li, T.; Graham, D. E.; Stathopoulos, C.; Haney, P. J.; Kim, H. S.; Vothknecht, U.; Kitabatake, M.; Hong, K. W.; Eggertsson, G.; Curnow, A. W.; Lin, W.; Celic, I.; Whitman, W.; Soll, D. *FEBS Lett.* **1999**, *462*, 302-306.

(12) Ambrogelly, A.; Ahel, I.; Polycarpo, C.; Bunjun-Srihari, S.; Krett, B.; Jacquin-Becker, C.; Ruan, B.; Kohrer, C.; Stathopoulos, C.; RajBhandary, U. L.; Soll, D. *J. Biol. Chem.* **2002**, *277*, 34749-34754.

(13) Lipman, R. S.; Beuning, P. J.; Musier-Forsyth, K.; Hou, Y. M. *J. Mol. Biol.* **2002**, *316*, 421-427.

(14) Sethi, A.; O'Donoghue, P.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 4045-4050.

(15) Sauerwald, A.; Zhu, W.; Major, T. A.; Roy, H.; Palioura, S.; Jahn, D.; Whitman, W. B.; Yates, J. R., 3rd; Ibba, M.; Soll, D. *Science* **2005**, *307*, 1969-1972.

(16) Wilcox, M.; Nirenberg, M. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *61*, 229-236.

(17) White, B. N.; Bayley, S. T. *Can. J. Biochem.* **1972**, *50*, 600-609.

(18) Tumbula, D.; Vothknecht, U. C.; Kim, H. S.; Ibba, M.; Min, B.; Li, T.; Pelaschier, J.; Stathopoulos, C.; Becker, H.; Soll, D. *Genetics* **1999**, *152*, 1269-1276.

(19) Ibba, M.; Soll, D. *Genes Dev.* **2004**, *18*, 731-738.

(20) Lapointe, J.; Duplain, L.; Proulx, M. *J. Bacteriol.* **1986**, *165*, 88-93.

(21) Salazar, J. C.; Ahel, I.; Orellana, O.; Tumbula-Hansen, D.; Krieger, R.; Daniels, L.; Soll, D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13863-13868.

(22) Curnow, A. W.; Hong, K.; Yuan, R.; Kim, S.; Martins, O.; Winkler, W.; Henkin, T. M.; Soll, D. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11819-11826.

(23) Sylvers, L. A.; Rogers, K. C.; Shimizu, M.; Ohtsuka, E.; Soll, D. *Biochemistry* **1993**, *32*, 3836-3841.

(24) Numata, T.; Ikeuchi, Y.; Fukai, S.; Suzuki, T.; Nureki, O. *Nature* **2006**, *442*, 419-424.

(25) Sekine, S.; Nureki, O.; Shimada, A.; Vassylyev, D. G.; Yokoyama, S. *Nat. Struct. Biol.* **2001**, *8*, 203-206.

(26) Tumbula, D. L.; Becker, H. D.; Chang, W. Z.; Soll, D. *Nature* **2000**, *407*, 106-110.

(27) Feng, L.; Yuan, J.; Toogood, H.; Tumbula-Hansen, D.; Soll, D. *J. Biol. Chem.* **2005**, *280*, 20638-20641.

(28) Curnow, A. W.; Ibba, M.; Soll, D. *Nature* **1996**, *382*, 589-590.

(29) Cathopoulis, T.; Chuawong, P.; Hendrickson, T. L. *Mol. Biosyst.* **2007**, *3*, 408-418.

(30) Becker, H. D.; Kern, D. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12832-12837.

(31) Min, B.; Pelaschier, J. T.; Graham, D. E.; Tumbula-Hansen, D.; Soll, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2678-2683.

(32) Woese, C. R.; Olsen, G. J.; Ibba, M.; Soll, D. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 202-236.

(33) Bairoch, A. *Nucleic Acids Res.* **2000**, *28*, 304-305.

(34) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 7342-7372.

(35) Cone, J. E.; Del Rio, R. M.; Davis, J. N.; Stadtman, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2659-2663.

(36) Burke, S. A.; Lo, S. L.; Krzycki, J. A. *J. Bacteriol.* **1998**, *180*, 3432-3440.

(37) Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. *Science* **2002**, *296*, 1462-1466.

(38) Johansson, L.; Gafvelin, G.; Arner, E. S. *Biochim. Biophys. Acta.* **2005**, *1726*, 1-13.

(39) Chambers, I.; Frampton, J.; Goldfarb, P.; Affara, N.; McBain, W.; Harrison, P. R. *EMBO J.* **1986**, *5*, 1221-1227.

(40) Zinoni, F.; Birkmann, A.; Stadtman, T. C.; Bock, A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4650-4654.

(41) Hatfield, D.; Choi, I. S.; Mischke, S.; Owens, L. D. *Biochem. Biophys. Res. Commun.* **1992**, *184*, 254-259.

(42) Kryukov, G. V.; Castellano, S.; Novoselov, S. V.; Lobanov, A. V.; Zehtab, O.; Guigo, R.; Gladyshev, V. N. *Science* **2003**, *300*, 1439-1443.

(43) Leinfelder, W.; Zehelein, E.; Mandrand-Berthelot, M. A.; Bock, A. *Nature* **1988**, *331*, 723-725.

(44) Forchhammer, K.; Bock, A. *J. Biol. Chem.* **1991**, *266*, 6324-6328.

(45) Forchhammer, K.; Boesmiller, K.; Bock, A. *Biochimie.* **1991**, *73*, 1481-1486.

(46) Liu, Z.; Reches, M.; Groisman, I.; Engelberg-Kulka, H. *Nucleic. Acids Res.* **1998**, *26*, 896-902.

(47) Carlson, B. A.; Xu, X. M.; Kryukov, G. V.; Rao, M.; Berry, M. J.; Gladyshev, V. N.; Hatfield, D. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12848-12853.

(48) Yuan, J.; Palioura, S.; Salazar, J. C.; Su, D.; O'Donoghue, P.; Hohn, M. J.; Cardoso, A. M.; Whitman, W. B.; Soll, D. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18923-18927.

(49) Copeland, P. R.; Fletcher, J. E.; Carlson, B. A.; Hatfield, D. L.; Driscoll, D. M. *EMBO J.* **2000**, *19*, 306-314.

(50) Chavatte, L.; Brown, B. A.; Driscoll, D. M. *Nat. Struct. Mol. Biol.* **2005**, *12*, 408-416.

(51) Rother, M.; Resch, A.; Wilting, R.; Bock, A. *Biofactors* **2001**, *14*, 75-83.

(52) Hao, B.; Zhao, G.; Kang, P. T.; Soares, J. A.; Ferguson, T. K.; Gallucci, J.; Krzycki, J. A.; Chan, M. K. *Chem. Biol.* **2004**, *11*, 1317-1324.

(53) Thauer, R. K.; Kaster, A. K.; Seedorf, H.; Buckel, W.; Hedderich, R. *Nat. Rev. Microbiol.* **2008**, *6*, 579-591.

(54) Paul, L.; Ferguson, D. J., Jr.; Krzycki, J. A. *J. Bacteriol.* **2000**, *182*, 2520-2529.

(55) Soares, J. A.; Zhang, L.; Pitsch, R. L.; Kleinholz, N. M.; Jones, R. B.; Wolff, J. J.; Amster, J.; Green-Church, K. B.; Krzycki, J. A. *J. Biol. Chem.* **2005**, *280*, 36962-36969.

(56) Krzycki, J. A. *Curr. Opin. Chem. Biol.* **2004**, *8*, 484-491.

(57) Ferguson, T.; Soares, J. A.; Lienard, T.; Gottschalk, G.; Krzycki, J. A. *J. Biol. Chem.* **2009**, *284*, 2285-2295.

(58) Ferguson, D. J., Jr.; Krzycki, J. A.; Grahame, D. A. *J. Biol. Chem.* **1996**, *271*, 5189-5194.

(59) Srinivasan, G.; James, C. M.; Krzycki, J. A. *Science* **2002**, *296*, 1459-1462.

(60) Polycarpo, C.; Ambrogelly, A.; Ruan, B.; Tumbula-Hansen, D.; Ataide, S. F.; Ishitani, R.; Yokoyama, S.; Nureki, O.; Ibba, M.; Soll, D. *Mol. Cell* **2003**, *12*, 287-294.

(61) 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*, 333-335.

(62) Theobald-Dietrich, A.; Frugier, M.; Giege, R.; Rudinger-Thirion, J. *Nucleic Acids Res.* **2004**, *32*, 1091-1096.

(63) Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11268-11273.

(64) Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Araiso, Y.; Ishitani, R.; Umehara, T.; Soll, D.; Nureki, O. *Nature* **2009**, *457*, 1163-1167.

(65) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *J. Mol. Biol.* **2008**, *378*, 634-652.

(66) Lee, M. M.; Jiang, R.; Jain, R.; Larue, R. C.; Krzycki, J.; Chan, M. K. *Biochem. Biophys. Res. Commun.* **2008**, *374*, 470-474.

(67) Namy, O.; Zhou, Y.; Gundllapalli, S.; Polycarpo, C. R.; Denise, A.; Rousset, J. P.; Soll, D.; Ambrogelly, A. *FEBS Lett.* **2007**, *581*, 5282-5288.

(68) Longstaff, C.; Whitton, C.; Thelwell, C.; Belgrave, D. *J. Thromb. Haemost.* **2007**, *5*, 412-414.

(69) Longstaff, D. G.; Larue, R. C.; Faust, J. E.; Mahapatra, A.; Zhang, L.; Green-Church, K. B.; Krzycki, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1021-1026.

(70) Gaston, M. A.; Zhang, L.; Green-Church, K. B.; Krzycki, J. A. *Nature* **2011**, *471*, 647-650.

(71) Baldwin, A. N.; Berg, P. *J. Biol. Chem.* **1966**, *241*, 839-845.

(72) Silvian, L. F.; Wang, J.; Steitz, T. A. *Science* **1999**, *285*, 1074-1077.

(73) Ambrogelly, A.; Gundllapalli, S.; Herring, S.; Polycarpo, C.; Frauer, C.; Soll, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 3141-3146.

(74) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. *Cell* **2007**, *128*, 1003-1012.

(75) Schwarzer, D.; Cole, P. A. *Curr. Opin. Chem. Biol.* **2005**, *9*, 561-569.

(76) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498-500.

(77) Wang, L.; Schultz, P. G. *Chem. Biol.* **2001**, *8*, 883-890.

(78) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11020-11024.

(79) Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *J. Am. Chem. Soc.* **2002**, *124*, 9026-9027.

(80) Wang, L.; Schultz, P. G. *Chem. Commun. (Camb)* **2002**, 1-11.

(81) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964-967.

(82) Wang, L.; Xie, J.; Deniz, A. A.; Schultz, P. G. *J. Org. Chem.* **2003**, *68*, 174-176.

(83) Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. *Nucleic Acids Res.* **2002**, *30*, 4692-4699.

(84) Wang, J.; Xie, J.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 8738-8739.

(85) Furter, R. *Protein Sci.* **1998**, *7*, 419-426.

(86) Ou, W.; Uno, T.; Chiu, H. P.; Grunewald, J.; Cellitti, S. E.; Crossgrove, T.; Hao, X.; Fan, Q.; Quinn, L. L.; Patterson, P.; Okach, L.; Jones, D. H.; Lesley, S. A.; Brock, A.; Geierstanger, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10437-10442.

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

(88) Hao, Z.; Song, Y.; Lin, S.; Yang, M.; Liang, Y.; Wang, J.; Chen, P. R. *Chem. Commun. (Camb)* **2011**, *47*, 4502-4504.

(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*, 4052-4055.

(90) 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*, 1557-1560.

(91) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 878-880.

(92) Eliot, A. C.; Kirsch, J. F. *Annu. Rev. Biochem.* **2004**, *73*, 383-415.

(93) Geoghegan, K. F.; Emery, M. J.; Martin, W. H.; McColl, A. S.; Daumy, G. O. *Bioconjug. Chem.* **1993**, *4*, 537-544.

(94) Mahal, L. K.; Yarema, K. J.; Bertozzi, C. R. *Science* **1997**, *276*, 1125-1128.

(95) Hang, H. C.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2001**, *123*, 1242-1243.

(96) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3*, 321-322.

(97) Cornish, V. W.; Mendel, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 8150-8151.

(98) Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56-61.

(99) Brustad, E. M.; Lemke, E. A.; Schultz, P. G.; Deniz, A. A. *J. Am. Chem. Soc.* **2008**, *130*, 17664-17665.

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

(101) Jones, P.; Altamura, S.; Chakravarty, P. K.; Cecchetti, O.; De Francesco, R.; Gallinari, P.; Ingenito, R.; Meinke, P. T.; Petrocchi, A.; Rowley, M.; Scarpelli, R.; Serafini, S.; Steinkuhler, C. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5948-5952.

(102) Huang, Y.; Russell, W. K.; Wan, W.; Pai, P. J.; Russell, D. H.; Liu, W. *Mol. Biosyst.* **2010**, *6*, 683-686.

(103) Wang, L.; Schultz, P. G. *Angew. Chem. Int. Ed. Engl.* **2004**, *44*, 34-66.

(104) Liu, C. C.; Schultz, P. G. *Nat. Biotechnol.* **2006**, *24*, 1436-1440.

(105) Wiley, J., *Fundamentals of biochemistry* 3rd ed.; 2006.

(106) Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Biotechnol.* **2007**, *25*, 770-777.

(107) Lall, S. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1110-1115.

(108) Wimberly, B. T.; Guymon, R.; McCutcheon, J. P.; White, S. W.; Ramakrishnan, V. *Cell* **1999**, *97*, 491-502.

(109) Van Dyke, N.; Murgola, E. J. *J. Mol. Biol.* **2003**, *330*, 9-13.

(110) Bouakaz, L.; Bouakaz, E.; Murgola, E. J.; Ehrenberg, M.; Sanyal, S. *J. Biol. Chem.* **2006**, *281*, 4548-4556.

(111) Kruse, J. P.; Gu, W. *Cell* **2009**, *137*, 609-622.

(112) Yang, X. J.; Seto, E. *Mol. Cell* **2008**, *31*, 449-461.

(113) Service, R. F. *Science* **2005**, *308*, 44.

(114) 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*, 3211-3214.

(115) Xie, J.; Schultz, P. G. *Methods* **2005**, *36*, 227-238.

(116) Wang, L.; Xie, J.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225-249.

(117) Fekner, T.; Li, X.; Lee, M. M.; Chan, M. K. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 1633-1635.

(118) Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S. *Chem. Biol.* **2008**, *15*, 1187-1197.

(119) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 8720-8721.

(120) Huang, Y.; Russell, W. K.; Wan, W.; Pai, P. J.; Russell, D. H.; Liu, W. *Mol. Biosyst.* **2010**, *6*, 683-686.

(121) Crisp, G. T.; Gore, J. *Tetrahedron* **1997**, *53*, 1523-1544.

(122) Talaga, P.; Benezra, C.; Stampf, J. L. *Bioorg. Chem.* **1990**, *18*, 199-206.

(123) Izdebski, J.; Yamashiro, D.; Ng, T. B.; Li, C. H. *Int. J. Pept. Protein Res.* **1982**, *19*, 327-333.

(124) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W. *Bioorg. Med. Chem. Lett. 20*, 878-880.

(125) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol. 395*, 361-374.

(126) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775-782.

(127) Li, X.; Fekner, T.; Ottesen, J. J.; Chan, M. K. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9184-9187.

(128) Nguyen, D. P.; Lusic, H.; Neumann, H.; Kapadnis, P. B.; Deiters, A.; Chin, J. W. *J. of Am. Chem. Soc.* **2009**, *131*, 8720-8721.

(129) Nguyen, D. P.; Alai, M. M. G.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. *J. of Am. Chem. Soc.* **2009**, *131*, 14194-12195.

(130) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W. *Bioorg. Med. Chem.* **2010**, *20*, 878-880.

(131) Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G. *J. Mol. Biol.* **2009**.

(132) Liu, W.; Brock, A.; Chen, S.; Schultz, P. G. *Nat. Methods* **2007**, *4*, 239-244.

(133) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 19-24.

(134) Polycarpo, C. R.; Herring, S.; Berube, A.; Wood, J. L.; Soll, D.; Ambrogelly, A. *FEBS Lett.* **2006**, *580*, 6695-6700.

(135) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Intl. Ed.* **2001**, *40*, 2004-2005.

(136) Wang, Q.; Parrish, A. R.; Wang, L. *Chem. Biol.* **2009**, *16*, 323-336.

(137) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. *Nat. Biotechnol.* **2002**, *20*, 1044-1048.

(138) Melancon, C. E., 3rd; Schultz, P. G. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3845-3847.

(139) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z.; Russell, D. H.; Liu, W. *Bioorg. Med. Chem. Lett. 20*, 878-880.

(140) Nurbekov, M. K.; Rasulov, M. M.; Voronkov, M. G.; Bobkova, S. N.; Belikova, O. A. *Dokl. Biochem. Biophys.* **2011**, *438*, 131-133.

(141) Ibba, M.; Soll, D. *Curr. Biol.* **2002**, *12*, R464-466.

(142) Polycarpo, C.; Ambrogelly, A.; Berube, A.; Winbush, S. M.; McCloskey, J. A.; Crain, P. F.; Wood, J. L.; Soll, D. *Proc. Natl. Acad. Sci. U.S..A.* **2004**, *101*, 12450-12454.

(143) Herring, S.; Ambrogelly, A.; Polycarpo, C. R.; Soll, D. *Nucleic Acids Res.* **2007**, *35*, 1270-1278.

(144) Zhang, Z.; Smith, B. A.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 6735-6746.

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

(146) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 9785-9789.

VITA

Name: Ying Huang

Address: 700 Dominik Dr. Apt. 2405

College Station, TX, 77840

Email Address: Ying.huang2006@gmail.com

Education: B.A., Chemistry, Peking University, 2007

B.A., Economics, Peking Universisy, 2007