A LEUCYL-tRNA SYNTHETASE WITHOUT THE EDITING DOMAIN AND ITS DIRECTED EVOLUTION FOR UNNATURAL PROTEIN ENGINEERING

By

Emily Fung

Submitted in partial fulfillment of the

Requirements for Departmental Honors in the

Department of Chemistry and Biochemistry

Texas Christian University

Fort Worth, Texas

May 2, 2016

A LEUCYL-tRNA SYNTHETASE WITHOUT THE EDITING DOMAIN AND ITS DIRECTED EVOLUTION FOR UNNATURAL PROTEIN ENGINEERING

Project Approved:

Supervising Professor: Youngha Ryu, Ph.D.

Department of Chemistry and Biochemistry

Kayla Green, Ph.D.

Department of Chemistry and Biochemistry

Giridhar Akkaraju, Ph.D.

Department of Biology

ABSTRACT

This project aims to develop a functional leucine-tRNA synthetase (LeuRS) without the editing domain and perform its directed evolution to incorporate various unnatural amino acids with desirable properties in *Escherichia coli*. This method will be useful to generate proteins with new or enhanced properties and will also serve as a powerful tool to study structure and function of proteins. We generated billions of different variants of the *Methanobacterium thermoautotrophium* LeuRS (MLRS) whose editing domain was replaced by peptide linkers containing six randomized amino acid residues. A genetic selection was subsequently performed to obtain a functional LeuRS without the editing domain (MLRS ΔCP1). The best variant was used in the subsequent directed evolution experiment. From the structure-based sequence alignment with a bacterial LeuRS, the leucine binding site of the MLRS was identified. The five amino acids found in the leucine binding pocket were randomized to generate billions of different variants of MLRS ΔCP1. In order to select MLRS ΔCP1 variants specific for unnatural amino acids, a dual genetic screen will be performed. In the positive selection, the bacteria cells containing MLRS ΔCP1 capable of attaching any natural or unnatural amino acid onto the tRNA can survive in the presence of chloramphenicol. In negative selection in the absence of the unnatural amino acid, the cells containing the MLRS ΔCP1 variants that add any natural amino acids to the tRNA die in the presence of 5-Fluorouracil. The cells harboring the MLRS ΔCP1 variants that attach only unnatural amino acids onto the tRNA can survive on 5-FU. These successful variants will be

tested for the effective incorporation of the unnatural amino acid with our model protein called the Z-domain and for specific intended applications with various proteins.

TABLE OF CONTENTS

INTRODUCTION

The incorporation of unnatural amino acids into proteins can provide a tool to manipulate the structure and function of proteins or a method to simplify their analysis such as with the use of fluorescent amino acids. Other types of amino acids with potentially useful properties include those which are metal binding, photo crosslinking, or photocaged. Unnatural amino acids have been incorporated into proteins through methods including peptide synthesis, modification of amino acid side chains, aminoacylation of tRNA or modification of tRNA synthetase specificity.[1] The expansion of the genetic code beyond the twenty naturally occurring amino acids using these methods has been achieved in a growing number of organisms including *Escherichia coli*, yeast, and mammalian cells, although these methods can still be improved as these techniques suffer from inefficiency from low yields.[2]

During translation, a tRNA is charged with an amino acid by a tRNA synthetase. Then the tRNA can decode a codon on an mRNA with its anticodon sequence and attach the amino acid charged to the tRNA to a growing peptide chain. The most prevalent method for the incorporation of unnatural amino acids involves the addition of the genes for a modified tRNA and tRNA synthetase pair from another species into the target organism's genome. Many of the experiments use an orthogonal pair of these tRNA and tRNA synthetases specific for the addition of tyrosine.

The incorporation of unnatural amino acids into a protein typically uses a suppressor tRNA that recognizes a stop codon. An orthogonal pair of aminoacyl tRNA synthetase and a tRNA derived from a different species must not interact with the host organism's native tRNA synthetases, tRNAs, or amino acids. The orthogonal pair is

introduced into the cell through transformation of the corresponding genes. In *E. coli*, all sixty four triplet codons are used in translation. However, three of these are stop codons which terminate the polypeptide chain growth. A suppressor tRNA has an anticodon that can pair with the amber stop codon and suppress the termination of translation.[2] In *E. coli*, UAG, the amber codon is used less than the other two stop codons (UAA and UGA), so an orthogonal pair that decodes the amber codon may interfere less with the bacteria's genome decoding during translation.[1] The amino acid specificity of the orthogonal tRNA synthetase can be adjusted through directed evolution to attach a desired or targeted unnatural amino acid to its corresponding amber suppressor tRNA.

In another method, an existing *E. coli*'s tRNA was modified to lose its affinity for its natural cognate synthetase. A synthetase was then generated to charge an unnatural amino acid to the modified tRNA.[3]

It is possible to increase in unnatural amino acid incorporation in *E. coli* by knocking out the release factor 1 (RF1) that recognizes both the UAA and UAG stop codons. This knockout can survive, if the RF2 gene is expressed. Since RF2 does not recognize the UAG codon, the unnatural amino acids can be efficiently incorporated in response to the UAG codon. It also allows for multiple unnatural amino acids to be efficiently added at multiple UAG sites.[4]

Typically, orthogonal tRNA synthetase and tRNA pairs derived from archaea such as *Methanocaldococcus jannaschii* have been used in *E. coli*. The amino acid binding sites of the orthogonal tRNA synthetases are mutated to generate a library. Then, positive and negative selections are carried out in order to to isolate the synthetase variants responsible for the incorporation of unnatural amino acids. [5]

The incorporation of two or more unique unnatural amino acids instead of only one can provide even more benefits. For example, the incorporation of two different fluorescent amino acids can be used in a fluorescence resonance energy transfer study to monitor protein structures. The addition of multiple unnatural amino acids can be achieved by transforming the bacterial cells with two orthogonal pairs containing the amber and the opal suppressor tRNA and their corresponding tRNA synthetases.[6]

MATERIAL AND METHODS

General.

XL1-Blue and NEBα *E. coli* cells were used for cloning and maintaining plasmids. GH371 *E. c*oli cells were used for the genetic selection. iProof HF DNA polymerase (Bio-Rad) was used for polymerase chain reaction (PCR).

Construction of MLRS ⊿CP1 Library

The plasmid pSupK-MLRS-HL(TAG) was amplified by PCR with the primers 5'-GGAAGGGCGCCNNKNNKNNKNNKNNKNNKGAGGACCAGTGGTTCATGAAGTAC-3' and 5'-GGGCAGACGCGTTCCAAGGC-3'. The resulting PCR product was inserted between the KasI and MluI sites of pSupK-MLRS-HL(TAG) to produce a plasmid library containing MLRS ΔCP1 variants, in which the CP1 domain was replaced by a linker of randomized six triplet codons.

Selection of a functional MLRS ΔCP1 variant

The plasmid containing the MLRS ΔCP1 library was transformed into GH371 *E. coli* cells harboring the pBREP plasmid. The cells were grown in the GMML plate containing chloramphenicol (50 μg/mL). The plasmid DNA was isolated from a few highly fluorescent colonies and sequenced. A best variant was used in the subsequent experiment.

Construction of MLRS ΔCP1 N-Terminal Library

The plasmid pSupK-MLRS-HL(TAG) ΔCP1 was amplified by two separate PCR reactions – one reaction with the primers 5'-TTTACGCTTTGAGGAATCCCATATG-3' (P1) and 5'-GCATCGCACCACTGGGGTAGGGMNNMNNGACTGTGAGGAATA-TCTTTTCTCTGTC-3' (P2); the other reaction with the primers 5'-CCCTACCCCAGTGGTGCGATGC-3' (P3) and 5'-GGGCAGACGCGTTCCAAGGC-3' (P4). The purified products from two PCR reactions were mixed and used in the subsequent PCR with the primers P1 and P4. The resulting PCR product was inserted between the NdeI and MluI sites of pSupK-MLRS-HL(TAG) ΔCP1 and used to transform NEBα E. coli cells. The transformed cells are plated onto a LB plate containing kanamycin (50 μg/mL). After overnight incubation, the colonies are collected and the plasmid DNA are isolated.

RESULTS AND DISCUSSION

A pair of *Methanobacterium thermoautotrophium* leucine tRNA synthetase (MLRS) and *Halcobacterium NSP-1* leucyl-tRNA was proven orthogonal in *E. coli*.

However, this pair derived from two different archaea species did not successfully incorporate unnatural amino acids, probably due to the proofreading capability of the MLRS. Once the catalytic domain of the synthetase adds an amino acid to the corresponding tRNA, the CP1 editing domain of the MLRS validates the identity of amino acid on the tRNA and removes any amino acid other than leucine from the tRNA (Figure 1). Therefore, the MLRS without its editing domain could add a variety of amino acids to its tRNA.[7,8]

The CP1 domain of the MLRS was identified by its sequence homology with *P. horikoshii* leucyl-tRNA synthetase, another archaea synthetase whose three-dimensional structure is available in the literature (Figure 1). The DNA sequence encoding the CP1 domain was replaced by a linker of six triplet codons, all with randomized sequences of NNK, to generate a library of 10^9 possible variants of the MLRS without its editing domain (MLRS ΔCP1) (Figure 2).

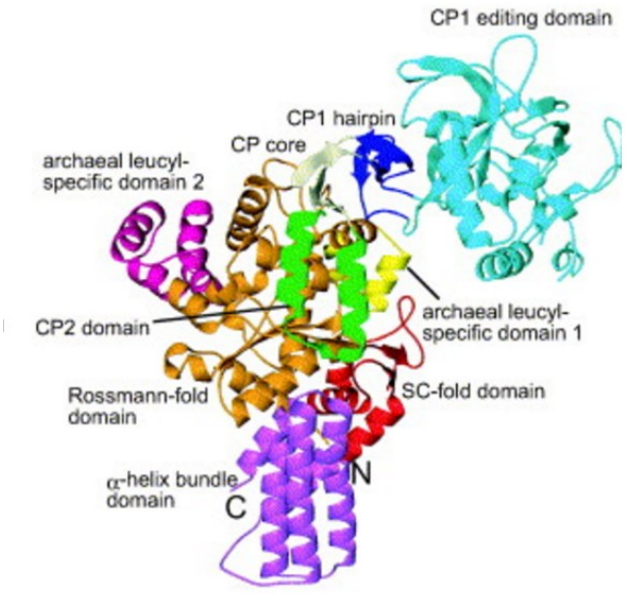


Figure 1: The structure of *P. horikoshii* LeuRS[8]

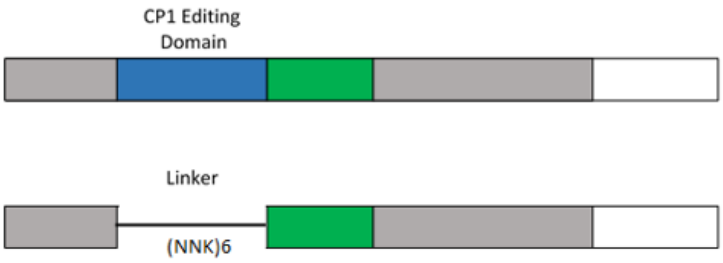


Figure 2: CP1 editing domain replaced by a randomized linker in the LeuRS gene

The library was constructed by polymerase chain reactions. The library was inserted into the pSupK-MLRS-HL(TAG) plasmid. As shown in Figure 3, the plasmid also contains the amber suppressor tRNA genes. This plasmid was then transformed into GH371 *E. coli* cells harboring the pBREP plasmid. The pBREP plasmid contains a chloramphenicol resistance gene with an amber stop codon in a permissible position (Asp112), T7 RNA polymerase gene with two amber stop codons, and a GFP gene under control of the T7 promoter (Figure 4).

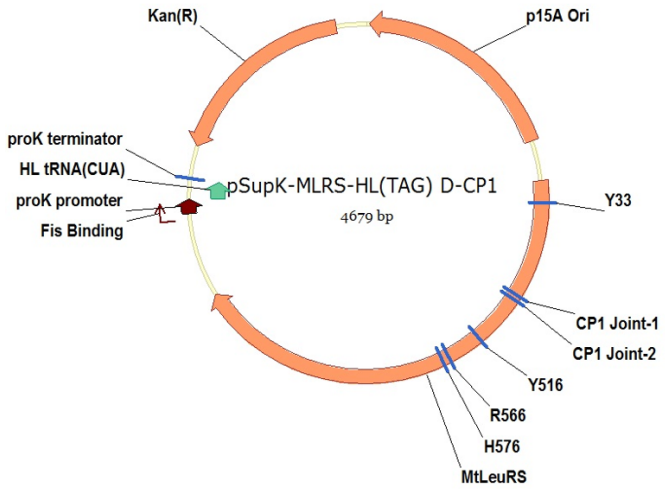


Figure 3: Plasmid map of pSupK-MLRS-HL(TAG)

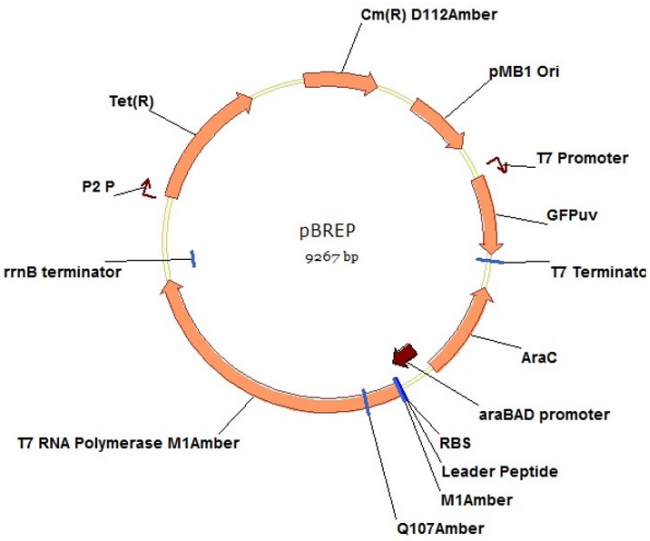


Figure 4: Plasmid map of pBREP

The T7 RNA polymerase transcribes the GFP under control of the T7 promoter. When the amber stop codons of the chloramphenicol resistance and T7 RNA polymerase genes are suppressed by the mutant LeuRS, cells survive in the presence of chloramphenicol and exhibit green fluorescence.

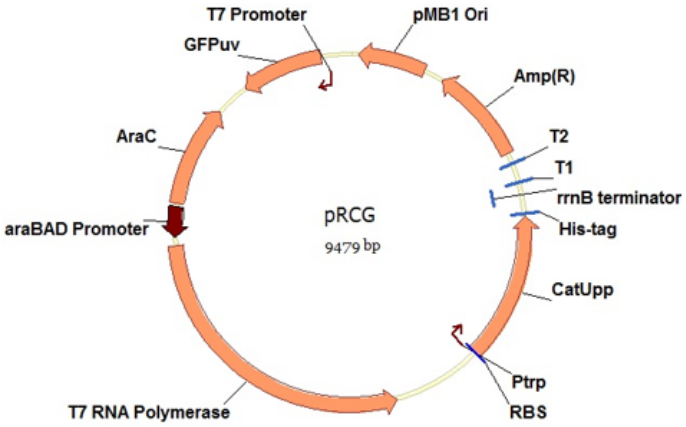


Figure 5: Plasmid map of pRCG

To select GH371 *E. coli* cells with functional LeuRS ΔCP1 genes, they were grown on chloramphenicol and tested for green fluorescence. A best variant was isolated and sequenced. The linker sequence of the LeuRS ΔCP1 gene was Leu-Tyr-His-Ala-Val-Tyr.

The MLRS ΔCP1 gene was further tested with the pRCG plasmid in the GH371 *E. coli* cells, which will be used in the selection of MLRS ΔCP1 mutants for the unnatural amino acid incorporation. The pRCG plasmid contains the CAT-UPP fusion gene in addition to the T7 RNA polymerase gene with the amber stop codons and the GFP gene under control of the T7 promoter (Figure 6). The CAT-UPP fusion gene is useful for the dual genetic selection, because the chloramphenicol resistance is tested in the positive selection, and the 5-FU sensitivity is used in the negative selection.

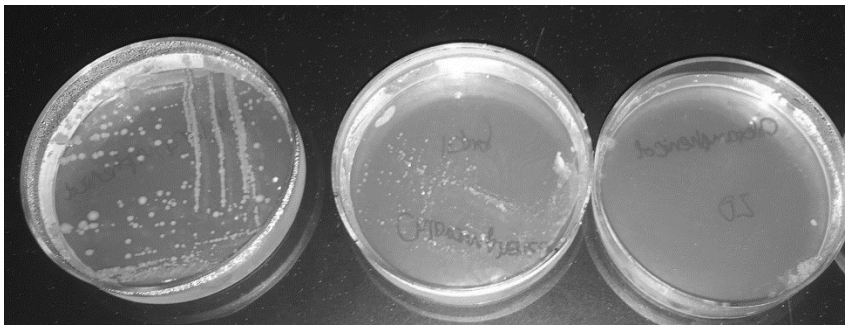


Figure 6: Minimal media plates grown with transformed GH371 cells and chloramphenicol concentrations 0, 10, and 20 μg/mL

In the positive selection, bacteria which produce functional MLRS will survive in the presence of chloramphenicol, because the chloramphenicol resistance genes contains the UAG stop codons. Without the functional MLRS, cells will die in the presence of chloramphenicol. The GH371 *E. coli* cells transformed with the pSupK-MLRS-HL(TAG) and pRCG plasmids were grown in minimal media containing different concentrations of chloramphenicol. The composition of the media is shown in Table 1.

The transformed E. coli cells GH371 showed resistance to up to 10 μg/mL of chloramphenicol (Figure 6). Therefore, any further selection experiments will be done in this concentration of chloramphenicol.

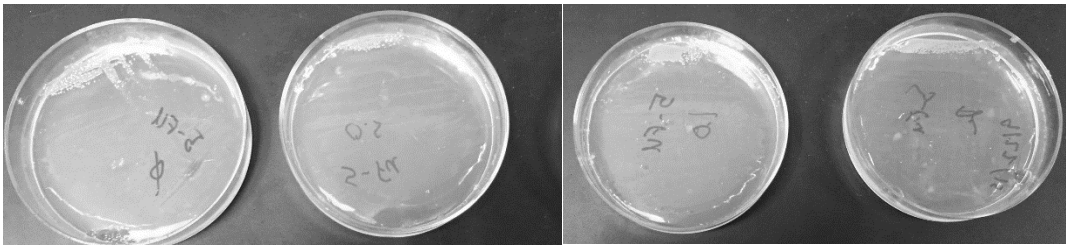
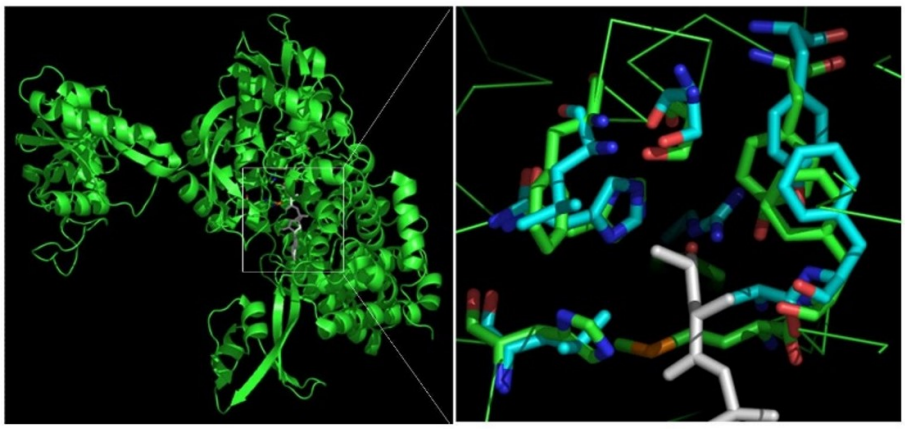


Figure 7: Minimal media plates grown with transformed GH371 cells and 5-FU concentrations 0, 0.5, 1, and 2 μg/mL (from left)

In the negative selection, the bacteria which produce MLRS ΔCP1 that charges its tRNA with unnatural amino acids will survive. However, any cells containing the MLRS ΔCP1 incorporating leucine will produce the full-length CAT-UPP fusion protein which converts 5-FU to a toxic compound leading to cell death. Since the cells are grown in the presence of natural amino acids, but not unnatural ones, those cells incorporating unnatural amino acids cannot produce the CAT-UPP fusion protein, as it is prematurely terminated. The GH371 E. coli cells containing the pRCG and pSupK MLRS-HL(TAG) plasmids were tested with concentrations up to 3 μg/mL 5-FU in minimal media and showed sensitivity at 2 μg/mL of 5-FU or higher (Figure 7).

The active site of the MLRS has been identified from the structural homology between *Thermus thermopilus* and *Pyrococcus horikoshii* leucyl-tRNA synthetases as shown in Figure 8.[8,9] To create the library of MLRS ΔCP1 variants that can incorporate unnatural amino acids, polymerase chain reactions were performed to randomize the

five amino acids in the leucine binding site. The randomization of the amino acid residues in the N-terminus and C-terminus was separately performed (Figure 9).



X-ray crystal structure of the *Thermus thermophilus* LeuRS

Structural homology in the leucine binding site between *Thermus thermophilus* LeuRS and *Pyrococcus horikoshi* LeuRS

Figure 8

For the N-terminal library, PCR was performed to randomize the A32 and Y33 positions, as these amino acids were shown to be involved in the leucine binding site (Figure 8). The N-terminal library which is about 700 base pairs long was synthesized using two PCR reactions to randomize both codons whose products are around 160 and 600 base pairs long (Figure 10) and then another PCR reaction which synthesizes the complete N-terminal library (Figure 11).

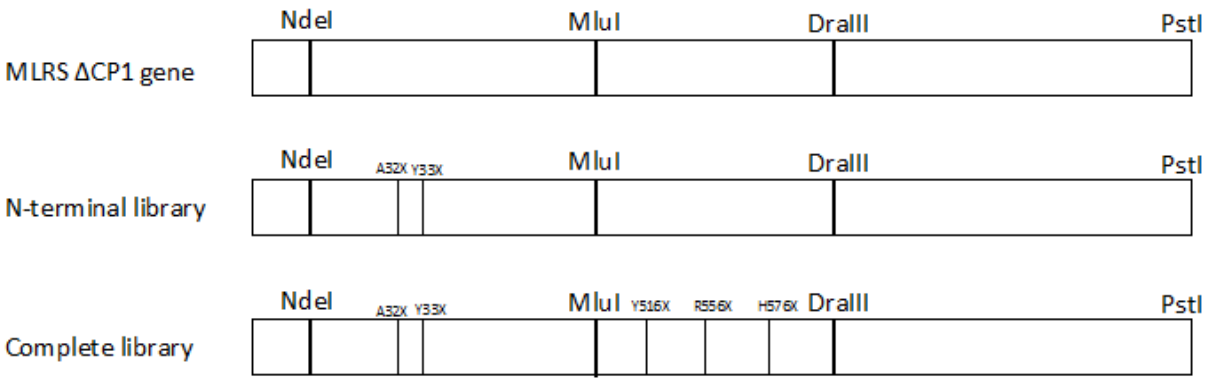


Figure 9: MLRS gene and relative locations of restriction sites and target amino acids

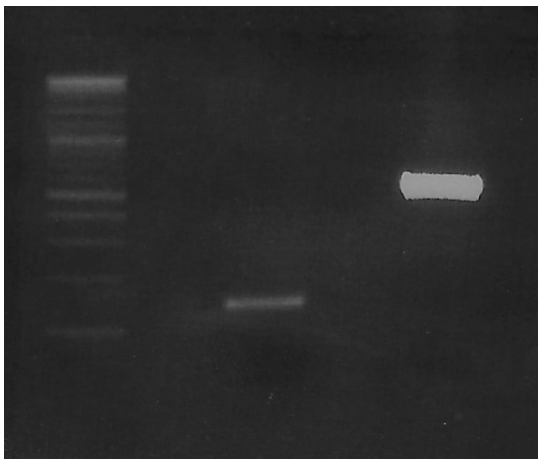


Figure 10: PCR products of the first reactions for the N-terminal library

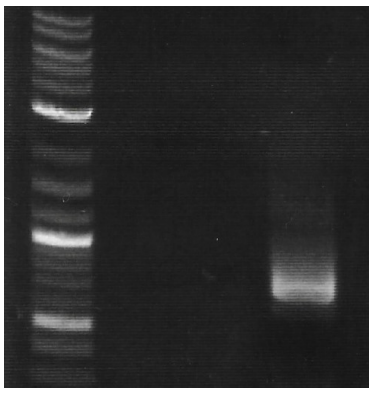


Figure 11: N-terminal library PCR products

FUTURE WORK

The N-terminal library, about 700 base pairs, is inserted between the NdeI and MluI restriction sites of pSupK-MLRS-HL(TAG) ΔCP1, which are subsequently transformed into NEB 5α cells and plated on LB agar containing kanamycin (50 μg/mL). After overnight incubation, the plasmid DNAs containing the MLRS ΔCP1 genes with the N-terminal randomization are isolated from the colonies. Three amino acids also found in the leucine binding site of LeuRS (Y516, R556, and H576) will be randomized to generate a C-terminal library, which will be inserted between the MluI and DraIII restriction sites of the N-terminal library plasmid, to generate a complete library containing five randomized triplet codons (NNK). The library containing about $3x10^7$ different MLRS ΔCP1 variants will be transformed into GH371 *E. coli* cells containing pRCG.

After the positive and negative selection for mutants, the MLRS ΔCP1 variants that can incorporate unnatural amino acids will be isolated and sequenced. The successful variants will be further tested for their efficiency in incorporating unnatural amino acids into a model protein called the Z-domain and other applications with specific proteins.

REFERENCES

1. Noren, C., Anthony-Cahill, S., Griffith, M., & Schultz, P. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. *Science, 244*, 182-188.
2. Wals K and Ovaa H (2014) Unnatural amino acid incorporation in *E. coli*: current and future applications in the design of therapeutic proteins. *Front. Chem*. **2**:15.
3. Wang, L., & Schultz, P. G. (2001). A general approach for the generation of orthogonal tRNAs. *Chemistry & Biology, 8*, 883-890.
4. Johnson, D. B. F., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Wang, L. (2011). RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nature Chemical Biology, 7*, 779.
5. Chin, J. W. (2014). Expanding and reprogramming the genetic code of cells and animals. *Annual Review of Biochemistry, 83*, 379-408.
6. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H., & Schultz, P. G. (2013). A versatile platform for single- and multiple-unnatural amino acid mutagenesis in escherichia coli. *Biochemistry, 52*, 1828-1837.
7. Yaremchuk, A., Tukalo, M., & Cusack, S. (2000). The 2 Å crystal structure of leucyl-tRNA synthetase and its complex with a leucyl-adenylate analogue. *The EMBO Journal,19*, 2351-2361.
8. Fukunaga, R., & Yokoyama, S. (2005). Crystal structure of leucyl-tRNA synthetase from the archaeon pyrococcus horikoshii reveals a novel editing domain orientation. *Journal of Molecular Biology, 346*, 57-71.