CONSTRUCTION OF MUTATED LEUCYL-tRNA SYNTHETASE AND
THE INCORPORATION OF UNNATURAL
AMINO ACIDS

By
Courtney Wright

Submitted in partial fulfillment of the
Requirements for Departmental Honors in
The Department of Chemistry & Biochemistry
Texas Christian University
Fort Worth, Texas

May 8th, 2017
CONSTRUCTION OF MUTATED LEUCYL-tRNA SYNTHETASE AND

THE INCORPORATION OF UNNATURAL

AMINO ACIDS

Project Approved:

Supervising Professor: Youngha Ryu, Ph.D.
Department of Chemistry & Biochemistry

Kayla Green, Ph.D.
Department of Chemistry & Biochemistry

Giridhar Akkaraju, Ph.D.
Department of Biology
ABSTRACT

The genetic code normally uses the canonical twenty amino acids in order to construct proteins and facilitate life. The process of translation involves an RNA template and codons that will be read and matched to corresponding tRNA molecules carrying charged amino acids. An aminoacyl tRNA synthetase (aaRS) specific to each amino acid is responsible for loading and charging the correct amino acid to the tRNA. In recent years, a few orthogonal pairs of the tRNA and aaRS have been utilized to expand the genetic code past the natural 20 amino acids. Expanding the genetic code can provide new insight into protein function, structure, and interactions within the cell. The introduction of new amino acids could lead to proteins with new chemical or biological activity and even advantageously alter function leading to evolutionary events. In our research, we attempt to incorporate unnatural amino acids using an orthogonal pair of Methanobacterium thermoautotrophicum leucyl-tRNA synthetase (MLRS) and Halobacterium sp. Leucyl tRNA. A mutant MLRS lacking an editing domain (MLRS ΔCP1) was generated. The best variant was isolated and sequenced. The leucine binding site, determined from structural homology, was randomized at five positions to create a library of mutants. In the positive selection, only the cells containing the MLRS ΔCP1 variants that add an amino acid to the tRNA will survive in the presence of chloramphenicol. In the negative selection, the cells containing the variants that add natural amino acids to the tRNA will die in the presence of 5-fluorouracil. The library can then be used for further experiments to determine how effectively unnatural amino acids are incorporated.
# Table of Contents

Introduction 4  
Methods and Materials 8  
Results and Conclusions 10  
Future Work 20  
References 21
INTRODUCTION

There has been extensive research and interest in the last decades at looking at expanding the genetic code. By incorporating amino acids past the canonical 20 amino acids, we can gain insight on protein function, structure and folding, and even generate cells with enhanced properties.¹ The incorporation of unnatural amino acids (UAAs) can also be used in fluorescence, photo crosslinking, metal binding, destructive chemical moieties, or photocaging.¹ Originally, the only alteration to the genetic code was site-specific mutations to a single amino acid, such as acetylation or various reactions with side chains, commonly cysteine and lysine because of the selective modifications of thiol and amine groups, respectively.² This was somewhat limited in scope, and the next step was to expand the genetic code by designing a cell which utilized an orthogonal pair specific to an UAA. Recently, over 70 UAAs have been successfully incorporated into various proteins using orthogonal pairs of tRNAs and their corresponding aminoacyl-tRNA synthetases (aaRS'). The tRNA molecule brings an amino acid to the ribosome and recognizes the mRNA codon. The aaRS is responsible for attaching the correct amino acid to the corresponding tRNA. To be orthogonal, the tRNA/aaRS pair must be very specific and not interact with other tRNAs and aaRS molecules present in the host organisms. In this method, all components must be metabolically stable and tolerated by the ribosome and EF-Tu, the translational factor responsible for bringing the aminoacyl-tRNA to the ribosome. It is most critical to establish specificity between the UAA, the aaRS/tRNA, and its codon.³

An aaRS/tRNA pair derived from archaea or eukaryotes could be used as an orthogonal pair in bacteria (e.g. *Escherichia coli*) because of the differences in the transcription/translation systems between the three different domains of life. An orthogonal pair that has been used to
successfully incorporate a number of different UAAs is the tyrosyl tRNA/aaRS pair from *Methanococcus jannaschii*, an archaea species. The tyrosyl tRNA has a minimalist anticodon loop and the tyrosyl-tRNA synthetase lacks the editing domain. The anticodon loop of the tRNA could be changed so that it suppresses a stop codon; in the case of *M. jannaschii* tyrosyl-tRNA, it was mutated to CUA to recognize the amber stop codon (UAG). Then, selection steps were taken to obtain the mutant with improved orthogonality. In the negative selections, the mutants recognized and aminoacylated by *E. coli* synthetases were removed, and in the positive selections only those aminoacylated by the *M. jannaschii* TyrRS were selected. The aaRS was then changed in a similar manner in order to recognize the UAA. The several amino acid residues in the tyrosine binding site was randomized to generate a library of aaRS variants. The positive and negative selections of the library yielded the aaRS variants that only charged the tRNA with the UAA of interest.\(^4\) Using this method, O-methyltyrosine was successfully incorporated into proteins in *E. coli* for the first time.\(^5\)

The amber stop codon was especially useful. In normal translation, this codon does not encode an amino acid, but instead leads to translational termination. By using one of the stop codons, interference with natural codons and the health of the host cell is minimized. Of the three stop codons, the amber codon is used least in the genome. In addition, it has been known that naturally occurring amber suppressor tRNAs can “read through” the amber stop codon to prevent termination and incorporate an amino acid.\(^6\) However, the read-through efficiency at the stop codons has been typically low because the orthogonal pairs must compete with the release factors that recognize stop codons and terminate translation. Therefore, the addition of multiple UAAs at multiple stop codons is especially challenging. However, it was recently
discovered that \textit{E. coli} were able to survive and function normally when the \textit{prfA} gene was removed. This gene encodes release factor 1 (RF1), which is responsible for termination in response to the amber stop codon. This allowed for UAAs to be incorporated at multiple sites in the mutant strain, JX33.\textsuperscript{6}

Recently, a leucine aaRS/tRNA pair derived from archaea has been proposed as a potential orthogonal pair for the incorporation of UAAs in \textit{E. coli}. The leucyl-tRNA synthetase (LeuRS), a class Ia aaRS, is one of three aminoacyl tRNA synthetases with editing mechanisms to remove any mischarged amino acid from the tRNA. Interestingly, the bacterial LeuRS does not recognize the long arm or the anticodon triplet of tRNA. Instead, it recognizes the D arm and the A73 discriminator nucleotide of tRNA. However, the archaea or eukaryotic LeuRS recognizes the A73 and the long arm of their corresponding tRNAs.\textsuperscript{7} The CPI domain, about 200 residues, of the LeuRS has been identified as the distinct editing domain in archaea LeuRS molecules.

For further expansion of the genetic code, it is desirable for the orthogonal pairs to recognize codons beyond the amber codon (UAG) such as four-base codons or the opal stop codon (UGA). It was demonstrated that the orthogonal pair of the LeuRS from \textit{Methanobacterium thermoautotrophicum} and the leucyl tRNAs from \textit{Halobacterium sp. NRC-1}, could recognize the amber and opal stop codons, as well as four-base codons.\textsuperscript{8}
Methods and Materials

General

GH371 and DH10B *E. coli* cells were used to for cloning and maintaining plasmids.

GH371 *E. coli* cells were used in the genetic selections. iProof HF DNA Polymerase (Bio-Rad) was used for polymerase chain reactions (PCR), NEB restriction enzymes were used for digestion, and NEB T4 DNA Ligase was used for ligation.

Construction and Selection of functional MLRS △CP1 variant

The MLRS △CP1 variant used in construction of the N and C-terminal libraries was constructed by PCR using the primers 5’-

GGAAGGGCGCCNNKNNKNNKNKNNKNNKNNKNNKNNKNGAGGACCAGTGGTTCATGAAGTAC-3’ and 5’-

GGGCAGACGCGTTCCAAGGC-3’.

This established a linker of randomized six triplet codons in the CPI domain of the plasmid pSupK-MLRS-HL(TAG). The PCR product was then inserted between the KasI and MluI restriction sites of the plasmid. To determine the best mutant the MLRS △CP1 library in the pSupK-MLRS-HL(TAG) plasmid was transformed into GH371 *E. coli* cells containing the pBREP plasmid, grown on GMML plates containing chloramphenicol (50 μg/mL). Highly fluorescent colonies were selected and sequenced.9

Construction of the MLRS △CP1 N-Terminal Library

In order to generate the N-Terminal library the pSupK-MLRS-HL(TAG) △CP1 plasmid was amplified in a two-step PCR process. The first step involved two reactions: one using primers 5’-

TTTACGCTTTTGAGGAATCCCATG-3’ (P1) and 5’-
GCATCGCACCAGGGTGAGGGMNNNGACTGTGAGGAATA-TCTTTTCTCTGTC-3’ (P2), and the second using primers 5’-CCCTACCCCAGTGGTGCAGTGC-3’ (P3) and 5’-GGGCAGACGCCTCAGGCG-3’ (P4). The PCR products of the two reactions were purified and mixed in a third round of PCR using the original P1 and P4 primers. The PCR product was purified and inserted into the pSupK-MLRS-HL(TAG)ΔCP1 between the NdeI and MluI restriction sites. The plasmid was then transformed into DH10B E. coli cells and plated on agar plates containing kanamycin (50 μg/mL). After incubation at 37 °C overnight the colonies were collected and the plasmid DNA was isolated.

Construction of the MLRS ΔCP1 C-Terminal Library

To construct the C-terminal insert a two-step PCR procedure was used. In the first reaction, there were four primers used as the template: MLRS OEPCR MluI P1 5’-CTTGGGAACGCGTCTGCCGGATG-3’, MLRS OEPCR Y516 P2 5’-CATCGAACCCCTCACAGACTCAACAA-3’, MLRS P3 5’-CAAACAACTCATCGTCCATCTCCGGCAA-3’, and MLRS P4 5’-CCGGAATTCCTCCCTGAGATCCTCAAGG-3’. In the second reaction four primers were also used as a template to generate the product: MLRS P5 5’-GGATCTCAGGGAGGAATTCCGTACTGGT-3’, MLRS P6 5’-CTCTGCAAAGGACCTCATAGGCAATNNKC-3’, MLRS P7 5’-CCACAGCACCACCGTGCGCCACCCCTGACTCA-3’, and MLRS P8 5’-CCACAGCACCACCGTGCCACC-3’. The resulting PCR products were purified and used as template for a subsequent round of PCR using primers P1 and P8. The product was purified and inserted between the restriction sites MluI and DralI.
Another method was also used to generate the insert using alternate primers. The first reaction used pSupK-MLRS-HL(TAG) ΔCP1 as template with primers MLRS PCR DraIII F 5’-
GTCAGGGTGCCACGCGGTGCTGGTGG-3’ and MLRS PCR AflII R 5’-
CCCTCAACACGCTTAAGGTAGTGG-3’. The PCR product was purified and used in a subsequent reaction as template along with the C-terminal insert created. P1 and MLRS PCR AflII R were used as primers. The PCR product was purified, digested with MluI and AflII, and inserted into pSupK-MLRS-HL(TAG) ΔCP1 N-terminal library. The plasmid was transformed into GH371 E.coli cells. The cells were spread onto agar plates containing kanamycin (50 μg/mL) and allowed to incubate overnight.

Results and Conclusions

The structural homology of two LeuRS from *T. thermophilus* and *P. horikoshii* was used to identify the CPI editing domain (Figure 1). This was crucial because the two LeuRS’ are not sequentially homologous. From the sequence homology with *P. horikoshii* LeuRS, the CP1 domain of *Methanobacterium thermoautotrophium* LeuRS (MLRS) was subsequently identified. Without the CPI editing domain the LeuRS should theoretically be able to add numerous UAAs without removing the “incorrect” amino acid. Therefore, the CPI editing domain of the MLRS gene in the pSupK-MLRS-HL(TAG) plasmid (Figure 2) was replaced with a six-codon linker by PCR to create a library of 10⁹ variants (MLRS ΔCP1) (Figure 3). The pSupK-MLRS-HL(TAG) plasmid also contains the amber suppressor tRNA genes and a kanamycin resistance gene. The plasmid library was 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, a T7 RNA polymerase gene containing an amber stop codon, and a GFP gene under the control of the T7 promoter. If the amber codons are successfully suppressed in both genes, the cells are resistant to chloramphenicol and express T7 RNA polymerase, which can then transcribe the GFP gene leading to fluorescent colonies. The selection process was repeated until the chloramphenicol resistant and fluorescent colonies were isolated and sequenced. The result was a linker sequence of Leu-Tyr-His-Ala-Val-Tyr. This variant was then used to generate the N-terminal and C-terminal libraries.

Figure 1: Structural homology between *T. thermophilus* and *P. horikoshi* leucyl-tRNA synthetases.
Figure 2: The map of the pSupK-MLRS-HL(TAG) ΔCP1 plasmid.

Figure 3: Six codon linker sequence used to randomize the CP1 editing domain

Using structural homology between *T. thermophilus* and *P. horikoshi* leucyl-tRNA synthetases the leucine binding site was also identified (Figure 1). The N-terminal library was first generated by randomizing the A32 and Y33 positions. This was done by overlapping extension PCR using synthetic oligonucleotides containing the intended changes. The N-
terminal library PCR product should be about 700 bp in length and contains the Ndel and Mlul restriction sites (Figure 5). The PCR product was inserted into pSupK-MLRS HL(TAG) ΔCP1 between the Ndel and Mlul restriction sites (Figure 6) using T4 DNA ligase. The resulting product was transformed into DH10B E. coli cells to generate the number of colonies to cover all $10^3$ possible variants since a NNK sequence was used to randomize each of the two amino acid positions (Figure 7). The plasmid DNA was isolated from the colonies and further used to generate the C-terminal library.

Figure 4: PCR products of the first round of reactions for N-terminal library.
Figure 5: Second round PCR product for N-terminal library.

Figure 6: Digestion of pSupK-MLRS-HL(TAG) ΔCPI and N-terminal insert with MluI and Ndel.
Figure 7: N-Terminal library in DH10B E. coli cells.

<table>
<thead>
<tr>
<th>MLRS ΔCP1 gene</th>
<th>Ndel</th>
<th>MluI</th>
<th>DraIII</th>
<th>AflII</th>
<th>PstI</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-terminal library</td>
<td>Ndel</td>
<td>MluI</td>
<td>DraIII</td>
<td>AflII</td>
<td>PstI</td>
</tr>
<tr>
<td>Complete library</td>
<td>Ndel</td>
<td>MluI</td>
<td>DraIII</td>
<td>AflII</td>
<td>PstI</td>
</tr>
</tbody>
</table>

Image 8: Mutation and restriction sites of MLRS ΔCP1.

The C-terminal library was constructed by overlapping extension PCR using oligonucleotides containing the intended mutations. The library was randomized at the Y516,
R556, and H576 positions between the MluI and DraIII sites. The C-terminal fragment about 300 bp was generated using the overlapping extension PCR (Figure 11). The PCR product was then inserted between the MluI and DraIII restriction sites using T4 DNA ligase. The ligation product was transformed into GH371 E. coli cells harboring the pRCG plasmid.

The pRCG plasmid (Figure 9) contains the CatUpp fusion gene containing an amber stop codon in a permissible position, the T7 RNA polymerase gene with amber stop codons, and the GFP gene under the control of the T7 promotort. In the positive selection, the MLRS ΔCP1 variants that effectively suppress the amber stop codon within the chloramphenicol resistance gene will survive, and those that lack functional MLRS will die in the presence of chloramphenicol. Those that suppress the amber stop codons also express T7 RNA polymerase which can then transcribe the GFP gene making the colonies fluoresce. When grown in minimal media, the GH371 E. coli cells showed resistance up to 10 \( \mu \text{g/mL} \) of chloramphenicol. Therefore, this concentration will be used in further experiments.

---

**Figure 9:** pRCG plasmid used in positive and negative selections.
In the negative selection only natural amino acids are included in the media. If the MLRS charges a natural amino acid to the tRNA instead of an UAA, the full CatUpp fusion protein is produced. In the presence of 5-fluorouracil (5-FU), the Upp protein converts it to a toxic compound leading to cell death. Therefore, only MLRS ΔCP1 that charge UAAs will survive because the CatUpp fusion protein will not be produced. The cells containing the MLRS plasmid were tested with minimal media and varying 5-FU concentrations. The GH371 E. coli cells were sensitive at concentrations of 2 µg/mL of 5-FU or higher.⁹

Initially, no colonies were found with both electroporation and chemical transformation of the ligation mixture. A DNA band about 300 bp, the anticipated size of the insert, was not visible on the gel (Figure 12). Test digestions were also done in a stepwise manner on both the pSupK MLRS HL(TAG) ΔCP1 and N-terminal library (Figure 13). The plasmids were digested with MluI, purified, and digested with DraIII. There were still no bands visible at the expected length.

![Image](image.jpg)

**Figure 10:** PCR products of first round reactions in generating C-terminal library.
Figure 11: Second round PCR product to generate C-terminal mutations.

Figure 12: Digestion of N-terminal library and C-terminal inserts with Mlul and Dralll.
Figure 13: Test stepwise digestion of pSupk-MLRS-HL(TAG) ΔCP1 and N-terminal library with MluI and DraIII.

Due to the small size of the insert (about 300 bp), PCR was used to extend the fragment length to utilize the AflII restriction site. PCR was first performed with the DraIII F primer and AflII R primer to generate a fragment of approximately 400 bp. The fragment was used as a template along with the C-terminal PCR fragment in a 1:1 ratio. This resulted in a final PCR product, about 700 bp in length, containing the MluI and AflII restriction sites (Figure 14). The longer C-terminal library was inserted into N-Terminal library plasmid at the MluI and AflII sites and transformed into GH371 electrocompetent *E. coli* cells.
Figure 14: PCR product using C-terminal and DraIII/AflII fragment as template.

Figure 15: N and C-Terminal Library in GH371 E. coli on kanamycin plates.
In conclusion, we produced a library of LeuRS mutants which lacked a CPI editing domain and were randomized in five positions, A32, Y33, Y516, R556, and H576, in the active site. With this molecular library of MLRS ΔCP1 variants we will begin the selection experiment to identify the MLRS variants that can incorporate UAAs into proteins in *E. coli*, using the established selection conditions. The incorporation of UAAs will be first tested with a model proteins and will be further applied to various important proteins to study their structural and functional changes.

**Future Work**

The transformation efficiency of the C-terminal library needs to be improved. It is important that the transformation efficiency is increased so that the number of colonies are large enough to cover all possible variants. Since the active site is mutated at five sites, each with the NNK sequence, the number of colonies expected to cover should be about 33 million. In order to do this, fresh electrocompetent GH371 *E.coli* will be made. The co-precipitation with yeast tRNA extract in ethanol will be also used. The library will be collected and selected for using chloramphenicol and 5-FU. The MLRS ΔCPI variants that effectively charge unnatural amino acids and suppress the amber stop codons will survive. The best mutants will be selected and sequenced and tested further by using them in the synthesis of a model protein called the Z-domain.
References


