STUDIES ON PYRROLYSINE – THE 22ND AMINO ACID

BY

JAYLONG WONG THONGKHAM

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree

With Honors in

Biochemistry and Molecular Biophysics

THE UNIVERSITY OF ARIZONA

MAY 2010

Approved by:

V. Bandarian

Dr. Vahe Bandarian

Department of Biochemistry and Molecular Biophysics

The University of Arizona Electronic Theses and Dissertations
Reproduction and Distribution Rights Form

Name (Last, First, Middle) Thongkham, Jaylong Wong	
Degree title (eg BA, BS, BSE, BSB, BFA): BS	
Honors area (eg Molecular and Cellular Biology, English, Studio Art): Biochemistry & Molecular Biophysics	
Date thesis submitted to Honors College: 5/5/2010	
Title of Honors thesis: Studies on Pyrrolysine - the 22nd Amino Acid	
:The University of Arizona Library Release	I hereby grant to the University of Arizona Library the nonexclusive worldwide right to reproduce and distribute my dissertation or thesis and abstract (herein, the "licensed materials"), in whole or in part, in any and all media of distribution and in any format in existence now or developed in the future. I represent and warrant to the University of Arizona that the licensed materials are my original work, that I am the sole owner of all rights in and to the licensed materials, and that none of the licensed materials infringe or violate the rights of others. I further represent that I have obtained all necessary rights to permit the University of Arizona Library to reproduce and distribute any nonpublic third party software necessary to access, display, run or print my dissertation or thesis. I acknowledge that University of Arizona Library may elect not to distribute my dissertation or thesis in digital format if, in its reasonable judgment, it believes all such rights have not been secured. Signed: _____ Date: 5/5/2010

Last updated: Nov 15, 2009

ABSTRACT

Recently, a 22nd amino acid has been discovered, pyrrolysine. Pyrrolysine (pyl) is coded by the translation of a amber (UAG) codon. Pyl was first found in the active site for genes encoding for methylamine methyltransferases in *Methanosarcinaceae*. New discoveries have now identified pyl in the Gram-positive *Desulfitobacterium hafniense*, but the pylS gene is divided in half – one corresponding to the C-terminal and one corresponding to the N-terminal of archaeal pylS. The pylBCD genes are responsible for the biosynthesis of pyrrolysine. Each pyl gene was cloned into *E. coli* to overexpress the protein in attempts to run kinetic reactions to determine the biosynthetic mechanism of pyrrolysine. The pylB and pylC genes have been successfully cloned into *E. coli* and proteins will be overexpressed. However, the pylD gene could not be ligated into pET28a vector. The pylB and pylC proteins will be expressed and purified.

INTRODUCTION

A naturally occurring 22nd amino acid, pyrrolysine, has been found in *Methanosarcinaceae*. One of the first signs to the existence of pyrrolysine was the translation of an in-frame amber codon in genes that codes for a methylamine methyltransferases in *Methanosarcina barkeri*. The pyrrolysine-tRNA is different from the original twenty amino acid tRNAs. Rather than being encoded by a sense codon, pyrrolysine has been shown to be translated by the UAG stop codon (1, 2).

Pyrrolysine has been documented in methylamine methyltransferase reactions, which initiate methane formation in Achaea. The archaeal methanogens that possess the methylamine methyltransferase genes require five *pyl* genes for the biosynthesis of pyrrolysine (3). In *Methanosarcinaceae*, pylT encodes tRNA^{pyl}, pylS encodes the pyrrolysyl-tRNA synthetase, and pylB, pylC, and pylD encode proteins for the biosynthesis of pyrrolysine. The pylBCD genes in *M. acetivorans* are believed to have similarities to radical SAM proteins, amide forming proteins, and amino acid dehydrogenases. Knock out experiments have shown that pylBCD are all essential to the biosynthesis of pyrrolysine (4).

The most recent species of bacterium to contain pyrrolysine was a Gram-positive *Desulfitobacterium hafniense*. This bacterial specie possesses pyl gene homologs to that of the *Methanosarcinaceae*, signifying a horizontal gene transfer involving the entire pyl gene cassette (3, 4). The main difference between archaeal and bacterial pyrrolysyl-tRNA synthetase (pylS) is in *D. hafniense*, the pylS gene is split into pylSc and pylSn. The usual pylS gene is located downstream of pylT but upstream pylBCD. The *D. hafnienses* pylSc is located between pylT and pylB and pylSn is located after pylBCD. Sequencing has shown that pylSc has a ~60% similarity

to the C-terminal of archaeal pylS, and pylSn has a ~50% similarity to the N-terminal of archaeal pylS (3).

Pyrrolysine is in the active site of methylamine methyltransferases. The role of methylamine methyltransferases is to methylate cognate corrinoid proteins. The best known group of corrinoid proteins is the cobalamins, such as vitamin B12 (5). In bacteria, cobalamins are used as cofactors for methionine synthase. Cobalamins could also be used as intermediates for methyl transfers (6).

There is significance in exploring pyrrolysine because it is the 22nd amino acid. There is potential to use pylS nd tRNA^{pyl} as an orthologous pair in other bacteria to make recombinant proteins with vast new properties (3). There is also value in understanding the mechanistic biosynthesis of pyrrolysine to determine whether the mechanism involves the use of radical SAM proteins or other types of radical chemistry.

MATERIALS AND METHODS

Plasmid Construction

Using polymerase chain reaction, the genes pylB, pylC, and pylD were amplified from *M. acetivorans* C2A genome. In the *M. acetivorans* genome, pylB and pylC have start codons beginning with TTG; therefore, the start codon had to be changed to ATG to be viable for *E. coli*. In addition to changing the start codon, bases CAT were added to the forward primer to create an NdeI restriction enzyme site (Table 1). The reverse primer was also modified by adding an additional TTA stop codon and adding AAGC for the terminal HindIII restriction enzyme site. Each primer was prepared to be approximately thirty bases long with the primer ending with a guanine or cytosine. Each corresponding forward and reverse primer were lengthened to have similar annealing points (within 3-5˚C). The 50µl PCR reaction included 5µl dimethyl sulfoxide (DMSO).

Gene transformation into E. coli

Each of the pyl genes were transformed into *E. coli* cells separately. Each pyl gene was first ligated into pGEM-T Easy Vector then transformed into XL-1 Blue electrocompetent cells for blue-white colony screening. The cells were plated on LB/agar with 100mg/ml Ampicillin (AMP), using X-Gal and Isopropyl β-D-1-thiogalactopyranoside (IPTG) to inoculate the cells to grow overnight at 37˚C. Isolated white colonies were picked and grown in 4mL LB/AMP overnight at 37˚C. The 4ml growth was mini-prepped to extract the plasmid DNA. The plasmids were digested with NdeI and HindIII to screen for the colonies that contained the correct pyl gene. The correct DNA fragment was excised and used to continue the gene translation.

The pyl genes were then ligated into pET28a in different ratios of insert:vector. The ligations were transformed into JM109 electrocompetent cells and XL-1 Blue electrocompetent cells and grown on a plate with LB, agar, and 100mg/ml Kanamyacin (KAN) antibiotics. Isolated colonies were picked and grown in 4ml LB and KAN overnight in 37˚C. The cells were mini-prepped to extract the DNA and screened to verify the presence of the pyl genes.

Making the pET28a Vector

Using stock pET28a plasmid, NdeI and HindIII restriction enzymes were used to cut adhesive ends that correlate to the ends of the pyl gene insert DNA. After exposing the plasmid to the restriction enzymes, calf intestine Alkaline Phosphatase (AP) to remove the 5' phosphates to inhibit any chances or re-annealing. Another set of pET28a stock was made by removing a known insert from stock DNA from previously cloned genes. The inserts were removed with NdeI and HindIII restriction enzyme digests and Alkaline Phosphatase treatment.

Expression of the pyl Genes

After picking one colony from the lot of successfully ligated pyl gene in pET28a, 1µl of the plasmid was transformed into BL21(DE3) electrocompetent cells for protein expression. After transformation, only a loop full of the cells was spread on a plate containing LB, agar, and 100mg/ml KAN. One colony was picked and grown in 5ml LB and 5µl KAN overnight at 37˚C to prepare for a 50ml growth. Enough of the 5ml overnight was added to the 50ml liquid LB with KAN antibiotics to have a starting OD_{600} = 0.05 for the growth experiment. Once the OD_{600} = 0.5, then the cells were induced with IPTG. Cell samples were collected at designated times of 0hrs, 2hrs, 4hrs, 8hrs, and overnight. The collected cells were spun down, discarding the supernatant, and stored in the freezer. The cell pellets were resuspended in 10mM Tris buffer

pH= 8.0 and then sonicated to lyse the cells. A SDS-Page gel was run to verify the expression of the target pyl protein.

When the presence of soluble protein is detected, a 6L growth is performed to express the pyl protein and to purify each protein. Protein purification is done by running the lysed cell solution through a HisTrap HP, 1ml nickel column. The binding buffer used was 20mM sodium phosphate, 0.5M NaCl, 20-40mM imidazole, pH 7.4 and the elution buffer was 20mM sodium phosphate, 0.5M NaCl, 50mM imidazole, pH 7.4. A nickel column is used because the pyl proteins are His-tagged.

RESULTS

Using the forward and reverse primers for each pyl gene, the PCR products are shown in Figure 1. The band for pylB is at approximately 1.05kbp, corresponding to the 1052bp, the actual length of the pylB gene. The pylC band is at approximately 1.1kbp, corresponding to 1091bp, the actual length of the pylC gene. The pylD band is approximately 0.8kbp, corresponding to 779bp, the actual length of the pylD gene. The bands noted with the black dots were excised and extracted with QiaGen gel extraction kit to yield viable plasmids for ligations.

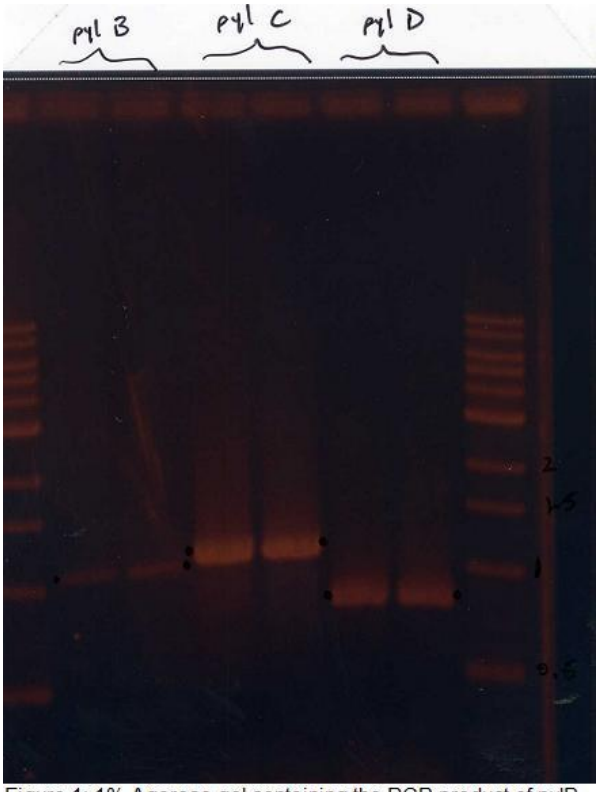

Figure 1: 1% Agarose gel containing the PCR product of pylB, pylC, and pylD genes amplified from M. acetivorans C2A.

Each PCR product was extended with a poly-Adenosine before ligating each pyl insert into pGEM-T Easy Vector. After ligating the pyl genes into the pGEM vector, the cells were transformed into XL-1 Blue electrocompetent cells for blue and white screening. The first time the transformation was performed, no colonies grew on any of the plates. When using a new batch of XL-1 Blue cells, the same transformations was performed, and many colonies were grew on the plates containing AMP/X-gal/IPTG. The plate with pylB transformed into XL-1 Blue had 20 white colonies and 19 blue colonies. The transformation of pylC into XL-1 Blue had 15 white colonies and 25 blue colonies. The pylD transformation into XL-1 Blue had 27 white colonies and 23 blue colonies. Ten white colonies from each plate were picked and grew in 4ml

LB/KAN. After mini-prepping the cells, the plasmid DNA was digested with NdeI and HindIII to verify the presense of pyl gene insert. The highest band correlates to a singly cut pylB + pGEM vector containing ~4kbp. The middle band correlates to only the pGEM vector which is ~3kbp. The lowest band is the pyl gene insert varying in length. The pylB agarose gel shows that lanes 1, 2, 4, 6, 9, 10, 11, and 12 all have the pylB gene insert (Figure 2). The pylC agarose gel shows that lanes 1, 2, 3, 6, 7, 8, 10, 11, and 12 contain the pylC gene insert in the gGEM plasmid (Figure 3). The pylD agarose gel shows that lanes 1, 3, 4, 5, 7, 8, 9, 10, and 11 contain the pylD gene insert (Figure 4).

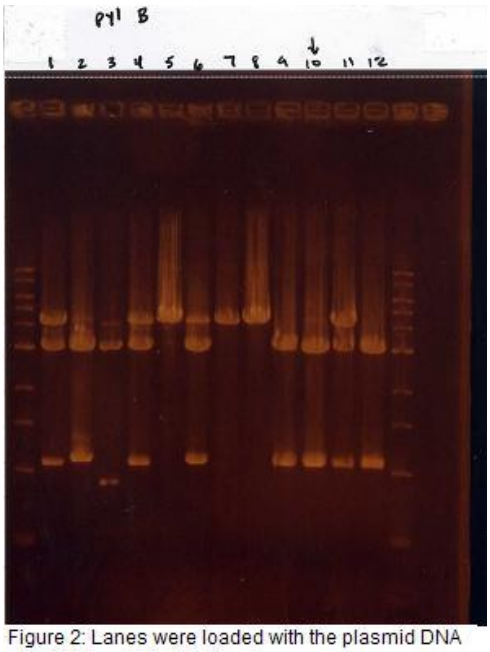

Figure 2: Lanes were loaded with the plasmid DNA after NdeI and HindIII digest.

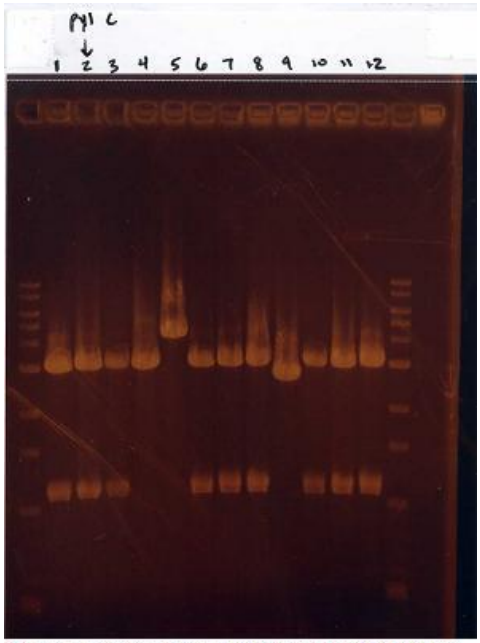

Figure 3: Agarose gel after NdeI and HindIII digest of the plasmid DNA of the pylC ligation into pGEM.

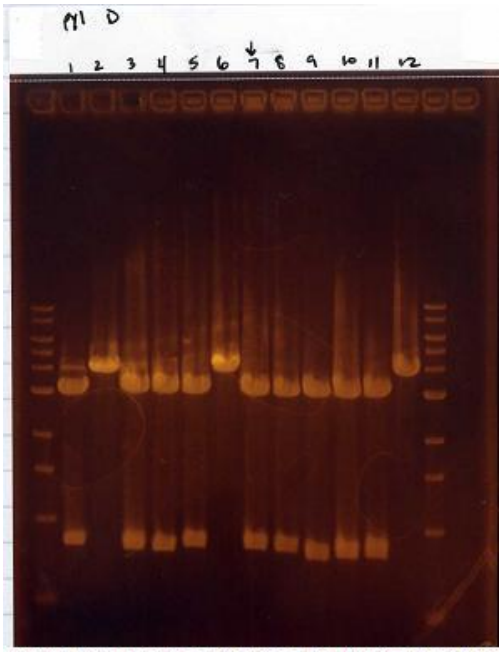

Figure 4: Agarose gel after NdeI and HindIII digest of the plasmid DNA of the pylD ligation into pGEM.

The plasmids identified by the small black arrow represented the plasmid that was used to carry out the rest of the experiment. Lane 10 from pylB, lane 2 from pylC, lane 7 from pylD were used for the ligation into pET28a.

A large scale digest was performed to use as the stock of pyl gene insert for further reactions. The pyl genes were then ligated into pET28a in insert:vector ratios of 1:1 and 1:2. After transforming the plasmids into XL-1 Blue cell and plating, a lawn of cells grew that did not resemble that of *E. coli*. With possible contamination issues with the XL-1 Blue strain, an alternative strain, JM109 electrocompetent cells, were used instead. The pyl genes were transformed into the JM109 cells, but no colonies grew on the LB plates with KAN antibiotics.

After many trials and errors, a new stock of XL-1 Blue cells was used for the pyl gene ligation into pET28a. The cells containing the pylB gene had ten colonies grow on the LB/KAN plate. All ten colonies were picked and grown in 4ml LB/KAN then mini-prepped and screened on an agarose gel. The mini-prep plasmids were digested with NdeI and HindIII. Lanes 1, 2, 4, 5, 6, 7, 8, and 9 all contained the pylB gene in the pET28a plasmid (Figure 5).

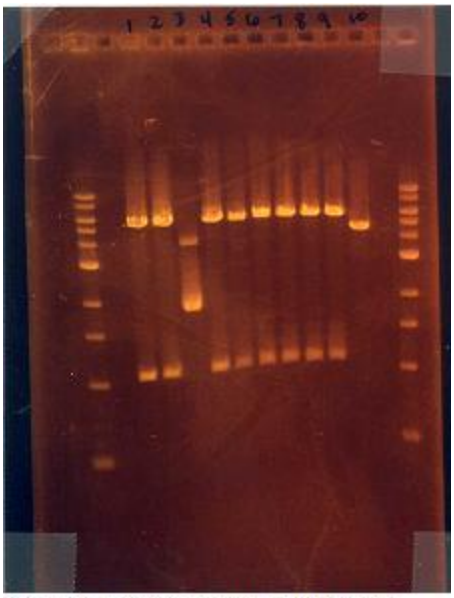

Figure 5: Agarose gel after NdeI and HindIII digest of pylB insert and pET28a vector.

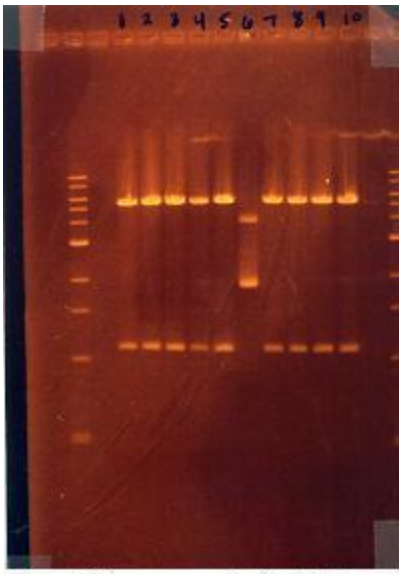

Figure 6: Agarose gel after NdeI and HindIII digest of pylC gene ligated in pET28a vector.

The ligated pylC+pET28a plasmid was transformed into the new XL-1 Blue cells and yielded 14 colonies on the LB/KAN plate. Ten colonies were picked and digested with NdeI and HindIII to screen for the presence of pylC in the pET28a vector. Lanes 1, 2, 3, 4, 6, 7, 8, and 9 all contained the ligated pylC gene (Figure 6).

The pylD ligation into pET28a vector was not as successful as the ones exhibited by pylB and pylC. The first attempt of ligation and transformation only yielded one colony, but the screen revealed no sign of the pylD gene. Different insert:vector ratios were utilized, and different vectors were used in attempts to find a successful transformation. None of the combinations resulted in a successful ligation. Figure 7 shows two different ratios of pylD insert: pET28a vector, all having unsuccessful ligations. Since the ligation of pylD into pET28a was not successful, the expression of pylD was unable to be carried out.

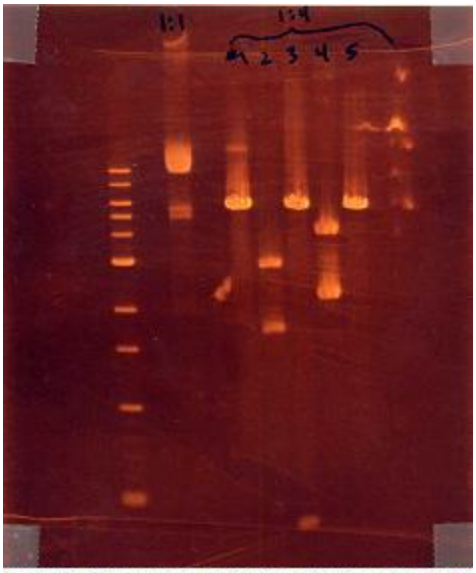

Figure 7: Agarose gel after NdeI and HindIII digestion of the attempted ligation of the pylD gene and the pET28a vector.

Protein expression was carried out with the pylB and pylC genes. The plasmids from lane 2 of the pylB ligation and lane 1 of the pylC ligation were used for transformation into BL21(DE3) cells. Two loops full of the transformed cells were plated on LB/KAN and a lawn was observed for both the pylB and pylC transformations. A single, isolated colony from each plate was picked and used for protein over-expression of both pylB and pylC proteins. A 50ml growth was done to determine if the protein was soluble. The pylB protein was observed to express after only being induced by IPTG, but the protein was not soluble (Figure 8). The weight of the pylB protein is 39513Da, but this pylB protein has a His-Tag, increasing the weight slightly.

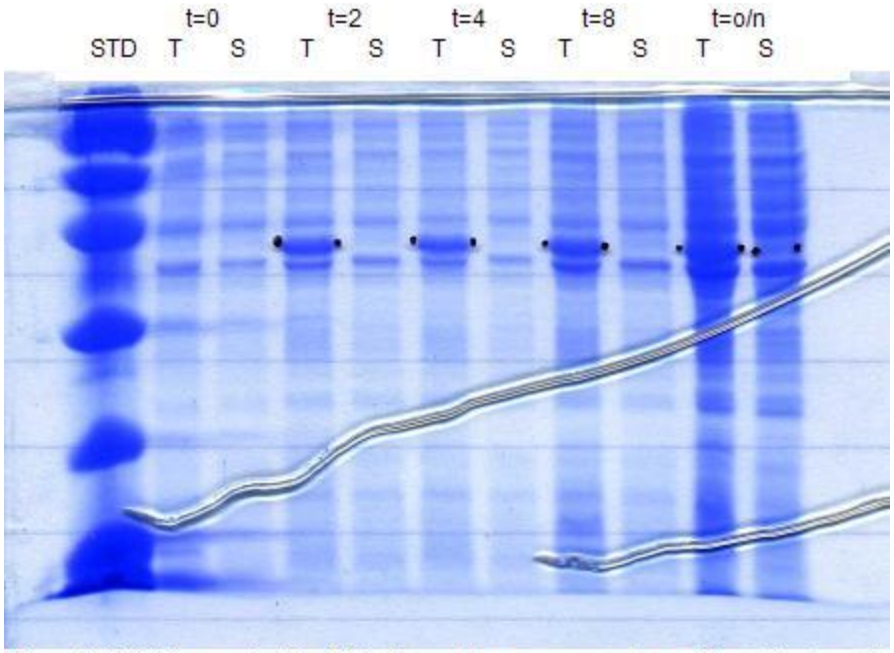

Figure 8: SDS-Page gel of the first attempt to overexpress the pylB protein by only induction with IPTG. T is for total lysed cell solution, and S is for only soluble proteins.

The pylB protein is indicated by the black dots, and it was only observed in the T (total) portion of the lysed solution. There are trace amounts found in the soluble portion only in the overnight

growth. A second growth was performed, but $Fe^{III}Cl_3$ was added to the growth media when the cells were induced with IPTG. With Fe^{III} in the growth media, there was observable pylB protein in the soluble portion of the expression (Figure 9).

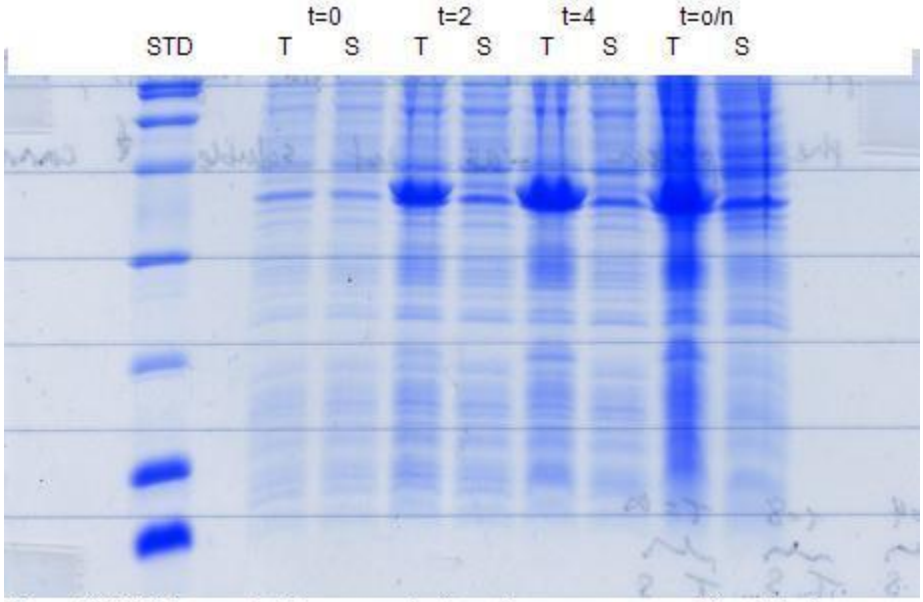

Figure 9: SDS-Page gel of the second attempt to overexpress pylB protein by inducing the cells with IPTG and adding Fe(III)Cl3 to the growth media. T is for total lysed solution, and S is for only soluble protein.

Over-expression of pylC was carried out by only inducing the cells with IPTG. The pylC protein is indicated by the black dots. The pylC protein was found to be in both the total and the soluble portions (Figure 10). The weight of pylC is 37546Da, and because the pylC has a His-Tag, it is slightly heavier.

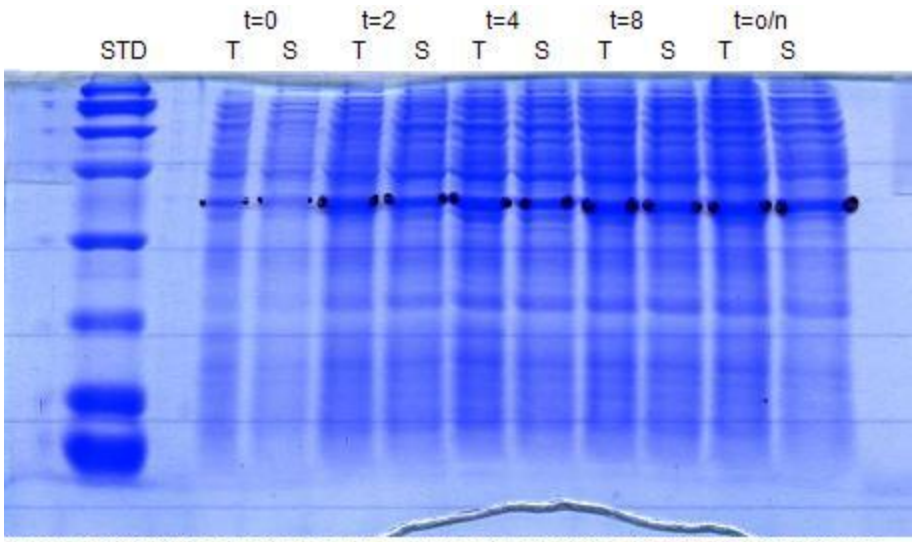

Figure 10: SDS-Page gel of the overexpression of pylC in BL21(DE3) by inducing the cells with IPTG. T is for the total lysed solution, and S is for only soluble protein.

Table 1

pylB:

for: *CAT**ATG***ATCAAAAAAATGGCAACCGAGGACCTTGACAGG T_M=70.0˚C

rev: *AAGCTT***CTACTA**GCACCTCAGGACAGTTTCGAATTCTGCCTG T_M=72.5˚C

pylC:

for: *CAT**ATG***AAAACCATATGCCTTATAGGCGGGAAACTCCAG T_M=70.0˚C

rev: *AAGC**TT***ATTACACAGCCGCTCCGAAGCGGTCTTTCAG T_M=72.4˚C

pylD:

for: *CAT**ATG***GCACTTTTAACCCCGGAAGACCTGGAAAAC T_M=70.4˚C

rev: *AAGC**TT***ATTACAGGATAGAATACAGCATGGATGCCGTCCC T_M=70.8˚C

italics = restriction enzyme site (NdeI for forward primer, HindIII for reverse primer)

bold = for start and stop codons

DISCUSSION

Using the synthetic primers, a good amount of PCR product was formed. Even though the pylB bands in Figure 1 were faint in comparison to pylC and pylD, there were still substantial amounts of DNA for ligation. Since the annealing temperatures for each of the primers were high, 5µl DMSO was added to weaken the high ratio of G-C bonds. After each gene was amplified, the DNA had a poly-(A) chain added to the 3' end.

The first time each pyl gene was ligated into pGEM-T Easy Vector and transformed into XL-1 Blue electrocompetent cells, no colonies grew on the AMP/X-Gal/IPTG plates. The cells were induced with IPTG because the pGEM-T Easy Vector contains a lacZ operon where the pyl genes were inserted. The pGEM-T Easy Vector was used because past experiments have shown that using the pGEM-T vector as an intermediate help in the ligation into pET28a vector. At first, it was thought that the ligation reaction did not work properly. Thus, the ligation was re-done with new T4 DNA ligase. After having no colonies a second time, it was assumed that the cells were not competent to accept the transformed ligated plasmid. As a result, a new batch of XL-1 Blue electrocompetent cells was made. The new cells were competent and many white and blue colonies grew on the plates.

After picking twelve isolated white colonies from each plate, 4ml growths were mini-prepped. For the pylB gene, plasmid from lane #10 was picked to use as the starting point because it had the cleanest bands. The lane showed purely double cuts of NdeI and HindIII. For the pylC gene, plasmid from land #2 was picked because the restriction digests appeared to be the cleanest and the pylC band appeared to be most concentrated with the pylC gene (Figure 3).

For the pylD gene, plasmid from lane #7 was picked using the same reasoning for pylC (Figure 4).

After extracting the pyl genes from the pGEM-T Easy Vectors, they were ligated into pET28a vectors. The first attempt was transforming the new ligations into the stock of XL-1 Blue cells. The result was a lawn of yellowish-white cells covering the LB/KAN plates. The cells did not resemble that of typical *E. coli* colonies, and the cells also had a fruity scent. A few colonies for each pyl gene were screened to verify the non-existence of the pyl genes in the colonies. After seeing nothing on an agarose gel of a NdeI and HindIII digest, it concluded that there was an issue with the pET28a vector or a contamination issue. A series of tests, isolating variables, were performed to identify the source of contamination. In the end, it was concluded that the contamination was due to the new batch of XL-1 Blue cells, the same ones used for the blue-white screening during the pGEM ligation. By having the pGEM transformation work and the pET28a not work meant that even though the XL-1 Blue cells were contaminated, *E. coli* could out compete the contaminate (which was thought to be yeast) if the ligation was properly transformed and incorporated into the XL-1 Blue cells.

In attempts to avoid making a new stock of XL-1 Blue cells, a new batch of pET28a vector was made by excising a known gene out of an existing pET28a+gene plasmid. NdeI and HindIII restriction enzymes were used to excise the known gene and also to create the corresponding adhesive ends to ligate with the pyl genes. After the restriction enzyme digests, the linear DNA was treated with calf intestine Alkaline Phosphatase to remove the 5’-phosphate to prevent any re-annealing from occurring.

While searching for the contaminant and making new vectors, the pyl ligations into pET28a were transformed into JM109 electrocompetent cells instead. No colonies grew on the pylB, pylC, or pylD ligations. Both the new pET28a vector and an old working stock were tested, but neither grew colonies. It was concluded that the JM109 cells were not electrocompetent enough for the transformed ligations.

Upon receiving a new batch of XL-1 Blue cells, the pylB, C, and D ligations were transformed and plated on LB/KAN. There were finally colonies that grew on each plate. For pylB, ten colonies grew; for pylC, 14 colonies grew; for pylD, only one colony grew. Ten colonies from pylB and pylC were picked and screened, and the lone colony from pylD was picked and screened. Eight out of ten colonies screened for pylB yielded successful ligations, and plasmid from lane #1 was used to create a 500µl plasmid stock (Figure 5). Nine out of the ten colonies screened for pylC yielded successful ligations, and plasmid from lane #1 was used to make a 500µl plasmid stock (Figure 6).

The lone colony from pylD did not yield a successful ligation as the band excised from the pET28a vector was too big. Therefore, new ligations with insert:vector ratios of 1:1, 1:2, and 1:4 were tested, transformed, and plated. Only one colony grew on the 1:1 plate, zero colonies on the 1:2 plate, and five colonies on the 1:4 plate. All the colonies were picked and screened. After the NdeI and HindIII digests, none of the colonies appeared to have had a successful ligation nor did they appear to even incorporate the pylD gene (Figure 7). Due to time constraints, pylD was not ligated into pET28a vector and protein over-expression was not performed.

1µl of the stock plasmid for pylB and pylC were transformed into BL21(DE3) electrocompetent cells for protein over-expression. A 50ml growth was performed to see when

and if soluble pylB and pylC protein were expressed. The first expression of pylB showed that the pylB protein was only expressed in the total volume of the lysed cells and not soluble (Figure 8). If the protein is not soluble, then there is no chance to isolate and purify it. This is possibly due to the fact that pylB is a protein that requires Fe and has residues of the radical SAM family, which utilizes iron-sulfur clusters $[Fe\text{-}S]_4$. Since there was no soluble protein in the first expression, $Fe^{III}Cl_3$ was added into the growth media in the second attempt of pylB over-expression. The results show that by incorporating Fe^{III} into the growth media yields soluble pylB protein (Figure 9). Even though the amount of pylB protein that is soluble is much less than that in the total, it is still able to be overexpressed and purified.

The pylC protein over-expression was successful because there appeared to be relatively the same amount of pylC protein in both the total volume of the lysed cells and the soluble proteins (Figure 10). Since pylC is soluble, a 6L growth was done to overexpress the pylC protein.

For future experiments, the cloning of pylD needs to be completed. The pylB and pylC proteins need to be overexpressed and purified. An alternative method of expressing the pylB protein could be co-transforming PVD1282 with pylB because PVD1282 codes for proteins involved in the $[Fe\text{-}S]_4$. Once all the pyl genes are cloned and pyl proteins expressed, kinetic reactions will be performed in attempts to discover the intermediates and the mechanism for the formation of pyrrolysine.

ACKNOWLEDGEMENTS

A special thanks to Dr. Vahe Bandarian and the Department of Biochemistry and Molecular Biophysics at The University of Arizona

REFERENCES

1. Herring S, Ambrogelly A, Gundllapalli S, O'Donoghue P, Polycarpo CR, Soll D. The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable *in vitro* but required for *in vivo* activity. NIH 2007; 581(17): 3197-3203.
2. Kavran J, Gundllapalli S, O'Donoghue P, Englert M, Söll D, Steitz TA. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. PNAS 2007; 104(27): 11268-11273.
3. Lee MM, Jiang R, Jain R, Larue RC, Krzycki J, Chan MK. Structure of *Desulfitobacterium hafniense* PylSc, a pyrrolysyl tRNA synthetase. NIH 2008; 374(3): 470-474.
4. Longstaff DG, Larue RC, Faust JE, Mahapatra A, Zhang L, Green-Church KB, Krzycki JA. A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. PNAS 2007; 104: 1021-1026.
5. Krzycki JA. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. Current Opinion in Chemical Biology 2004; 8(5): 484-491.
6. Matthews RG, Koutmos M, Datta S. Cobalamin and Cobaminde-dependent Methyltransferases. NIH 2008; 18(6): 658-666.