2019

Expansion of the Genetic Code to Include Acylated Lysine Derivatives and Photocaged Histidine

William D. Kinney
Virginia Commonwealth University

Expansion of the Genetic Code to Include Acylated Lysine Derivatives and Photocaged Histidine

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By
William D. Kinney
B.S. Chemistry, 2012 James Madison University

Director: T. Ashton Cropp, Ph.D.
Associate Professor, Department of Chemistry

Virginia Commonwealth University
Richmond, Virginia
May 2019

Dedication

To my supportive wife, family and friends

Acknowledgements

Thank you to those who have supported me through my doctoral experience. Obtaining a Ph.D. has proven to be a difficult task that would not have been made possible without the support of very special people in my life. Starting from high school, my wife has always been on board with my life decisions. Through thick and thin, she has been there to help me see things through. Thank you to my parents and grandparents who I am eternally grateful for selflessly giving me every opportunity in life to succeed. Without them, I am not the man I am today.

Thank you to my principal investigator, Dr. T. Ashton Cropp, and the entire Cropp research group. I appreciate Ashton's unwavering confidence, insights, and consistent availability. It was a pleasure working on the project with him. Thank you to those in the Cropp group who helped me along the way. Dr. Christine Ring, Will Knight, and Dr. Courtney Braxton were very supportive from my first days in the lab. Thank you to Sahan Galbada for collaborating with me on the project. He was very insightful and professional in his approach to the work. Additionally, I would like to thank Dr. Tsemre Tessema. T was my best friend in the program and acted accordingly.

Thank you to our collaborators in Dr. Alex Deiter's research group at the University of Pittsburgh. I would like to acknowledge the funding received from the National Science Foundation. Lastly, I would like to acknowledge the chemistry Ph.D. program at Virginia Commonwealth University for extending this opportunity.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

o-NB	*o*-nitro benzyl
DIPEA	diisopropylethylamine
THF	tetrahydrofuran
IVKRS	isovaleryl lysine aminoacyl-tRNA synthetase
PCHis	*o*-NB-histidine
PCHisRS	photocaged histidine aminoacyl-tRNA synthetase
RNAP	RNA polymerase
Ara-bad	ararbinose-induced pBad promoter
P_{T7}	T7 promoter
EtOAc	ethyl acetate
TEA	triethylamine
IVK	N^{ε}-isovaleryl lysine
IBK	N^{ε}-isobutyryl lysine
tRNA	transfer RNA
HBK	N^{ε}-β-hydroxybutyryl lysine
mRNA	messenger RNA
MjTyrRS	*Methanococcus janaschii* tyrosyl-tRNA synthetase
JY^{cua}	*Methanococcus janaschii* tyrosyl-tRNA
EcTyrRS	*E. coli* tyrosyl-tRNA synthetase

EcLeuRS	*E. coli* leucyl-tRNA synthetase
PylT	pyrrolysyl $tRNA^{cua}$
PylRS	pyrrolysyl-tRNA synthetase
PylK	Pyrrolysine (4R,5R)-4-methyl-pyrroline-5- carboxylate
BocK	tert-butoxycarbonyl-L-lysine
CAT	chloramphenicol acetyl transferase
GFPuv	green fluorescent protein UV
FACS	fluorescence activated cell sorting
Ac-CoA	acyl-coenzyme A
KDAC	lysine deacylases
KAT	lysine acyl transferases
IVK	N^{ε}-isovaleryl lysine
IBK	N^{ε}-isobutyryl lysine
Ub	ubiquitin
PTM	post-translational modification
kDa	kilodaltons
ATP	adenosine 5'-triphosphate
DNA	deoxyribonucleic acid
UAA	unnatural amino acid
AA	amino acids
tRNA	transfer ribonucleic acid

aaRS	aminoacyl-tRNA synthetase
BocLys	N^{ε}-Boc-$_L$-lysine
wtUb	wild type ubiquitin
TFA	trifluoroacetic acid
UV	ultra violet
CBD	chitin binding domain
MjTyRS	*Methanococcus jannaschii* tyrosyl-tRNA synthetase
$tRNA^{tyr}$	tyrosyl tRNA
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
bp	base pairs
EDTA	ethylenediaminetetraacetic acid
LB	Luria-Bertani broth
SOC	super optimal catobilite-repression broth
EDTA	ethylenediaminetetraacetic acid
PMSF	phenylmethylsulfonyl fluoride
IPTG	isopropyl-L-D-1-galactopyronaside
OD	optical density
DTT	dithiothreitol
MS	mass spectrometry

PylT	pyrrolysyl tRNA
PylRS	pyrrolysyl aminoacyl-tRNA synthetase
sfGFP	superfolder GFP
DIPEA	*N,N*-diisopropylethylamine
ESI-MS	electrospray ionization mass spectrometry
Mb	*Methanosarcina bakeri*
Mm	*Methanosarcina mazei*
Amp	ampicillin
Cm	chloramphenicol
Kan	kanamycin
Tet	tetracycline
Arab	arabinose

Abstract

By William D. Kinney

A dissertation submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2019

Major Director: T. Ashton Cropp. Ph.D., Associate Professor
Department of Chemistry

The genetic code of all known organisms is comprised of the 20 proteinogenic amino acids that serve as building blocks on a peptide chain to form a vast array of proteins. Proteins are responsible for virtually every biological process in all organisms; however, the 20 amino acids contain a limited number of functional groups that often leaves much to be desired. The lack of diversity addresses the need to increase the genetic repertoire of living cells to include a variety of amino acids with novel structural, chemical, and physical properties not found in the common 20 amino acids. In order to expand the chemical scope of the genetic code beyond the functionalities that can be directly genetically encoded, unnatural amino acids must be added to the proteome. The ability to incorporate unnatural amino acids (UAAs) into proteins at defined sites has a direct impact on the ability of scientists to study biological processes that are difficult or impossible to address by more classical methods.

The UUAs of interest are acylated lysine derivatives (isovaleryl, isobutyryl, and β-hydroxybutyryl) and photocaged histidine. Acylation of histone lysine has been linked to epigenetic regulation of metabolism.[1] A means to site-specifically incorporate each acylated lysine derivative would help study the effect of acylated lysine in epigenetic regulation. Likewise, in order to elucidate the role of histidine in specific protein functions, one can replace a critical histidine with a photocaged histidine. Photocaged amino acids are those that possess a

photo-cleavable, aromatic caged group. Light-induced protein activation allows for the biological activity of the protein to be spatiotemporally regulated under non-invasive external control.[2]

The site-specific *in vivo* incorporation of unnatural amino acids is made possible by amber codon suppression by an orthogonal suppressor aminoacyl-tRNA synthetase (aaRS)/tRNA pair.[3] In amber codon suppression the amber stop codon is decoded for an UAA by a suppressor aaRS/tRNA pair. To accept the UAA, the aaRS must be evolved to achieve orthogonal activity with specific UUAs. The pyrrolysyl aaRS/tRNA (PylRS/PylT) pair from *M. barkeri* and *M. mazei* was used to construct multiple, large-scale aaRS mutant libraries where critical residues within the active site of PylRS are mutated via site-saturated mutagenesis.[4] The libraries were subjected to directed evolution through a series of positive and negative selections to enrich aaRS variants that exclusively bind to acylated lysine derivatives and photocaged histidine as substrates.[5] The PylRS selection survivors were screened for UAA activity and identified successful clones underwent a fluorescent activity assay. The active aaRS were used for amber codon suppression to express the respective UAA in ubiquitin and green fluorescent protein constructs.

Chapter 1: Introduction

1.1 The Genetic Code

The 20 proteinogenic amino acids make up the standard genetic code of all organisms.[3,6] The genetic code is contained within the genetic material that serves as a template for protein biosynthesis. In protein translation, the single stranded messenger ribonucleic acid (mRNA) serves as a template and encodes a protein as a series of codons.[7,8] Every three base pairs is a codon that codes for a particular amino acid (Figure 1). Of the 64 triplet codons of the genetic code, 61 sense codons are assigned to a specific standard amino acid.

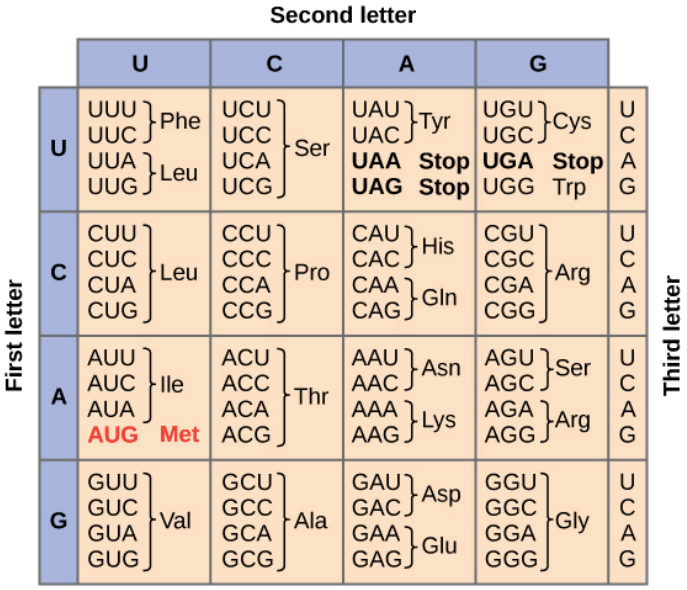


Figure 1.1. mRNA codon table of the genetic code[9]

The sequential order of the codons dictates the order the amino acids are arranged in the protein. The ribosome binds to the mRNA and decodes each codon. A transfer ribonucleic acid (tRNA) bound to the corresponding amino acid is recruited to the ribosome bound mRNA in response to the decoded codon. The anticodon of the tRNA then binds to the three bases of the

codon. The anticodon tRNA is the reverse complimentary sequence to its corresponding codon found on the anticodon loop of the tRNA. On the 3' end of the tRNA is a covalent attachment to the amino acid that is coded for by the particular codon. Once the tRNA is bound to the mRNA, the ribosome catalyzes the peptidyl transfer of the amino acid to the nascent polypeptide chain (Figure 1.2). The ribosome continues decoding mRNA and catalyzing the peptidyl transfer of each amino acid until it reaches a stop codon. There are three stop codons in the genetic code. The stop codon signals for protein translation termination. There does not exist a tRNA in the traditional translational machinery capable of binding to the stop codon. Instead, release factors bind to the stop codon, the ribosome dissociates from the mRNA and the polypeptide is released. The polypeptide folds into its native conformation, localizes in the cell, and performs its native function.

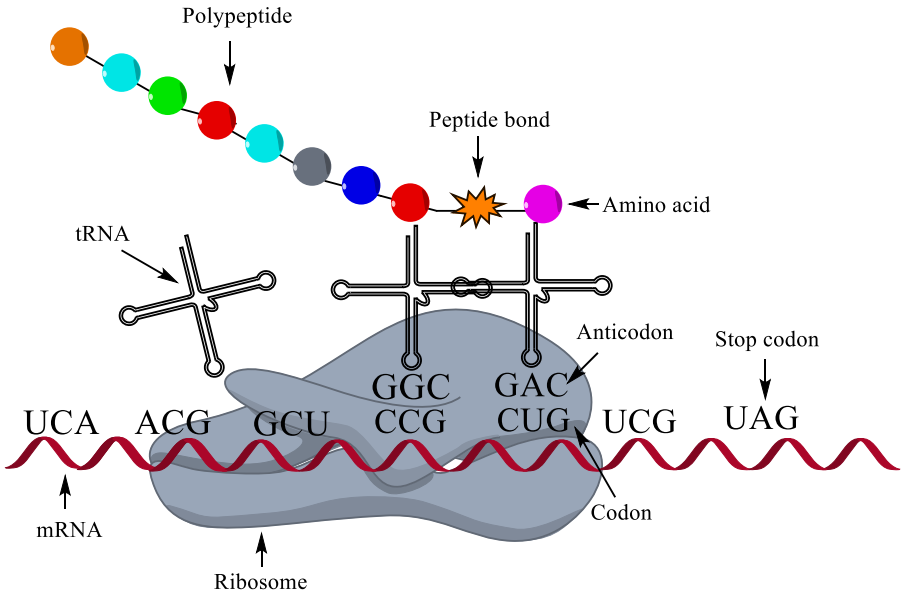


Figure 1.2. Ribosome facilitated translation of a coded gene.

1.2. Genetic Code Expansion

Proteins are responsible for virtually every biological process in all organisms from catalysis and gene regulation to immune response and signal transduction. However, the 20 amino acids contain a limited number of functional groups that often lead more to be desired. The lack of diversity addresses the need to increase the genetic repertoire of living cells to include a variety of amino acids with novel structural, chemical, and physical properties not found in the common 20 amino acids.

In order to expand the chemical scope of the genetic code beyond the functionalities that can be directly genetically encoded, unnatural amino acids must be added to the proteomic repertoire. The ability to incorporate unnatural amino acids into proteins at defined sites has a direct impact on the ability of scientists to study biological processes that are difficult or impossible to address by more classical methods. Comprehension of these functional roles of these modifications requires homogenously modified proteins that are usually difficult to purify from their natural sources. *In vivo* incorporation of UAAs with unique modifications at defined sites creates new opportunities to generate proteins with novel properties, elucidate the roles of established post-translational modifications, and utilize biochemical or biophysical probes for structural and functional analysis.[10–13]

1.3 Unnatural Amino Acids

Genetic code expansion now allows the site-specific incorporation of unnatural amino acids into proteins in *E. coli*,[11,14–16] *Saccharomyces cerevisiae*,[17] mammalian cells,[14–16,18,19] *Caenorhabditis elegans*,[14] *Drosphila melanogaster*,[20] and *Arabidopsis thaliana*.[14] To date, around 200 unnatural amino acids have been added to the genetic code. Some examples of the

UAAs added to the genetic code are highlighted in Figure 1.3 By incorporating photocross-linkers, it has been possible to map weak, transient, or pH sensitive protein interactions, and to assign protein interactions to functional states.[16,21–23] The incorporation of post-translationally modified amino acids has allowed the synthesis of homogeneously and site-specifically modified proteins.[24–28] This has revealed the function of the modifications or the proteins that regulate them.[29] The incorporation of photocaged amino acids has allowed specific functions of proteins to be controlled with light and forms the basis of time-resolved studies of signaling and transport processes in cells.[10,19,24,30,31] An emerging area is the site-specific incorporation of unnatural amino acids that contain functional groups that allow subsequent chemoselective and rapid labeling at a single site in the protein.[32]

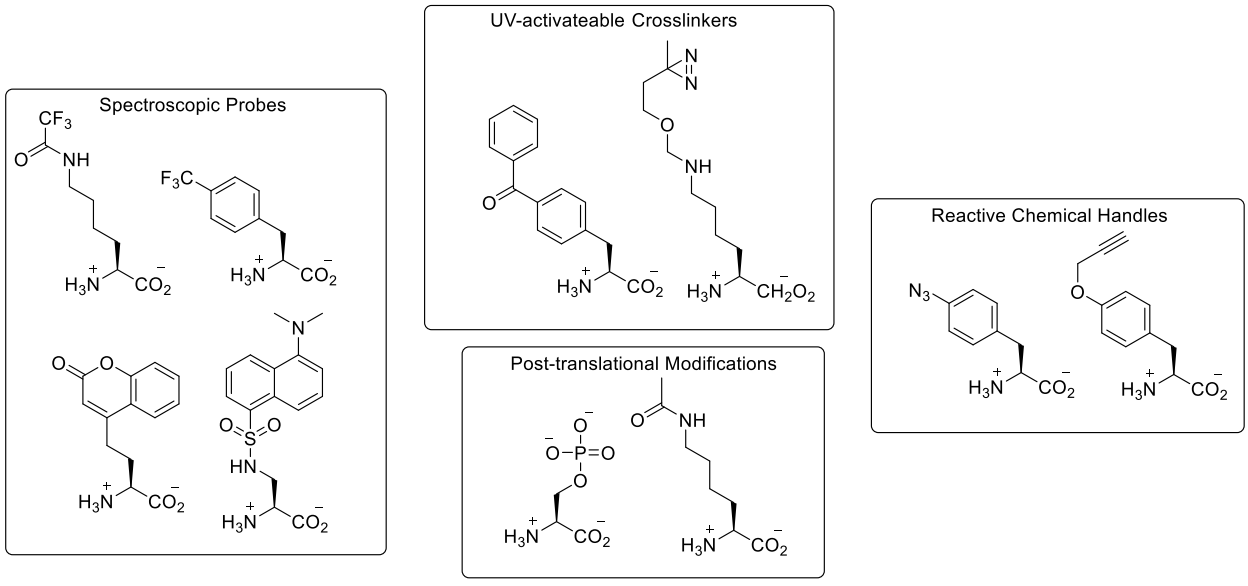


Figure 1.3. Examples of unnatural amino acids added to the genetic code.

1.4 Unnatural Amino Acid Incorporation Methods

Methods have been developed that have made possible the systematic expansion of the genetic code to not only study protein structure-function but to enable the evolution of proteins

with new or enhanced properties. Numerous *in vitro* and *in vivo* approaches have been implemented to capture a robust method for unnatural amino acid incorporation into proteins. Pioneered by Merrifield *et al* in 1963, stepwise solid-phase peptide synthesis has facilitated the synthesis of peptides and small proteins (< 100 amino acids) containing unnatural amino acids.[33] To overcome the size limitations of SPPS, efficient strategies have been developed to ligate synthetic peptides together to make larger proteins through enzymatic and chemical ligation.[34,35] Intramolecular native chemical ligation has also been used to create cyclic peptides and proteins.[36,37] Chemical peptide synthesis is useful for incorporating amino acid analogues that are toxic to cells or incompatible with the translational machinery, however, it can be problematic with peptides or proteins that have poor solubility, and becomes tedious, low yielding, and expensive when applied to larger proteins.

Semi-synthesis is used as an alternative method to generate larger proteins containing unnatural amino acids. In this method, larger peptide fragments recombinantly expressed or derived from the cleavage of natural proteins are ligated to synthetic peptides.[38] While semisynthetic methods improve upon the size of the proteins generated, it is hindered by the requirement for appropriate sites for cleavage and ligation. The ability to incorporate unnatural amino acids directly into proteins *in vivo* has several advantages over *in vitro* methods including high yields of mutant proteins, technical ease, and the potential to study proteins in cells or multicellular organisms.

1.5 Unnatural Amino Acid Incorporation via Amber Codon Suppression

A genomic code expansion approach to introduce UAAs *in vivo* involves site-specific cotranslational incorporation through amber stop codon suppression (Figure 1.4). During

translation, an amino acid is amino acylated onto its tRNA by an enzyme called aminoacyl-tRNA synthetase (aaRS). An aaRS catalyzes the esterification of its cognate amino acid onto the 3' end of the cognate tRNA in a two-step process (Figure 1.4). First the amino acid is activated with adenosine triphosphate (ATP) to form an amino acid-adentylate. Then the amino acid-adentylate is transferred to the hydroxyl group of the 3'-terminal adenosine of its cognate tRNA. An elongation factor then binds the aminoacyl-tRNA and delivers it to the ribosome. In amber codon suppression, the targeted protein is expressed from an mRNA template containing an internal amber stop codon that serves as a blank codon reassigned for the UAA. An UAA specific aaRS aminoacylates a cognate suppressor tRNA devoted to decoding the amber codon. The suppressor tRNA contains the anticodon, CUA, that anneals to the amber stop codon. Such a suppressor tRNA aminoacylated with the UAA competes with release factors for the binding to the amber stop codon at the ribosome. If the suppressor tRNA can outcompete the release factors, the UAA is then incorporated in response to the internal amber stop codon and added to the growing polypeptide by ribosome catalyzed peptidyl transfer.

Recoding of the amber codon for UAA incorporation is preferred over the ochre and opal stop codons. Opposed to the opal and ochre stop codons, the amber codon is the least used stop codon and has not been linked to critical genes in the genome of *E. coli*.[39] A significant advantage of using amber codon suppression for UAA incorporation is that the amber codon can replace any codon(s) within the gene of the protein by mutagenesis. This capability allows for site specific incorporation of the UAA.

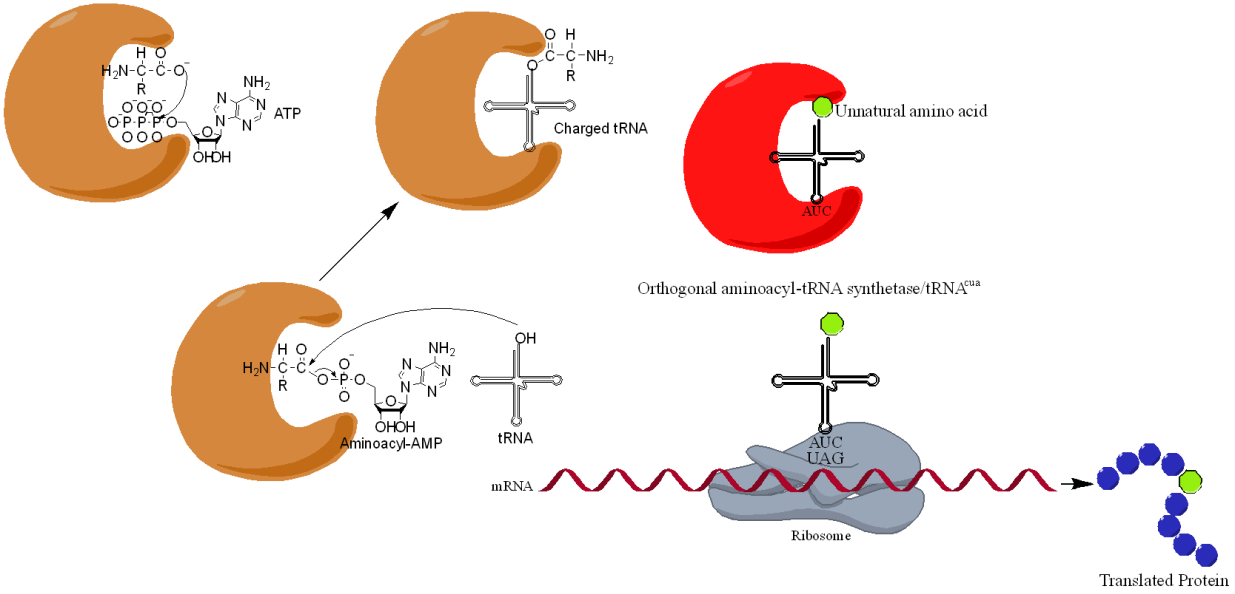


Figure 1.4. Amber codon suppression by the amino of an unnatural amino acid onto tRNA by an aminoacyl tRNA synthetase.

1.6 Amber Codon Suppression by Pyrrolysyl tRNA/Synthetase

The success of UAA incorporation strongly depends on the efficiency of the orthogonal aaRS/tRNA pair. The efficiency of each pair is determined by its orthogonality, the ability of the aaRS to recognize the cognate suppressor tRNA, and the flexibility of the tRNA synthetase active site. The exogenous suppressor tRNA and aaRS must be orthogonal to the host's translational machinery. An orthogonal aminoacyl-tRNA synthetase does not aminoacylate any

endogenous tRNAs in the host cell, but specifically aminoacylates its cognate orthogonal tRNA. Furthermore, an orthogonal tRNA is not a substrate for endogenous aminoacyl-tRNA synthetases, but is a specific substrate for the orthogonal synthetase, and directs the translational incorporation of amino acid substrates of the orthogonal aminoacyl-tRNA synthetase in response to an amber codon in mRNA.

Four main orthogonal aminoacyl-tRNA synthetases have been developed for genetic code expansion: the *Methanococcus janaschii* tyrosyl-tRNA synthetase (MjTyrRS)/ $tRNA^{CUA}$ pair, the *E. coli* tyrosyl-tRNA synthetase (EcTyrRS)/$tRNA^{CUA}$ pair, the *E. coli* leucyl-tRNA synthetase (EcLeuRS)/ $tRNA^{CUA}$ pair, and pyrrolysyl-tRNA synthetase (PylRS)/$tRNA^{CUA}$) pairs from certain *Methanosarcina*. Herein, the PylRS/$tRNA^{CUA}$ was chosen as the orthogonal aminoacyl-tRNA synthetase pair due to its inherent advantages over the other pairs. The pyrrolysyl aaRS/tRNA pair belong to a family of methanogens called methanosarcinales in the archaea domain. The suppressor pair is found in two strains of the *methanosarcina* family-*barkeri* and *mazei*. By sequencing clustered genes in *Methanosarcina barkeri*, it was discovered by Krzycki et al. that the open reading frame of monomethylamine methyltransferase contains an in-frame amber stop codon that does not prevent truncation of the full-length protein.[4] This proved that the in-frame amber codon is reassigned to code an amino acid that is critical for functional conservation among methylamine methyltransferases. The identity of this modification was determined by crystal structure of monomethylamine methyltransferase to be (4R,5R)-4-methyl-pyrroline-5- carboxylate (Figure 1.5).[40] This novel amino acid was coined as pyrrolysine (PylK), the 22nd naturally occurring genetically encoded amino acid to be discovered. PylK contains a methylated pyrroline ring covalently bound to the epsilon nitrogen of lysine. Along with PylRS, Krzycki et al. revealed a unique tRNA gene, pyrrolysyl tRNA

(PylT), near a methylamine methyltransferase gene cluster.[8] PylT codes a special tRNA with a CUA anticodon making it a natural suppressor tRNA (Figure 1.5).

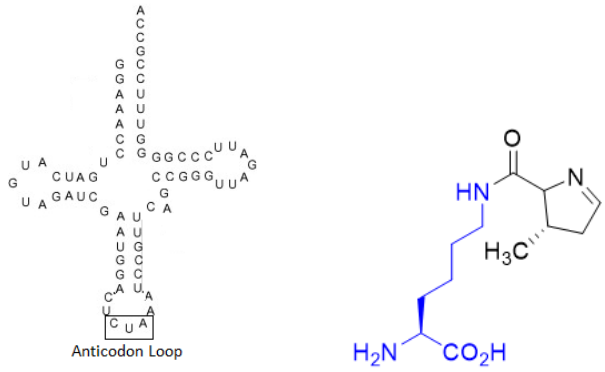


Figure 1.5. (Left) mRNA sequence of PylT. Anticodon loop shown in box. (Right) Structure of PylK. Lysine is shown in blue.

The pyrrolysyl aaRS/tRNA amber codon suppression system affords many advantages over other available stop codon suppression cotranslational pairs. First, PylRS and $tRNA^{Pyl}$ constitute an excellent orthogonal aaRS/tRNA pair. Herein, the expression host used for this project was *E. coli* and the tRNA/aaRS pairs of archaeal origin are likely to be orthogonal because the substantial sequence divergence from their bacterial counterparts has been shown to result in low cross-reactivity.[41] The pyrrolysyl system is the lone naturally occurring suppressor orthogonal in *E. coli*, *S. cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* and mammalian cells.[8,42,43,44,45] PylRS has shown a high degree of substrate side chain promiscuity and shows almost no recognition of the tRNA anticodon (Figure 1.6).[46] Unlike the majority of aaRSs, PylRS has a low selectivity toward the tRNA anticodon.[47] Anticodons in tRNAs are typically used as recognition elements of AARSs for strictly pairing amino acids with their corresponding codons to avoid misincorporation during translation. Lastly, pylT contains the innate anticodon CUA capable of being coded and bound by the UAG, amber codon.

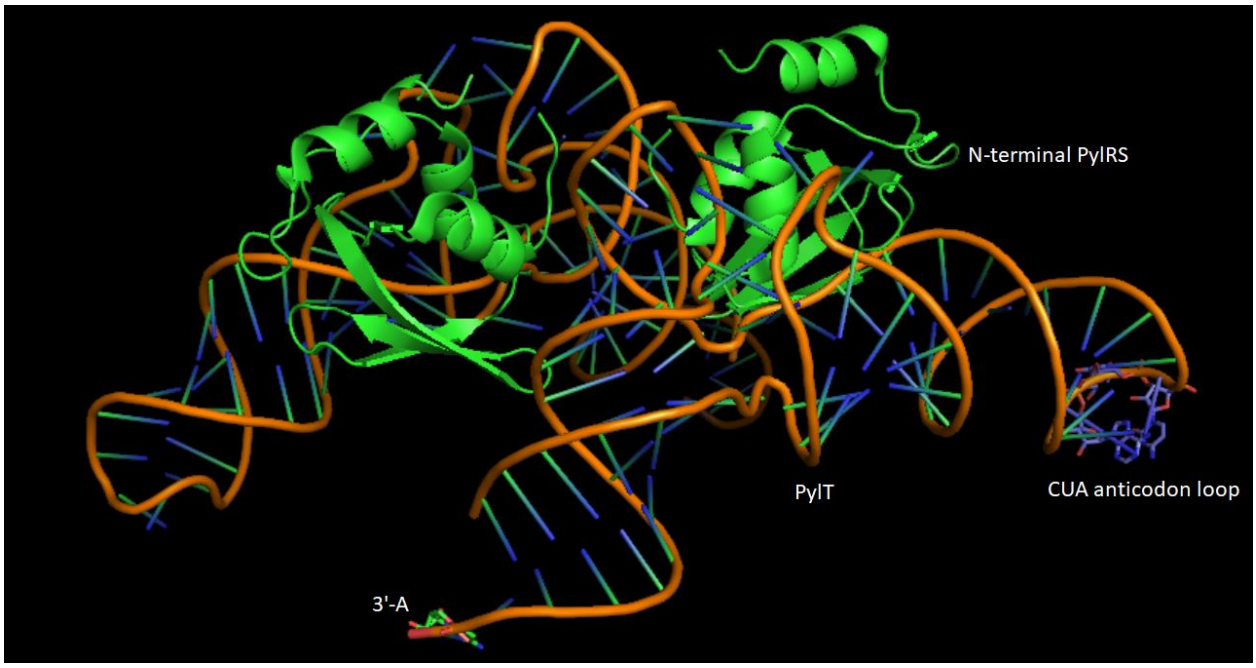


Figure 1.6. Crystal structure of the N-terminal tRNA binding domain of PylRS bound to PylT. PDB code 5UD5.[46]

Although the C-terminal catalytic domain of PylRS binds tightly to Pyl, PylRS displays high tolerance toward variations of the substrate side chain at the pyrrole ring (Figure 1.7).[48] One of the major reasons for the high tolerability is due to the lack of an editing domain. Due to the exceptionally large size of PylK's structure, the active site is larger than most as well. For instance, the deep amino acid binding pocket is twice as deep as the L-lysine-binding pocket in lysyl-tRNA synthetase (Figure 1.7).[49,50]

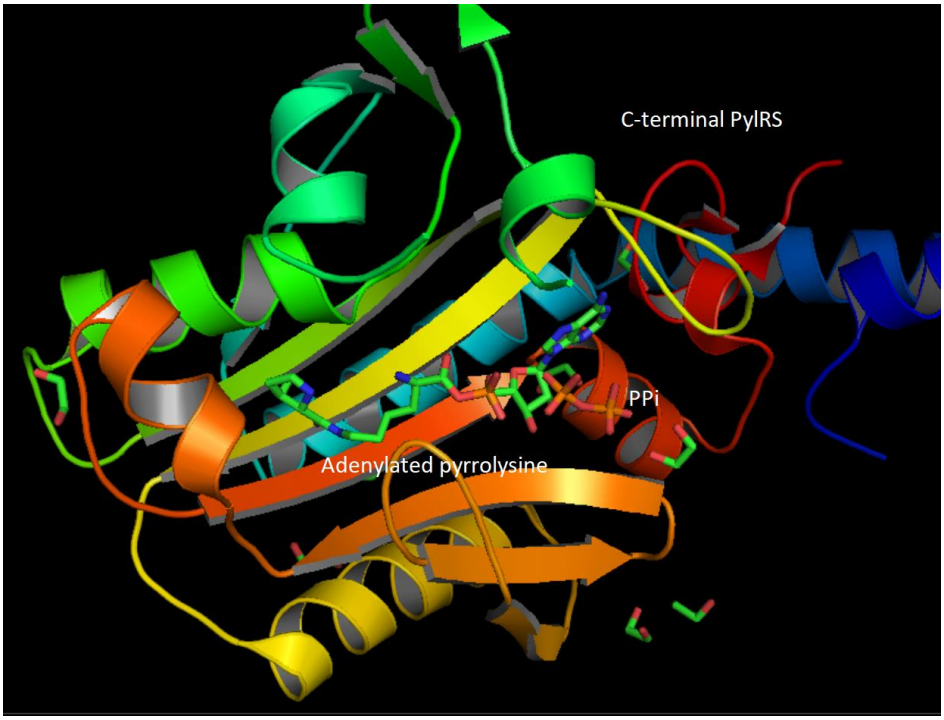


Figure 1.7. Crystal structure of C-terminal (101 amino acids) bound to adenylated pyrrolysine and pyrophosphate. PDB code 2Q7H[50]

Chapter 2. PylRS Library Construction

Chapter 2.1. PylRS Library Residue Selection

Due to the aforementioned reasons, the PylRS/PylT amber codon suppression system was chosen for directed evolution to actively incorporate unnatural amino acids into the proteome of *E. coli*. In order to accept the UAAs of interest, the active site needed to be engineered (Figure 2.1). PylRS is likely to recognize the substrate amino acids broadly by their characteristics, such as a hydrophobic moiety with an appropriate size and bulkiness accommodated in the hydrophobic pocket and an adjacent Nε hydrogen-bond acceptors/donor. The pyrrolysyl moiety of the enzyme-bound PylK and pyrrolysyladenylate (pyrrolysyl-AMP) are recognized within the deep catalytic pocket of PylRS. [49,50]

Previously reported PylRS active site architectures were consulted along with examination of the crystal structure to determine the critical residues to undergo site-saturated mutagenesis.[49] Both the PylRS from *M. barkeri* and *M. mazei* were used in the library construction. The active site of the PylRS derived from each strain are nearly identical, however, the amino acid numbering is different. The residue numbering is reported as *mazei* (*barkeri*). Upon inspection of the binding pocket from the crystal structure of *M. mazei* PylRS (c270) coordinated with adenylated PylK, five residues were carefully selected. The residues selected were N346(311), C348(313), V401(366), W417(382), and G421(386) (Figure 2.1). These sites were selected for amino acid reassignement due to their proximity and reported interactions with the native substrate. Structure based mutational analysis by Yokoyama et al. revealed that hydrophobic residues V401(366) and W417(382) in the amino acid-binding tunnel are important for accommodating the PylK side chain and that N346(311) is essential for anchoring the side-chain carbonyl and α-amino groups of PylK.[49] N346(311) forms two hydrogen-bonding

interactions: the side-chain amino and carbonyl groups of N346(311) interact with the side-chain carbonyl group and the α-amino group, respectively, of PylK. C348(313) also binds to the carbonyl of PylK.[51] Hydrophobic residues in the catalytic cleft V401(366) and W417(382) represent the hydrophobic residues that accommodate and trap the bulky methyl-pyrroline ring. Though the sidechain-less G421(386) is relatively inert, its close proximity to the coordinated adenylated pyrrolysine warranted inclusion in the library.

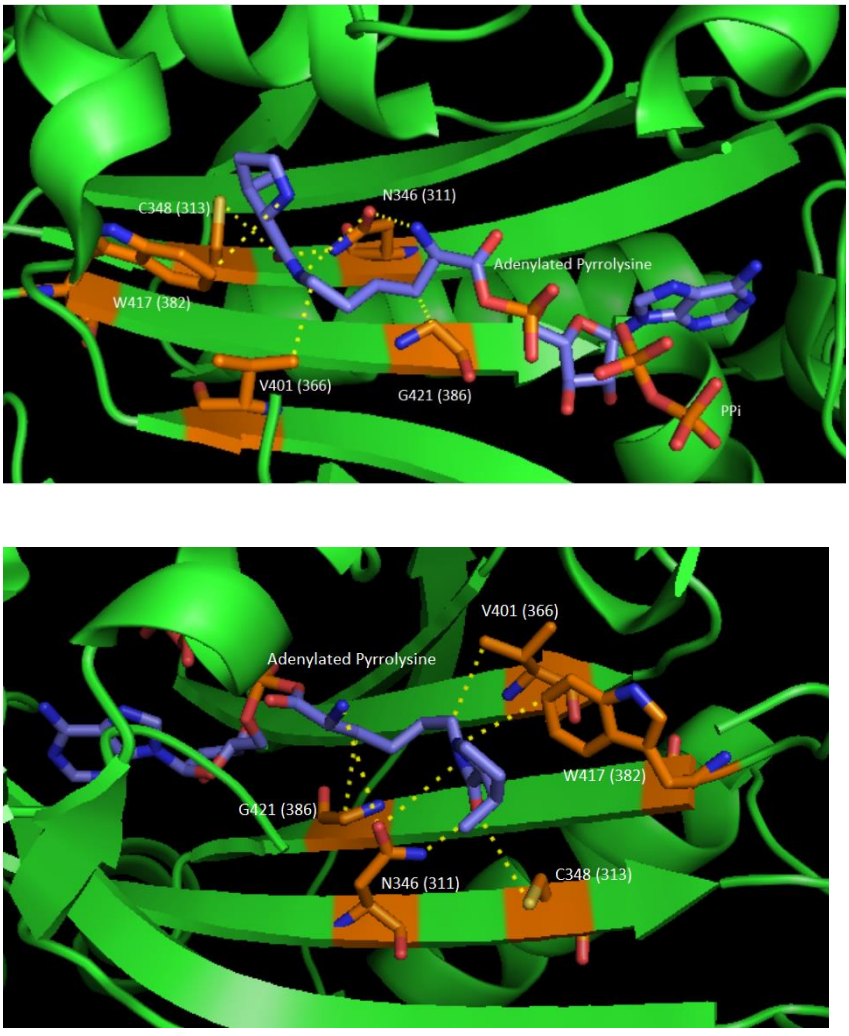


Figure 2.1. Frontside (top) and backside (bottom) view of the crystal structure of pyrrolysyl-tRNA synthetase (green, C-terminal residues 288-454) bound to adenylated PylK (blue).

Residues selected for site-saturated mutagensis are shown in orange. The numbering of the residues is *mazei* (*barkeri*). PDB code: 2Q7H[50]

Herein, the close structural analog of PylK, tert-butoxycarbonyl-L-lysine (BocK), was used as positive control in place of PylK (Figure 2.2). PylRS was reported to recognize the natural lysine derivative as well as many lysine analogs including BocK with diverse side chain sizes and structures.[52] The synthesis of PylK has proven to be quite difficult and commercially expensive. For this reason, the much less expensive BocK was used as the substrate with wild-type PylRS as the positive control. PylRS has been shown to have comparable activity with BocK and PylK. When comparing *barkeri* derived synthetases to the positive control, wild-type PylRS was used. When comparing *mazei* derived synthetases to the positive control, PylRS Y349F was used.

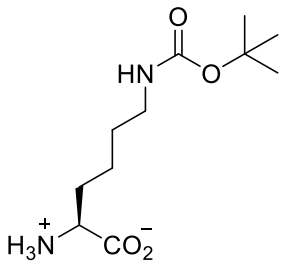


Figure 2.2. Chemical structure of BocK.

2.2. Library Construction Strategies

The *Methanosarcina barkeri* and *mazei* libraries consisting of codon randomization of residues N346(311), C348(313), V401(366), W417(382), and G421(386) was built using separate methods of inverse polymerase chain reaction (PCR) by golden gate assembly (Figure 2.3) and Gibson assembly (Figure 2.4). In inverse PCR by golden gate assembly, primers are designed that contain the degenerate codon, NNK, for site-saturated mutagenesis and the

nucleotide recognition sequence of the endonuclease, BsaI. Site-saturated mutagenesis mutates each chosen residue to any of the codons contained within the degenerate codon NNK. NNK codes for 32 of the 64 codons in the genetic code. The N represents any base pair while the K represents guanosine or thymine (G or T). Thus 32 codons are used by NNK to code for any of the 20 amino acids along with the stop codon UAG. The degenerate codon NNK was used to accomplish site-saturated mutagenesis while minimizing the number of stop codons that can be coded for in the library. Mutation to a stop codon at any residue would result in an inactive variant. The designed primers annealed to the PylRS gene on plasmid pBK PylRS in a back-to-back fashion. During inverse PCR, the DNA polymerase amplified the vector backbone affording a linear PCR product containing the primer designed BsaI recognition sites on the termini. BsaI is a type IIa endonuclease that cuts outside its DNA recognition site. This off-site DNA cleavage results in a scar-less (BsaI recognition site removed) product with complimentary 3' four base pair overhangs. BsaI-HF V2 (NEB) was used for the digestion. BsaI-HF V2 has been optimized to be active in ligation buffer. This allowed for the digestion and ligation reactions to occur simultaneously in the same buffer which eliminated the need for purification between reactions. Purification of DNA between steps was found to severely affect the ultimate transformation diversity. The plasmid containing the degenerate codons is recircularized by ligation and transformed into DH-10B *E. coli* cells. When each step in the approach was optimized, the transformation diversity yielded 10^4-10^5 colonies per transformation.

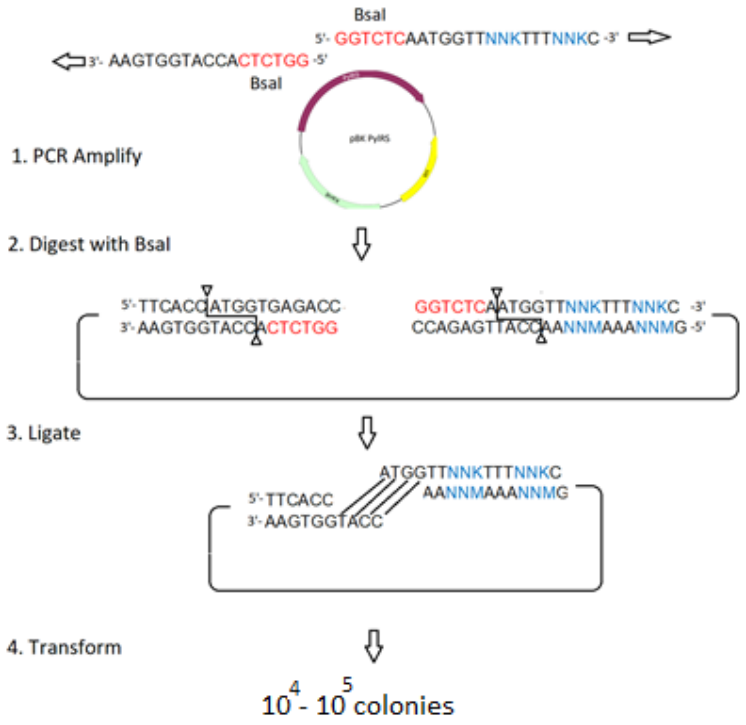


Figure 2.3. Site-saturated mutagenesis by inverse PCR via golden gate assembly for site-saturated mutagenesis of PylRS residues in the construction of a PylRS library.

Gibson assembly was the second approach used for construction of the libraries.[53] Gibson assembly efficiently joins multiple overlapping DNA fragments in a single-tube isothermal reaction (Figure 2.4). Primers were designed that anneal to PylRS and contain the degenerate codons for site-saturated mutagenesis. The primers were used to introduce the degenerate codons and amplify the PylRS gene with homologous ends to the amplified vector. Likewise, primers were designed to amplify the vector with homologous ends to the amplified insert. The PCR generated fragments were then added to a master mix that includes three different enzymes active in the same single buffer. First, the exonuclease created single-stranded 3' overhangs that facilitated the annealing of the two fragments that shared complementarity at both ends (overlap region). Next, the DNA polymerase filled the gaps within the annealed fragments. Finally, the

DNA ligase sealed nicks in the assembled DNA. The pBK PylRS plasmid was recircularized with introduced site-saturated mutations and transformed into DH10B *E. coli* cells. When each step in the approach was optimized, the transformation diversity yielded 10^4-10^5 colonies per transformation.

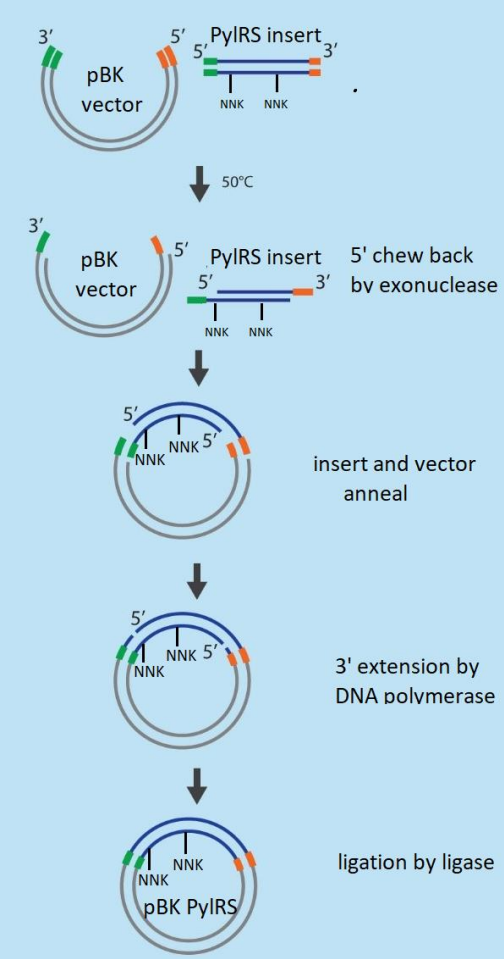


Figure 2.4. Gibson assembly of fragments with PCR generated homologous ends. Figure adapted from SGI-DNA.

2.3 *M. barkeri* Library Construction

Using the PylRS gene derived from *M. barkeri*, mutations were completed by two rounds of PCR generated site-saturated mutagenesis in order to randomize residues to all possible amino acids through degenerate codon NNK randomization (Methods 2.6.1). The two sub-libraries

N311, C313 and V366, W382, G386 were built using inverse PCR by golden gate assembly with the product being used as the template for subsequent full library construction. The primers CL1648 and CL1649 were used in the inverse PCR reaction to generate the N311, C313 sub-library. The primers CL1646 and CL1648 were used in the inverse PCR reaction to generate the V366, W382, G386 sub-library. The PCR product for each sub-library was digested with the typeIIs endonuclease, BsaI. The digested products were recircularized by ligation and used to transform into electrocompetent DH10B cells. The diversity of each sub-library had at least 10-fold coverage upon transformation. The isolated plasmids served as templates for the second round of mutagenesis.

The second round of PCR reactions to generate the library with codon randomization at residues N311, C313, V366, W382, and G386 was performed using sub-library V366, W382, G386 as template for a subsequent Gibson assembly reaction (Methods 2.6.1). Primers CL1892 and CL1893 were used to create the insert and primers CL1894 and CL1895 were used to create the vector. Gibson assembly was used to ligate the two fragments. The circularized plasmid was used to transform electrocompetent DH10B cells to generate full library N311, C313, V366, W382, G386 at a diversity of 5.0×10^6 colonies. DNA isolated from the pool of cells was isolated and sequenced to check for codon bias (Figure 2.5). Additionally, ten individual clones were sequenced (not shown) and no apparent codon bias was found.

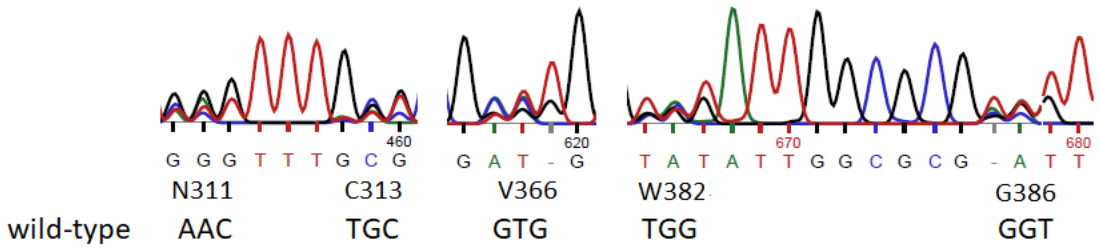


Figure 2.5. Trace sequence analysis of the pooled *M. barkeri* library generated from Gibson assembly. Red=T, Blue=C, Green=A, Black=G

Additionally, inverse PCR by golden gate assembly was used as a method to generate the library with codon randomization at residues N311, C313, V366, W382, and G386 (Methods 2.6.1). The primers CL1646 and CL1647 were used in the inverse PCR reaction with sub-library N311, C313 as template. The PCR product in the second round was digested with BsaI, re-circulized by ligation and used to transform electrocompetent DH10B cells to generate full library N311, C313, V366, W382, G386 at a diversity of 3.5 x 10^6 colonies. DNA isolated from the pool of cells was isolated and sequenced to check for codon bias (Figure 2.6). Additionally, ten individual clones were sequenced (not shown) and no apparent codon bias was found.

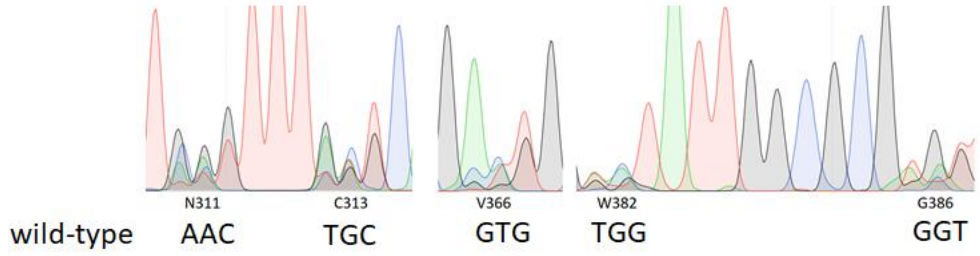


Figure 2.6. Trace sequence analysis of the pooled *M. barkeri* library generated from inverse PCR by golden gate assembly. Red=T, Blue=C, Green=A, Black=G

2.4 *M. mazei* Library Construction

The *Methanosarcina mazei* library was built using two rounds of Gibson assembly to randomize the codons of residues N346, C348, V401, W417, and G421, with the product of one round acting as a template for the next round (Methods 2.6.2). Primers CL1974 and CL1893 were used in the first round of PCR reactions to create the insert and primers CL1894 and CL1975 were used to create the vector. Gibson assembly was used to ligate the two fragments to generate sub-library N346, C348. Primers CL1984 and CL1985 were used to create the insert and primers CL1986 and CL1987 were used to create the vector in the second round of PCR reactions. Gibson assembly was used to ligate the two fragments to generate the full library N346, C348, V401, W417, G421 at a diversity of 4.0×10^6 diversity. DNA isolated from the pool of cells was isolated and sequenced to check for codon bias (Figure 2.7). The pooled DNA sequencing results showed that there was bias for the codon CCC (proline) at sites 417 and 421. Likewise, there was a general lack of thymine observed in the first two base pairs at sites 346 and 401. Ten individual clones were sequenced (not shown) and no apparent codon bias was found though the sequencing of more clones could elaborate on the apparent bias observed in the pooled DNA sequencing analysis.

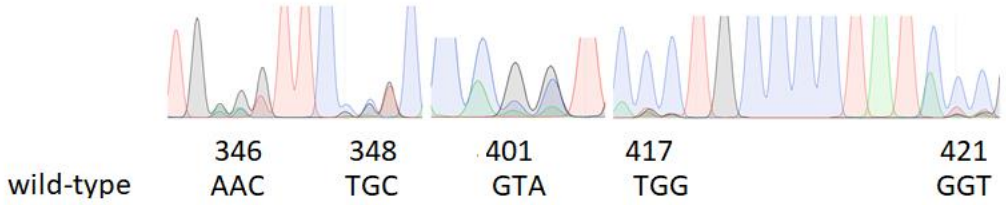


Figure 2.7. Trace sequence analysis of pooled *M. mazei* library generated by Gibson assembly. Red=T, Blue=C, Green=A, Black=G

2.5 Additional Libraries

Project collaborators from the Deiter's group at University of Pittsburgh (formerly North Carolina State) sent DNA of two mixed *M. barkeri* libraries. The first library was derived from previous selections by the group for activity with lysine analogs and had the following mutations: L274A, Y348F, A267NNK, Y271NNK, and C313NNK. The second library was derived from previous selections by the group for activity with tyrosine analogs and had the following mutations: N311A, C313A, Y271NNK, L274NNK, Y349NNK. Together, these mixed *M. barkeri* libraries were referred to simply as mixed *barkeri* libraries.

2.6 Methods and Materials

2.6.1 *M. barkeri* library construction

The *Methanosarcina barkeri* libraries consisting of codon randomization of residues N311, C313, V366, W382, and G386 were built using separate methods of Gibson assembly (NEB #E2611S) and inverse PCR by Golden Gate Assembly (Figure 2.8). The two sub-libaries N311, C313 and V366, W382, G386 were built using inverse PCR by Golden Gate Assembly with the product being used as the template for subsequent full library construction. The primers CL1648 and CL1649 were used in the inverse PCR reaction to generate the N311, C313 sub-library. The primers CL1646 and CL164R were used in the inverse PCR reaction to generate the V366, W382, G386 sub-library. The PCR reactions were prepared in 160µL containing 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (NEB #N0447), 0.5 µM of each primer (Eurofins), 8 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl aliquots using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec

at 68°C, 2 min at 68°C), 5 min at 72°C. The PCR products were purified following the Zymo DNA clean and concentrator kit (Zymo D4013). The purified PCR products were digested with 2 U DpnI (NEB #R0176) to remove methylated template and 10 U BsaI (NEB #R03535) overnight at 37°C to ensure complete digestion. The digested products were purified following the Zymo DNA clean and concentrator kit (Zymo D4013). The purified digestion products were re-circularized by ligation using 400 U T4 ligase (NEB #M020) in 20 µL reactions and used to transform 100 µL electrocompetent DH10B cells. The transformed cells were recovered in 900 µL super optimum broth (SOC). The recovered cells were plated at 1:1, 1:100, 1:1,000 and 1:10,000 dilutions (Kan50) to determine total diversity. The sub-library DNA was isolated from 100 mL cultures (Kan50) by following the Promega midiprep wizard kit (Promega A7640). The sub-library DNA was used as template in subsequent PCR reactions.

The second round of PCR reactions to generate the library with codon randomization at residues N311, C313, V366, W382, and G386 was performed using sub-library V366, W382, G386 as template for a subsequent Gibson assembly reaction. Primers CL1892 and CL1893 were used in the first round of PCR reactions to create the insert and primers CL1894 and CL1895 were used to create the vector. The PCR reactions were prepared in 160µL containing 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (NEB #N0447), 0.5 µM of each primer (Eurofins), 8 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl aliquots using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec at 68°C, 2 min at 68°C), 5 min at 72°C. The PCR products were digested with 2 U DpnI for 2 hours at 37°C. The digested PCR products were purified following the Zymo DNA clean and

concentrator kit (Zymo D4013). 50 ng of insert and 100 ng of vector were added to a 20 µL Gibson assembly (NEB #E2611) reaction and incubated for 1 hr at 50°C. 2 µL of the Gibson reaction was used to transform into 100 µL electrocompetent DH10B cells. The transformed cells were recovered in 900 µL super optimum broth (SOC). The recovered cells were plated at 1:1, 1:1,000 and 1:10,000 dilutions (Kan50) to determine total diversity. This procedure was repeated several times until an acceptable diversity was achieved (5.0×10^6).

Additionally, inverse PCR was used as a method to generate the library with codon randomization at residues N311, C313, V366, W382, and G386. The primers CL1646 and CL1647 were used in the inverse PCR reaction with sub-library N311, C313 as template. The PCR reaction conditions, purification, digestion, and transformation followed the same procedure as the inverse PCR sub-libraries as described. The PCR product in the second round was first digested with DpnI (NEB #R0176) and BsaI (NEB #R03535), re-circularized by ligation using T4 ligase (NEB #M020) and used to transform electrocompetent DH10B cells to generate full library N311, C313, V366, W382, G386. This procedure was repeated several times until an acceptable diversity was achieved (3.5×10^6). A flow chart of the construction of the *M. barkeri* library can be seen in Figure 2.8.

2.6.2 *M. mazei* library construction

The *Methanosarcina mazei* library was built using two rounds of Gibson assembly to randomize the codons of residues N346, C348, V401, W417, and G421, with the product of one round acting as a template for the next round (Figure 2.8). Primers CL1974 and CL1893 were used in the first round of PCR reactions to create the insert with N346NNK and C348 NNK mutations and primers CL1894 and CL1975 were used to create the vector. The PCR reactions were prepared in 160µL containing 1x Phusion High-Fidelity DNA polymerase buffer (NEB

#B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (NEB #N0447), 0.5 µM of each primer (Eurofins), 8 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl aliquots using the following temperature program: 30 sec at 98°C, 30x (15 sec at 98°C, 30 sec at 66°C, 2 min at 72°C), 5 min at 72°C. The PCR products were digested with 2 U DpnI for 2 hours at 37°C. The digested PCR products were purified following the Zymo DNA clean and concentrator kit (Zymo D4013). Gibson assembly (NEB #E2611) was used to ligate the two fragments to generate sub-library N346, C348. 50 ng of insert and 100 ng of vector were added to Gibson assembly (NEB #E2611) reaction and incubated for 1 hr at 50°C. 2 µL of the Gibson reaction was used to transform into 100 µL electrocompetent DH10B cells. The transformed cells were recovered in 900 µL super optimum broth (SOC). The recovered cells were plated at 1:1, 1:100, 1:1,000 and 1:10,000 dilutions (Kan50) to determine total diversity.

The second round *M. mazei* library construction followed the Gibson assembly procedure as described for the first round. Primers CL1984 and CL1985 were used to create the insert and primers CL1986 and CL1987 were used to create the vector in the second round of PCR reactions. Gibson assembly (NEB #E2611) was used to ligate the two fragments to generate the full library. 2 µL of the Gibson reaction was used to transform into 100 µL electrocompetent DH10B cells. The transformed cells were recovered in 900 µL super optimum broth (SOC). The recovered cells were plated at 1:1, 1:1,000 and 1:10,000 dilutions (Kan50) to determine total diversity. This procedure was repeated several times until an acceptable diversity was achieved (4.0×10^6). A flow chart of the construction of the *M. mazei* library can be seen in Figure 2.8.

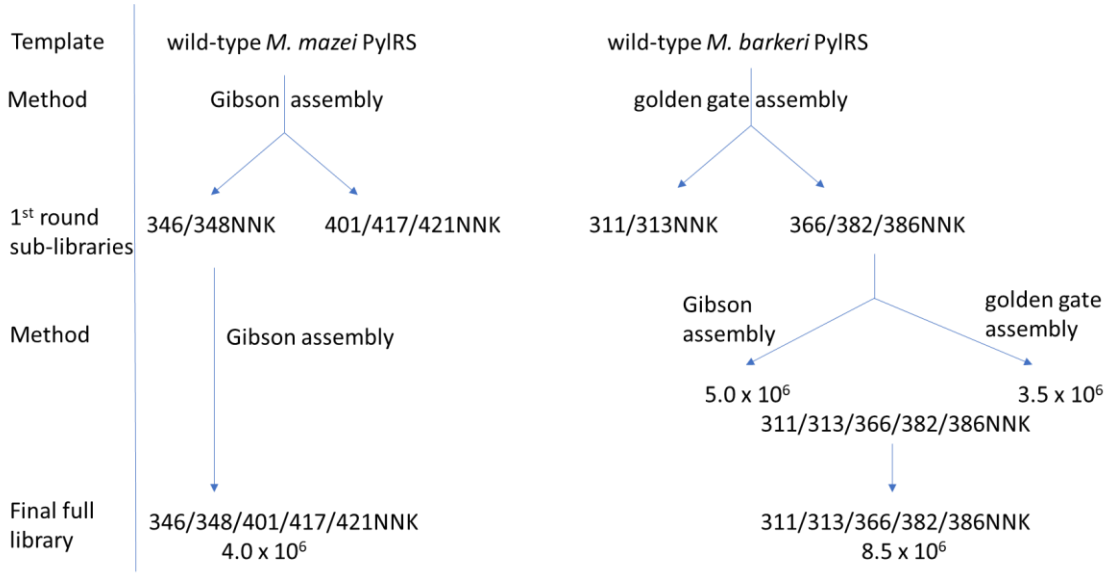


Figure 2.8 Flowchart of the construction of the *M. barkeri* and *mazei* libraries

Chapter 3. Directed Evolution

Chapter 3.1 Selection Strategy

In order to identify synthetase variants from the library that exclusively incorporate the UAAs of interest, a selection system was used to enrich the clones expressing active synthetase(s). The enrichment was conducted in rounds of positive and negative selections (Figure 3.1). First, the entire population of clones are subjected to selection media containing the UAA of interest. Each clone expresses a PylRS mutant variant. The first-round positive selection selects for clones that express an active synthetase. The clones survive the first-round positive selection by expressing a PylRS variant that is active with the UAA or any of the endogenous amino acids. Clones that survive the first positive selection are subjected to a negative selection where they are grown in the absence of the UAA to select against clones that express a synthetase that is active with any of the endogenous amino acids. Surviving clones of the negative selection, in principle, are those that express a PylRS variant that is exclusively active with the UAA of interest.

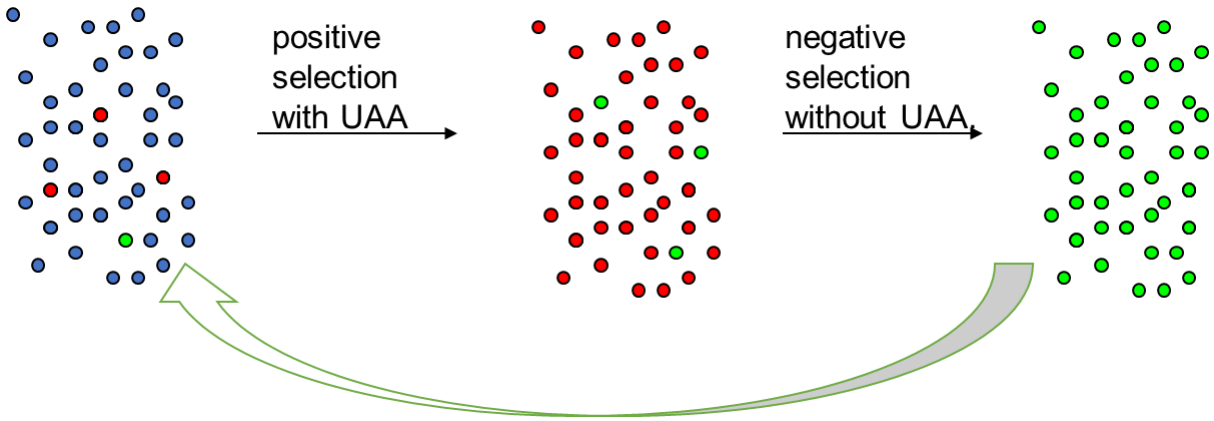


Figure 3.1 General scheme for directed evolution of PylRS variants using rounds of positive and negative selection. Blue= inactive PylRS. Red=PylRS active with endogenous amino acids. Green= PylRS active with UAA

Chapter 3.2 Positive Selection Strategy

Positive selection is achieved by transforming the library of mutant PylRS variants into *E. coli cells* that contain the positive selection plasmid. The positive selection plasmid is derived from the previously reported selection plasmid, pRep_JY^{CUA}, where the *Methanococcus jannaschii* tyrosyl tRNA had been cloned out in favor of PylT affording pRep PylT (Methods 3.5.1).[5] pRep PylT encodeca proK promoter-driven PylT and a gene on a constitutive promoter encoding chloramphenicol acetyltransferase (CAT, CmR) with an in-frame UAG stop codon mutation at the permissive site D111TAG (Figure 3.2). A permissive site is an amino acid residue where mutation to any other amino acid does not disrupt the tertiary structure and subsequent activity of the protein. Expression of chloramphenicol acetyltransferase confers cells resistant to the toxic antibiotic chloramphenicol by means of acetylation of chloramphenicol. Cells grown in the selection media containing the UAA and chloramphenicol will only survive if the full-length chloramphenicol resistance gene is expressed. Full-length CAT is only expressed if the in-frame amber stop codon is suppressed by the binding of PylT causing readthrough of the

stop codon. After the positive selection only clones containing PylRS mutants capable of aminoacylating PylT with the UAA and/or endogenous amino acids to cause expression of CAT by the readthrough of the amber codon will survive.

To demonstrate the robustness of the positive selection scheme, cells containing wild-type PylRS and the positive selection plasmid, pRep PylT, were assayed for growth on a range of chloramphenicol (Cm) concentrations in media containing 2 mM BocK and media without the UAA (Figure 3.2) (Methods 3.5.2). The assay showed that cells grown in the absence of BocK were killed in as low as 35 µg/mL Cm. Cells grown in media containing BocK were able to survive in media containing up to 200 µg/mL Cm. The phenotype determined that the optimal Cm concentration range for selection were 35-75 µg/mL. The concentration selected for selections was 50 µg/mL Cm.

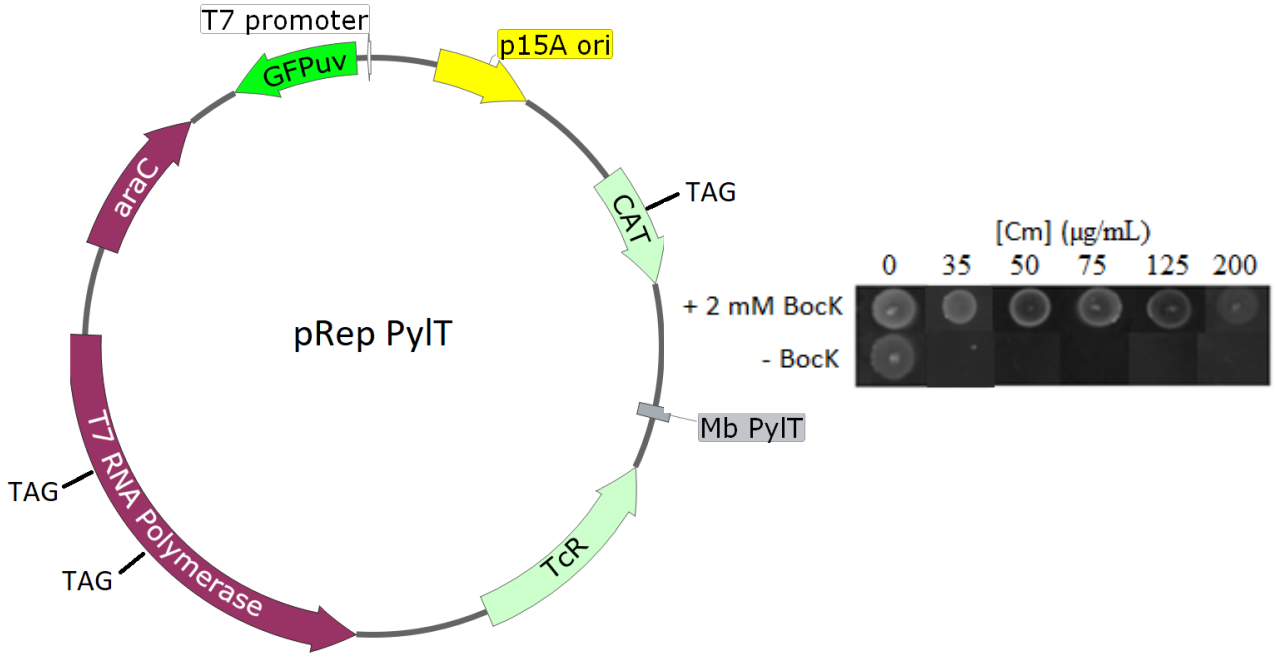


Figure 3.2. Left- Plasmid map of pRep PylT. Right- Positive selection growth assay.

3.3 Negative Selection Strategy

Surviving clones are then transformed into *E. coli* cells harboring the negative selection plasmid pBar_PylT (Figure 3.3). pBar_PylT expresses PylT and the gene coding for the toxic protein barnase with in-frame amber stop codon mutations at permissive sites Q2TAG and D44TAG. Barnase is a ribonuclease that is toxic to the cell when expressed. The expression of barnase is dependent upon readthrough of the in-frame stop codons by amber codon suppression. In this negative selection step, clones that express PylRS variants that are active with the endogenous amino acids are selected against by being grown in the absence of the UAA. The ability of PylRS variants to non-specifically suppress amber codons in the toxic barnase gene in the absence of the UAA, results in the expression of the toxic protein and subsequent bacterial cell death and elimination of that clone. Therefore all clones whose mutant PylRs aminoacylates endogenous amino acids onto PylT will be removed from the selection process. In the end of the round of positive and negative selection, clones that express PylRS variants that suppress amber codons exclusively with the UAA remain.

To demonstrate the robust phenotype of the negative selection, cells expressing wild-type PylRS and pBar PylT were assayed for growth on a range of D-arabinose concentrations (% w/v) in the presence and absence of 2 mM BocK (Figure 3.3) (Methods 3.5.3). The barnase gene is on an arabinose induced pBad promoter. The inducible promoter adds tunability to the expression levels of barnase. The phenotype demonstrated that there was cell death from the media containing the UAA opposed to the media without the UAA due to read through of the barnase in-frame amber codon. The growth assay determined that the optimal arabinose concentration was 0.02% as the cells grown in the media with UAA were killed off in as low as

0.2%. Interestingly, cells grown in the media supplemented with UAA exhibited an inhibition of growth in 0% arabinose. This is owed to the leaky expression of the barnase gene.

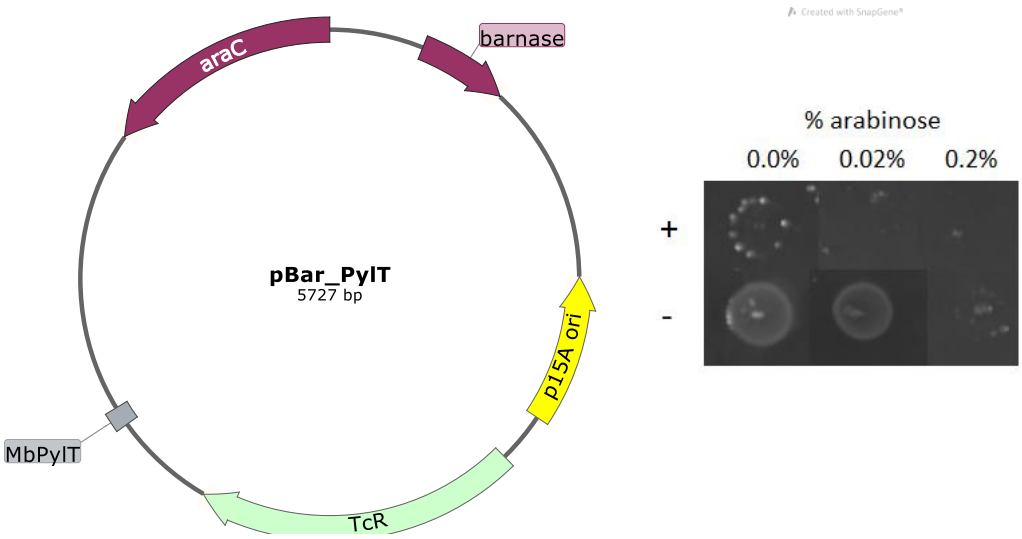


Figure 3.3. Left- Plasmid map of pBar_PylT. Right- Negative selection growth assay.

3.4 Screening Strategy

After the first round of positive/negative selections, the survivors are subjected to another positive selection followed by a screening (Figure 3.4) (Methods 4.14.1-2). If hit(s) are not obtained from the first screening, the surviving clones from the second positive selection are subject to another round of negative/positing selections followed by a second screening. The screening process is performed by inoculating individual wells on a 96-well plate with a single colony from the previous positive selection plate. Each monoclonal culture is grown to saturation and then assayed for its growth on 50 μg/mL Cm supplemented selective media that either contains the UAA or does not. A clone is considered a 'hit' if there is evident growth on the media containing the UAA and minimal to no growth on the media with the UAA omitted.

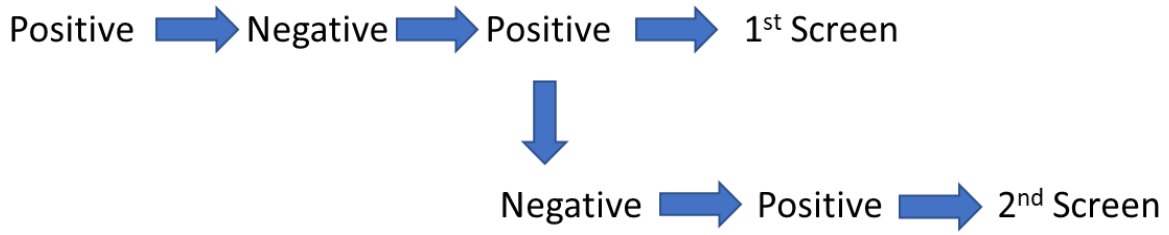


Figure 3.4. Sequence of events for the selection and screening of UAA incorporation by PylRS.

To identify survivors that have circumvented the chloramphenicol selection scheme, pRep_PylT also contains the reporter gene encoding green fluorescent protein UV mutant (GFPuv) to act as a secondary screen/selection. The expression of GFPuv operates under a cascade system. Expression of GFPuv is driven by the T7 promoter thus GFP expression is conditional on the expression of T7 polymerase. pRep PylT contains a gene under the control of an arabinose induced pbad promoter that encodes T7 RNA polymerase with the permissive stop codon mutations of M1TAG and Q107TAG. Successful amber codon suppression and subsequent expression of the T7 RNA polymerase ensues GFPuv expression. When irradiated with ultraviolet (UV) light around 395 nm, the cell population emits green light at 509 nm *in vivo*.

The advantage of using GFPuv as a secondary reporter gene in the selection scheme is its utilization as a screening tool. Unlike the CAT protein in the positive selection, the clone's survival is not contingent on the amber codon suppression dependent expression of GFPuv. In this way, false positive clones can be identified that survive in the Cm supplemented positive selection media by means other than amber codon suppression of CAT such as stop codon back mutation. Clones that fit this criterion will survive the rounds of positive selection, however,

will be easily identified by their lack of light emission under UV light and subsequently omitted from the screening process.

The second advantage of using GFPuv is that it offers an easy method to assay hits obtained after screening. The UAA dependent activity of each PylRS variant hit can be quantitatively assessed using the fluorescence intensity detected by a fluorimeter. This feature allows for a quick PylRS/UAA activity comparison from variant to variant.

The third advantage of using GFPuv as a reporter gene in the selection scheme is that it can be utilized as a dual positive or negative selection reporter. GFPuv offers the option of selecting for active amber codon suppression PylRS variants by utilizing fluorescence activated cell sorting (FACS) on a flow cytometer. Taken in tandem with the Cm positive selection, FACS allows for selecting against false positive survivors as well. On the other hand, FACS sorting can be utilized as a negative selection to select against clones containing PylRS variants that express GFPuv in the absence of the UAA. In addition, cell sorting allows for tunable sorting stringency by setting fluorescent intensity gates.

3.5 Methods and Materials

3.5.1 pRep_PylT Cloning

Plasmid pRep_PylT was generated by replacing the *Methanococcus jannaschii* tyrosyl tRNA (JY^{CUA}) of plasmid pRep_2YC_JYCUA with the pyrrolysyl tRNA (PylT) from *Methanosarcina barkeri* from plasmid pCATUPP_PylT.[5] pRep_2YC_JY^{CUA} (2 μg) was digested with 10 U PshAI overnight. The digest was separated by mass on a 1% agarose gel and the vector fragment (~9.1 kB) was isolated by gel extraction following a Qiagen gel extraction kit (Qiagen #28706). pCATUPP_PylT was used as a template in a PCR reaction to amplify an insert fragment containing the PylT gene with homologous ends to the vector fragment. This

PCR used the primers CL2226 and CL2227. The PCR reactions were prepared in 160µL containing 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (NEB #N0447), 0.5 µM of each primer (Eurofins), 8 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl aliquots using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec at 68°C, 2 min at 68°C), 5 min at 72°C. 100 ng of the vector and 10 ng of the insert were added to a Gibson assembly (NEB #E2611) reaction and incubated for 1 hour at 50°C to ligate the two fragments. 2 uL of the Gibson assembly reaction was transformed into 100 µL of DH10B cells and recovered in 900 µL SOC. The recovered cells were plated on LB agar (Tet15) and incubated overnight at 37°C. Ten colonies were picked and inoculated into individual 5 mL LB cultures (Tet15). The cultures were grown to saturation and miniprepped following a NEB monarch plasmid mini prep kit (T1010L). The DNA extracted from each culture was digested with XbaI (unique to PylT) and SacI to determine successful cloning of the PylT insert into the pRep vector. Successful cloning was also confirmed by sequencing.

3.5.2 Positive Selection Growth Assay

pBK_PylRS (wild-type) was transformed into DH10B cells harboring the positive selection pRep_PylT plasmid and plated onto LB agar (Tet15/Kan50). Upon growth, a colony was picked from the plate and inoculated into a 5 mL LB culture (Tet15/Kan50). The culture was grown to saturation. 5 µL was spotted onto 6 plates (Tet15/Kan50/arab 0.2%) containing 2 mM BocK and increasing amount of Cm of 0, 35, 50, 75, 125, 200 µg/mL. 5 µL was also spotted onto 6 plates (Tet15/Kan50/arab 0.2%) containing no BocK and increasing amount of Cm of 0, 35, 50, 75, 125, 200 µg/mL. The plates were incubated at 37°C overnight.

3.5.3 Negative Selection Growth Assay

pBK_PylRS (wild-type) was transformed into DH10B cells harboring the negative selection pBar_PylT plasmid and plated onto LB agar (Tet15/Kan50). Upon growth, a colony was picked from the plate and inoculated into a 5 mL LB culture (Tet15/Kan50). The culture was grown to saturation. 5 µL was spotted onto 3 plates (Tet15/Kan50) containing 2 mM BocK and increasing amount of arabinose of 0, 0.02, and 0.2% (w/v). 5 µL was also spotted onto 3 plates (Tet15/Kan50) containing no BocK and increasing amount of arab 0, 0.02, 0.2%. The plates were incubated at 37°C overnight.

Chapter 4. Acylated Lysine Derivatives

4.1 Post-translational Modifications

Post-translational modifications (PTM) are nature's escape from genetic imprisonment. Precise control of protein function is essential for the organization and function of biological systems. Among different regulatory processes, PTMs provide this mechanism to govern extensive protein functions. PTMs of proteins modulate and extend the range of possible protein functions by covalently attaching small chemical moieties to selected amino acid residues. Proteomes contain different types of PTMs such as phosphorylation, glycosylation, methylation, acylation, ubiquitination and sumoylation. PTMs confer novel properties to the modified proteins, including changes in enzymatic activity, subcellular localization, interaction partners, protein stability and DNA binding. Most PTMs are important pillars of cell signaling and enable the cell to react specifically to internal and external stress conditions.[54] These modifications greatly expand the coding capacity especially of eukaryotic genomes and lead to proteomes much more complex than the encoding genomes would predict.

4.2 Lysine Acylation

Lysine acylation is the PTM of focus due to its prominence and prevalence in the proteome. Many discovered acylated proteins have been involved in metabolism, translational regulation or stress response. Protein acylation is an evolutionarily conserved, reversible, and tightly regulated PTM occurring in both prokaryotes and eukaryotes, in which an acyl group is transferred to a specific lysine on a polypeptide chain (Figure 4.1). Lysine acyl transferases (KATs) catalyze the enzymatic transfer of an acyl group from acyl-coenzyme A (Ac-CoA) to the εN-lysine side chain. The deacylation of lysine residues is catalyzed by lysine deacylases

(KDACs). The tight balance between protein acylation and deacylation events play a critical role in the regulation of gene expression and signal pathways, affecting a range of cellular processes. Under physiological conditions, cells are constantly exposed to variations in nutrient availability and environmental stress. It is believed that histone acylation provides a means of adapting to these challenges to maintain physiological homeostasis.[55] Consequently, malfunctioning acylation machinery can lead to a variety of diseases such as cancer, neurodegenerative diseases and cardiovascular disorders.[56–59]

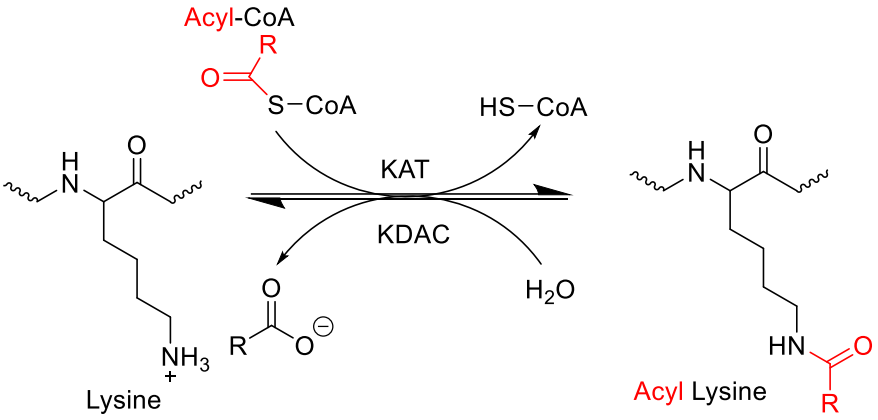


Figure 4.1. Reversible lysine acylation by lysine acyl transferase and lysine deacylase.

4.3 Lysine Acylation in Epigenetics

The acylation of histone proteins discovered by Vincent Allfrey and colleagues in 1964 became the first well established example of biologically functional protein acylation. [58,60] The histone molecules are modified by various PTMs, of which phosphorylation, methylation and acylation play the most crucial roles for DNA accessibility and transcriptional regulation. The PTMs, and especially histone acylation, determine the histone assembling as well as the folding and compactness of the DNA-histone interaction. Current evidence supports the idea that certain site-specific acylation of histone tails is sufficient to alter nucleosome dynamics and chromatin

folding.[61,62] The histone tails insert themselves in the minor grooves of the DNA and extend through the double helix, which leaves them open for modifications involved in transcriptional activation.[63] Histone tail acylation destabilizes the DNA-histone interaction because an acylated lysine side chain loses its positive charge and thus the ability to form salt bridges with the negatively charged phosphate backbone of DNA.[64] In this case, acylation plays a large role in epigenetic regulation, acting in part to 'open up' chromatin into the euchromatin form for appropriate transcriptional machinery to access the DNA template resulting in transcriptional activation (Figure 4.2).[61] Conversely, histones found in a hypoacetylated state maintain the compact heterochromatin form rendering the DNA template transcriptionally repressed.[65] After binding to the DNA, histones are deacylated, and reacylated again in a different pattern, which is essential for chromatin-based transcriptional regulation.[66,67]

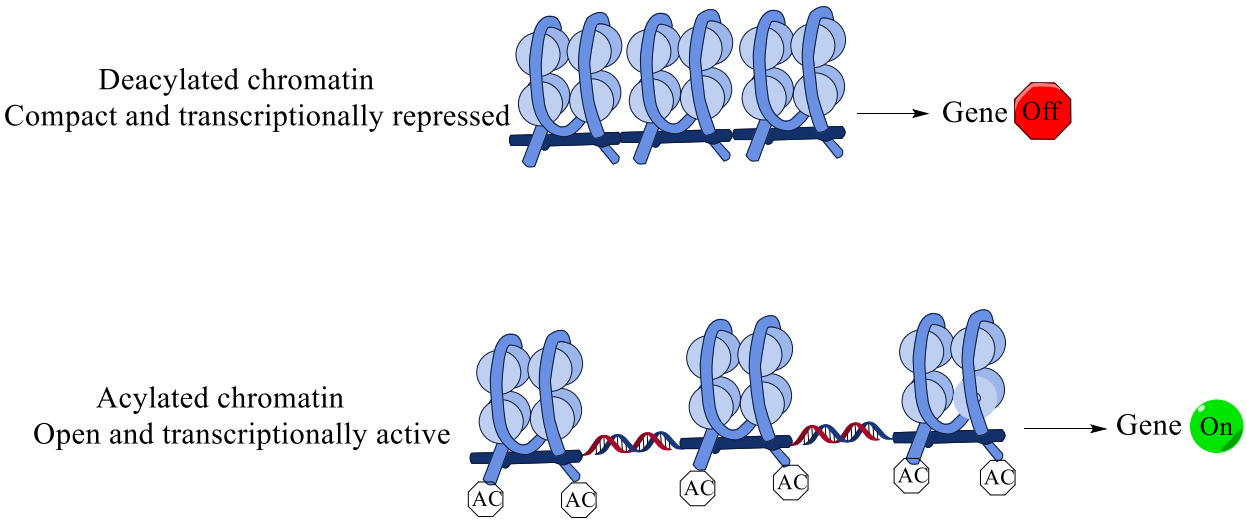


Figure 4.2 Epigenetic regulation by reversible lysine acylation on histone tails.

4.4 Acylated Lysine Derivatives

Lysine has been shown to be modified by several different acyl groups which are often donated from the corresponding acyl-CoA. [68697071] Acetylation, succinylation,[69] propionylation,[72] butyrylation,[73,74] malonylation,[75] glutarylation,[76] myristolyation,[77] and crotonylation[78] are among lysine acyl modifications that have been discovered in proteins (Figure 4.3).[79] Out of those discovered modifications, acetylated,[12] propionylated,[13] butyrylated,[80] and crotonylated[80] lysine have been added to the genetic code. All such acylations are most likely derived from their CoA derivatives, possibly through a nonenzymatic mechanism.[81–83] Other acyl-CoAs present in the mitochondria are also likely to modify lysine, therefore, there is potential that other acylations on lysine residues have yet to be discovered. Specific acylations may be enriched on metabolic enzymes belonging to the pathway that generates their respective acyl-CoA. For example, isobutyrylation (a valine derivative), isovalerylation (a leucine derivative), and methylbutyrylation (an isoleucine derivative) may be found on enzymes involved in branched-chain amino acid catabolism. These known and potential lysine modifications that are derived from intracellular acyl-CoA metabolites provide a wide spectrum of epigenetic control of gene expression. The extent to which these modifications are actively added and removed by enzymes is a current interest in deciphering the “histone code.”

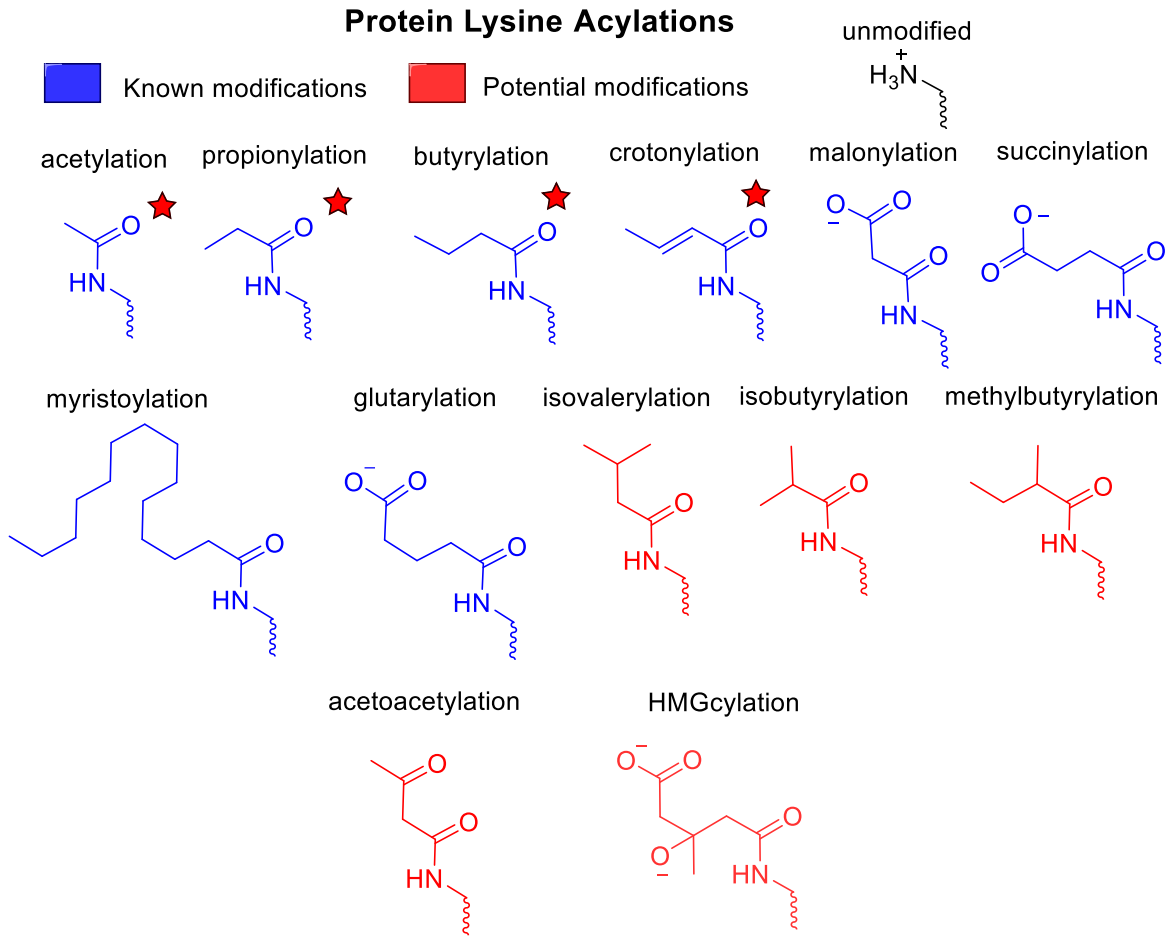


Figure 4.3 Known acylated lysine derivatives found in proteins (blue). Potential acylated lysine derivatives yet to be found in proteins (red). Starred modification indicates that the modification of lysine has been added to the genetic code.

4.5 β-hydroxybutyrylation

The connection between acylated histone lysine and metabolism is of particular interest as it may provide new viewpoints to understand cellular function. Because the activity of histone lysine acylation can be regulated by cellular metabolites, fluctuations in levels of these metabolites may up- or downregulate gene expression by modulating chromatin states.[84,85] In

2016 Zhao et al. discovered the evolutionary conserved PTM of histone N^{ε}-β-hydroxybutyryl lysine (HBK) among diverse eukaryotic species.[86] They found that β-hydroxybutyrylation serves as an important mechanism by which cells adapt to changes in cellular energy sources by regulating gene expression. The group linked the significant increase of HBK-containing histones levels under conditions of starvation to elevated β-hydroxybutyrate levels. During prolonged fasting, carbohydrates are scarce, and consequently energy is derived from fatty acid metabolism. Under these conditions, a large amount of acetyl-CoA generated from fatty acid oxidation is converted into ketone bodies.[85] Ketone bodies consisting of β-hydroxybutyrate provide an energy source during starvation.

Zhao et al. reported that the increased HBK on histone H3 lysine 9 (H3K9HBK) is associated with upregulation of genes involved in starvation-responsive pathways. The metabolite directed histone modification of β-hydroxybutyrylation was proven to have a critical role in gene expression regulation during prolonged fasting, however, the direct mechanism remains unknown. Despite the importance of HBK in homeostatic pathways, homogeneous recombinant proteins that contain HBK at defined sites have not been produced. The ability to site-specifically incorporate HBK into histones would allow one to study its effect on chromatin remodeling and subsequent gene regulation. For this reason, HBK is the main acylated lysine derivative of interest to add to the genetic code.

At the time, synthesis of HBK by Sahan Galbada of the Cropp research group proved to be quite difficult. In order to circumvent this difficulty while still being able to run directed evolution of the PylRS libraries to accept HBK, structural analogs served as surrogates. The close structural analogs, N^{ε}-isovaleryl lysine (IVK) and N^{ε}-isobutyryl lysine (IBK), were synthesized by Sahan and used in place of HBK for selecting for an active synthetase with HBK.

IVK and IBK are both potential acylated lysine modifications that have not yet been found in proteins, however, their metabolite co-factors have been discovered. Along with acting as surrogates in the selection for an active synthetase with HBK, IVK and IBK both were appealing to add to the genetic code for their potential roles in protein function (Figure 4.4).

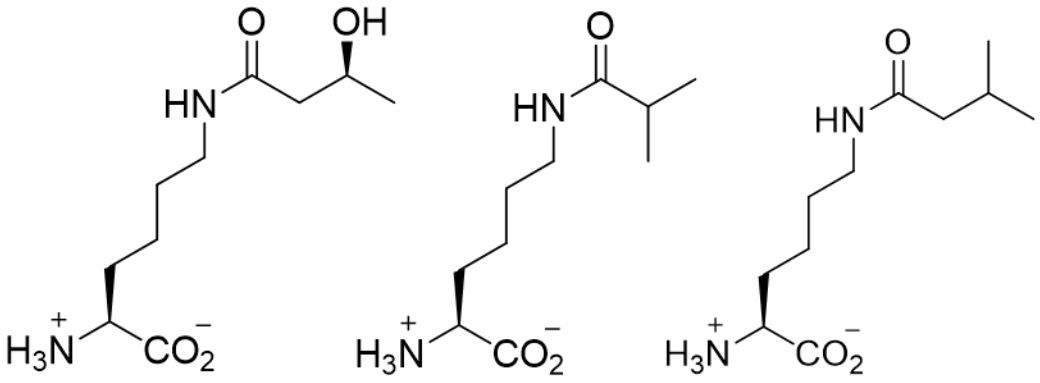


Figure 4.4. Chemical structures of HBK (left), IBK (middle), and IVK (right).

4.6 IBK Selections & Screening

Directed evolution for an active PylRS variant with IBK was performed using the positive/negative selection scheme along with the following libraries: the *barkeri* derived mixed lysine analog and tyrosine analog libraries 1. L274A/Y348F/A267NNK/Y271NNK_/C313NNK 2. N311A/C313A/Y271NNK/ L274NNK/Y349NNK, the *mazei* derived 346NNK/348NNK/401NNK/417NNK/421NNK, and the *barkeri* derived library 311NNK/313NNK/366NNK/382NNK/386NNK (Methods 4.14.1). For full confidence the library diversity was covered, ten times the diversity of each library was subjected to the positive selection media containing 50 μg/mL Cm and 2 mM IBK. The library plasmid was isolated from the thousands of surviving clones and transformed into cells harboring the negative selection plasmid, pBar_PylT. The recovered cells were plated on negative selection media containing the absence of IBK and 0.02% arabinose. Again, the library plasmid was isolated from the

thousands of surviving clones and transformed into cells harboring the positive selection plasmid, pRep_PylT. The recovered cells were plated onto positive selection media containing 50 μg/mL Cm and 2 mM IBK and let incubate for ~40 hours. Interestingly, the colonies need substantial more incubation time for growth after the second positive selection. Individual fluorescent clones from the plate were picked to inoculate the wells of a 96-well culture block. The 96-well block incubated till growth saturation. Each clone was pintooled from the block and spotted onto screening plates containing 50 μg/mL and +/- 2 mM IBK (Figure 4.5) (Methods 4.14.2). Examination of the screening plates showed that 32/96 spotted colonies were hits.

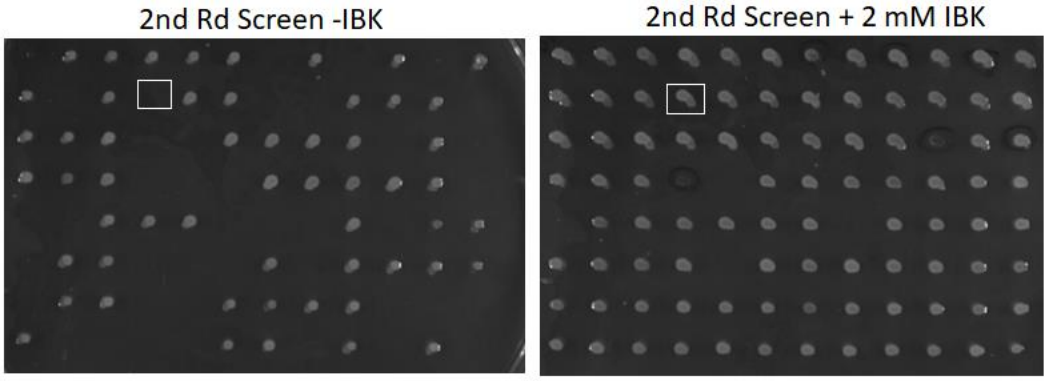


Figure 4.5. Left- IBK screening plate without IBK. Right- IBK screening plate with 2 mM IBK. White rectangle represents a hit for example.

4.7. IVK Selections & Screening

Directed evolution for an active PylRS variant with IVK was performed using the positive/negative selection scheme along with the following libraries: the *barkeri* derived mixed lysine analog and tyrosine analog libraries 1. L274A/Y348F/A267NNK/Y271NNK_/C313NNK 2. N311A/C313A/Y271NNK/ L274NNK/Y349NNK, the *mazei* derived 346NNK/348NNK/401NNK/417NNK/421NNK, and the *barkeri* derived library 311NNK/313NNK/366NNK/382NNK/386NNK (Methods 4.14.1). For full confidence the

library diversity was covered, ten times the diversity of each library was subjected to the positive selection media containing 50 μg/mL Cm and 2 mM IVK. The library plasmid was isolated from the thousands of surviving clones and transformed into cells harboring the negative selection plasmid, pBar_PylT. The recovered cells were plated on negative selection media containing the absence of IVK and 0.02% arabinose. Again, the library plasmid was isolated from the thousands of surviving clones and transformed into cells harboring the positive selection plasmid, pRep_PylT. The recovered cells were plated onto positive selection media containing 50 μg/mL Cm and 2 mM IVK and let incubate for ~40 hours. Individual fluorescent clones from the plate were picked to inoculate the wells of a 96-well culture block. The 96-well block incubated till growth saturation. Each clone was pintooled from the block and spotted onto screening plates containing 50 μg/mL and +/- 2 mM IVK (Figure 4.6) (Methods 4.14.2). Examination of the screening plates showed that 4/60 spotted colonies were hits.

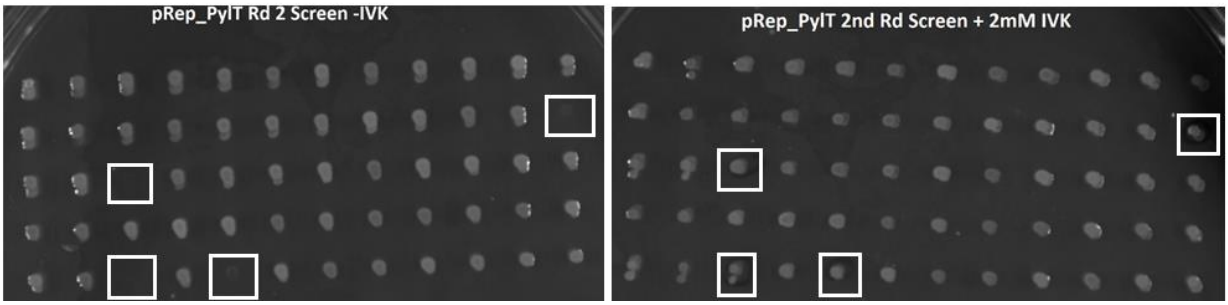


Figure 4.6. Left- IVK screening plate without IVK. Right- IVK screening plate with 2 mM IVK. White rectangles represent hits.

4.8. IBK and IVK Active Synthetases Examination

From the IBK and IVK screening pates, half the IBK hits (16) and all of the IVK hits (4) had the DNA isolated to identify which *Methanosarcina* strain each hit was derived from. The

DNA from each clone was digested with HindIII which cuts *mazei* derived PylRS twice (2.0 and 1.4 kb), *barkeri* derived PylRS once (3.3 kb) and pRep_PylT twice (8.8 and 1.6 kb). The digested products were separated by mass on a 1% agarose by gel electrophoresis (Figure 4.7).

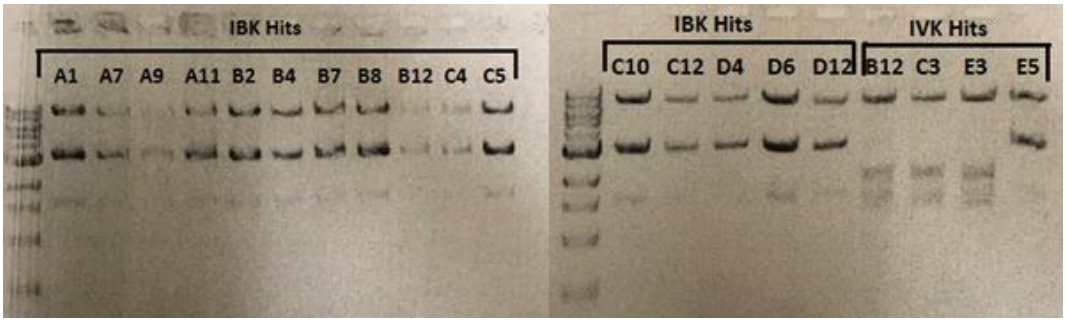


Figure 4.7. HindIII digested IBK and IVK hits separated by mass on an agarose gel.

Every IBK hit was derived from the *barkeri* strain and ¾ of the IVK hits were derived from the *mazei* strain. Each hit was sequenced to identify the amino acid content of the synthetase (Table 4.1). All 16 sequenced IBK hits converged on the wild-type synthetase identity. The fact that every observed wild-type IBK hit contained the same exact codon at the mutated sites as wild-type suggested that there was a contamination of wild-type PylRS in the *barkeri library*. Nonetheless, the enrichment of the wild-type contamination in the IBK selections demonstrated that wild-type is the superior synthetase for aminoacylating IBK onto PylT. Two IVK hits from the *mazei* strain, B12 and E3, also converged on the wild-type synthetase identity. IVK hits, C3 and E5, converged on the same synthetase identity from the *mazei* and *barkeri* strain, respectively. They both had the subtle mutation of a cysteine to a threonine. Interestingly, two synthetases from two different libraries derived from different strains both converged on the same mutation. This hetero library convergence confirmed that the selections worked. It also proved that wild-type PylRS is active with both IBK and IVK.

Table 4.1. Sequenced IBK and IVK hits.

	346 (311)	348 (313)	401 (366)	417 (382)	421 (386)	Strain
Wild-type	**N**	**C**	**V**	**W**	**G**	
All IBK Hits	**N**	**C**	**V**	**W**	**G**	**Mb**
IVK C3	**N**	**T**	**V**	**W**	**G**	**Mm**
IVK B12/E3	**N**	**C**	**V**	**W**	**G**	**Mm**
IVK E5	**N**	**T**	**V**	**W**	**G**	**Mb**

A fluorescence assay was performed on the converged IBK hit and the three unique IVK hits (Figure 4.8) (Methods 4.14.3) The expression of GFPuv of each clone grown in the presence and the absence of the respective UAA was quantified by a fluorimeter. Each culture was excited with 395 nm light and emission was detected at 515 nm. The fluorescence assay revealed that the activity of wild-type with IVK and IBK was comparable to that of wild-type with BocK. Also, it was demonstrated that GFPuv expression was directly dependent on the inclusion of each UAA in the media where absence of the UAA resulted in minimal GFPuv expression. Hit IVK E5 was chosen to be the synthetase used in subsequent IVK incorporation studies and was named IVKRS.

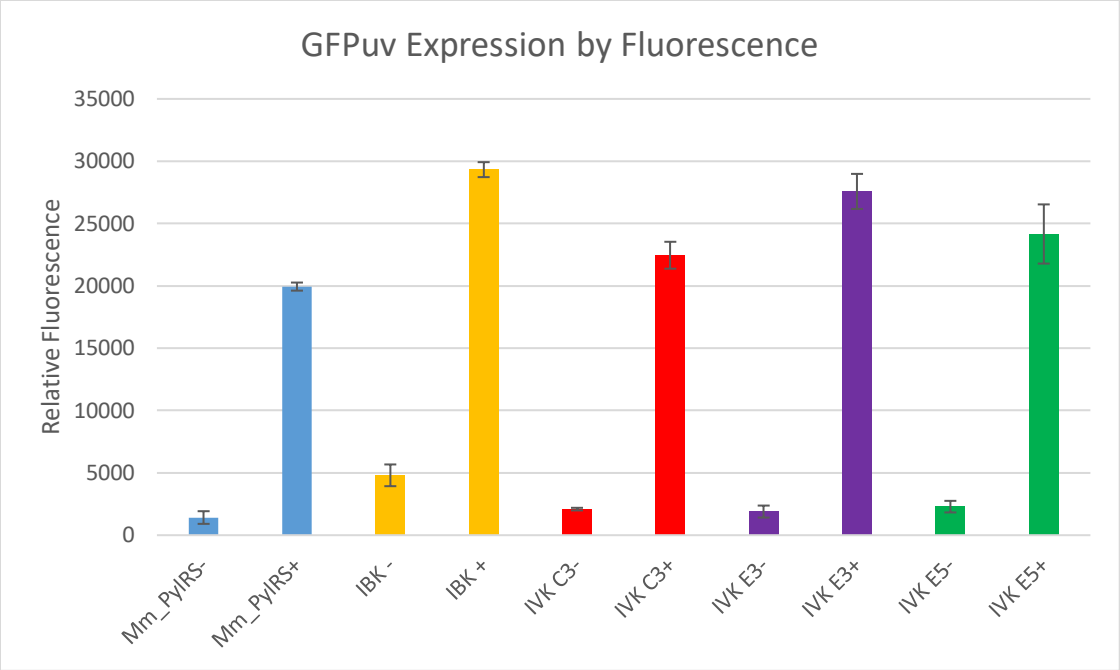


Figure 4.8. GFPuv expression by relative fluorescence for IBK and IVK hits. Error bars show standard deviation of expression cultures performed in triplicate.

4.9 IBK and IVK Incorporation into Protein

In order to incorporate IBK and IVK into protein, the model protein ubiquitin (Ub) was used. Ubiquitin is a small (8.6 kDa) regulatory protein that is used as a PTM for cell signaling, protein-protein interactions, and protein function alteration. Ubiquitin was used as a model protein because previous work in the Cropp group used ubiquitin for amber codon suppression to incorporate UV-activatable cross linkers to study protein-protein interactions. The ubiquitin gene on the pCNB_synUBK11TAG plasmid used contained a K11TAG in-frame amber stop codon mutation. It was also used as a model protein because it has shown to 'fly well' in the mass spectrometer to obtain the mass that confirms UAA incorporation.

Cells containing IVKRS and pCNB_synUBK11TAG were grown in three 25 mL cultures where each culture was induced with 0.2% arabinose and 1. no UAA 2. 2 mM IBK and 3. 2 mM

IVK. It was determined by fluorescence assay that IVKRS was also active with IBK (Figure 4.9) (Methods 4.14.3). The ubiquitin protein is expressed with an intein-chitin binding domain (CBD) fusion. The fusion protein was immobilized on a chitin resin affinity matrix and purified by intein-mediated cleavage of the fusion protein chitin binding domain (Methods 4.14.4).

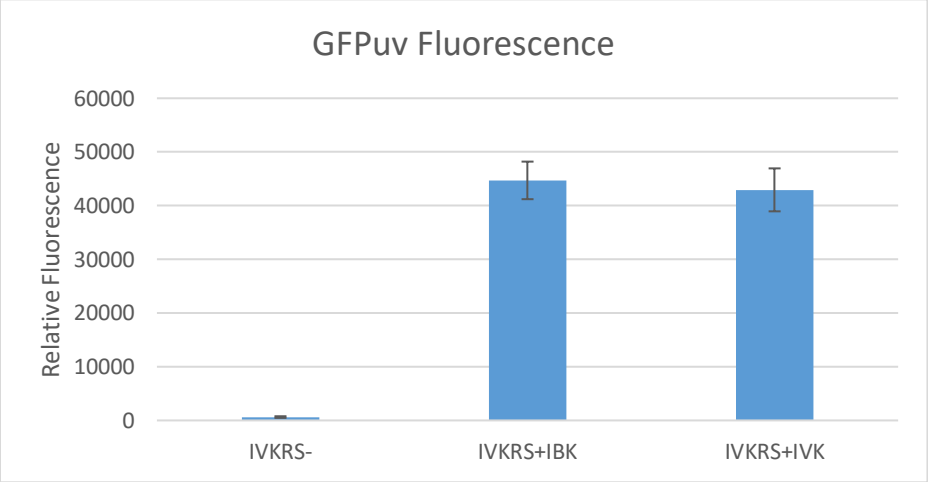


Figure 4.9. GFPuv relative fluorescence of IVKRS without UAA, IBK, and IVK.

The purified ubiquitin construct from each protein expression was separated by mass on a sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (SDS-PAGE) protein gel (Figure 4.10). The protein gels showed that ubiquitin expression was dependent on each UAA in the media for read through of the in-frame amber codon in the ubiquitin construct. When no UAA was supplemented in the expression culture, ubiquitin expression was minimal.

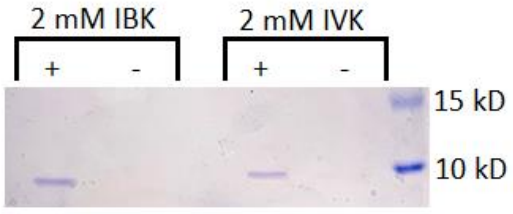


Figure 4.10. Inset of SDS-PAGE gels of Ub expression with IBK and IVK.

Incorporation of IBK and IVK into the model protein ubiquitin was confirmed by mass spectrometry. Each purified modified ubiquitin protein was separated by HPLC and the monoisotopic mass was determined by a mass spectrometer. The mass spectrum of IBK showed two major peaks (Figure 4.11). The smaller of the two peaks at 8,636 Da was the expected mass of intact ubiquitin with IBK incorporated at the K11TAG site. The larger of the two peaks at 8,522 Da was the mass of intact ubiquitin with IBK incorporated at the K11TAG site but with the c-terminal glycines cleaved. The difference of 114 Da from the removal of the two glycines accounted for the difference in mass. The cleavage of the c-terminal glycines by a suspected protease has been demonstrated before in work with this ubiquitin construct by the Cropp group.

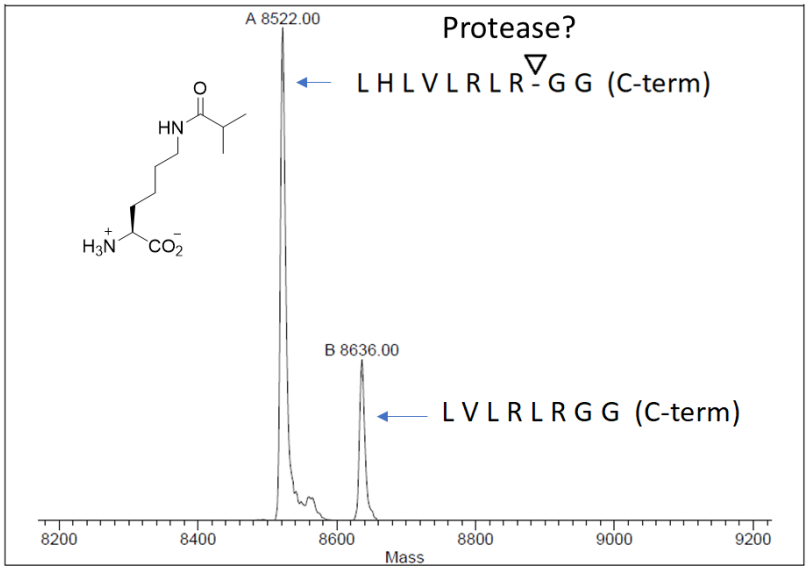


Figure 4.11. Deconvoluted monoisotopic mass spectrum of IBK incorporation into ubiquitin. Calculated M/Z^+= 8,636.3 Da. Observed M/Z^+=8,522.0 and 8,636.0 Da.

The mass spectrum of IVK incorporated into ubiquitin showed two major peaks as well (Figure 4.12). The larger of the two peaks at 8,650 Da was the monoisotopic mass from intact ubiquitin with IVK incorporated at the K11TAG site. The smaller of the two peaks at 8,556 Da

was derived from native ubiquitin. Deacylation of IVK at position 11 resulted in native ubiquitin.

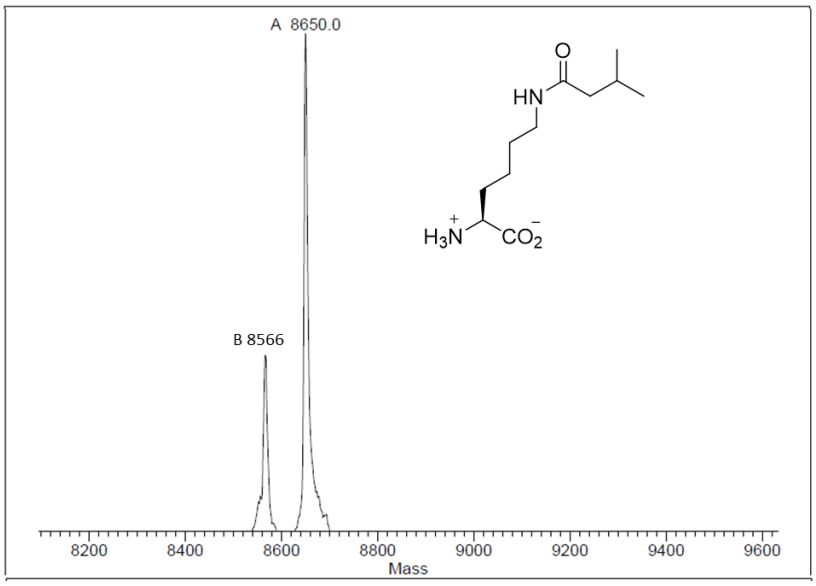


Figure 4.12. Deconvoluted monoisotopic mass spectrum of IVK incorporation into ubiquitin. Calculated M/Z^+= 8,650.3 Da. Observed M/Z^+=8,566.0 and 8,650.0 Da.

4.10 Hydroxybutyryl lysine selections

HBK was synthesized by Sahan Galbada as a methyl ester protected racemic mixture (Figure 4.13), therefore, if there was equal enantiomer bias, the media was supplemented with 1 mM of each enantiomer (Appendix III.II). The methyl ester protecting group was expected to be hydrolyzed affording the free amino acid once introduced into the cell. Selection for an active PylRS variant with HBK was performed using the positive/negative selection scheme along with the following libraries: the *barkeri* derived mixed lysine analog and tyrosine analog libraries 1. L274A/Y348F/A267NNK/Y271NNK_/C313NNK 2. N311A/C313A/Y271NNK/ L274NNK/Y349NNK, the *mazei* derived 346NNK/348NNK/401NNK/417NNK/421NNK, and

the *barkeri* derived library 311NNK/313NNK/366NNK/382NNK/386NNK (Methods 4.14.1). For full confidence the library diversity was covered, ten times the diversity of each library was subjected to the positive selection media containing 50 μg/mL Cm and 2 mM HBK. The library plasmid was isolated from the thousands of surviving clones and transformed into cells harboring the negative selection plasmid, pBar_PylT. The recovered cells were plated on negative selection media containing the absence of IVK and 0.02% arabinose. Again, the library plasmid was isolated from the thousands of surviving clones and transformed into cells harboring the positive selection plasmid, pRep_PylT. The recovered cells were plated onto positive selection media containing 50 μg/mL Cm and 2 mM HBK and let incubate for ~40 hours. Individual fluorescent clones from the plate were picked to inoculate the wells of a 96-well culture block. The 96-well block incubated till growth saturation. Each clone was pintooled from the block and spotted onto screening plates containing 50 μg/mL and +/- 2 mM HBK (Figure 4.13) (Methods 4.14.2).

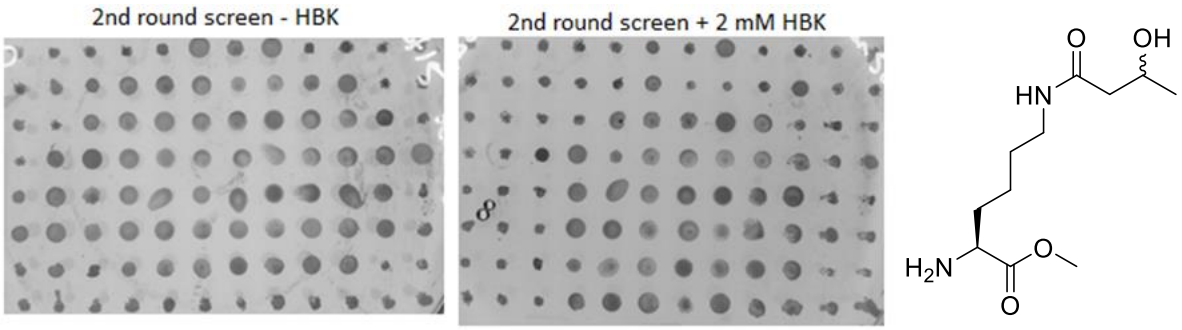


Figure 4.13. Unsuccessful screening for HBK synthetase after 2nd round positive selection. Left- Screening plate without HBK. Middle- Screening plate with 2 mM IBK. Right- Chemical structure of racemic methyl ester hydroxybutyryl lysine.

The clone screening demonstrated that the picked clones were not exclusively active with HBK as there was equal growth on the screening plates. Two more 96-well blocks worth of clones were picked and screened in the same fashion (Methods 4.14.1-2). The screening plates showed similar results where there was equal growth on spotted plates without HBK as the spotted plates with 2 mM HBK. Amber codon suppression of the in-frame amber codon on CAT by PylRS variants charging PylT with any of the endogenous amino acids was thought to be the cause of growth on the screening plate without HBK supplemented. With no successful hits, the surviving clones from the second positive selection were subjected to another round of negative selection followed by a final round of positive selection. This was done in order to further select against PylRS variants that were active with the endogenous amino acids. From the last round positive selection plate, 192 clones were picked and screened. Much like the screening plates in Figure 4.13, there were no clones that expressed a PylRS variant that exhibited HBK exclusive amber codon suppression.

4.11 Additional library construction

Multiple rounds of selections and screening of hundreds of clones had proved that the current libraries did not offer a synthetase variant active with HBK. The lack of success in being able to select for an active synthetase variant caused reevaluation of the libraries being used in the selections. The fact that HBK is a close structural analog of both IBK and IVK, both active with wild-type PylRS, gave cause that a PylRS variant could accept HBK, however, it was clearly not contained in the current libraries.

The crystal structure of *M. mazei* coordinated with adenylated PylK was reexamined as well as literature was consulted of other research groups who have previously made libraries

from PylRS. The idea was to create an additional library with new mutations that could be added to one of the sub-libraries. Upon examination of the crystal structure, three new residues, L305(270), Y306(271), and L309(274) were selected for site-saturated mutagenesis (Figure 4.14). These three residues are found in the hydrophobic cleft deep in the active site cavity and are in close proximity to the pyrroline ring of pyrrolysine. Modifications such as the pyrroline ring to the ε-N of lysine can be predicted to occupy this space within the cavity. Therefore, mutations to these three residues could prove to better accommodate UAAs with these modifications.

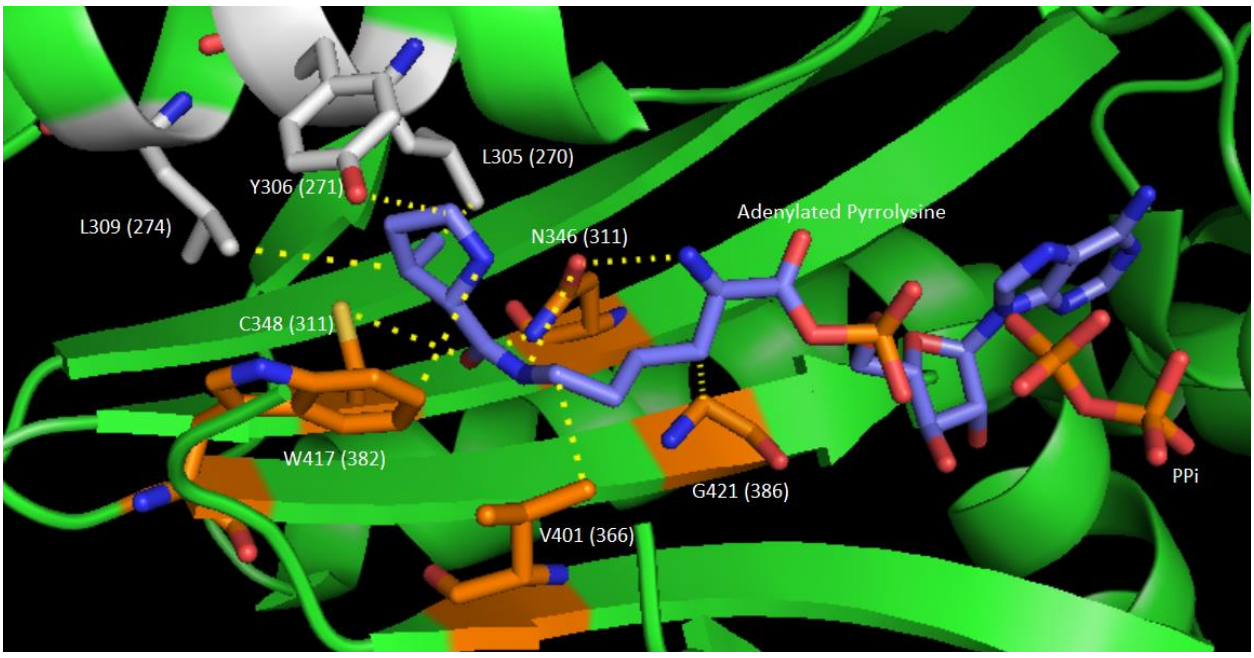


Figure 4.14. Crystal structure of *M. mazei* PylRS (green) coordinated with adenylated pyrrolysine (blue). Original library mutations are shown in orange. New library mutations are shown in white. PDB code 2Q7H.[50]

The three selected residues were chosen to be added to a library with the N311/C313 mutations. Of the five mutations used in the original library, N311/C313 are thought to be more

effective for PylRS evolution compared to the V366,W382, and G386 mutations. Therefore, a *barkeri* derived L270NNK/Y271NNK/L274NNK sub-library was constructed using site-saturated mutagenesis by inverse PCR (Methods 4.14.5). The *M. barkeri* L270/Y271/L274N sub-library was used as a template to generate the L270/Y271/L274/N311/C313 full library using both the Gibson assembly and inverse PCR by golden gate assembly approaches (Methods 4.14.5). The Gibson assembly method yielded 2.5 x 10^6 variants while the inverse PCR by golden gate assembly yielded 7.5 x 10^6 variants. The total population of clones were pooled together to form a 1.0 x 107-member library (Figure 4.15).

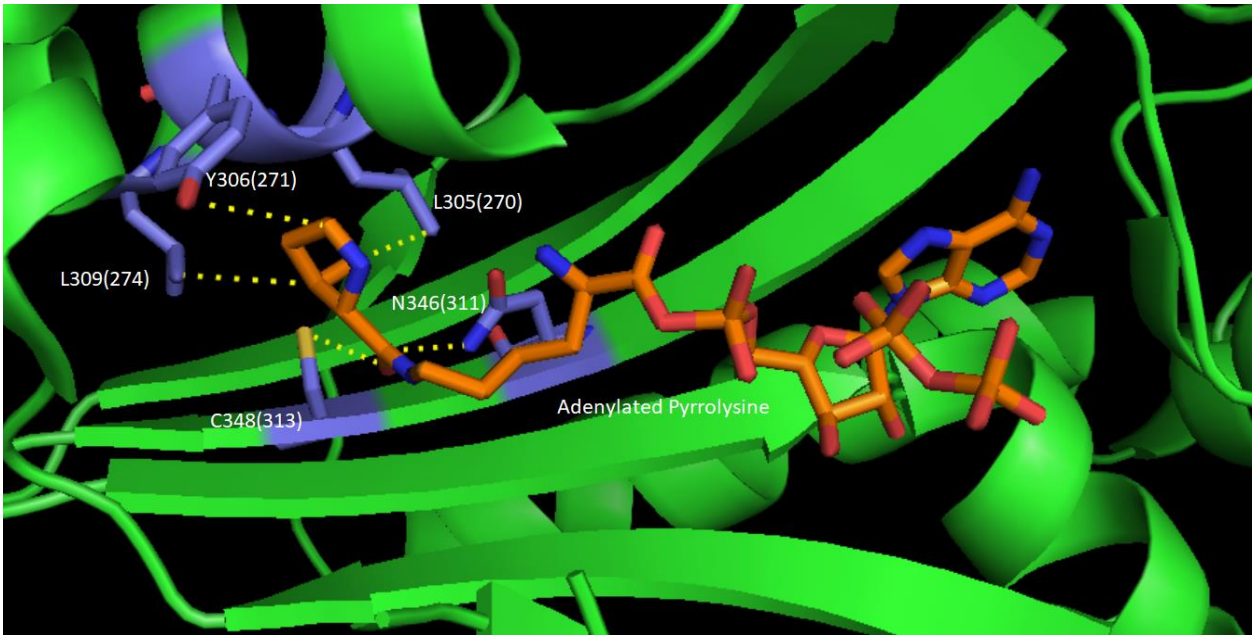


Figure 4.15. Crystal structure of *M. mazei* PylRS (green) coordinated with adenylated pyrrolysine (orange). Shown in blue are the residues used in site-saturated mutagenesis. PDB code 2Q7H[50]

4.12 Hydroxybutyryl lysine selections with additional library

The second attempt to select for an active PylRS variant with racemic methyl ester HBK was performed using the positive/negative selection scheme along with the following libraries: the *barkeri* derived mixed lysine analog and tyrosine analog libraries 1. L274A/Y348F/A267NNK/Y271NNK/C313NNK 2. N311A/C313A/Y271NNK/ L274NNK/Y349NNK, the *mazei* derived 346NNK/348NNK/401NNK/417NNK/421NNK, the *barkeri* derived library 311NNK/313NNK/366NNK/382NNK/386NNK and the newly constructed *barkeri* derived L270NNK/Y271NNK/L274NNK/N311NNK/C313NNK (Methods 4.14.1). For full confidence the library diversity was covered, ten times the diversity of each library was subjected to the positive selection media containing 50 μg/mL Cm and 2 mM HBK. The library plasmid was isolated from the thousands of surviving clones and transformed into cells harboring the negative selection plasmid, pBar_PylT. The recovered cells were plated on negative selection media containing the absence of IVK and 0.02% arabinose. Again, the library plasmid was isolated from the thousands of surviving clones and transformed into cells harboring the positive selection plasmid, pRep_PylT. The recovered cells were plated onto positive selection media containing 50 μg/mL Cm and 2 mM HBK and let incubate for ~40 hours. Individual fluorescent clones from the plate were picked to inoculate the wells of a 96-well culture block. The 96-well block incubated till growth saturation. Each clone was pintooled from the block and spotted onto screening plates containing 50 μg/mL and +/- 2 mM HBK (Figure 4.16) (Methods 4.14.2).

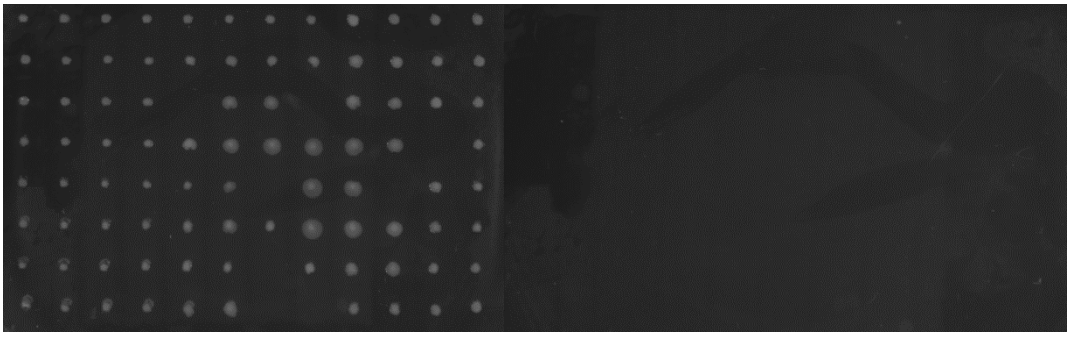


Figure 4.16. (Left) Hydroxybutyryl lysine screening plate containing 2 mM HBK. (Right) Hydroxybutyryl lysine screening plate without HBK.

The screening plate without HBK supplementation resulted in zero growth for every spotted clone. The screening plate supplemented with 2 mM HBK had growth at nearly every spotted clone. Fifteen clones were picked from the HBK supplemented plate to be evaluated. To identify whether each clone was *mazei* or *barkeri* library derived, a diagnostic digest of the DNA with an endonuclease (PvuII) unique to the *barkeri* gene was performed (not shown). All clones proved to be *barkeri* derived. The identity of the mutated residues in each variant was identified by sanger sequencing by Eurofins Genomics. Eleven (73%) of the clones converged on the same residue content named as HBKRS1 (Table 4.2). The other four (27%) of the clones converged on the same residue makeup named as HBKRS2. The variants were produced by the newly constructed *barkeri* derived library 270NNK/271NNK/274NNK/311NNK/313NNK as this was the only library able to produce mutants at the specific amino acid sites.

Table 4.2. Comparison of residue identity between wild type PylRS and HBKRS hits.

	270	271	274	311	313	366	382	386	Non-programmed	Codon Difference
WT PylRS	L	Y	L	N	C	V	W	G		
HBKRS1	L	M	A	N	A	V	W	G	Y349F	
HBKRS2	L	G	A	N	M	V	W	G	Y349F	A267 optimized

The sequencing results indicated that 3/5 of the mutated residues from the library were mutated (271, 274, 313). It was reported by Yokoyama *et al.* that mutation of Y271 is critical to expand the hydrophobic binding pocket.[49] The Y271G mutation of HBKRS2 allows for the expansion of the hydrophobic binding pocket as the larger tyrosine is replaced by the much smaller glycine. The hydroxyl group of HBK is bound to the γ carbon to the carbonyl whereas the amine group of PylK is bound to the β carbon to the carbonyl. This slight difference in size could warrant the need for an expanded pocket. The Y271M mutation of HBKRS1 slightly contradicts this notion. Methionine is only slightly smaller than tyrosine, but much larger than the sidechain-less glycine. The converged mutation of L274A by both variants substituted a larger aliphatic amino acid in leucine to the smaller aliphatic amino acid in alanine creating more space in the binding pocket as well. Without the co-crystallization of HBK coordinated in each PylRS variant active site, it is impossible to pinpoint how the residue mutations are oriented and interact with the substrate. Interestingly, each variant contained a non-programmed mutation of Y349F that has been reported to increase the wild-type PylRS activity with BocK and PylK.[87] Y349F has been reported to make specific hydrogen-bonding interactions with both the pyrrol nitrogen and the α-amino group of PylK.[50] The non-programmed mutation showed a purposeful, adaptive mutation for survival by each clone containing HBKRS1 or HBKRS2. Additionally, HBKRS2 contained a non-programmed A267 codon optimization where the native GCC codon (25%) was mutated to the more abundant GCG codon (34%). The genetic code of *E. coli* is biased towards GCG codon usage over the other three synonymous codons that also code for alanine. Codon biases reflect a natural selection for translational optimization. Optimal codons in fast-growing microorganisms like *E. coli*, reflect the composition of its genomic tRNA pool.[88]

Codon optimization is believed to help to achieve faster translation and higher accuracy thus increased expression. The A267 residue is contained within the hydrophobic catalytic cleft.[49] While this particular mutation in the binding pocket was not included amongst the libraries, the non-programmed codon optimization of A267 revealed that it's translation efficiency in PylRS is important.

A fluorescence assay was conducted to determine which of the two HBK synthetases had superior activity with HBK between the two (Figure 4.17) (Methods 4.14.3). The assay revealed that HBKRS1 was nearly twice as active with HBK compared to HBKRS2. Both synthetase mutants showed that there was minimal GFPuv expression from read through of the amber codons on T7 RNA Polymerase when the cells were grown in the absence of HBK. The fluorescence assay proved that each synthetase exclusively binds to HBK and that HBKRS1 is the superior synthetase. HBKRS1 has previously been reported to be a synthetase capable of catalyzing the aminoacylation of *o*-nitrobenzyl lysine (Appendix .[89] HBKRS2 has not been reported in the literature.

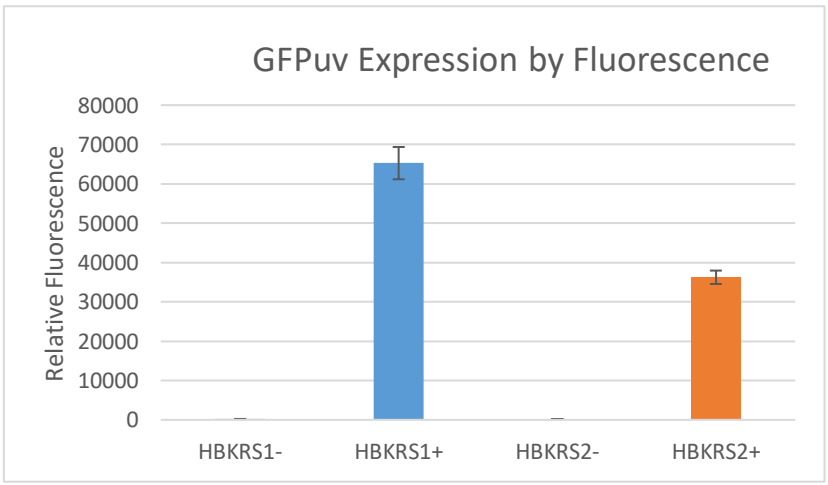


Figure 4.17. GFPuv fluorescence assay comparing the expression of HBKRS1 to HBKRS2 with HBK.

4.13 Incorporation of HBK into Protein

HBK was successfully incorporated into two model proteins, GFP and ubiquitin. GFP serves as an outstanding reporter protein because its expression is even visibly noticeable without instrumentation, however, its use as a model protein for UAA incorporation is limited as it does not mass spec well. The plasmid pTrcHisA-sfGFPV150TAG was used for GFP expression. GFP contains an in-frame amber stop codon at the permissive site V150. Because the expression vector of pTrcHisA-sfGFPV150TAG is incompatible with the pBK vector that HBKRS is coded on, they cannot be copropagated into the same cells. This is due to confliction in the copy numbers from each respective origin of replication. Rather, the HBKRS gene was cloned into a pSup vector affording pSup_HBKRS (Methods 4.14.6). Cells cotransformed with pSup_HBKRS and pTrcHisA-sfGFPV150TAG were grown in expression cultures supplemented with and without 2 mM HBK. The recombinant his tagged GFP protein was purified by immobilized metal affinity chromatography using nickel resin (Methods 4.14.7). The purified protein from each expression culture was separated by mass on a SDS-PAGE gel (Figure 4.17). The gel demonstrated that GFP expression was dependent on HBK in the media for read through of the in-frame amber codon. When HBK was not supplemented in the media, minimal to no GFP expression was observed.

Likewise, cells expressing HBKRS1 and harboring the pCNB_UbK11TAG plasmid were grown in expression cultures supplemented with and without 2 mM HBK. The recombinant ubiquitin protein was immobilized on a chitin resin affinity matrix and purified by intein-mediated cleavage of the fusion protein chitin binding domain (Methods 4.14.4). The purified proteins were separated by mass on a SDS-PAGE gel (Figure 4.18). The gel confirmed the

dependence of HBK media supplementation for ubiquitin expression that was also observed with the GFP expression.

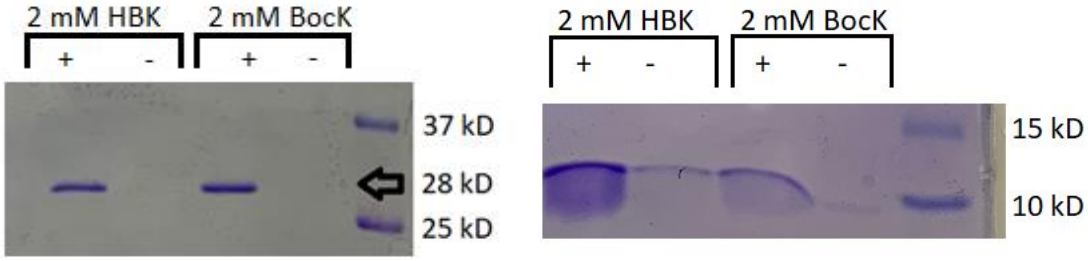


Figure 4.18. Inlet of SDS-PAGE gel of GFP (left) and ubiquitin (right) +/- HBK

The purified ubiquitin protein was separated by HPLC and the monoisotopic mass was detected by a mass spectrometer (Figure 4.19). There were two major peaks observed on the spectrum. The larger of the two peaks at 8,699 Da was larger than the expected mass of 8,652 Da. The mass difference of 47 Da was perplexing. The closest salt adduct that could account for the mass difference would be the coordination of the protein with two Na^+ ions with the abstraction of a proton (45 Da). Other possibilities could be if HBK was not what was being incorporated in response to the amber codon and rather a metabolized derivative of HBK was used as the substrate. In order to elucidate what accounts for the mass difference, tryptic digest of the protein followed by LC/MS/MS would need to be performed. The masses of the tryptic peptide fragments can be used to figure out what exactly was incorporated in response to the K11TAG mutation. The smaller of the two peaks had a monoisotopic mass of 8,566 Da. This mass was believed to arise from the deacylation of HBK to native lysine as previously seen with IVK incorporation into the UBK11TAG construct.

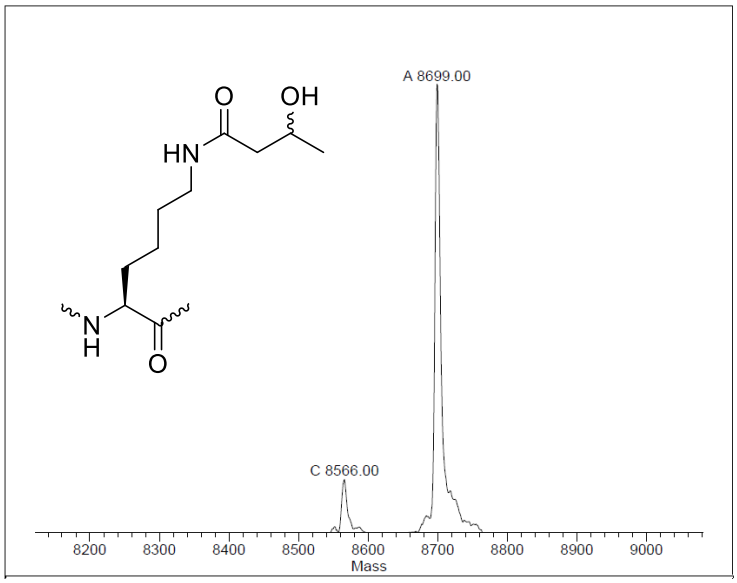


Figure 4.19 Deconvoluted monoisotopic mass spectrum of HBK incorporated into ubiquitin. Calculated M/Z^+= 8,652.3 Da. Observed M/Z^+=8,566.0 and 8,699.0 Da.

4.14 Methods and Materials

4.14.1 Selections

10x diversity of each library included in the selections was added to a plate containing the LB agar positive selection media (Tet15/Kan50/0.2% arab/Cm50/2 mM UAA). The plate was incubated at 37°C overnight. The surviving colonies from the positive selection were then scraped off the plate by addition of 2 mL LB. Depending on the amount of growth on the plate, the resuspended colonies were either directly midiprepped following a Promega wizard midiprep kit (Promega A7640) or the resuspended colonies were inoculated into a 100 mL LB culture to amplify the growth followed by a midiprep. The extracted DNA (2 μg) was separated by mass on a 1% gel agarose gel to separate the pRep_PylT selection plasmid (~10 kb) from the pBK_PylRS library plasmid (~3.3 kb) by mass. The library plasmid was gel extracted following a Qiagen gel extraction kit (Qiagen 28706). The library plasmid was then transformed into 100 μL DH10B

cells harboring the negative selection plasmid pBar_PylT. The cells were recovered in 900 μL LB and plated on LB agar negative selection media (Tet15/Kan50/0.02% arab). The plate was incubated at 37°C overnight. The surviving colonies from the negative selection were then scraped off the plate by addition of 2 mL LB. The resuspended colonies were midiprepped following a Promega midiprep kit. The extracted DNA (2 μg) was run on a 1% agarose gel to separate the negative selection pBar_PylT plasmid (~7.0 kb) from the library plasmid (~3.3 kb) by mass. The library plasmid was gel extracted following a Qiagen gel extraction kit. The library plasmid was then transformed into 100 μL DH10B cells harboring the positive selection pRep_PylT plasmid. The cells were recovered in 900 μL LB and plated on LB agar positive selection media (Tet15/Kan50/0.2% arab/Cm50/2 mM UAA). The plate was incubated at 37°C overnight. After incubation, clones are picked for screening (Methods 4.14.2). If screening attempts were unsuccessful in finding an active synthetase with the respective UAA, another round of negative selection and positive selection were performed as described in this protocol.

4.14.2 Screenings

Following the overnight incubation of the last round of positive selection, there was growth of hundreds to thousands of fluorescent and non-fluorescent colonies. The fluorescent colonies were carefully picked and inoculated into the wells of a 96-well block. The 96-well block contained LB media (Tet15/Kan50). The 96-well block was incubated at 37°C overnight for cell growth saturation. The cells in each well were spotted onto two screening plates using a pintool (replicator). One screening plate contained LB agar (Tet15/Kan50/0.2% arab/Cm50/2 mM UAA) and the other screening plate contained LB agar (Tet15/Kan50/0.2% arab/Cm50). The plates were incubated at 37°C overnight. After incubation, the plates were analyzed for hits displaying the

phenotype of UAA dependent in-frame amber codon suppression of the CAT protein (growth on plate with UAA/no growth on plate without UAA).

4.14.3 GFPuv Fluorescence Assays

GFPuv fluorescence assays were performed by inoculating a 5 mL LB (Tet15/Kan50) culture with the respective synthetase hit from the 96-well block. The cultures were grown to an $OD600_{nm}$~0.6 and were induced with 0.2% arabinose. Each 5 mL culture was split into 2x 2.5 mL culture. One culture was supplemented with 2mM of the respective UAA and the other was not. The expression cultures grew to saturation. Each culture was spun down at 8,000 RPM for 5 min. The pelleted cells were washed with 1 mL of phosphate buffered saline buffer (PBS, 137 mM NaCL, 10 mM phosphate, 2.7 mM KCl, pH=7.4) and spun down again at 8,000 RPM for 5 min. The washed pellets were resuspended in 250 µL PBS buffer and added to the wells of a clear bottom 96 well plate. The $OD600_{nm}$ of each sample was detected on the fluorimeter for normalization. The 515 nm emission of each sample was detected by the fluorimeter upon excitation at 395 nm.

4.14.4 Protein expression and batch mode purification under native conditions of PCHis, IBK, IVK, HBK, and BocK modified Ubiquitin using intein-mediated cleavage of UB-CBD fusion protein:

Plasmids pBK_PCHisRS, pBK_IVKRs, pBK_HBKRS, and pBK_PylRS were each cotransformed with pCNB_UbK11TAG_PylT into DH10B cells and plated on LB agar (Tet15/Kan50). A single colony was picked from each respective plate to grow a 5 mL overnight starter culture in LB media (Tet 15/Kan 50). A 50 mL flask of LB media (Tet 15/ Kan 50) was

inoculated with 2.0 mL of starter culture. In the case of IVKRS, a 75 mL flask was inoculated with 3.0 mL of starter culture. The cells were grown to an $OD600_{nm}$~0.6 and were induced with 20% arab (0.2% final conc.). The 50 mL culture was split into two 25 mL expression cultures (the 75 mL IVKRS culture split into three 25 mL). One culture was supplemented with 2 mM of the respective UAA; the other culture was not. The 25 mL cultures containing cells expressing IVKRS were induced with 2 mM IVK and 2 mM IBK each. The 25 mL culture containing PCHis was induced with 4 mM PCHis. The expression cultures grew overnight at 37°C incubation. Cells were harvested at 5,000 rpm for 10 min, washed with 1 mL of distilled water, and spun at 5,000 rpm for 10 min. The cell pellet was washed with 1 mL of column buffer (20 mM Hepes, pH 8.5, 0.5 M NaCl, 1 mM EDTA, 0.02% NaN_3) and spun at 5,000 rpm for 10 min. The cell pellet was washed again with 1 mL of lysis buffer (column buffer + 1 mM PMSF). The cells were resuspended in 2 mL lysis buffer and were freeze/thawed three times by liquid nitrogen and room temperature water bath. The cells were then placed on ice and sonicated with pulse for 1.5 min on and 1.0 min off. The cells were centrifuged at 12,000 rpm for 20 min to isolate the soluble proteins in the lysate. A batch mode purification of the intein-mediated cleavage of the CBD/Ub fusion protein was performed. 500 uL of chitin beads were equilibrated by 2.5 mL of column buffer. The cell lysate was added to the chitin beads, spun down at 5,000 rpm for 1 min. The flow through was added to the chitin beads again and spun down at 5,000 rpm for 1 min. The chitin beads were washed five times with 10 mL of column buffer. On column cleavage was induced by flushing the beads with 1 mL of elution buffer (20 mM Hepes, pH 8.5, 0.5 M NaCl, 1 mM EDTA, 0.02% NaN_3, and 50 mM DTT). The beads were incubated at 37°C for 48 hrs. Protein was eluted three times with 500 µL fresh elution buffer.

4.14.5 Additional Library Construction

The additional *Methanosarcina barkeri* consisting of codon randomization of residues L270, Y271, L274, N311, and C313 were built using separate methods of Gibson assembly (NEB #E2611S) and inverse PCR by Golden Gate Assembly. Primers CL2202 and CL2203 were used in the inverse PCR reaction to generate the L270, Y271, L274 sub-library. The PCR reactions were prepared in 160µL containing 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (NEB #N0447), 0.5 µM of each primer (Eurofins), 8 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl aliquots using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec at 68°C, 2 min at 68°C), 5 min at 72°C. The PCR product was purified following the Zymo DNA clean and concentrator kit (Zymo D4013). The purified PCR products were digested with 2 U DpnI (NEB #R0176) to remove methylated template and 10 U BsaI (NEB #R03535) overnight at 37°C to ensure complete digestion. The digested product was purified following the Zymo DNA clean and concentrator kit (Zymo D4013). The purified digestion products were re-circularized by ligation using 400 U T4 ligase (NEB #M020) in 20 µL reactions and used to transform 100 µL electrocompetent DH10B cells. The transformed cells were recovered in 900 µL super optimum broth (SOC). The recovered cells were plated at 1:1, 1:100, 1:1,000 and 1:10,000 dilutions (Kan50) to determine total diversity. The sub-library DNA was isolated from 100 mL cultures (Kan50) by following the Promega midiprep wizard kit (Promega A7640). The sub-library DNA was used as template in subsequent PCR reactions.

The second round of PCR reactions to generate the library with codon randomization at residues L270, Y271, L274, N311, C313 was performed using sub-library L270, Y271, L274 as

template for subsequent Gibson assembly. Primers CL1892 and CL1893 were used to create the insert and primers CL1894 and CL1895 were used to create the vector. Primers CL1892 and CL1893 were used in the first round of PCR reactions to create the insert and primers CL1894 and CL1895 were used to create the vector. The PCR reactions were prepared in 160µL containing 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (NEB #N0447), 0.5 µM of each primer (Eurofins), 8 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl aliquots using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec at 68°C, 2 min at 68°C), 5 min at 72°C. The PCR products were digested with 2 U DpnI for 2 hours at 37°C. The digested PCR products were purified following the Zymo DNA clean and concentrator kit (Zymo D4013). 50 ng of insert and 100 ng of vector were added to a 20 µL Gibson assembly (NEB #E2611) reaction and incubated for 1 hr at 50°C. 2 µL of the Gibson reaction was used to transform into 100 µL electrocompetent DH10B cells. The transformed cells were recovered in 900 µL super optimum broth (SOC). The recovered cells were plated at 1:1, 1:1,000 and 1:10,000 dilutions (Kan50) to determine total diversity. This procedure was repeated several times until an acceptable diversity was achieved (2.5×10^6).

Additionally, inverse PCR by golden gate assembly was used as a method to generate the library with codon randomization at residues L270, Y271, L274, N311, and C313. The primers CL1646 and CL1647 were used in the inverse PCR reaction with sub-library L270, Y271, L274 as template. The PCR reactions were prepared in 160µL containing 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (NEB #N0447), 0.5 µM of each primer (Eurofins), 8 ng template and 2 U

Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl aliquots using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec at 68°C, 2 min at 68°C), 5 min at 72°C. The digestion and ligation of the second-round inverse PCR product was different than the inverse PCR by golden gate assembly reported in Methods 2.6.1-2. The PCR product in the second round was first digested with DpnI (NEB #R0176) and purified by following the Zymo DNA clean and concentrator kit (Zymo D4013). The purified digest (100 ng) was added to a 20 µL ligation/digestion mix with 400 U ligase and 30 U BsaI HF-V2 (NEB #R3733S). BsaI HF-V2 has been engineered for improved performance in golden gate assembly. BsaI HF-V2 maintains activity in ligase buffer for an extended time at elevated temperatures. Because the endonuclease is active in the ligase buffer with ligase, the reaction is a one-tube dynamic cutting/re-ligating process. The reaction was run on the following temperature program: 5 min at 37°C, 5 min at 16°C) for 30 cycles followed by 5 min at 60°C. 2 µL of the ligation/digestion reaction was transformed into Electromax DH10B cells (Invitrogen 18290015) that report <1.0 x 10^{10} cfu/µg transformation efficiency. The transformed cells were recovered in 900 µL SOC. The recovered cells were plated at 1:1, 1:1,000 and 1:10,000 dilutions (Kan50) to determine total diversity. This procedure was repeated six times until an acceptable diversity was achieved (7.5 x 10^{6}).

4.14.6 Cloning PylRS Variants into pSup Vector

PylRS variants were cloned into the pSup vector by Gibson assembly. To amplify the PylRS variant gene from pBK_PylRS, the following was added to a 20 µL PCR reaction: 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM

MgCl2 (NEB #M0510), 200 µM of each dNTP (CL1884 and CL1885) (NEB #N0447), 0.5 µM of each primer (CL1884 and CL1885) (Eurofins), 1 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were donein 20 µl using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec at 68°C, 2 min at 68°C), 5 min at 72°C.

To amplify the pSup vector from the plasmid pSup_Mm_BocLys, the following was added to a 20 uL PCR reaction: 1x Phusion High-Fidelity DNA polymerase buffer (NEB #B0518S) supplemented with 2.5 mM MgCl2 (NEB #M0510), 200 µM of each dNTP (CL1884 and CL1885) (NEB #N0447), 0.5 µM of each primer (CL1886 and CL1887) (Eurofins), 1 ng template and 2 U Phusion High-Fidelity Polymerase (NEB #M0530). PCR reactions were done in 20 µl using the following temperature program: 30 sec at 98°C, 6x(15 sec at 98°C, 30 sec at 65°C [-2°C/cycle], 2 min at 72°C), 30x(15 sec at 98°C, 30 sec at 68°C, 2 min at 68°C), 5 min at 72°C.

The PCR products were digested with 2 U DpnI for 2 hours at 37°C. The digested PCR products were purified following the Zymo DNA clean and concentrator kit (Zymo D4013). 50 ng of insert and 100 ng of vector were added to a 20 µL Gibson assembly (NEB #E2611) reaction and incubated for 1 hr at 50°C. 2 µL of the Gibson reaction was used to transform into 100 µL electrocompetent DH10B cells. The recovered cells were plated on LB agar (Cm35) and incubated overnight at 37°C. Ten colonies were picked and inoculated into individual 5 mL LB cultures (Cm35). The cultures were grown to saturation and miniprepped following a NEB monarch plasmid mini prep kit (T1010L). The DNA extracted from each culture was digested with 10 U NcoI for 2 hour at 37°C. The NcoI recognition site is found in *barkeri* PylRS and not in *mazei* PylRS. Each digest was separated by mass on a 1% agarose gel to determine successful cloning of the mutant PylRS gene into the pSup vector.

4.14.7 Nickel Purification of His-Tagged Proteins under Native Conditions

Plasmid pSup_HBKRS was cotransformed with expression plasmid pTrcHisAsfGFPV150TAG into BL21(DE3) cells. BL21(DE3) is a T7 expression strain that contains the λDE3 lysogen that carries the gene for T7 RNA polymerase under control of the lacUV5 promoter. IPTG is required to maximally induce expression of the T7 RNA polymerase in order to express recombinant genes cloned downstream of a T7 promoter. A colony was picked from the selection plate (Amp100/Cm35) and inoculated into a 5 mL overnight starter culture. To a 200 mL culture, 4 mL of the starter culture was added. The culture was grown to $OD600_{nm}$~0.6 and induced with 1 mM IPTG. The culture was split into two 100 mL cultures where one was supplemented with 2 mM HBK and the other was not. Both cultures were incubated overnight at 37°C. The expressions were harvested at 5, 000 rpm for 10min at 4ºC. The cells were resuspended in 5 mL of binding buffer (50 mM NaH_2PO_4, pH 8.00, 300 mM NaCl, 10mM imidazole, and 1 mM PMSF). The cells were lysed by lysozyme (0.1 mg/mL binding buffer), supplemented with 1 mM PMSF, and incubated in ice for 30 min. The cells were further lysed by sonication by six 10 second bursts with a 10 second cooling period between each burst. The cells were centrifuged at 12, 100 rpm for 20 min at 4ºC to obtain the soluble protein. A batch mode nickel purification was performed. The cellular lysate was added to 500 μL of washed (50 mM NaH_2PO_4, pH 8.00, 300 mM NaCl, and 20mM imidazole) Promega HisLink resin (Promega, Madison, WI) and incubated 1 hr at 4ºC rotating. The resin was washed five times with 14 mL of bind buffer and then eluted four times with 500 μL of elution buffer (50 mM NaH_2PO_4, pH 8.00, and 300 mM NaCl, and 500mM imidazole). The eluted protein was desalted and concentrated using a Spin-X UF concentrator (5, 000 MWCO).

Chapter 5. Photocaged Histidine

5.1 Photocaged Amino Acids

To better understand the function, mechanism of action, and regulation of a specific protein, it is desirable to be able to non-invasively modulate the protein activity and observe the effects produced by its activation and deactivation. Photocaged proteins are modified proteins in which biological activity can be controlled by photolytic conversion from an inactive to an active form by light. Caging groups are light-removable protecting groups typically comprised of aromatic rings that are attached to amino acid residues that give researchers the ability to control the function of a specific protein using light irradiation. This is particularly useful as irradiation is a relatively noninvasive method that allows both spatial and temporal control of protein activity in complex cellular systems.[90] Caging groups are covalently bound to specific amino acid residues in such a way that their natural activities are inhibited. When irradiated, these molecules undergo photolytic cleavage to expose the 'decaged' amino acid in its natural form thereby either inactivating or activating protein function (Figure 5.1).

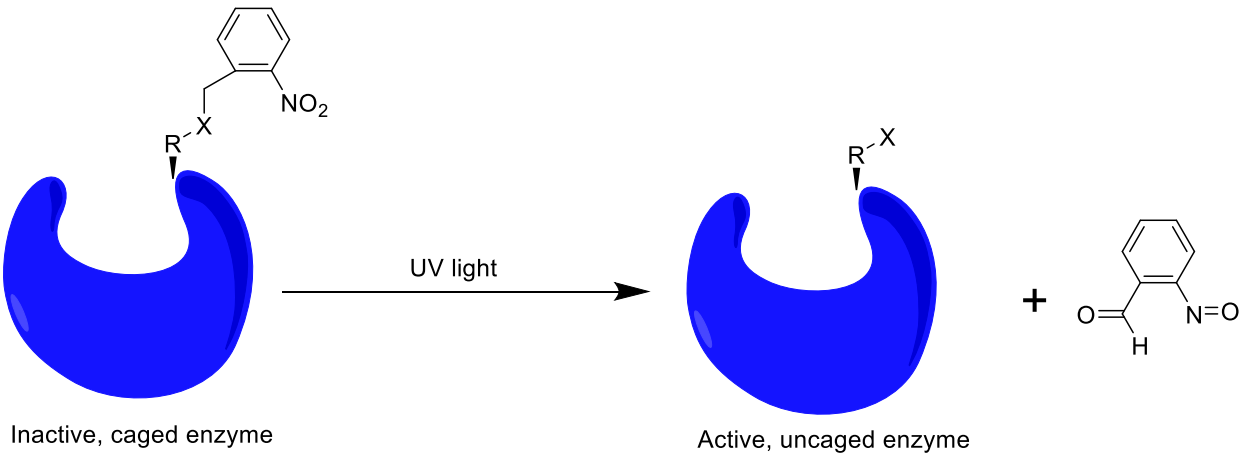
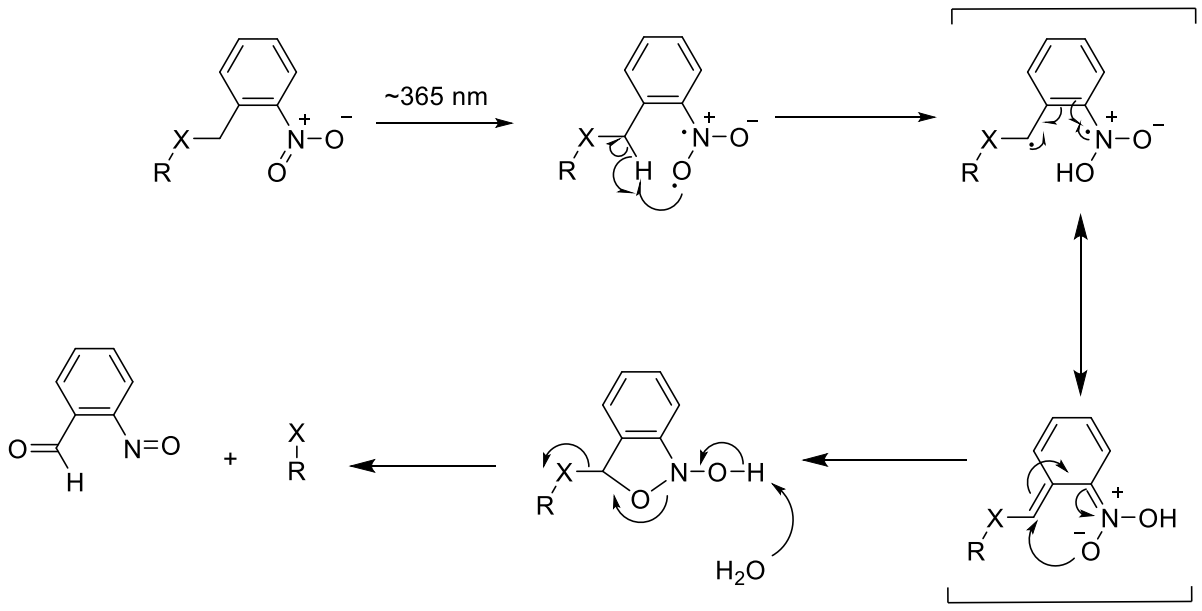


Figure 5.1. Photoactivation of a protein upon irradiation with light.

The most common photocaging groups are *o*-nitrobenzyl groups and their derivatives, which can be installed on hydroxy, carboxy, thio, or amino groups of proteins and are readily cleaved upon irradiation with 365-nm light.[91] Other caging groups with enhanced photochemical properties have been developed as well.[92–94] To be useful in biological experiments this group must satisfy several criteria. First, it should render the biomolecule inert to the biological system used. It should release the biomolecule in high yield at sufficient speed by photolysis at wavelengths of light that are non-detrimental to the biological preparation. Lastly, any photoproducts other than the desired biomolecule should not interact or interfere with the biological system.[93,95] The o-NB-caged compound undergoes decaging according to a Norrish type II mechanism, affording a free substrate and a nitroso-benzaldehyde (Scheme 5.1).[90,96]



Scheme 5.1. Light induced decomposition of a 2-nitrobenzyl-group protected amino acid by a Norrish type II mechanism.

Photocaged amino acids have been used as probes of protein function *in vitro* and in cells. In 2006, Schultz *et al*. successfully incorporated the first photocaged amino acid of photocaged tyrosine (*o*-nitrobenzyl-*O*-tyrosine) into protein in *E. coli* using the *Methanococcus jannaschii* tyrosyl system.[97] Since then, photocaged cysteine,[10] serine,[24] and lysine[19] have been site-specifically incorporated into proteins to allow the photodynamic regulation of their activity (Figure 5.2). In one example, the localization of the yeast transcription factor Pho4 was followed in living cells by selectively blocking its phosphorylation with a photocaged serine. Irradiation with low-energy blue light uncaged the serine and allowed nuclear export of the phosphorylated Pho4 to be followed in real time.[24]

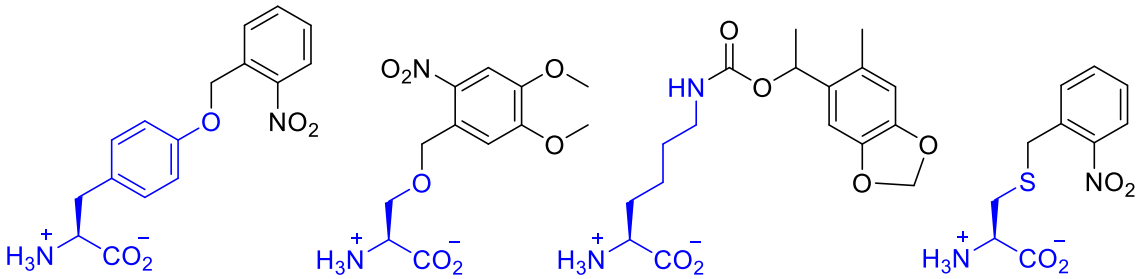


Figure 5.2. Photocaged tyrosine, serine, lysine, and cysteine added to the genetic code. Native amino acid shown in blue. Caging group shown in black.

5.2 Photocaged Histidine

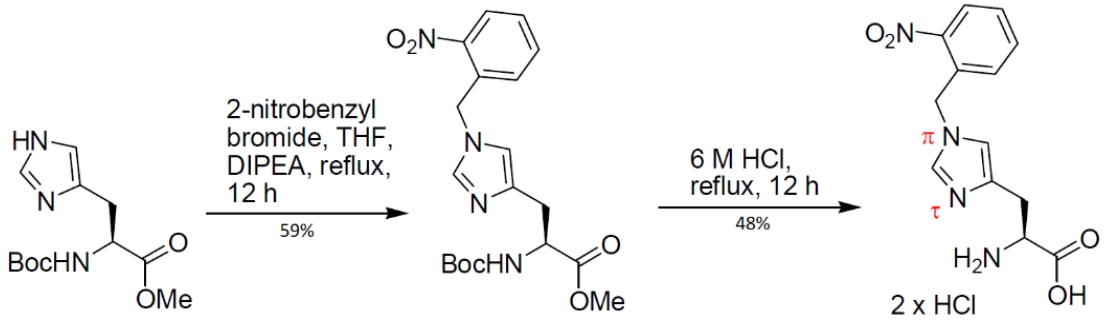
The photocaged amino acid of interest to incorporate into proteins to elucidate its method of action was photocaged histidine (PCHis). Among the twenty natural amino acids, histidine may be the most versatile in protein architectures and bioactivities.[98–101] The versatility of histidine in molecular interactions arises from its unique molecular structure. The imidazole side chain is an ionizable aromatic ring, a coordinating ligand for metallic cations, and a hydrogen bond donor and acceptor.[102] The unique structure of histidine makes it involved in multiple roles

in molecular interactions. The ability to switch between protonated and unprotonated states in physiological conditions allows histidine to participate in acid-base catalysis. In its neutral form, histidine participates in cation-π interactions in which the histidine acts as the aromatic π-motif with metallic cations or organic cations such as protonated lysine or arginine.[103–105] On the other hand, the protonated form is an organic cation, which can join the cation-π interactions as an organic cation with other aromatic amino acids.[106–109] The imidazole ring is heavily involved in hydrogen bonding interactions as the polar hydrogen atom of the imidazole is a hydrogen-bond donor, and the basic nitrogen atom is a hydrogen-bond acceptor.[110–112] Lastly histidine is involved in coordinate interactions between histidine and metallic cations.[113]

Due to its unique functional versatility and role within the active site of enzymes, a photocaged histidine was of particular interest for incorporation into the proteome (Figure 5.3). In order to elucidate the role of histidine in specific protein functions, one can replace a presumed critical histidine codon with an amber codon to code for photocaged histidine. Replacement of histidine with photocaged histidine in critical site(s) within the protein renders the protein inactive.[2] This alteration of the active site disrupts docking of the natural substrate by direct inhibition of the interaction of the amino acid residue with the substrate or by indirect inhibition such as inducing a change in protein folding. Activity of the enzyme is restored upon light induced activation which allows biological activity of the protein to be regulated under non-invasive external control.[2]

5.3 Photocaged Histidine Synthesis

Synthesis was accomplished following the protocol reported in the thesis of Hrvoje Lusic of our collaborator Dr. Alex Deiter's research group from University of Pittsburgh (formerly North Carolina State).[114] Synthesis of *o*-NB caged histidine (PCHis) was accomplished by heating Boc-(L)His-OMe and *o*-nitrobenzylbromide in tetrahydrofuran (THF), in the presence of N,N-Diisopropylethylamine (DIPEA) as a proton scavenger, affording a protected *o*-NB caged histidine in 59% yield (Scheme 5.2) (Methods 5.8.1). The π and τ caged histidines were easily separated by silica gel chromatography after the caging step. The π substituted histidine was found to be the less polar of the two compounds, and thus eluted off the column first. The π and τ caged derivatives of histidine were observed as products in a 95:5 ratio. Removal of the Boc group and hydrolysis of the methyl ester was accomplished in one step by heating protected *o*-NB caged histidine under reflux in 6 M HCl, to give the HCl salt *o*-NB caged histidine in 48% yield (Scheme 5.2).



Scheme 5.2. Synthesis of *o*-NB caged histidine.

5.4 Photocaged Histidine Selections

Selection for an active PylRS variant with photocaged histidine was performed using the positive/negative selection scheme along with the following libraries: the *barkeri* derived mixed lysine analog and tyrosine analog libraries 1. L274A/Y348F/A267NNK/Y271NNK_/C313NNK

2. N311A/C313A/Y271NNK/ L274NNK/Y349NNK, the *mazei* derived 346NNK/348NNK/401NNK/417NNK/421NNK, the *barkeri* derived library 311NNK/313NNK/366NNK/382NNK/386NNK and the *barkeri* derived library 270NNK/271NNK/274NNK/311NNK/313NNK (Methods 4.14.1). For full confidence the library diversity was covered, ten times the diversity of each library was subjected to the positive selection media containing 50 μg/mL Cm and 2 mM PCHis. The surviving clones were scraped from the plate, the DNA was isolated, and separated by mass on a 1% agarose gel (Figure 5.3).

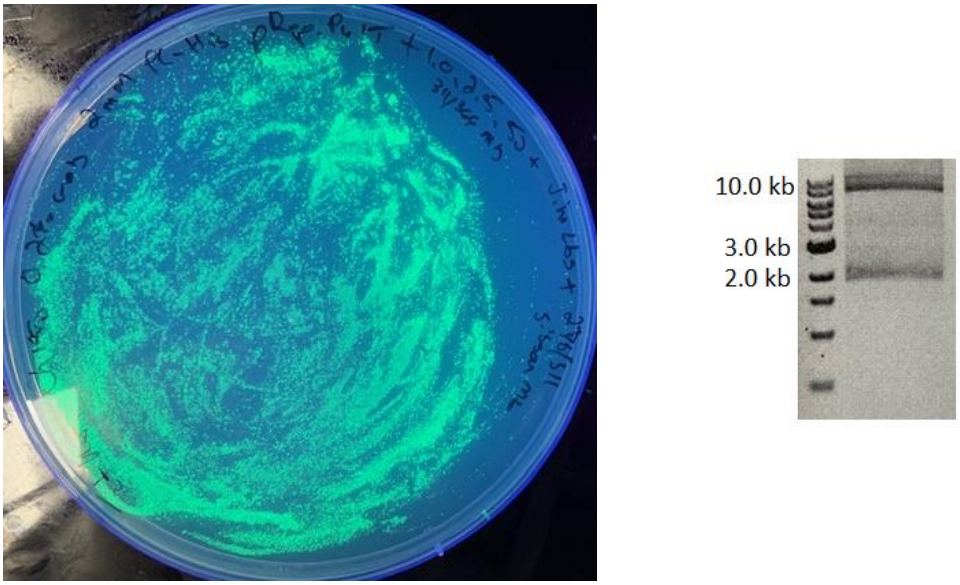


Figure 5.3. (Left) First round positive selection plate growth for PCHis. (Right) DNA from the surviving clones from the first-round positive selection separated by mass on a 1% agarose gel.

The library plasmid band (~2 kb) was isolated from selection plasmid (~10 kb) by gel extraction, purified, and transformed into cells containing the negative selection plasmid, pBar_PylT. The recovered cells were plated on negative selection media containing the absence of PCHis and 0.02% arabinose and incubated ~16 hours. Again, the surviving clones were scraped from the plate, DNA was isolated, and separated by mass on 1% agarose gel. The

library plasmid band (~2 kb) was isolated from the selection plasmid (~7 kb) by gel extaction, purified, and transformed into cells containing the positive selection plasmid, pRep_PylT. The recovered cells were plated onto positive selection media containing 50 μg/mL Cm and 2 mM PCHis and let incubate for ~40 hours. Individual fluorescent clones from the plate (Figure 5.4) were picked to inoculate the wells of a 96-well culture block.

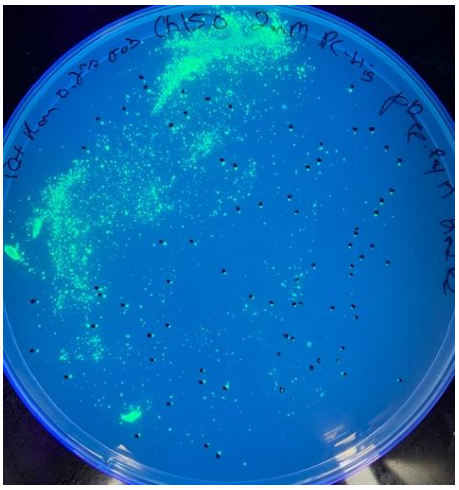


Figure 5.4. Second positive selection plate growth for PCHis.

The 96-well block incubated till growth saturation. Interestingly, only ~50% of the wells had growth. Each clone was pintooled from the block and spotted onto screening plates containing 50 μg/mL and +/- 2 mM PCHis (Figure 5.5). The screening plates were incubated ~16 hours.

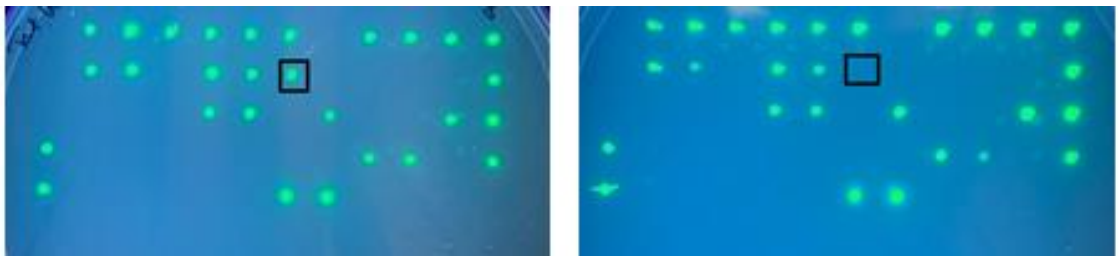


Figure 5.5. (Left) Photocaged histidine screening plate containing 2 mM PCHis. (Right) Photocaged histidine screening plate without PCHis. Square indicates a hit.

5.5 Photocaged Histidine Hit Evaluation

One clone at position B6 exhibited PCHis dependent growth and fluorescence hence known as PCHisRS. To verify that PCHisRS was indeed due to exclusive PCHIS induced amber codon suppression, the DNA was isolated from the clone, library plasmid isolated via gel extraction, and retransformed into DH10B cells containing the positive selection plasmid pRep_PylT. The phenotype was assayed by growing the cells on media containing +/- 2 mM PCHis and 50 μg/mL Cm (Figure 5.6).

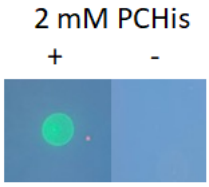


Figure 5.6. PCHis variant hit spotted on positive selection media.

The stark contract in growth between the inclusion and absence of PCHis indicated that PCHisRS exclusively incorporated PCHis into the reporter protein CAT. A fluorescence assay using the GFPuv expression system of pRep_PylT was performed to compare the amount of GFPuv expression from PCHisRS amber codon suppression by PCHis to the expression level of GFPuv from PylRS amber codon suppression by BocK (Figure 5.7). The cells of each were grown in the presence and absence of 2 mM of the respective UAA. The fluorescence assay showed that PCHisRS' activity with PCHis was about 25% as much as wild-type PylRS' activity with BocK. It also showed that GFPuv expression is dependent on PCHis in the media where there was minimal GFPuv expression when PCHis was omitted from the media.

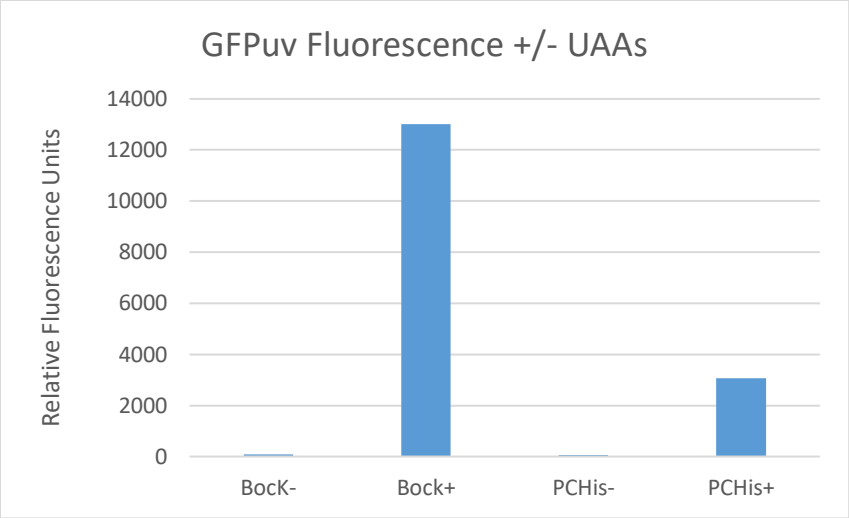


Figure 5.7. Relative GFPuv fluorescence from amber codon suppression by PCHisRS and PylRS with their respective UAAs.

To identify whether PCHisRS was *mazei* or *barkeri* derived, a diagnostic digest of the DNA with an endonuclease (PvuII) unique to the *barkeri* gene was performed (not shown). PCHisRS proved to be *barkeri* derived. The identity of the mutated residues in the PylRS variant, PCHisRS, were identified by sanger sequencing by Eurofins Genomics (Table 5.1).

Table 5.1. Comparison of residue identity between wild type PylRS and PCHisRS

	305 (270)	305 (271)	309 (274)	346 (311)	348 (313)	401 (366)	417 (382)	421 (386)
PylRS WT	L	Y	L	N	C	V	W	G
PCHisRS	L	Y	L	G	V	K	W	G

The sequencing results showed that the variant was derived from the *barkeri* library 311NNK/313NNK/366NNK/382NNK/386NNK as this is the sole library capable of producing this mutant identity. Within this library, 3/5 of the residues diverged from the wild type. The larger and more functional asparagine at site 311 was substituted with the much smaller glycine.

While the mutation of N311 disrupts the anchoring hydrogen-bonding interaction with the α-amino group, the smaller amino acids affords a wider binding pocket. The mutation from C313 to valine provided a subtle difference. When comparing the spatial structural arrangement of PCHis to PylK, the imidazole side chain of PCHis is covalently bound to the β carbon making it considerably wider than the aliphatic ε-carbon bound amine group of pyrrolysine. Because of this difference in size, PCHis would seem to benefit from a wider pocket that 'opens up' shallower in the pocket. Where PylK would benefit from a hydrophobic pocket (L305(270), Y306(271), L309(274), V401(366), and W417(382)) deeper into the tunnel of the binding pocket, PCHis would seem to need the bulky imidazole and nitrobenzyl group to be accommodated for in a shallower but wider hydrophobic pocket. This conjecture accounts for the shallow N311G and C313V mutations where the larger, polar amino acids are substituted for the smaller hydrophobic residues. Likewise, the smaller sized G386 located near the α carbon of the UAA remained unchanged. While the unchanged G386 and N311G C313V mutations seem to 'open up' a shallow hydrophobic pocket, the V366K mutation seems to minimize the size of the native hydrophobic pocket. The V366K substitutes a small hydrophobic amino acid with a much larger lysine residue. In addition, the ε amine group of K366 may hydrogen bond with the amine of the imidazole ring of PCHis.

5.6 Incorporation of PCHis into Protein

PCHis was incorporated into the UbK11TAG construct in response to the amber codon. Expression cultures supplemented with and without 4 mM PCHis expressed the ubiquitin construct from the plasmid pCNB_UbK11TAG_PylT. The recombinant proteins were purified by intein-mediated cleavage of the CBD fusion protein. The purified ubiquitin proteins were separated by mass on a SDS-PAGE gel (Figure 5.8). The protein gel displayed that ubiquitin

expression occurred when PCHis was supplemented in the media and minimal ubiquitin expression occurred when PCHis was omitted. This showed that PCHisRS exclusively charged PCHis onto PylT for amber codon suppression of the UBK11TAG construct.

Figure 5.8. Inset of SDS-PAGE gel of purified ubiquitin (8.7 kDa) expressed in media with and without 4 mM PCHis.

The purified ubiquitin protein was separated by HPLC and the monoisotopic mass was detected by a mass spectrometer (Figure 5.9). The major monoisotopic mass observed was 8,710 Da which corresponded to ubiquitin modified with PCHis in response to the K11TAG mutation. Interestingly, the mass of 8,680 Da was observed and corresponds to reduction of the nitro group to an amine on the *o*-nitro benzyl group.

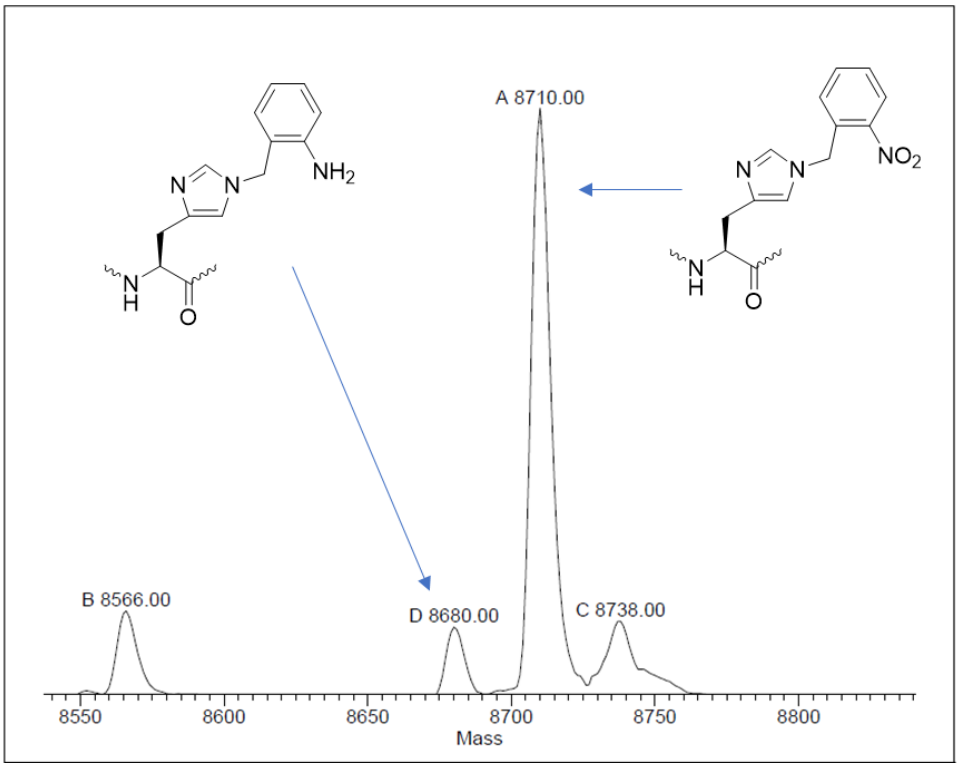


Figure 5.9. Deconvoluted monoisotopic mass spectrum of purified ubiquitin modified with PCHis. Calculated M/Z^{+}= 8,710.3 Da. Observed M/Z^{+}=8,566.0, 8680.0, 8710.0 and 8,738.0 Da.

5.7 Preliminary Decaging Studies

Time course decaging studies were performed on protected PCHis free amino acid (Figure 5.10). Decaging tests were done by irradiating protected PCHis in MeOH (15 mM) with a hand-held UV lamp (302 nm; 4 W) for 15, 30, 60, and 120 mins (Figure 5.11) and irradiating protected PCHis in MeOH (15 mM) with a hand-held UV lamp (365 nm; 24 W) for 15, 60, and 120 mins (Figure 5.12) (Methods 5.8.2). The separation of the reactions was done by HPLC/UV. A UV visible HPLC detector uses light to analyze samples. By measuring the sample's absorption of light at different wavelengths, the analyte can be identified. Separation by HPLC was monitored at the maximum absorption of histidine- 210 nm. The HPLC-UV chromatogram

showed that protected histidine (100% decaged) had a retention time around 4.5 mins. The results showed that the *o*-nitrobenzyl histidine had minimal decaging until 60 mins of irradiation at 302 nm. At 120 mins of exposure with 302 nm light, the compound showed partial decaging (free protected histidine) demonstrated by the appearance of a peak around 4.5 mins (Figure 5.11). Decaging by irradiation at the less damaging 365 nm light was not observed until 2 hours as well. Despite the difference in power between the 365 nm and 302 nm lights (23 W vs. 6 W, respectively), irradiation with 302 nm resulted in a greater amount of free protected histidine observed. It was reported by our collaborators in the Deiter's research group at the University of Pittsburgh that protected HBK showed 20% decaging efficiency.

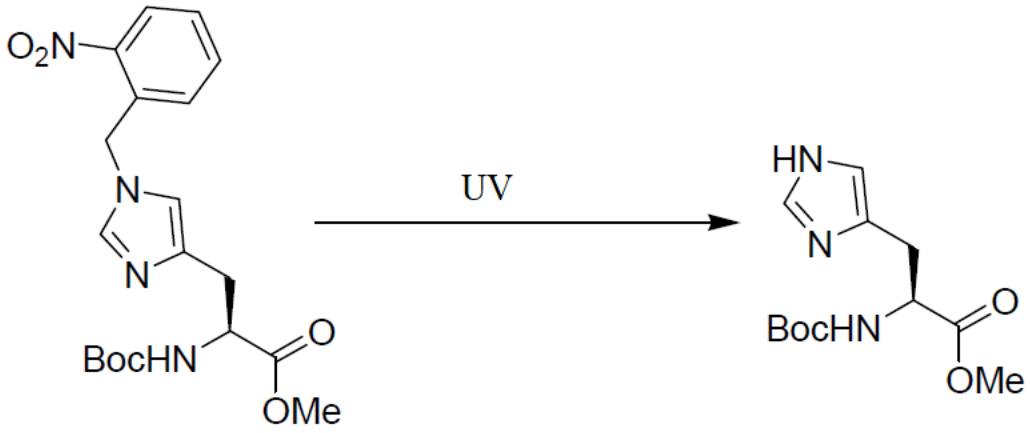


Figure 5.10 Light irradiated decaging of protected PCHis to protected histidine.

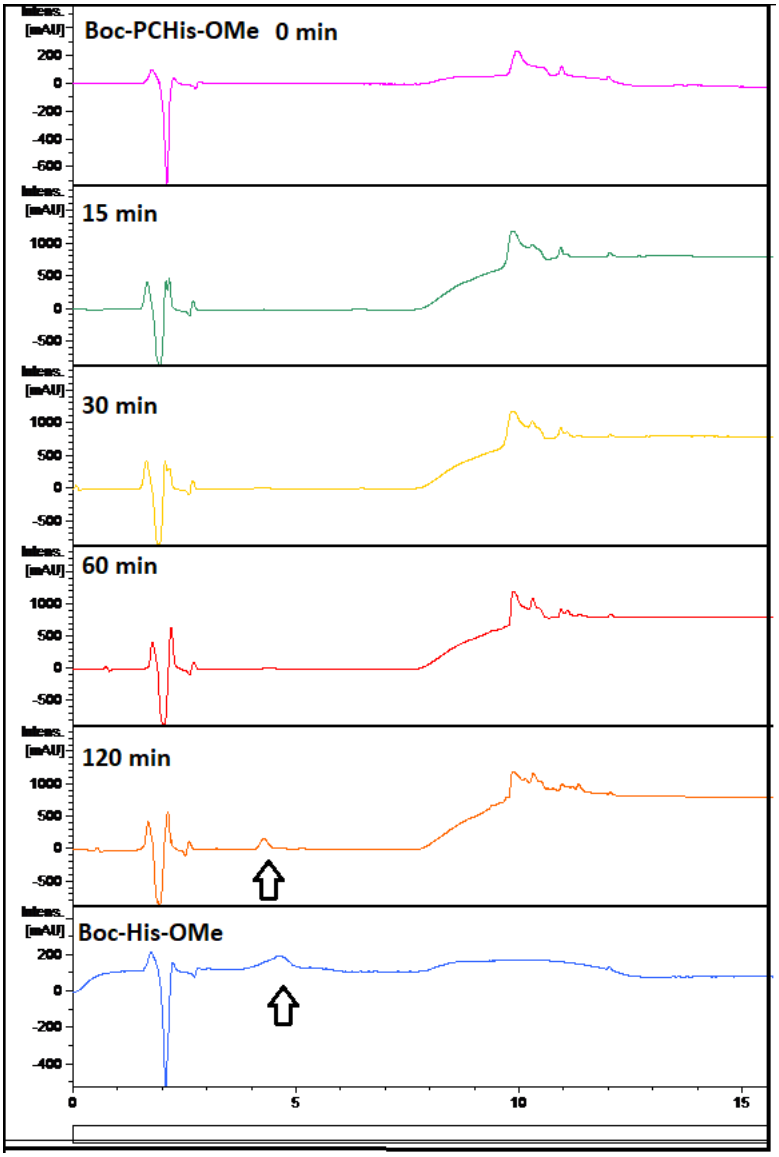


Figure 5.11. HPLC-UV chromatogram monitoring absorption at 210 nm for time course decaging of protected PCHis at 302 nm light irradiation. Arrow shows retention time of protected histidine.

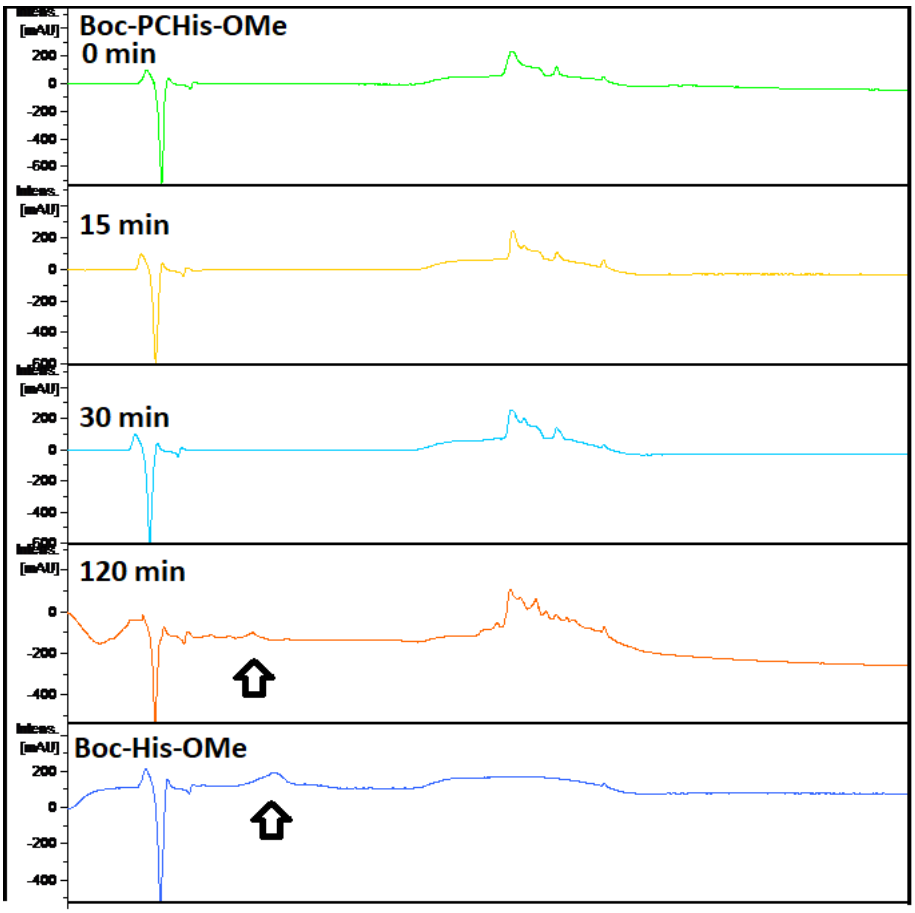
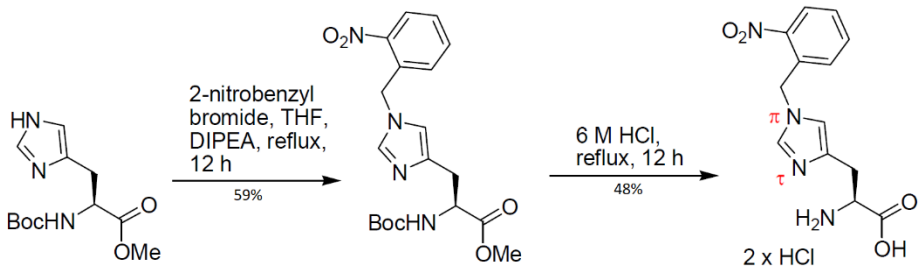


Figure 5.12. HPLC-UV chromatogram monitoring absorption at 210 nm for time course decaging of protected PCHis at 365 nm light irradiation. Arrow shows retention time of protected histidine.

5.8 Methods and Materials

5.8.1 Synthesis of Photocaged Histidine



(*S*)-Methyl 2-(*tert*-butoxycarbonylamino)-3-[1-(2-nitrobenzyl)-1*H*-imidazol-4-yl] propanoate

To a solution of Boc-(L)His-OMe (2.00 g, 7.4 mmol) containing DIPEA (2.11 g, 2.84 mL, 16.32 mmol) was added 2-nitrobenzyl bromide (1.44 g, 6.64 mmol). The reaction was subsequently heated to reflux for 12 h. Upon cooling, the reaction was taken up into ethyl acetate (EtOAc) (50 mL) and washed with saturated sodium bicarbonate (NaHCO3) solution (40 mL) and brine (40 mL). The organic layer was dried over MgSO4, filtered, and the volatiles were evaporated. The residue was purified by silica gel chromatography, using hexanes: EtOAc (1:2), containing 1% triethylamine (TEA), affording protected *o*-nitrobenzyl histidine as a golden oil in 59% yield (1.59 g, 3.93 mmol).

(S)-2-Amino-3-[1-(2-nitrobenzyl)-1H-imidazol-4-yl]propanoic acid dihydrochloric salt

Protected *o*-nitrobenzyl histidine (1.6 g, 3.93 mmol) was dissolved in 6 M HCl (20 mL) and heated to reflux for 12 h. Upon cooling to r.t., the aqueous layer was washed with Et2O (10 mL). The volatiles from the aqueous layer were evaporated and the residue was redissolved in

MeOH (10 mL) and precipitated into Et2O (250 mL), affording *o*-nitrobenzyl histidine as a brown oil in 48% yield (690 mg, 1.90 mmol) 1H NMR (400 MHz, D2O) δ = 3.38 (d, J = 7.2 Hz, 2 H), 4.25 (t, J = 6.8 Hz,1 H), 5.75 (s, 2 H), 7.49 (s, 1 H), 7.52 (d, J = 7.7 Hz, 1 H), 7.70 (t, J = 8.2 Hz, 1 H), 7.81 (t, J = 7.6 Hz, 1 H), 8.25 (d, J = 8.2 Hz, 1 H). 13C NMR (101 MHz, D2O) δ = 25.3, 50.4, 52.2, 121.1, 126.0, 127.7, 128.1, 130.9, 132.0, 135.2, 135.3, 147.4, 170.8. HRMS: m/z calcd for $C_{13}H_{14}N_4O_4$ [M+H]+: 291.1088; found: 291.1.

2nd Batch:

(*S*)-Methyl 2-(*tert*-butoxycarbonylamino)-3-[1-(2-nitrobenzyl)-1*H*-imidazol-4-yl] propanoate

To a solution of Boc-(L)His-OMe (3.00 g, 11.1 mmol) containing DIPEA (3.16 g, 4.26 mL, 24.48 mmol) was added 2-nitrobenzyl bromide (2.16 g, 9.96 mmol). The reaction was subsequently heated to reflux for 12 h. Upon cooling, the reaction was taken up into ethyl acetate (EtOAc) (50 mL) and washed with saturated sodium bicarbonate (NaHCO3) solution (40 mL) and brine (40 mL). The organic layer was dried over MgSO4, filtered, and the volatiles were evaporated. The residue was purified by silica gel chromatography, using hexanes: EtOAc (1:1), containing 1% TEA and hexanes: EtOAc (1:2), containing 1% TEA , affording protected *o*-nitrobenzyl histidine as a golden oil in 56% yield (2.2 g, 5.60 mmol).

(S)-2-Amino-3-[1-(2-nitrobenzyl)-1H-imidazol-4-yl]propanoic acid dihydrochloric salt

Protected *o*-nitrobenzyl histidine (2.27 g, 5.60 mmol) was dissolved in 6 M HCl (30 mL) and heated to reflux for 12 h. Upon cooling to r.t., the aqueous layer was washed with Et2O (30 mL). The volatiles from the aqueous layer were evaporated and the residue was redissolved in MeOH

(25 mL) and precipitated into Et2O (250 mL), affording *o*-nitrobenzyl histidine as a brown oil in 71% yield (1.45 g, 3.99 mmol)**.**

5.8.2 Decaging Studies of Protected PCHis

Decaging studies of protected PCHis were performed using 5 mM in MeOH of both protected PCHis and protected hisidtine (starting material) from the synthesis of PCHis (Methods 5.8.1). Protected PCHis was placed in a glass tube submerged in ice to prevent evaporation of the solvent from the heat given off by the lamps. A 302 nm (4 W) and 365 nm (23 W) lamps were used. Sample was removed at each determined time and separated by HPLC-UV. 5 mM of protected histidine was also separated by HPLC-UV to determine retention of free histidine in respect to 210 nm absorbance.

Chapter 6. Discussion

The PylRS libraries were constructed using the two methods of Gibson and golden gate assembly. Golden gate assembly was determined to be the superior of the two approaches because it resulted in greater PylRS mutant transformants per transformation. This greater transformation efficiency can be attributed to the difference in ligation efficiency between intramolecular (golden gate) versus intermolecular (Gibson). In the golden gate assembly, using the optimized BsaI endonuclease that is active in ligation buffer allowed for ligation and digestion to occur sequentially without the need for a purification step. This feature proved to have a significant difference in transformation efficiency as well. Additionally, a concerning lack of fidelity in the DNA sequence was found in clones generated from Gibson assembly. Unprogrammed inserts around the overlap region of the two PCR fragments were observed. This lack of fidelity is problematic for large-scale library construction; therefore, Gibson assembly is more suited for traditional cloning. A drawback to golden gate assembly is that the plasmid must not contain natural BsaI recognition sites. Overall, golden gate assembly was found to be the superior site-saturated mutagenesis method, however, Gibson cloning is the preferred method for traditional cloning.

Upon analysis of the successfully evolved active synthetases with the UAAs, it was revealed that all residues that were site-saturated mutated in the libraries were mutated in the evolved active hits except Y271, W417(382), and G421(386). While the selection sample size is rather small, conclusions may be drawn as to the effectiveness of these three amino acids in the evolution of PylRS variants. The reported hydrogen bonding interactions of Y271 may be too crucial for the coordination of a UAA in the active site that mutation is simply not observed. W417(382), and G421(386) represent the largest and smallest in size of all the amino acids,

respectively. Deviation from these native amino acids possibly poses too drastic of an effect on the shape and size of the active site cavity. Additionally, G421(386) is the closest in proximity to the adenylated pyrrolysine (Figure 2.1) of all selected residues for inclusion in the libraries, therefore, its small size seems necessary for proper binding of the synthetase to an UAA. Mutation to other amino acids potentially inhibits binding of an UAA substrate in the active site. Also, G421(386) is located close to the alpha amino group of pyrrolysine, so modifications on the epsilon amino group of lysine probably has little to no interaction with the residue, therefore, limiting its effectiveness in directed evolution of PylRS to accept ε-nitrogen modifications on lysine. If library residues were to be chosen again, G421(386) would not be included. Because only five, possibly six, residues are chosen for site-saturated mutagenesis due to the inherent large theoretical diversity, the efficiency and effectiveness of directed evolution of a protein library is significantly reliant upon the careful selection of residues for site-saturated mutagenesis.

When looking at the origin of the libraries each successful hit was derived from, it is shown that the *barkeri* libraries were favored over the *mazei* libraries. This bias may be the result of the unequal distribution of the total population of library members subjected to the selections. Because larger libraries of *barkeri* than *mazei* were made, there are more total *barkeri* library members than *mazei* derived library members being subjected to the selections. A successful active synthetase was derived from the *mazei* library, however. Both the *mazei* and *barkeri* libraries converged on the cysteine to threonine mutation found in IVKRS. This proved that the *mazei* library is being evolved in the selection process. Interestingly, no hits were derived from the mixed *barkeri* library despite the lysine analog library being reported to be effective in evolving PylRS for specific activity with lysine derivatives.

Selections with IBK and IVK resulted in wild-type PylRS and IVKRS which is very similar to wild-type. Only half of the hits observed in the IBK selections were sequenced to identify their residue content. Though all 15 that were sequenced converged on the wild-type identity, it would be interesting to see if any of the additional hits deviated from wild-type. IVK and IBK were not needed to serve as surrogates for the enrichment of a PylRS synthetase specifically active with HBK after all. Fortunately, when it was time to conduct selections on HBK, the synthesis was sorted out and optimized. The construction and addition of the *barkeri* library L270/Y271/L274/N311/C313 proved to be essential and successful when evolving a PylRS synthetase variant active with HBK. At the time, the design of the library was engineered with the structure of PCHis in mind to better accommodate PCHis when selections were to be performed. The lack of success in enriching a synthetase specifically active with HBK expedited the need for the additional library. It would be interesting to repeat the selections with IBK and IVK using the additional *barkeri* library to observe if a synthetase that exhibits superior binding to IBK and/or IBK can be selected for.

The selection and screening process using PCHis proved to be quite fortunate. A limited number of clones were screened before a hit was observed. It would be interesting to screen more clones for exclusive PCHis activity to potentially discover a superior synthetase. On the other hand, it was unfortunate that the decaging studies did not confirm the reported decaging efficiency reported by our collaborators in the Deiter's research group at the University of Pittsburgh. A caging efficiency of 20% was reported for protected PCHis after 15 minutes of exposure of light at 365 nm (23 W). However, partial decaging was observed from HPLC-UV studies at 2 hours of 365 nm (23 W) and 2 hours of 302 (4 W). Despite having 25% the power, the 302 nm exhibited greater decaging of protected PCHis free amino acid. This greater

decaging efficiency is because 302 nm is closer to the maximum absorption of 300 nm by nitrobenzyl therefore 302 nm exhibits a larger molar-absorption coefficient. The drawback is that around 300 nm, irradiation is more likely to be absorbed by (and possibly cause damage to) the biological environment.

Chapter 7. Conclusions and Current/Future Work

7.1 Conclusions

Large-scale PylRS libraries derived from both *M. barkeri* and *M. mazei* were successfully constructed by site-saturated mutagenesis using golden gate assembly and Gibson assembly. Growth assays demonstrated that the designed directed evolution of the PylRS libraries to specifically bind UAAs was a robust and viable strategy. The selection process validated that synthetases with activity toward the endogenous 20 amino acids and the UAA of interest were enriched in the positive selections utilizing in-frame stop codon suppression of chloramphenicol acetyl transferase. Likewise, the negative selection proved that synthetases that were active with the endogenous amino acids can be selected against using in-frame amber stop codon suppression of the toxic protein Barnase.

The PylRS libraries were subjected to directed evolution to enrich synthetases that were specifically active with the acylated lysine derivatives of isobutyryl lysine, isovaleryl lysine, and β-hydroxybutyryl lysine. Selections on isobutyryl lysine and isovaleryl lysine resulted in the convergence on wild-type PylRS for both acylated lysine derivatives as well as the mutation of cysteine to threonine in IVKRS for both *mazei* and *barkeri* synthetases. Selections on β-hydroxybutyryl lysine proved to be more difficult. After multiple rounds of selections and screenings, a synthetase variant exclusively active with β-hydroxybutyryl lysine was not obtained. Consequently, an additional PylRS library was constructed by rational design that was believed to be more effective in directed evolution of PylRS variants to accept unnatural amino acids that have modifications on the ε-nitrogen of lysine. The new library included site-saturated mutagenesis of amino acids found deep in the active site cavity in the hydrophobic cleft. When subjected to rounds of selection to select for a PylRS variant active with β-hydroxybutyryl

lysine, the newly constructed library proved to be extremely effective. Of the many hits that were screened, the synthetases converged on two synthetase identities derived from the new library.

To demonstrate that the acylated lysine derivatives can be incorporated into protein, high levels of expression of the model protein, ubiquitin, was accomplished. UAA dependent suppression of the in-frame amber stop codon in ubiquitin and subsequent exclusive incorporation of the acylated lysine derivatives into ubiquitin was demonstrated by separation on SDS-PAGE protein gels and the intact protein masses were confirmed by mass spectrometry.

The PylRS libraries were subjected to directed evolution to enrich synthetases that were specifically active with photocaged histidine as well. Photocaged histidine was successfully synthesized and used as a substrate in the selections to enrich a synthetase specifically active with the UAA. The selections on photocaged histidine were successful as a synthetase was enriched that exclusively binds to photocaged histidine. To demonstrate that photocaged histidine can be incorporated into protein, high levels of expression of the model protein, ubiquitin, was accomplished. UAA dependent suppression of the in-frame amber stop codon in ubiquitin and subsequent exclusive incorporation of photocaged histidne into ubiquitin was demonstrated by separation on a SDS-PAGE protein gel and the intact protein mass was confirmed by mass spectrometry. Decaging studies on protected photocaged histidine free amino acid revealed that successful decaging required a smaller wavelength at a longer exposure time than what has been reported.

Overall, construction of the large-scale PylRS libraries along with setting up the selection system allows for substantial future work in evolving the PylRS libraries to accept UAA for the Cropp research group. The Cropp lab has an array of synthesized unnatural amino acids

available and ready to be used in selecting for an active PylRS variant to ultimately incorporate the UAAs into proteins.

7.2 Current and Future Work

Currently, work is being done to identify what exactly was incorporated, if not HBK, into ubiquitin in response to the K11TAG amber stop codon. Tryptic digest followed by HPLC MS/MS of UBK11HBK is being performed to determine the difference in expected mass and the mass detected by mass spectrometry. A MASCOT search of fragmented peptides can be used to definitively see what has been incorporated. Tryptic digest HPLC MS/MS will demonstrate if racemic HBK was truly incorporated into ubiquitin.

Because HBK was synthesized as racemic mixture, it's difficult to determine which or if both enantiomers are substrates for HBKRS. The R-enantiomer was discovered in histone modifications in response to epigenetic regulation of metabolic genes during starvation. Both the R- and S-enantiomer are currently being synthesized and fluorescence assay studies will be performed to determine the activity with each HBKRS synthetase. If the fluorescence assay proves that HBKRS can specifically bind to the R- and/or S-enantiomer of HBK, then R- and/or S-HBK will be used to express recombinant histone subunit 3 modified in response to a K9TAG stop codon. First, the synthetic gene of H3 with a K9TAG must be cloned into the pCNB_UbK11TAG plasmid to replace the ubiquitin construct in the CBD-intein fusion protein. Once recombinant modified histone is expressed, histone/DNA binding studies can be performed to study the effects of hydroxybutyrylation vs hypoacylation. Though they have yet to be discovered in proteins, it may be prudent to incorporate IBK and/or IVK into the recombinant ubiquitin protein in response to the K9TAG amber stop codon as well.

Currently, decaging studies are being conducted on ubiquitin modified with PCHis in response to the K11TAG amber stop codon to determine the decaging efficiency. Decaging efficiency using a light at 302 nm is being monitored by mass spec. If decaging of UbK11PCHis proves to be efficient, then protein activation dependent upon decaging will be studied in model proteins. The two model proteins are luciferase and beta-galactosidase. Luciferase is a photoprotein that emits light when luciferase is bound to the luciferin substrate. Beta-galactosidase activity may be detected using an artificial chromogenic substrate, X-gal. β-galactosidase will cleave the glycosidic bond in X-gal and form galactose and an intense blue product that is easy to identify and quantify. Both enzymes contain critical histidine residues in their active site where replacement with PCHis should inactive each enzyme. Restoration of luciferease protein activity upon decaging will be monitored by fluorescence. Restoration of beta-galactosidase protein activity upon decaging will be monitored with an x-gal assay.

A derivative of PCHis (Figure 7.1) where the *o*-nitrobenzyl caged group contains an ether linkage (NBOM histidine) was obtained generously from our collaborators in the Deiter's group at the University of Pittsburgh. The group reported that protected histidine modified with this caging group undergoes 100% decaging upon irradiation of light at 365 nm for 15 mins. This red shifted decaging efficiency is ideal for protein activation studies as the longer wavelength is less damaging to the biological environment and 100% decaging is optimal. We were given a scarce amount of NBOM histidine, so selections have not been performed to evolve an active synthetase. HBKRS1 and 2 were assayed for their activity (not shown) with the derivative by fluorescence assay, however, both synthetases were not active with the derivative. While NBOM histidine has proven to be more efficient in decaging, PCHis was chosen for selections because it was thought to be a superior substrate for evolution of an active PylRS variant. Also,

protected PCHis free amino acid was reported to decage with 20% efficiency which is suitable for protein activity regulation studies. If PCHis in protein proves to have extremely poor decaging results, then the next route would be to synthesize a bulk quantity of NBOM histidine and run selections using the synthetic UAA.

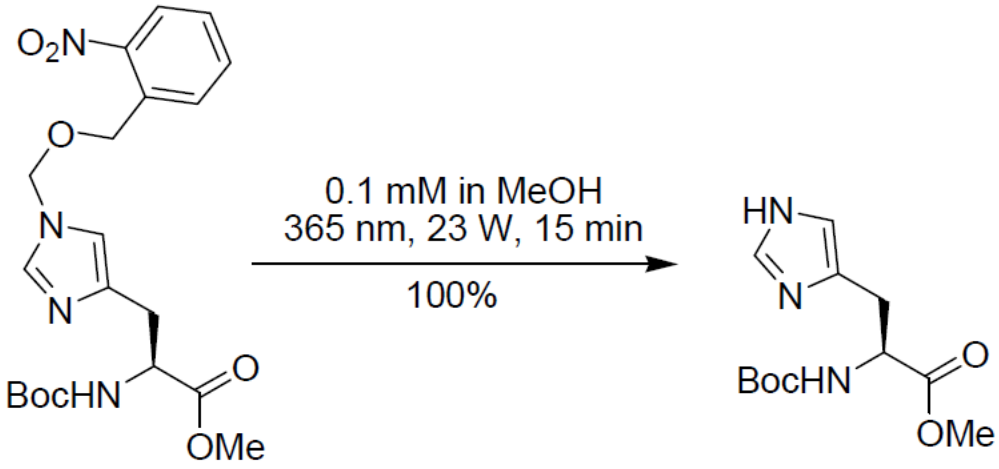


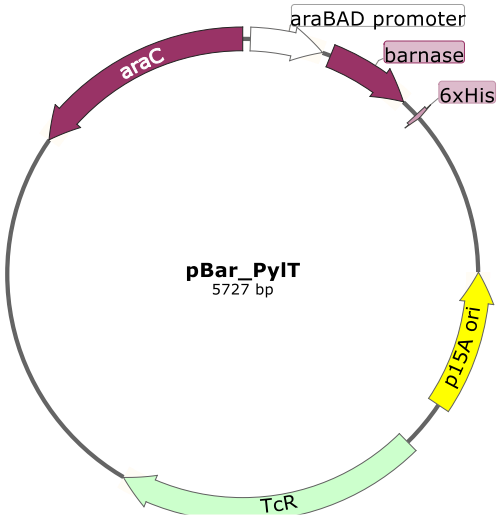
Figure 7.1. Photolytic cleavage of NBOM histidine.

Appendix I. Primer Sequences

Primer	Sequence
1646	GCGCAGGAAAGGTCTCACGTGAATGGGGCATTGATAAACCGNNKATTGGCGCGNNKTTTGGCCTGGAACGTCTGCTGAAAGTGATGC
1647	GCGCAGAGTAGGTCTCACACGATCCAGGCTAACCGGACCCACMNNCGCGCTGCTCAGTTCCAGATCGCCATGCATAATATCC
1648	GCGCAGGAAAGGTCTCAATGGTTNNKTTTNNKCAAATGGGCAGCGGCTGCACCCGTGAAA
1649	GCGCAGAGTAGGTCTCACCATGGTGAATTCTTCCAGGTGTTCTTTGCCATCG
1884	GTTATACGTTGTTTACGCTTTGAGGAATCCCATATGATGGATAAAAAACCGCTGGATGTG
1885	GCGCATCAGGCAATTTAGCGTTTGAAACTGCATTTACAGGTTCGTGCTAATGCC
1886	ATGGGATTCCTCAAAGCGTAAACAACGTATAAC
1887	ATGCAGTTTCAAACGCTAAATTGCCTGATGCGC
1892	AAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTNNKTTTNNKCA
1893	TGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATA
1894	TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGC
1895	TTCCAGGTGTTCTTTGCCATCGCTTTCTTTGCGATAGCACGGGCCCACTTCAAAAATTT
1984	CCCATACCGCTTGACCGGGAATGGGGTATTGATAAACCCNNKATAGGGGCANNKTTCGGG
1985	TGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGT
1986	AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCC
1987	GGGTTTATCAATACCCCATTCCCGGTCAAGCGGTATGGGTCCGACMNNTGCAGAGGAAAG
2202	TATGGTCTCCCCGACCNNKNNKAACTATNNKCGTAAACTGG
2203	ATAGGTCTCGTCGGGGCCAGCATCGGACGCAG
2226	TAAGTGCGGCGACGAtagtcatgccccgcgcccaccgg
2227	TGGGCGCGGGGCATGACTAtcgtcgccgcacttatgactgtc
2266	GCGTCCGATGCTGGCCCCGACCNNKNNKAACTATNNKCGTAAACTGG

II. Plasmid Maps and Sequences

pBar_PylT:



Note: This plasmid was given to our lab from our collaborators at the University of Pittsburgh in the Deiter's lab.

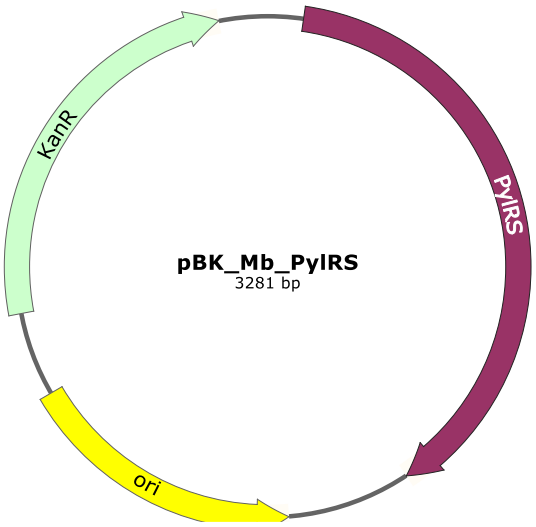
Barnase
AraBad promoter
Tet Efflux Protein
PylT
AraC
p15A Ori

CCATACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATTGTCCATATTGCATCA
GACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAACCGGTAACCCC
GCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTA
ACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCA
CACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTT
TTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATT
AACCATGGCATAGGTTATCAACACGTTTGACGGGGTTGCGGATTATCTTCAGACATA
TCATAAGCTACCTGATAATTACATTACAAAATCAGAAGCACAAGCCCTCGGCTGGGT
GGCATCAAAAGGGAACCTTGCATAGGTCGCTCCGGGGAAAAGCATCGGCGGAGACA
TCTTCTCAAACAGGGAAGGCAAACTCCCGGGCAAAAGCGGACGAACATGGCGTGAA
GCGGATATTAACTATACATCAGGCTTCAGAAATTCAGACCGGATTCTTTACTCAAGC
GACTGGCTGATTTACAAAACAACGGACCATTATCAGACCTTTACAAAAATCAGATA
ACATTGGGAATTCGAAGCTTGGGCCCGAACAAAAACTCATCTCAGAAGAGGATCTG
AATAGCGCCGTCGACCATCATCATCATCATCATTGAGTTTAAACGGTCTCCAGCTTG
GCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCA
GAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCT
GACCCCATGCCGAACTCAGAAGTGAAACGCCGTAgcgccgatggtagtgtggggtctccccatgcgaga
gtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctg

agtaggacaaatccgccgggagctgtccctcctgttcagctactgacggggtggtgcgtaacggcaaaagcaccgccggacatcagcgc
tagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgc
gtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatggctt
acgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcggcaaagccgtttttccataggc
tccgccccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttcccct
ggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacact
cagttccgggtaggcagttcgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtct
tgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccgg
ttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcga
aaaaccgccctgcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagata
aaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttat
catcgataagctttaatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgtgtATGAAATCTAACAATGCG
CTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTTATG
CCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCAC
TATGGCGTGCTGCTAGCGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCG
GAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTTGGAG
CCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCTCTACGCCG
GACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATC
GCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTG
TTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTC
CTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGGCCTCAACCTACTACTGGG
CTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAG
CCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCAC
TTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTGGG
TCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTG
CGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCA
AACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACGCGCTGGGC
TACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTC
TCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTGTCCAGGCAGGTAG
ATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTT
CGATCATTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGA
ACGGGTTGGCATGGATTGTAGGCGCCGCCCTATACCTTGTCTGCCTCCCCGCGTTGC
GTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCG
CTAACGGATTCACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGA
ATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCCATCTCCAGCAGCC
GCACGCGGCGCATCTCGGGCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCA
ACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAG
CGTCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTT
TTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTAAAAAAAATCCTTA
GCTTTCGCTAAGATCTGCAGTGGCGGAAACCCCGGGAATCTAACCCGGCTGAACGG
ATTTAGAGTCCATTCGATCTACATGATCAGGTTCCCGCGGCCGCGAATTCAGCGNNN
CNANNATTACACAAAGTTTTTTATGTTGAGAATATTTTTTTGATGGGGCGCCACTTAT
TTTTGATCGTTCGCTCAAAGAAGCGGCGCCAGGGTTGTTTTTCTTTTCACCAGTGAGA
CGGGCAACAGAACGCCATGAGCGGCCTCATTTCTTATTCTGAGTTACAACAGTCCGC
ACCGCTGCCGGTAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTT
ATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCCCAACAGT

CCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCCCTGCACCAT
TATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATC
TGTATTAACGAAGCGCTAACCGTTTTTATCATGCTCTGGGAGGCAGAATAAATGATC
ATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCATTT
GAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATA
GACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCT
TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCG
TTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGC
AGTTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTG
ACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAG
CTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCAGATCA
ATTCGCGCGCGAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGGACGAAGCA
GGGATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTCTGATTCGTTACC
AATTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAAC
TCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCG
TCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAG
CAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAA
CTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGA
CGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCT
CGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCC
GCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCC
CTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTT
CATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCAT
GCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCC
GGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATCACCCGG
TCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAATGGTGAG
ATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACC
GTTGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCG
GCAGCAGGGGATCATTTTGCGCTTCAGCCAT

pBK_Mb_PylRS

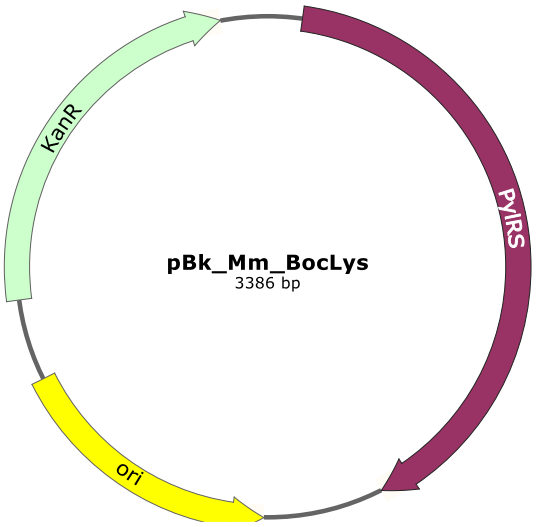


PylRS_WT
OriE1
KanR
Site Saturated Mutations (NNK):
N311, C313, V366, W382, G386
L270, Y271, L274

CTCGGGAGTTGTCAGCCTGTCCCGCTTATAAGATCATACGCCGTTATACGTTGTTTAC
GCTTTGAGGAATCCCCATATGATGGATAAAAAACCGCTGGATGTGCTGATTAGCGCG
ACCGGCCTGTGGATGAGCCGTACCGGCACCCTGCATAAAATCAAACATCATGAAGT
GAGCCGCAGCAAAATCTATATTGAAATGGCGTGCGGCGATCATCTGGTGGTGAACA
ACAGCCGTAGCTGCCGTACCGCGCGTGCGTTTCGTCATCATAAATACCGCAAAACCT
GCAAACGTTGCCGTGTGAGCGATGAAGATATCAACAACTTTCTGACCCGTAGCACCG
AAAGCAAAAACAGCGTGAAAGTGCGTGTGGTGAGCGCGCCGAAAGTGAAAAAAGC
GATGCCGAAAAGCGTGAGCCGTGCGCCGAAACCGCTGGAAAATAGCGTGAGCGCG
AAAGCGAGCACCAACACCAGCCGTAGCGTTCCGAGCCCGGCGAAAAGCACCCCGAA
CAGCAGCGTTCCGGCGTCTGCGCCGGCACCGAGCCTGACCCGCAGCCAGCTGGATC
GTGTGGAAGCGCTGCTGTCTCCGGAAGATAAAATTAGCCTGAACATGGCGAAACCG
TTTCGTGAACTGGAACCGGAACTGGTGACCCGTCGTAAAAACGATTTTCAGCGCCTG
TATACCAACGATCGTGAAGATTATCTGGGCAAACTGGAACGTGATATCACCAAATTT
TTTGTGGATCGCGGCTTTCTGGAAATTAAAAGCCCGATTCTGATTCCGGCGGAATAT
GTGGAACGTATGGGCATTAACAACGACACCGAACTGAGCAAACAAATTTTCCGCGT
GGATAAAAACCTGTGCCTGCGTCCGATGCTGGCCCCGACCCTGTATAACTATCTGCG
TAAACTGGATCGTATTCTGCCGGGTCCGATCAAAATTTTTGAAGTGGGCCCGTGCTA
TCGCAAAGAAAGCGATGGCAAAGAACACCTGGAAGAATTCACCATGGTTAACTTTT
GCCAAATGGGCAGCGGCTGCACCCGTGAAAACCTGGAAGCGCTGATCAAAGAATTC
CTGGATTATCTGGAAATCGACTTCGAAATTGTGGGCGATAGCTGCATGGTGTATGGC
GATACCCTGGATATTATGCATGGCGATCTGGAACTGAGCAGCGCGGTGGTGGGTCC
GGTTAGCCTGGATCGTGAATGGGGCATTGATAAACCGTGGATTGGCGCGGGTTTTGG
CCTGGAACGTCTGCTGAAAGTGATGCATGGCTTCAAAAACATTAAACGTGCGAGCC
GTAGCGAAAGCTACTATAACGGCATTAGCACGAACCTGTAACTGCAGTTTCAAACG

CTAAATTGCCTGATGCGCTACGCTTATCAGGCCTACATGATCTCTGCAACACACCGA
GCCCGCGTGCTTTTGCAGGCCGGATAAGGCGTTCGCGCCGCATCCGGCAAGAAACA
GCAAACAATCCAAAACGCCGCGTTCGGCGGCGTTTTTTCTGCTTTTCTTCGCGAATTA
ATTCCGCTTCGCAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAA
GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAA
TCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGT
TTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATA
CCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGG
TATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCC
GTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTA
AGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAG
GTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGT
TGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTG
CAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTT
CTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGA
GTTGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGA
ACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATT
CAACGGGAAACGTCTTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTA
TATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCG
ATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCG
TTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGC
CTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCAC
TGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTG
AAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTG
TAATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAAT
GAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGT
TGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGT
CACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGG
TTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCT
ATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATAT
GGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTT
TCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGAC
GGGACGGCGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATC

pBK_Mm_PylRS

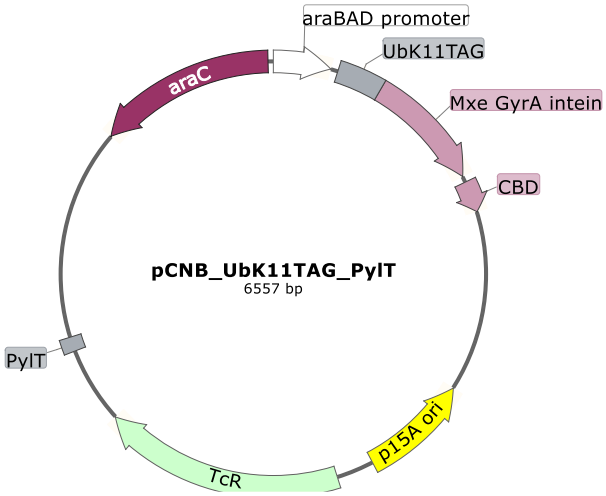


KanR
OriE1
PylRS
TTT= Y384F (BocK mutation)
Site Saturated Mutations (NNK)-
N311 (346), C313 (348), V366 (401), W382 (417), G386 (421)
Note: *barkeri* Residue # (*mazei* Residue #)

CTCGGGAGTTGTCAGCCTGTCCCGCTTATAAGATCATACGCCGTTATACGTTGTTTAC
GCTTTGAGGAATCCCATATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCG
GGCTCTGGATGTCCAGGACCGGAACAATTCATAAAATAAAACACCACGAAGTCTCT
CGAAGCAAAATCTATATTGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCC
AGGAGCAGCAGGACTGCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTGCA
AACGCTGCAGGGTTTCGGATGAGGATCTCAATAAGTTCCTCACAAAGGCAAACGAA
GACCAGACAAGCGTAAAAGTCAAGGTCGTTTCTGCCCCTACCAGAACGAAAAAGGC
AATGCCAAAATCCGTTGCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCAC
AGGCTCAACCTTCTGGATCTAAATTTTCACCTGCGATACCGGTTTCCACCCAAGAGT
CAGTTTCTGTCCCGGCATCTGTTTCAACATCAATATCAAGCATTTCTACAGGAGCAA
CTGCATCCGCACTGGTAAAAGGGAATACGAACCCCATTACATCCATGTCTGCCCCTG
TTCAGGCAAGTGCCCCCGCACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTG
TTAAACCCAAAAGATGAGATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAG
TCCGAATTGCTCTCTCGCAGAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAG
GGAGAATTATCTGGGGAAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGG
GTTTTCTGGAAATAAAATCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGG
GCATTGATAATGATACCGAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCT
GCCTGAGACCCATGCTTGCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGG
CCCTGCCTGATCCAATAAAAATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCG
ACGGCAAAGAACACCTCGAAGAGTTTACCATGCTGAACTTCTGCCAGATGGGATCG
GGATGCACACGGGAAAATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGG
AATTGATTTCAAGATCGTAGGCGATTCCTGCATGGTC**TTT**GGGGATACCCTTGATGT
AATGCACGGAGACCTGGAACTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCG
GGAATGGGGTATTGATAAACCCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCT

AAAGGTTAAACACGACTTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACT
ATAACGGGATTTCTACCAACCTGTAAATCTAGATTTCAAACGCTAAATTGCCTGATG
CGCTACGCTTATCAGGCCTACATGATCTCTGCAACACACCGAGCCCGCGTGCTTTTG
CAGGCCGGATAAGGCGTTCGCGCCGCATCCGGCAAGAAACAGCAAACAATCCAAAA
CGCCGCGTTCGGCGGCGTTTTTTCTGCTTTTCTTCGCGAATTAATTCCGCTTCGCAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGG
CGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTC
AGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGC
TCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC
TCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGT
GTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCG
CTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATC
GCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTG
CTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGAT
CCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTA
CGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACG
CTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGTTGTGTCTCAAAAT
CTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAACAATAAAACTGTCT
GCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTC
TTGCTCGAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATG
GGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGC
CCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTA
CAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCA
AGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGA
AAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATG
CGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAA
CAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGT
TGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAA
AGAAATGCATAAGCTTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTC
TCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGA
CGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGT
GAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTG
ATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGT
TAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGGCTTTGT
TGAATAAATCGAACTTTTGCTGAGTTGAAGGATC

pCNB_UBK11TAG_PylT

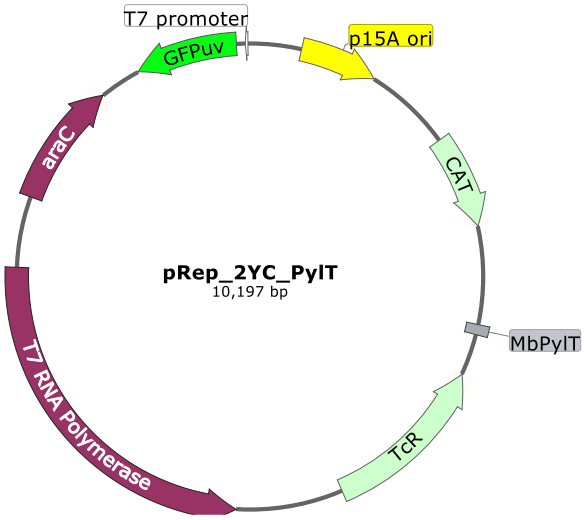


AACTACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATT
TTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCT
TCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTAT
TCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAA
GTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT
CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTCTCCAATGATGAG
CACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGA
GCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGT
CACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCA
TAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCG
AAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGT
TGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCC
TGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGC
TTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTC
TGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGC
GTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCG
TAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATC
GCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCA
TATATACTTTAGATTGATTTACCCCGGTTGATAATCAGAAAAGCCCCAAAAACAGGA
AGATTGTATAAGCAAATATTTAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTT
AAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCC
TTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAA
GAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATC
AGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGT
GCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGG
GGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGC
GCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGC
GCTTAATGCGCCGCTACAGGGCGCGTAAAAGGATCTAGGTGAAGATCCTTTTTGATA
ATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCG
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTT
GCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTAC

CAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCC
TTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACAT
ACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTC
TTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG
ATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGG
ACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCC
AGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGA
GCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCA
ACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCT
GCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACC
GCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCTATGGT
GCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCT
ATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGC
GCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTC
CGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGC
TGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCAT
CCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGC
GGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGG
GATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATA
CGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACT
GGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCCGA
ACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGATAATGGCCTGC
TTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTG
CAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGC
GGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATG
ATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAA
GGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAAT
GAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAA
ACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTG
CGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATT
GCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCC
CAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGT
CTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACT
CGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCA
GTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCA
CTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTAT
GCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGC
GCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCT
TCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAA
CGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCG
GATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTT
TACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTT
GATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGA
CTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACG
CGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCG

CAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCG
GCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGA
CTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTG
TCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAG
TAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCCGGCATGCCG
CCCTTTCGTCTTCAAGAATTAATTCCCAATTCCCCAGGCATCAAATAAAACGAAAGG
CTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCT
GAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAG
GGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAA
ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCG
GTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAA
GCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAAT
TCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTA
TCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATT
TGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACT
GCCAGGAATTAATTCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTG
GGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCG
CCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGAC
GCCCGCCATAAACTGCCAGGAATTAATTCCCCAGGCATCAAATAAAACGAAAGGCT
CAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGA
GTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGG
TGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTGGGGATCGGAATTAATTCC
CGGTTTAAACCGGGGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAAT
TGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
TATACATATGCAGATTTTTGTGAAAACCCTGACCGGCAAAACCATTACCCTGGAAGT
GGAACCGAGCGATACCATTGAAAATGTGAAAGCGAAAATTCAGGATAAAGAAGGC
ATTCCGCCGGATCAGCAGCGTCTGATTTTTGCGGGCAAACAGCTGGAAGATGGTCGT
ACCCTGAGCGATTACAATATTCAGAAAGAAAGCACCCTGCATCTGGTGCTGCGTCTG
CGTGGCGGCTGCATCACGGGAGATGCACTAGTTGCCCTACCCGAGGGCGAGTCGGT
ACGCATCGCCGACATCGTGCCGGGTGCGCGGCCCAACAGTGACAACGCCATCGACC
TGAAAGTCCTTGACCGGCATGGCAATCCCGTGCTCGCCGACCGGCTGTTCCACTCCG
GCGAGCATCCGGTGTACACGGTGCGTACGGTCGAAGGTCTGCGTGTGACGGGCACC
GCGAACCACCCGTTGTTGTGTTTGGTCGACGTCGCCGGGGTGCCGACCCTGCTGTGG
AAGCTGATCGACGAAATCAAGCCGGGCGATTACGCGGTGATTCAACGCAGCGCATT
CAGCGTCGACTGTGCAGGTTTTGCCCGCGGAAAACCCGAATTTGCGCCCACAACCTA
CACAGTCGGCGTCCCTGGACTGGTGCGTTTCTTGGAAGCACACCACCGAGACCCGG
ACGCCCAAGCTATCGCCGACGAGCTGACCGACGGGCGGTTCTACTACGCGAAAGTC
GCCAGTGTCACCGACGCCGGCGTGCAGCCGGTGTATAGCCTTCGTGTCGACACGGC
AGACCACGCGTTTATCACGAACGGGTTCGTCAGCCACGCTACTGGCCTCACCGGTCT
GAACTCAGGCCTCACGACAAATCCTGGTGTATCCGCTTGGCAGGTCAACACAGCTTA
TACTGCGGGACAATTGGTCACATATAACGGCAAGACGTATAAATGTTTGCAGCCCCA
CACCTCCTTGGCAGGATGGGAACCATCCAACGTTCCTGCCTTGTGGCAGCTTCAATG
ACTGCAGGAAGGGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTG
CTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGA
GGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

pRep_PylT:



p15A ori
CmR (D111TAG)
MbPylT
TcR
T7 RNA Polymerase (1_107TAG)
GFPUv

GCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCT
CGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATC
TACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGAT
AGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACT
TTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTT
GATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGAC
CCCTTAATAAGATGATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTG
AAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACT
CTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAG
TTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAGTG
GCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTA
CCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTT
GGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGC
CATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGA
GGGAGCCGCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC
ACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAA
ACGGCTTTGCCGCGGCCCTCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGG
AAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTCGCCGCAGTCGAACGACCGAGCGT
AGCGAGTCAGTGAGCGAGGAAGCGGAATATATCCTGTATCACATATTCTGCTGACG
CACCGGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTGACACCCTCATCA
GTGCCAACATAGTAAGCCAGTATACACTCCGCTAGCGCTGATGTCCGGCGGTGCTTT
TGCCGTTACGCACCACCCCGTCAGTAGCTGAACAGGAGGGACAGCTGATAGAAACA
GAAGCCACTGGAGCACCTCAAAAACACCATCATACACTAAATCAGTAAGTTGGCAG
CATCACCCGACGCACTTTGCGCCGAATAAATACCTGTGACGGAAGATCACTTCGCAG

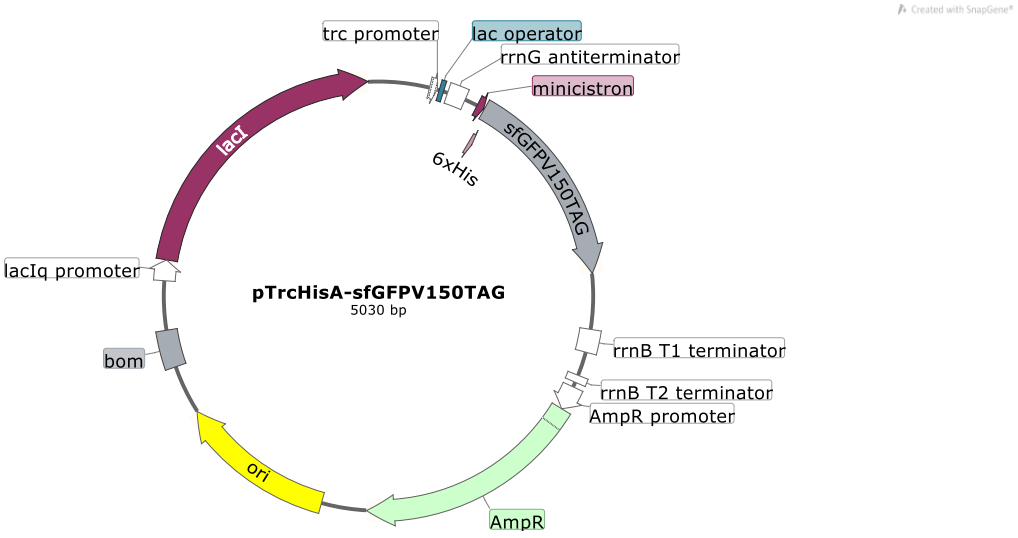
AATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCCAACTTTTGGC
GAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATA
AGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCT
AAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCG
TAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGT
TCAGCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTA
TCCGGCCTTTATTCACATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATG
GCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTTCACCCTTGTTACACCGTT
TTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACCTATTC
CGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCC
TATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGA
GTTTCACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTT
CACCATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTC
AGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGGCAGAATGCTTAATGAATTAC
AACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTATTGGT
GCCCTTAAACGCCTGGTGCTACGCCTGAATAAGTGATAATAAGCGGATGAATGGCA
GAAATTCGAAAGCAAATTCGACCCGGTCGTCGGTTCAGGGCAGGGTCGTTAAATAG
CCGCTTATGTCTATTGCTGGTTTACCGGTTTATTGACTACCGGAAGCAGTGTGACCGT
GTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAAT
TGACGATATGATCATTTATTCTGCCTCCCAGAGCCTGATAAAAACGGTTAGCGCTTC
GTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATC
CGGAACATAATGGTGCAGGGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGT
GGCCGGGGGACTGTTGGGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGA
AGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTA
CCGGACAGCGGTGCGGACTGTTGTAACTCAGAATAAGAAATGAGGCCGCTCATGGC
GTTGCGCCGCTTCTTTGAGCGAACGATCAAAAATAAGTGGCGCCCCATCAAAAAAA
TATTCTCAACATAAAAAACTTTGTGTAATACTTGTAACGCTGAATTCGGGAACCTGA
TCATGTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTC
CGCCAAGATCTTCCTTAGCGAAAGCTAAGGATTTTTTTTAAGCTTACGCTCTCCCTTA
TGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGC
CGCCGCAAGGAATGGTGCATGCAAGGAGCCCGAGATGCGCCGCGTGCGGCTGCTGG
AGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTCACAGTTC
TCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGGTGCC
GCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCGACGCAACGCGG
GGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGTTCCATGTG
CTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCAGTGATCGAAGTTAG
GCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTG
CCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAA
TCATAATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCC
AGCGCGTCGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTG
GCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAG
CGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCC
AGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGT
GCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGC
TCTCAAGGGCATCGGTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCC
CAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGG

AGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAA
CAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGC
GATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTC
CGGCGTAGAGGATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTG
GCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTG
CTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACG
CCATAGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAG
TACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGA
CGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACT
GTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTAC
AACTTATATCGTATGGGGCTGACTTCAGGTGCTACATTGCTCAAAGATGCAGGGGTA
AAAGCTAACCGCATCTTTACCGACAAGGCATCCGGCAGTTCAACAGATCGGGAAGG
GCTGGATTTGCTGAGGATGAAGGTGGAGGAAGGTGATGTCATTCTGGTGAAGAAGC
TCGACCGTCTTGGCCGCGACACCGCCGACATGATCCAACTGATAAAAGAGTTTGATG
CTCAGGGTGTAGCGGTTCGGTTTATTGACGACGGGATCAGTACCGACGGTGATATGG
GGCAAATGGTGGTCACCATCCTGTCGGCTGTGGCACAGGCTGAACGCCGGAGGATC
CTCCGGGCCGTTGCTTCGCAACGTTCAAATCCGCTCCCGGCGGATTTGTCCTACTCA
GGAGAGCGTTCACCGACAAACAACAGATAAAACGAAAGGCCCAGTCTTTCGACTGA
GCCTTTCGTTTTATTTGATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCCACA
CTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGGCATGGGGTCAGGTGGGACC
ACCGCGCTACTGCCGCCAGGCAAATTCTGTTTTATCAGACCGCTTCTGCGTTCTGATT
TAATCTGTATCAGGCTGAAAATCTTCTCTCATCCGACGTCTTAGGCGAAGGCGAAGT
CCGACTCTAAGATGTCACGGAGGTTCAAGTTACCTTTAGCCGGAAGTGCTGGCATTT
TGTCCAATTGAGACTCGTGCAACTGGTCAGCGAACTGGTCGTAGAAATCAGCCAGTA
CATCACAAGACTCATATGTGTCAACCATAGTTTCGCGCACTGCTTTGAACAGGTTCG
CAGCGTCAGCCGGAATGGTACCGAAGGAGTCGTGAATCAGTGCAAAAGATTCGATT
CCGTACTTCTCGTGTGCCCACACTACAGTCTTACGAAGGTGGCTACCGTCTTGGCTGT
GTACAAAGTTAGGAGCGATACCAGACTCCTGTTTGTGTGCATCAATCTCGCTATCTT
TGTTGGTGTTAATGGTAGGCTGTAAGCGGAACTGACCGAGGAACATCAGGTTCAAG
CGCGTCTGAATAGGCTTCTTGTATTCCTGCCACACAGGGAAACCATCAGGAGTTACC
CAATGCACAGCGCAACGCTTGCGAAGAATCTCTCCAGTCTTCTTATCTTTGACCTCA
GCAGCCAGCAGCTTAGCAGCAGACTTAAGCCAGTTCATTGCTTCAACCGCAGCTACC
ACCGTCACGCTCACAGATTCCCAAATCAGCTTAGCCATGTATCCAGCAGCCTGATTC
GGCTGAGTGAACATCAGACCCTTGCCGGAATCAATAGCTGGCTGAATGGTATCTTCC
AGCACTTGTTGACGGAAGCCGAACTCTTTGGACCCGTAAGCCAGCGTCATGACTGAA
CGCTTAGTCACACTGCGAGTAACACCGTAAGCCAGCCATTGACCAGCCAGTGCCTTA
GTGCCCAGCTTGACTTTCTCAGAGATTTCACCAGTGTTCTCATCGGTCACGGTAACTA
CTTCGTTATCGGTCCCATTGATTGCGTCTGCTTGTAGAATCTCGTTGACTTTCTTAGC
AACAATCCCGTAGATGTCCTGAACGGTTTCACTAGGAAGCAAGTTAACCGCGCGAC
CACCTACCTCATCTCGGAGCATCGCGGAGAAGTGCTGGATGCCAGAGCAAGACCCG
TCAAACGCCAGCGGAAGGGAGCAGTTATAGCTCAGGCCGTGGTGCTGTACCCCAGC
GTACTCAAAGCAGAACGCAAGGAAGCAGAACGGAGAATCTTGCTCAGCCCACCAAG
TGTTCTCCAGTGGAGACTTAGCGCAAGCCATGATGTTCTCGTGGTTTTCCTCAATGA
ACTTGATGCGCTCAGGGAACGGAACCTTATCGACACCCGCACAGTTTGCACCGTGG
ATTTTCAGCCAGTAGTAACCTTCCTTACCGATTGGTTTACCTTTCGCCAGCGTAAGCA
GTCCTTTGGTCATATCGTTACCTTGCGGGTTGAACATTGACACAGCGTAAACACGAC

CGCGCCAGTCCATGTTGTAAGGGAACCAGATGGCCTTATGGTTAGCAAACTTATTGG
CTTGCTCAAGCATGAACTCAAGGCTGATACGGCGAGACTTGCGAGCCTTGTCCTTGC
GGTACACAGCAGCGGCAGCACGTTTCCACGCGGTGAGAGCCTCAGGATTCATGTCG
ATGTCTTCCGGTTTCATCGGGAGTTCTTCACGCTCAATCGCAGGGATGTCCTCGACC
GGACAATGCTTCCACTTGGTGATTACGTTGGCGACCGCTAGGACTTTCTTGTTGATTT
TCCATGCGGTGTTTTGCGCAATGTTAATCGCTTTGTACACCTCAGGCATGTAAACGTC
TTCGTAGCGCATCAGTGCTTTCTTACTGTGAGTACGCACCAGCGCCAGAGGACGACG
ACCGTTAGCCCAATAGCCACCACCAGTAATGCCAGTCCACGGCTTAGGAGGAACTA
CGCAAGGTTGGAACATCGGAGAGATGCCAGCCAGCGCACCTGCACGGGTTGCGATA
GCCTCAGCGTATTCAGGTGCGAGTTCGATAGTCTCAGAGTCTTGACCTACTACGCCA
GCATTTTGGCGGTGTAAGCTAACCATTCCGGTTGACTCAATGAGCATCTCGATGCAG
CGTACTCCTACATGAATAGAGTCTTCCTTATGCCACGAAGACCACGCCTCGCCACCG
AGTAGACCCTTAGAGAGCATGTCAGCCTCGACAACTTGCATAAATGCTTTCTTGTAG
ACGTGCCCTACGCGCTTGTTGAGTTGTTCCTCAACGTTTTTCTTGAAGTGCTTAGCTT
CAAGGTCACGGATACGACCGAAGCGAGCCTCGTCCTCAATGGCCCGACCGATTGCG
CTTGCTACAGCCTGAACGGTTGTATTGTCAGCACTGGTTAGGCAAGCCAGAGTGGTC
TTAATGGTGATGTACGCTACGGCTTCCGGCTTGATTTCTTGCAGGAACTGGAAGGCT
GTCGGGCGCTTGCCGCGCTTAGCTTTCACTTCCTCAAACCAGTCGTTGATGCGTGCA
ATCATCTTAGGGAGTAGGGTAGTGATGAGAGGCTTGGCGGCAGCGTTATCCGCAAC
CTCACCAGCTTTAAGTTGACGCTCAAACATCTTGCGGAAGCGTGCTTCACCCATCTC
GTAAGACTCATGCTCAAGGGCCAACTGTTCGCGAGCTAAACGCTCACCGTAATGGTC
AGCCAGAGTGTTGAACGGGATAGCAGCCAGTTCGATGTCAGAGAAGTCGTTCTTAG
CAATGTTAATGGTATTCTAGTGCACGGTAATCATGGTCATGGTTAATTCCTCCTGTTA
GCCCAAAAAAACGGGTATGGAGAAACAGTAGAGAGTTGCGATAAAAAGCGTCAGG
TAGGATCCGCTAATCTTATGGATAAAAATGCTATGGCATAGCAAAGTGTGACGCCGT
GCAAATAATCAATGTGGACTTTTCTGCCGTGATTATAGACACTTTTGTTACGCGTTTT
TGTCATGGCTTTGGTCCCGCTTTGTTACAGAATGCTTTTAATAAGCGGGGTTACCGGT
TTGGTTAGCGAGAAGAGCCAGTAAAAGACGCAGTGACGGCAATGTCTGATGCAATA
TGGACAATTGGTTTCTTCTCTGAATGGCGGGAGTATGAAAAGTATGGCTGAAGCGCA
AAATGATCCCCTGCTGCCGGGATACTCGTTTAATGCCCATCTGGTGGCGGGTTTAAC
GCCGATTGAGGCCAACGGTTATCTCGATTTTTTTATCGACCGACCGCTGGGAATGAA
AGGTTATATTCTCAATCTCACCATTCGCGGTCAGGGGGTGGTGAAAAATCAGGGACG
AGAATTTGTTTGCCGACCGGGTGATATTTTGCTGTTCCCGCCAGGAGAGATTCATCA
CTACGGTCGTCATCCGGAGGCTCGCGAATGGTATCACCAGTGGGTTTACTTTCGTCC
GCGCGCCTACTGGCATGAATGGCTTAACTGGCCGTCAATATTTGCCAATACGGGGTT
CTTTCGCCCGGATGAAGCGCACCAGCCGCATTTCAGCGACCTGTTTGGGCAAATCAT
TAACGCCGGGCAAGGGGAAGGGCGCTATTCGGAGCTGCTGGCGATAAATCTGCTTG
AGCAATTGTTACTGCGGCGCATGGAAGCGATTAACGAGTCGCTCCATCCACCGATGG
ATAATCGGGTACGCGAGGCTTGTCAGTACATCAGCGATCACCTGGCAGACAGCAAT
TTTGATATCGCCAGCGTCGCACAGCATGTTTGCTTGTCGCCGTCGCGTCTGTCACATC
TTTTCCGCCAGCAGTTAGGGATTAGCGTCTTAAGCTGGCGCGAGGACCAACGTATCA
GCCAGGCGAAGCTGCTTTTGAGCACCACCCGGATGCCTATCGCCACCGTCGGTCGCA
ATGTTGGTTTTGACGATCAACTCTATTTCTCGCGGGTATTTAAAAAATGCACCGGGG
CCAGCCCGAGCGAGTTCCGTGCCGGTTGTGAAGAAAAAGTGAATGATGTAGCCGTC
AAGTTGTCATAATTGGTAACGAATCAGACAATTGACGGCTTGACGGAGTAGCATAG
GGTTTGCAGAATCCCTGCTTCGTCCATTTGACAGGCACATTATGCATGCCGCTTCGCC

TTCGCGCGCGAATTGATCTGCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTC
TGACACATGCAGCTCCCGGAGACGGTCACAGCCTGCAGCAAAAAACCCCTCAAGAC
CCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCCAA
CTCAGCTTCCTTTCGGGCTTTGTTATTTGTAGAGCTCATCCATGCCATGTGTAATCCC
AGCAGCAGTTACAAACTCAAGAAGGACCATGTGGTCACGCTTTTCGTTGGGATCTTT
CGAAAGGGCAGATTGTGTCGACAGGTAATGGTTGTCTGGTAAAAGGACAGGGCCAT
CGCCAATTGGAGTATTTTGTTGATAATGGTCTGCTAGTTGAACGGATCCATCTTCAAT
GTTGTGGCGAATTTTGAAGTTAGCTTTGATTCCATTCTTTTGTTTGTCTGCCGTGATG
TATACATTGTGTGAGTTATAGTTGTACTCGAGTTTGTGTCCGAGAATGTTTCCATCTT
CTTTAAAATCAATACCTTTTAACTCGATACGATTAACAAGGGTATCACCTTCAAACT
TGACTTCAGCACGCGTCTTGTAGTTCCCGTCATCTTTGAAAGATATAGTGCGTTCCTG
TACATAACCTTCGGGCATGGCACTCTTGAAAAAGTCATGCCGTTTCATATGATCCGG
ATAACGGGAAAAGCATTGAACACCATAAGAGAAAGTAGTGACAAGTGTTGGCCATG
GAACAGGTAGTTTTCCAGTAGTGCAAATAAATTTAAGGGTAAGTTTTCCGTATGTTG
CATCACCTTCACCCTCTCCACTGACAGAAAATTTGTGCCCATTAACATCACCATCTA
ATTCAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTACTCATATGTA
TATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGAAACCGTTGTGGTCTCCC
TATAGTGAGTCGTATTAATTTCGCGGGATCG

pTrcHisA-sfGFP (V150tag)



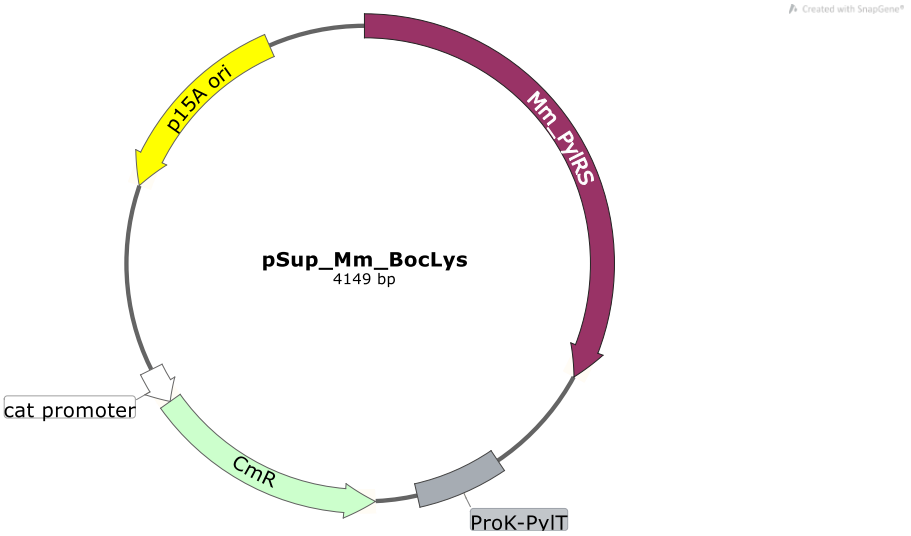
Highlighted in red is the ORF of sfGFP with an upstream fusion peptide. The coding region of sfGFP starts from the shaded “ATG”. Highlighted in blue is the V150tag mutation.

GTTTGACAGCTTATCATCGACTGCACTGGTGCACCAATGCTTCTGGCGTCAGGCAGC
CATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGC
TCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTG
GCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCCGGCTCGTATAATGTGTGG
AATTGTGAGCGGATAACAATTTCACACAGGAAACAGCGCCGCTGAGAAAAAGCGAA
GCGGCACTGCTCTTTAACAATTTATCAGACAATCTGTGTGGGCACTCGACCGGAATT

ATCGATTAACTTTATTATTAAAAATTAAAGAGGTATATATTAATGTATCGATTAAAT
AAGGAGGAATAAACCATGGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGC
ATGATGAGCAAAGGCGAAGAACTGTTTACCGGCGTGGTTCCGATTCTGGTGGAACT
GGATGGCGATGTAAATGGCCACAAGTTTAGCGTGCGTGGGGAAGGGGAGGGCGATG
CGACCAATGGCAAACTGACCCTGAAGTTTATTTGCACGACCGGGAAACTGCCGGTTC
CTTGGCCCACCCTTGTCACCACCCTGACGTATGGCGTGCAATGCTTTAGCCGTTACC
CGGACCACATGAAGCGGCATGACTTCTTCAAAAGCGCCATGCCTGAAGGCTATGTTC
AGGAACGGACGATCTCGTTTAAGGATGACGGCACCTATAAGACCCGTGCGGAGGTC
AAATTCGAAGGCGATACCCTGGTGAACCGCATTGAGCTGAAGGGCATCGACTTCAA
AGAGGATGGCAACATACTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAACta
cTACATCACCGCCGACAAGCAGAAGAACGGCATTAAGGCCAACTTCAAGATTCGGC
ACAATGTGGAGGACGGAAGCGTTCAGCTGGCGGATCATTATCAACAGAATACCCCC
ATTGGCGACGGTCCCGTGCTTCTGCCGGATAATCATTACTTGAGCACCCAGAGCGTG
CTGAGCAAGGACCCGAATGAGAAACGGGATCACATGGTGCTGCTGGAATTTGTGAC
CGCTGCGGGCATTACACATGGCATGGATGAACTGTATAAATAAAAGCTTGGCTGTTT
TGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCG
GTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCC
ATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCA
TGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGA
CTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAAT
CCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAG
GACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGAT
GGCCTTTTTGCGTTTCTACAAACTCTTTTTGTTTATTTTTCTAAATACATTCAAATATG
TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAA
GAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGC
CTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAG
TTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAG
AGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTG
GCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACT
ATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATG
GCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCG
GCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCAC
AACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGC
CATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGC
GCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACT
GGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCT
GGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAG
CACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTC
AGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATT
AAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAA
CTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCA
AAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCA
AAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCC
GAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCC
GTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCT

AATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGA
CTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGT
GCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGT
GAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGT
AAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC
TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGT
GATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTA
CGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGA
TTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCG
AACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGG
TATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTA
CAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGA
CTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGG
CTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCA
TGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCAGATCAATTCGC
GCGCGAAGGCGAAGCGGCATGCATTTACGTTGACACCATCGAATGGTGCAAAACCT
TTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTG
AAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTT
TCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGA
AGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGG
GCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGT
CGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGG
TGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTT
CTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCC
ATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACC
AGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGG
AGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTT
CTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAA
TTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAA
ACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGAT
CAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGC
GGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCC
GTCAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTT
GCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACT
GGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGT
TGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAG
TGAGCGCAACGCAATTAATGTGAGTTAGCGCGAATTGATCTG

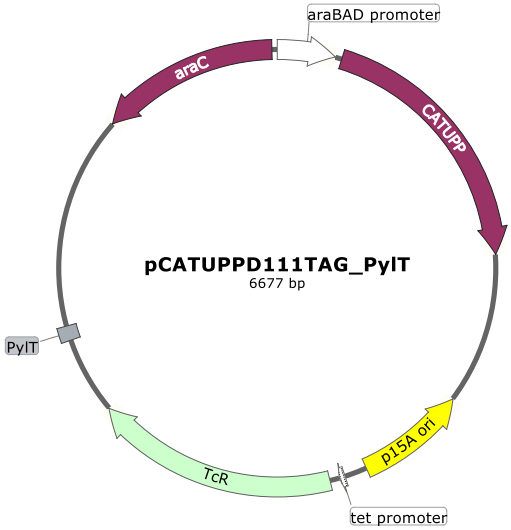
pSup_Mm_BocLys



ATGGATAAAAAACCACTAAACACTCTGATATCTGCAACCGGGCTCTGGATGTCCAG
GACCGGAACAATTCATAAAATAAAACACCACGAAGTCTCTCGAAGCAAAATCTATA
TTGAAATGGCATGCGGAGACCACCTTGTTGTAAACAACTCCAGGAGCAGCAGGACT
GCAAGAGCGCTCAGGCACCACAAATACAGGAAGACCTGCAAACGCTGCAGGGTTTC
GGATGAGGATCTCAATAAGTTCCTCACAAAGGCAAACGAAGACCAGACAAGCGTAA
AAGTCAAGGTCGTTTCTGCCCCTACCAGAACGAAAAAGGCAATGCCAAAATCCGTT
GCGAGAGCCCCGAAACCTCTTGAGAATACAGAAGCGGCACAGGCTCAACCTTCTGG
ATCTAAATTTTCACCTGCGATACCGGTTTCCACCCAAGAGTCAGTTTCTGTCCCGGCA
TCTGTTTCAACATCAATATCAAGCATTTCTACAGGAGCAACTGCATCCGCACTGGTA
AAAGGGAATACGAACCCCATTACATCCATGTCTGCCCCTGTTCAGGCAAGTGCCCCC
GCACTTACGAAGAGCCAGACTGACAGGCTTGAAGTCCTGTTAAACCCAAAAGATGA
GATTTCCCTGAATTCCGGCAAGCCTTTCAGGGAGCTTGAGTCCGAATTGCTCTCTCG
CAGAAAAAAAGACCTGCAGCAGATCTACGCGGAAGAAAGGGAGAATTATCTGGGG
AAACTCGAGCGTGAAATTACCAGGTTCTTTGTGGACAGGGGTTTTCTGGAAATAAAA
TCCCCGATCCTGATCCCTCTTGAGTATATCGAAAGGATGGGCATTGATAATGATACC
GAACTTTCAAAACAGATCTTCAGGGTTGACAAGAACTTCTGCCTGAGACCCATGCTT
GCTCCAAACCTTTACAACTACCTGCGCAAGCTTGACAGGGCCCTGCCTGATCCAATA
AAAATTTTTGAAATAGGCCCATGCTACAGAAAAGAGTCCGACGGCAAAGAACACCT
CGAAGAGTTTACCATGCTGAACTTCTGCCAGATGGGATCGGGATGCACACGGGAAA
ATCTTGAAAGCATAATTACGGACTTCCTGAACCACCTGGGAATTGATTTCAAGATCG
TAGGCGATTCCTGCATGGTCTTTGGGGATACCCTTGATGTAATGCACGGAGACCTGG
AACTTTCCTCTGCAGTAGTCGGACCCATACCGCTTGACCGGGAATGGGGTATTGATA
AACCCTGGATAGGGGCAGGTTTCGGGCTCGAACGCCTTCTAAAGGTTAAACACGAC
TTTAAAAATATCAAGAGAGCTGCAAGGTCCGAGTCTTACTATAACGGGATTTCTACC
AACCTGTAAATGCAGTTTCAAACGCTAAATTGCCTGATGCGCTACGCTTATCAGGCC
TACATGATCTCTGCAATATATTGAGTTTGCGTGCTTTTGTAGGCCGGATAAGGCGTTC
ACGCCGCATCCGGCAAGAAACAGCAAACAATCCAAAACGCCGCGTTCAGCGGCGTT
TTTTCTGCTTTTCTTCGCGAATTAATTCCGCTTCGCAACATGTGAGCACCGGTTTATT
GACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCA
GGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTCAAAAGCGGCCGCAAA
ACTAGTGGCAGCGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCAT

TTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTG
GGAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCC
CGGGGTTTCCGCCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGGTCGACA
AACTCGAGCAGCTCAGGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATT
ATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAA
AAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTG
CCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAG
CACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTT
GTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGA
GACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTA
ACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTC
ACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGT
GAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGAT
GAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTA
TTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG
TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATA
TATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGA
AAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAA
GTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGG
CTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGT
CACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTA
GTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTC
CTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGC
GCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGT
GCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATA
CAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCG
GCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGA
TACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCC
CCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACA
GGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTT
CCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCA
CGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAA
CCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAAC
CCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGG
AGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTG
ACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCT
TCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCA
GACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTT
CAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGT
TGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCGAGCTCCCGGTCA
TCAATCATCCCCATAATCCTTGTTAGATTATCAATTTTAAAAAACTAACAGTTGTCAG
CCTGTCCCGCTTTAATATCATACGCCGTTATACGTTGTTTACGCTTTGAGGAATCCCA

pCATUPP-PylT



AraC
PylT
TetR
p15a Ori
CATUPPD111TAG

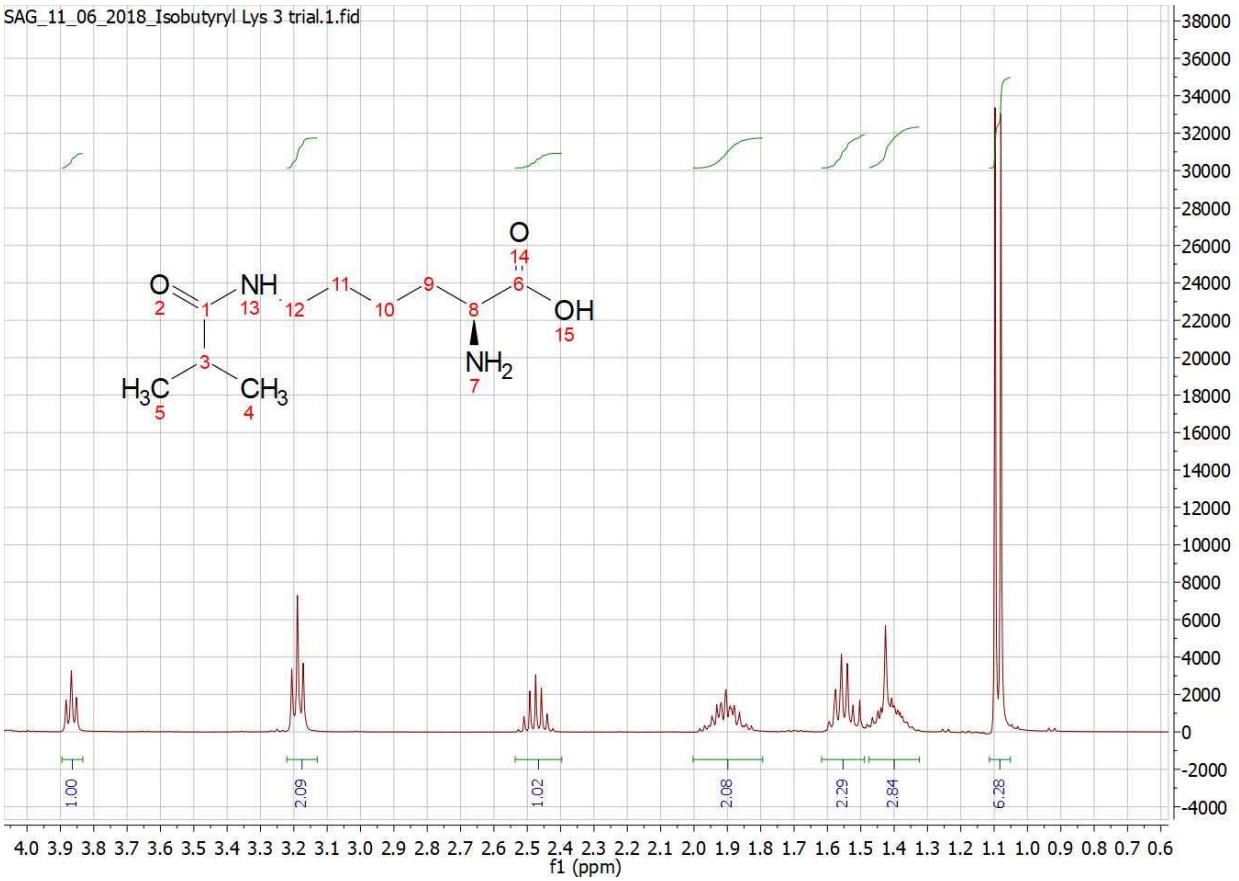
aagaaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactggctcttctcgctaaccaaaccggtaaccccgcttattaaaa
gcattctgtaacaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagtgtctataatcacggcagaaaagtccacattgattatttg
cacggcgtcacactttgctatgccatagcatttttatccataagattagcggatcctacctgacgcttttatcgcaactctctactgtttctccata
cccgtttttgggctaacaggaggaattaaccatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaac
attttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagc
acaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagctggtgata
tgggatagtgttcacccttgttacaccgttttccatgagcagactgaaacgttttcatcgctctggagtgaataccacTAGgatttccggcagt
ttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaat
ccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattacacgcaagg
cgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtac
tgcgatgagtggcagggcggggcggcaaagatcgtggaagtcaaacacccactcgtcaaacacaagctgggactgatgcgtgagcaag
atatcagcaccaagcgctttcgcgaactcgcttccgaagtgggtagcctgctgacttacgaagcgaccgccgacctcgaaacggaaaaag
taactatcgaaggctggaacggcccggtagaaatcgaccagatcaaaggtaagaaaattaccgttgtgccaattctgcgtgcgggtcttggt
atgatggacggtgtgctggaaaacgttccgagcgcgcgcatcagcgttgtcggtatgtaccgtaatgaagaaacgctggagccggtaccgt
acttccagaaactggtttctaacatcgatgagcgtatggcgctgatcgttgacccaatgctggcaaccggtggttccgttatcgcgaccatcg
acctgctgaaaaaagcgggctgcagcagcatcaaagttctggtgctggtagctgcgccagaaggtatcgctgcgctggaaaaagcgcac
ccggacgtcgaactgtataccgcatcgattgatcagggactgaacgagcacggatacattattccgggcctcggcgatgccggtgacaaa
atctttggtacgaaataatcgaccatatgggaattcgaagcttgggcccgaacaaaaactcatctcagaagaggatctgaatagcgccgtcg
accatcatcatcatcatcattgagtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcaga
acgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaac
gccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaaga
ctgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagctgtccctcctgttcagctactgacgg
ggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtg
aagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactc
gctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacag

ggaagtgagagggccgcggcaaagccgtttttccataggctccgccccccctgacaagcatcacgaaatctgacgctcaaatcagtggtgg
cgaaacccgacaggactataaagataccaggcgtttcccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcat
tccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaacccc
ccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccac
tggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagcca
gttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggttttttcgttttcagagcaagagattacgcg
cagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtc
agccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagctttaatgcggtagtttatcacagttaaattgctaacgcagtc
aggcaccgtgtatgaaatctaacaatgcgctcatcgtcatcctcggcaccgtcaccctggatgctgtaggcataggcttggttatgccggtac
tgccgggcctcttgcgggatatcgtccattccgacagcatcgccagtcactatggcgtgctgctagcgctatatgcgttgatgcaatttctatg
cgcacccgttctcggagcactgtccgaccgctttggccgccgcccagtcctgctcgcttcgctacttggagccactatcgactacgcgatca
tggcgaccacacccgtcctgtggatcctctacgccggacgcatcgtggccggcatcaccggcgccacaggtgcggttgctggcgcctata
tcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtgg
ccgggggactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttccta
atgcaggagtcgcataagggagagcgtcgaccgatgcccttgagagccttcaacccagtcagctccttccggtgggcgcggggcatgact
atcgtcgccgcacttatgactgtcttctttatcatgcaactcgtaggacaggtgccggcagcgctctgggtcattttcggcgaggaccgctttc
gctggagcgcgacgatgatcggcctgtcgcttgcggtattcggaatcttgcacgccctcgctcaagccttcgtcactggtcccgccaccaaa
cgtttcggcgagaagcaggccattatcgccggcatggcggccgacgcgctgggctacgtcttgctggcgttcgcgacgcgaggctggat
ggccttccccattatgattcttctcgcttccggcggcatcgggatgcccgcgttgcaggccatgctgtccaggcaggtagatgacgaccatc
agggacagcttcaaggatcgctcgcggctcttaccagcctaacttcgatcactggaccgctgatcgtcacggcgatttatgccgcctcggcg
agcacatggaacgggttggcatggattgtaggcgccgccctataccttgtctgcctccccgcgttgcgtcgcggtgcatggagccgggcca
cctcgacctgaatggaagccggcggcacctcgctaacggattcaccactccaagaattggagccaatcaattcttgcggagaactgtgaat
gcgcaaaccaacccttggcagaacatatccatcgcgtccgccatctccagcagccgcacgcggcgcatctcgggctccttgcatgcacca
ttccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtctggcgaaag
ggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttaaaaa
aaatccttagctttcgctaagatctgcagtggcggaaaccccgggaatctaacccggctgaacggatttagagtccattcgatctacatgatc
aggttcccgcggccgcgaattcagcgttacaagtattacacaaagtttttatgttgagaatatttttttgatggggcgccacttatttttgatcgtt
cgctcaaagaagcggcgccagggttgtttttctttttcaccagtgagacgggcaacagaacgccatgagcggcctcatttcttattctgagttac
aacagtccgcaccgctgccggtagctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgactgtcttctttatcatgcaact
cgtaggacaggtgccggcagcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgccctgcaccat
tatgttccggatctgcatcgcaggatgctgctggctaccctgtggaacacctacatctgtattaacgaagcgctaaccgtttttatcaggctctg
ggaggcagaataaatgatcatatcgtcaattattacctccacggggagagcctgagcaaactggcctcaggcatttgagaagcacacggtc
acactgcttccggtagtcaataaaccggtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgacgaccgggtcga
atttgctttcgaatttctgccattcatccgcttattatcacttattcaggcgtagcaccaggcgtttaagggcaccaataactgccttaaaaaaatt
acgccccgccctgccactcatcgcagtgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggct
tgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgc
gaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcataatgtgcctgtcaaatggacgaagcagggattctgcaaaccctat
gctactccgtcaagccgtcaattgtctgattcgttaccaattatgacaacttgacggctacatcattcacttttttcttcacaaccggcacggaact
cgctcgggctggccccggtgcatttttttaaatacccgcgagaaatagagttgatcgtcaaaaccaacattgcgaccgacggtggcgatagg
catccgggtggtgctcaaaagcagcttcgcctggctgatacgttggtcctcgcgccagcttaagacgctaatccctaactgctggcggaaaa
gatgtgacagacgcgacggcgacaagcaaacatgctgtgcgacgctggcgatatcaaaattgctgtctgccaggtgatcgctgatgtactg
acaagcctcgcgtacccgattatccatcggtggatggagcgactcgttaatcgcttccatgcgccgcagtaacaattgctcaagcagatttat
cgccagcagctccgaatagcgcccttcccccttgcccggcgttaatgatttgcccaaacaggtcgctgaaatgcggctggtgcgcttcatccg
ggcgaaagaaccccgtattggcaaatattgacggccagttaagccattcatgccagtaggcgcgcggacgaaagtaaacccactggtgat
accattcgcgagcctccggatgacgaccgtagtgatgaatctctcctggcgggaacagcaaaatatcacccggtcggcaaacaaattctcg
tccctgatttttcaccaccccctgaccgcgaatggtgagattgagaatataacctttcattcccagcggtcggtcgataaaaaaatcgagataa

ccgttggcctcaatcggcgttaaacccgccaccagatgggcattaaacgagtatcccggcagcaggggatcattttgcgcttcagccatact
tttcatactcccgccattcagag

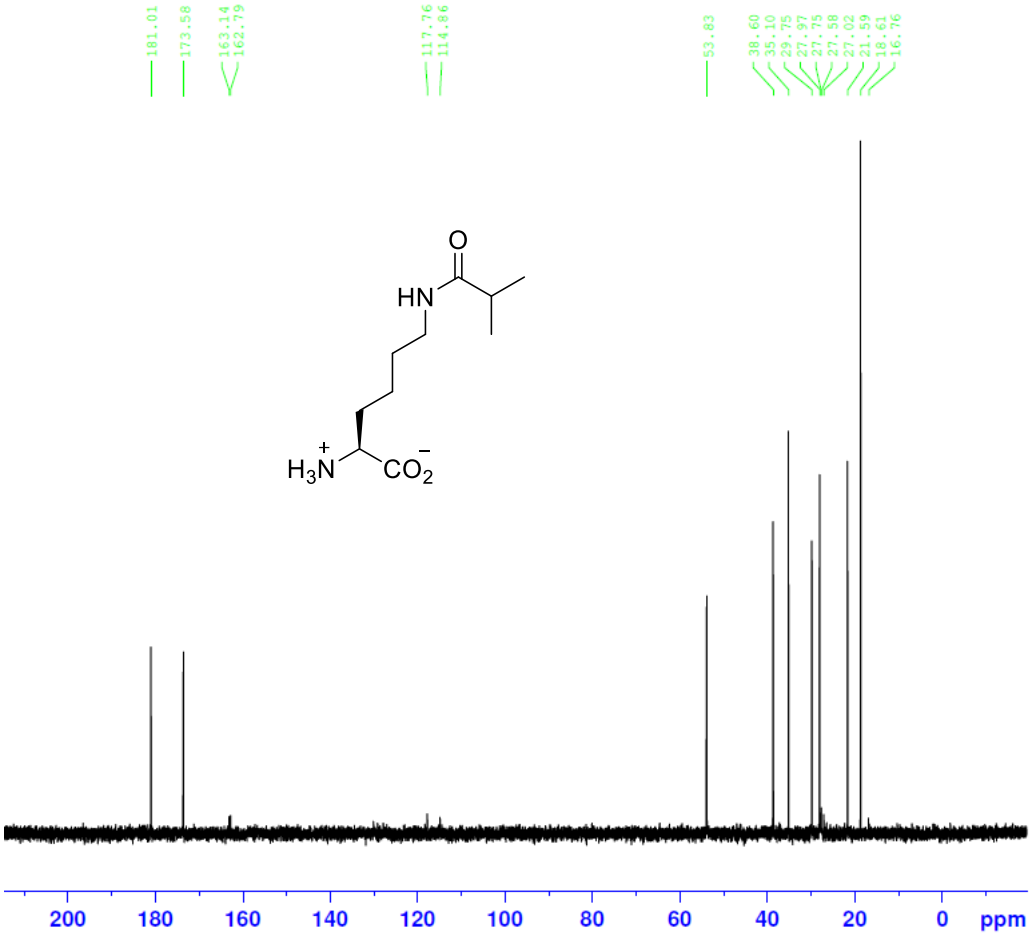
Appendix III. Unnatural Amino Acid Synthesis Characterization

III.I IBK Characterization



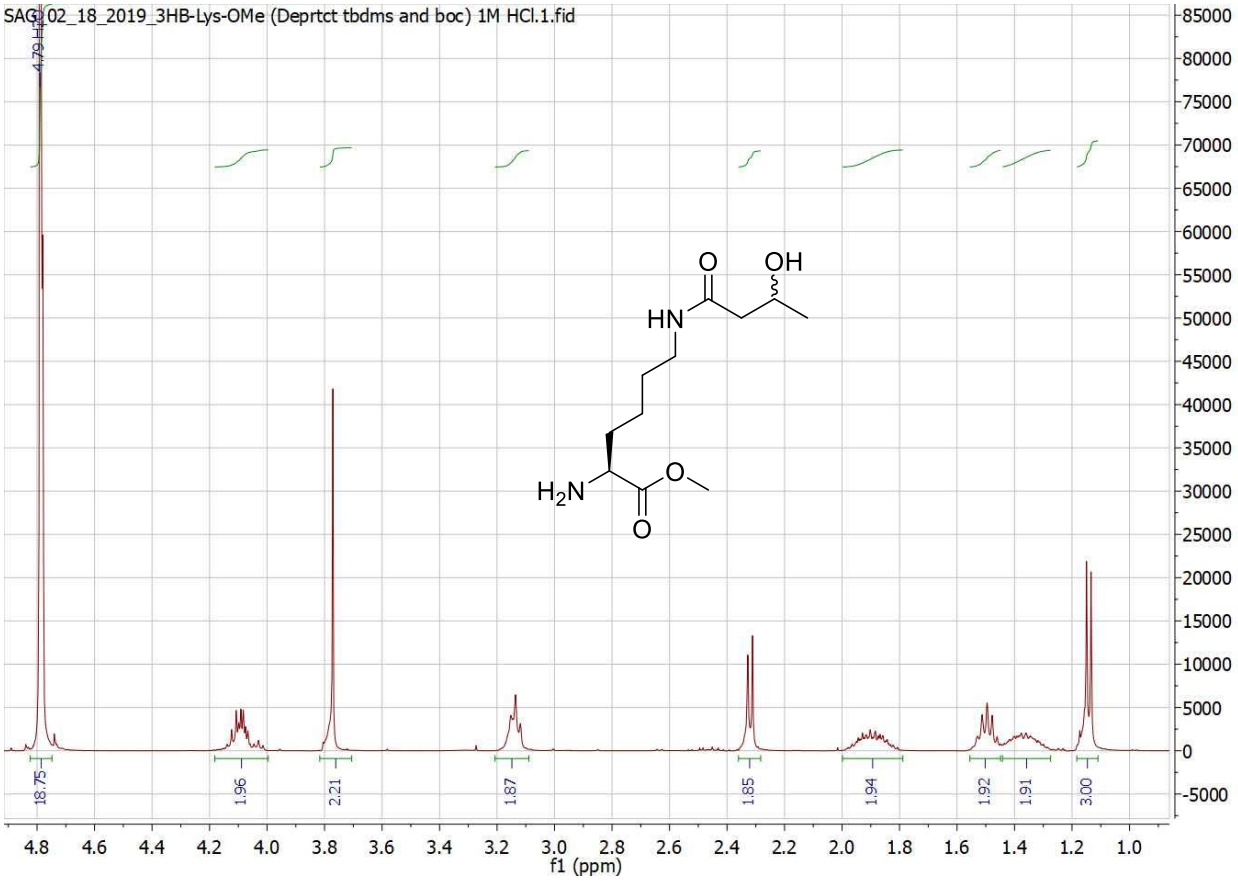
^{1}H NMR (400 MHz, D_2O) δ 1.09 (d, *J* = 6.9 Hz, 5H), 1.28 – 1.48 (m, 2H), 1.49 – 1.64 (m, 2H), 1.77 – 1.97 (m, 2H), 2.48 (p, *J* = 6.9 Hz, 1H), 3.19 (t, *J* = 6.9 Hz, 2H), 3.69 – 3.76 (m, 1H).

1

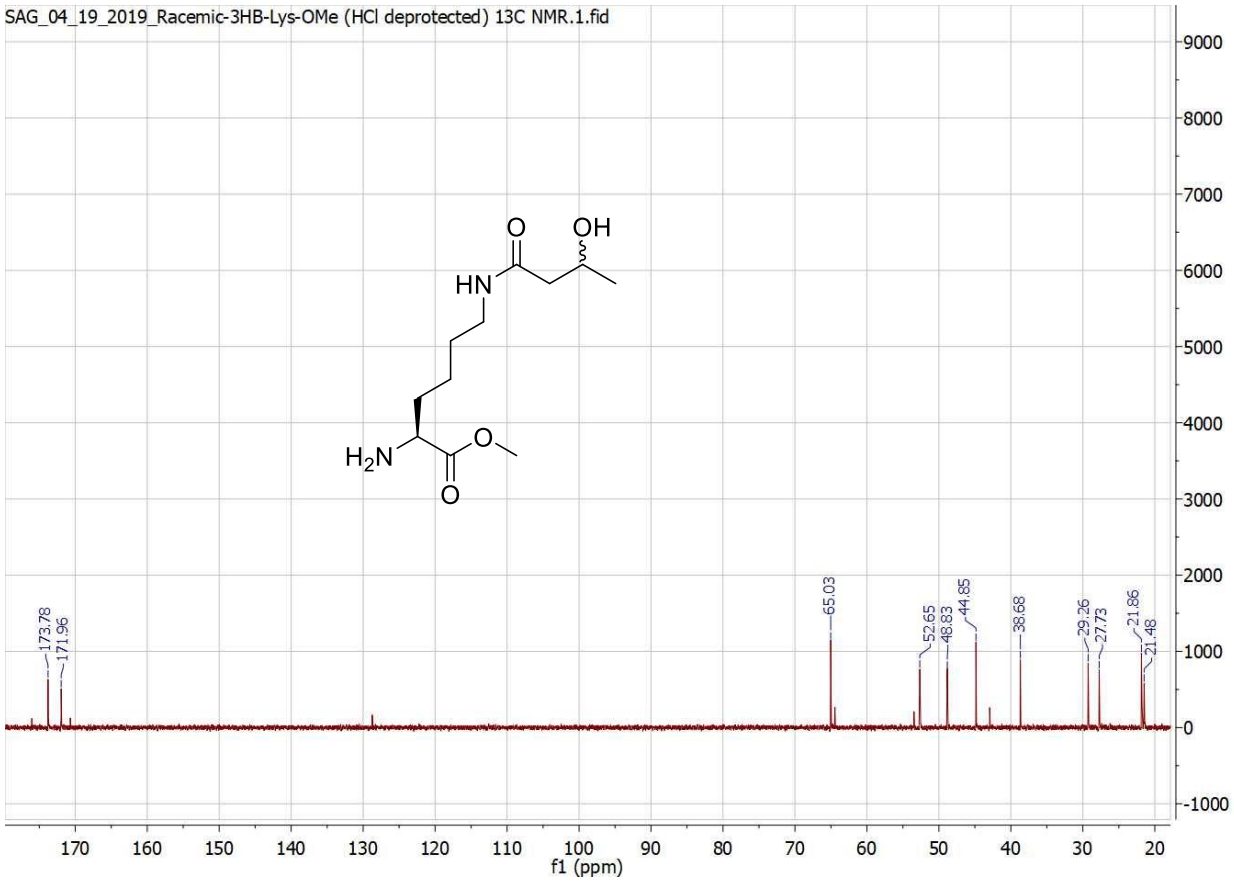


^{3}C NMR (101 MHz, D_2O) δ 180.99, 173.57, 53.82, 38.60, 35.10, 29.75, 27.97, 21.60, 18.61

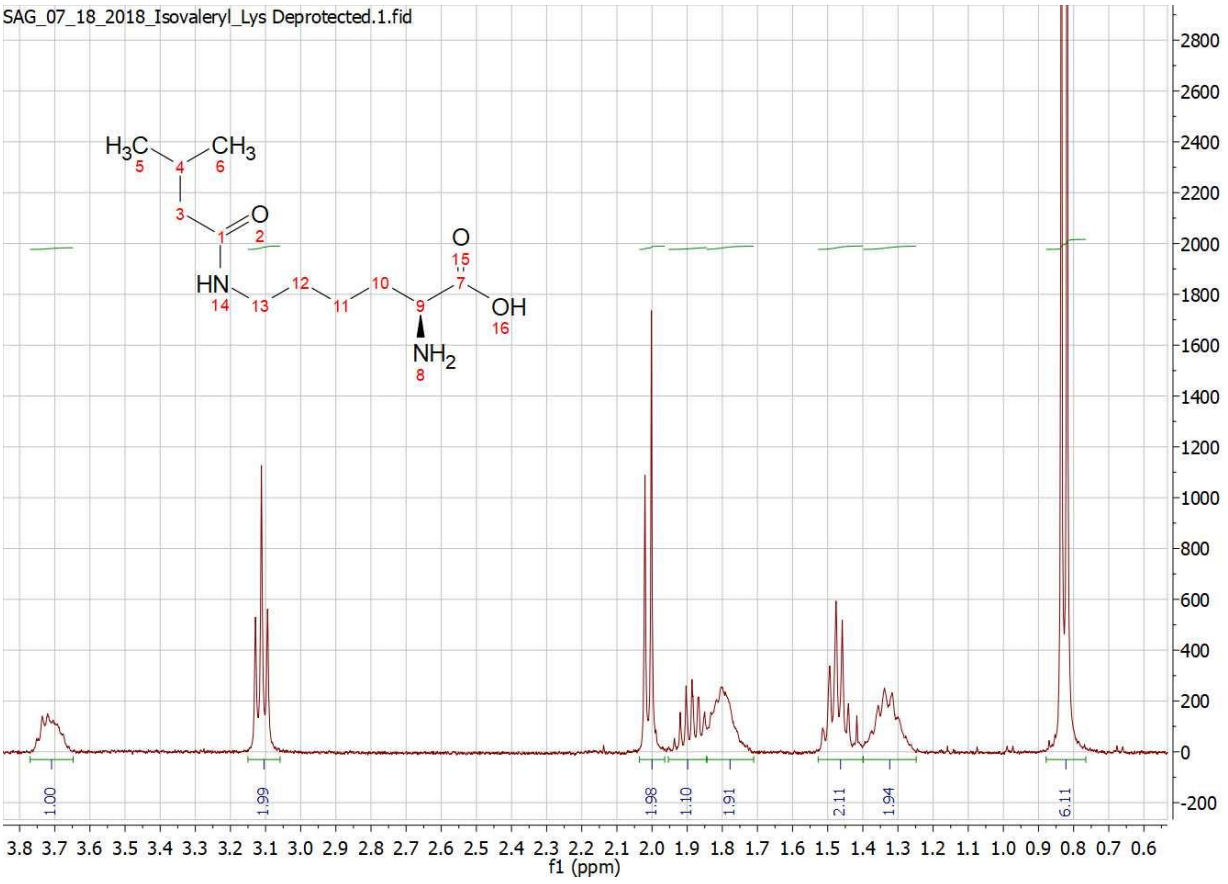
Appendix III.II HBK Characterization



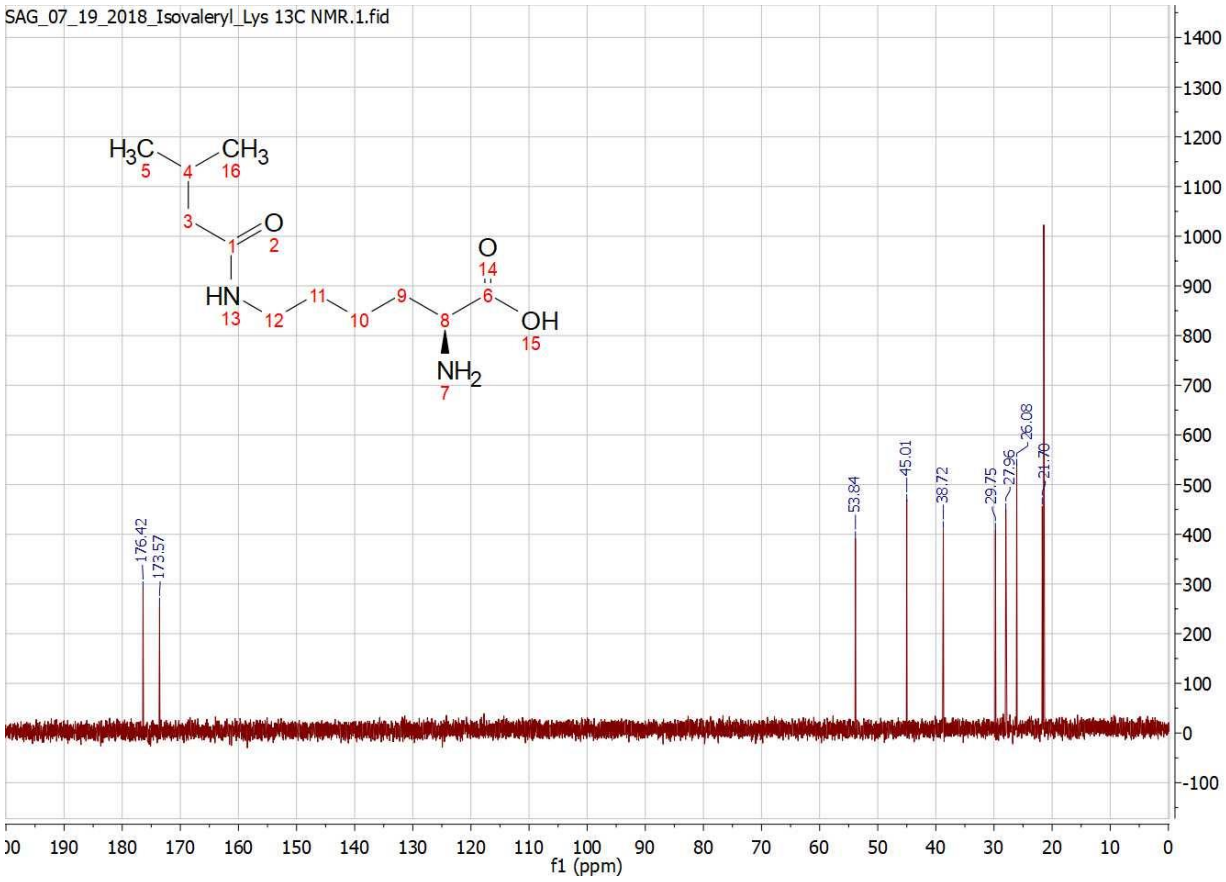
^{1}H NMR (400 MHz, Deuterium Oxide) δ 4.20 – 4.04 (m, 1H), 4.12 – 4.00 (m, 1H), 3.82 (s, 3H), 3.18 (t, J = 6.8 Hz, 2H), 2.36 (d, J = 6.4 Hz, 2H), 2.05 – 1.84 (m, 2H), 1.57 – 1.51 (m, 2H), 1.49 – 1.31 (m, 2H), 1.19 (d, J = 6.3 Hz, 3H).


SAG_04_19_2019_Racemic-3HB-Lys-OMe (HCl deprotected) 13C NMR.1.fid
O
OH
HN
H2N
O
O
173.78
171.96
65.03
52.65
48.83
44.85
38.68
29.26
27.73
21.86
21.48
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
f1 (ppm)

Appendix III.III IVK Characterization

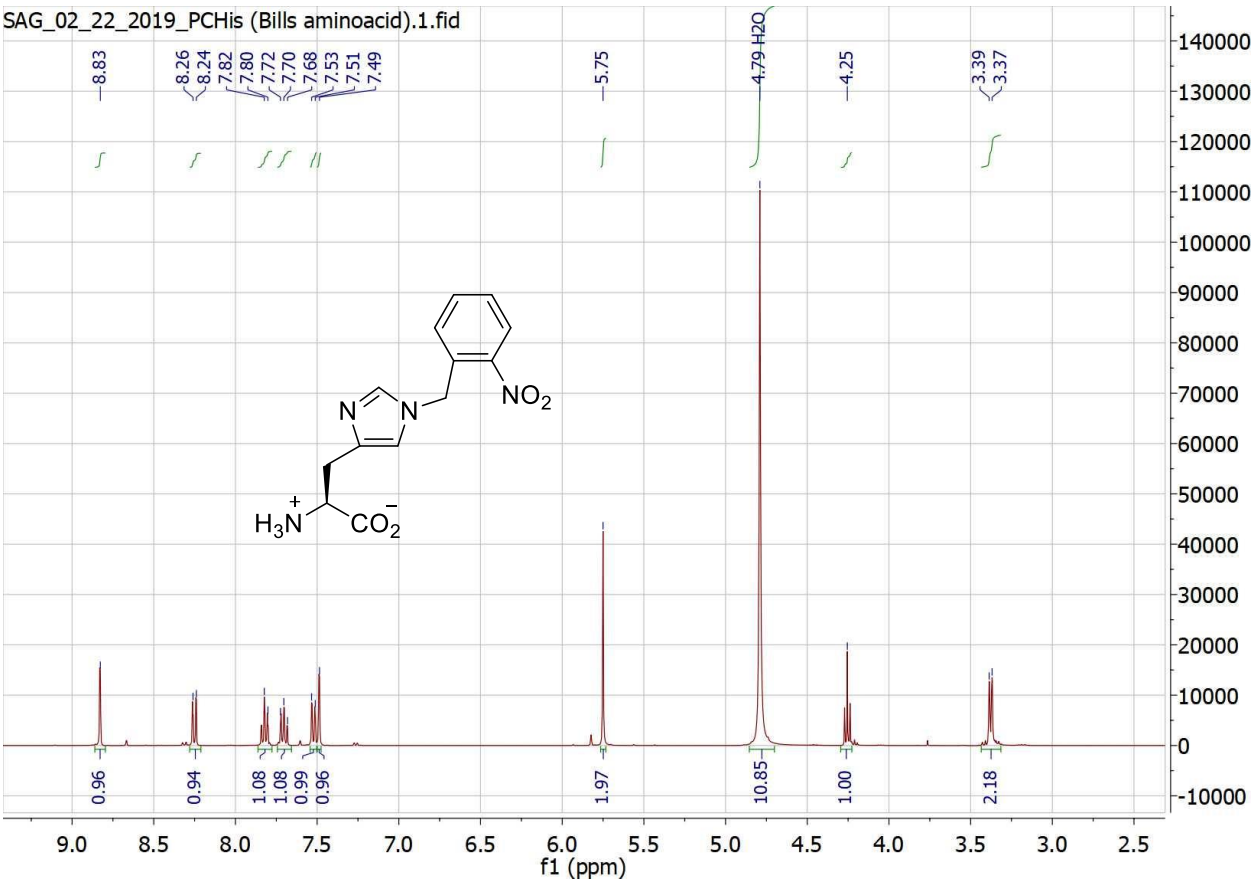


^{1}H NMR (400 MHz, Deuterium Oxide) δ 4.83 – 4.59 (m, 3H), 3.73 (d, *J* = 7.5 Hz, 1H), 3.11 (t, *J* = 6.9 Hz, 2H), 2.05 – 1.98 (m, 2H), 1.89 (dq, *J* = 13.9, 6.6 Hz, 1H), 1.80 (s, 2H), 1.54 – 1.39 (m, 2H), 1.35 (d, *J* = 7.4 Hz, 1H), 1.31 (s, 1H), 0.83 (d, *J* = 6.6 Hz, 5H).

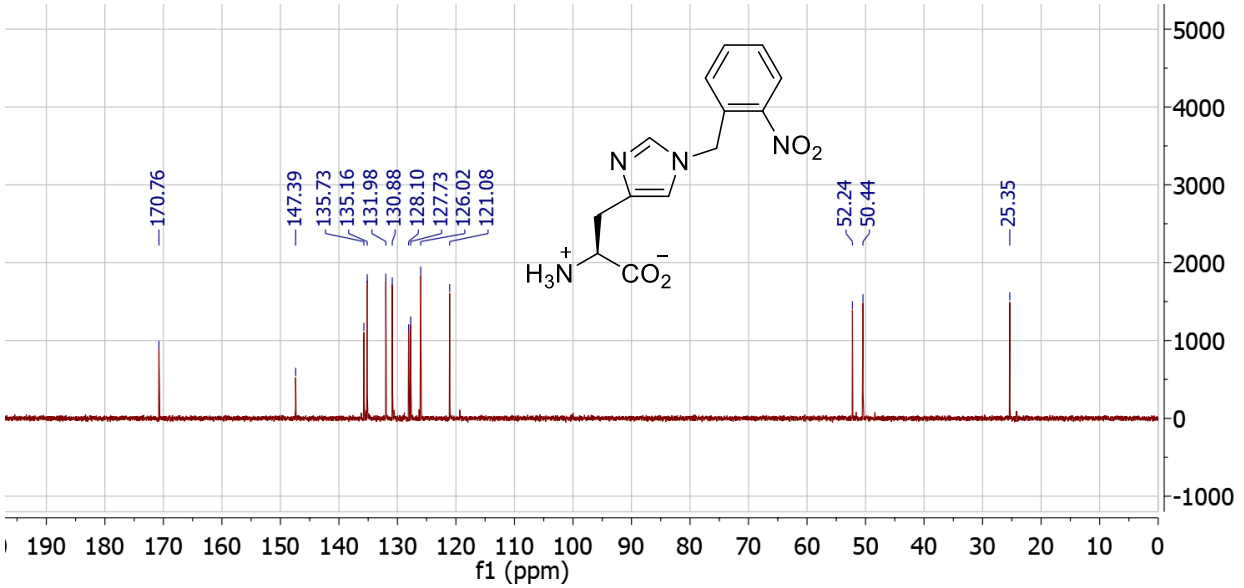


^{13}C NMR (101 MHz, D_2O) δ 176.42, 173.57, 53.84, 45.01, 38.72, 29.75, 27.96, 26.08, 21.70, 21.43

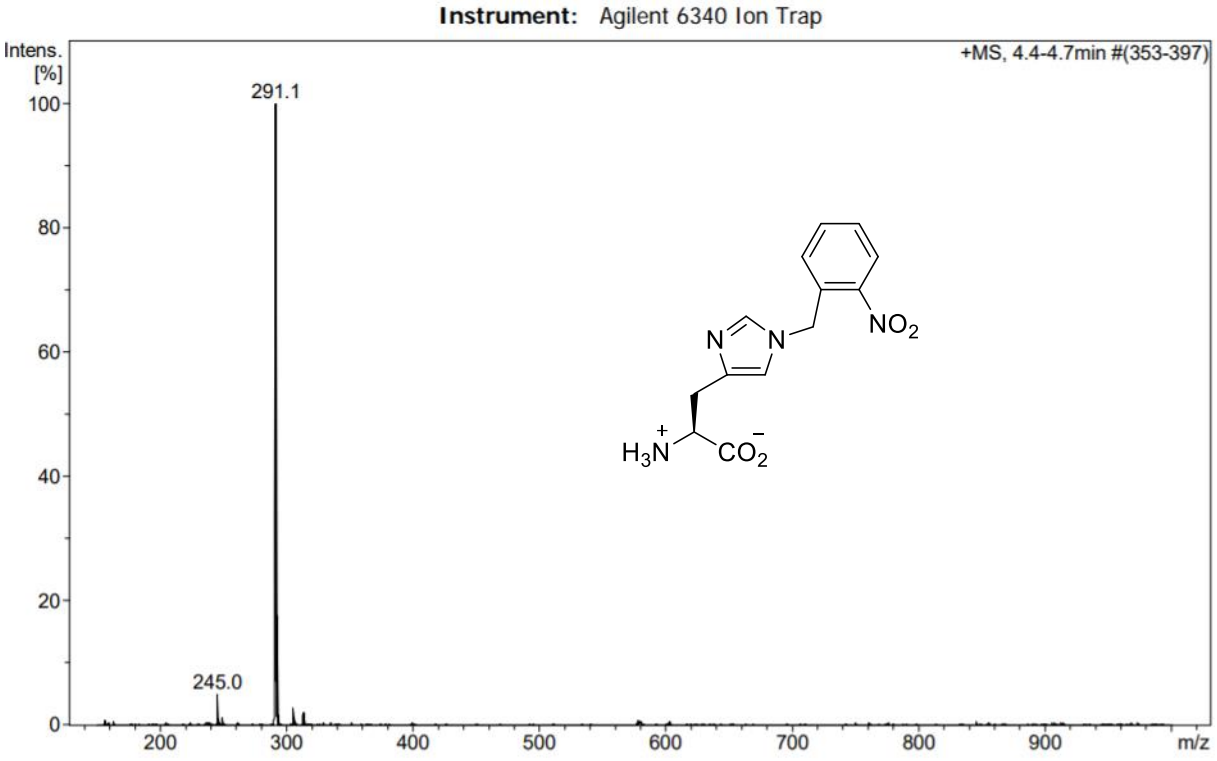
Appendix III.IV PCHis Characterization



1H NMR (400 MHz, D2O) δ = 3.38 (d, J = 7.2 Hz, 2 H), 4.25 (t, J = 6.8 Hz,1 H), 5.75 (s, 2 H), 7.49 (s, 1 H), 7.52 (d, J = 7.7 Hz, 1 H), 7.70 (t, J = 8.2 Hz, 1 H), 7.81 (t, J = 7.6 Hz, 1 H), 8.25 (d, J = 8.2 Hz, 1 H).



13C NMR (101 MHz, D2O) δ = 25.3, 50.4, 52.2, 121.1, 126.0, 127.7, 128.1, 130.9, 132.0, 135.2, 135.3, 147.4, 170.8.

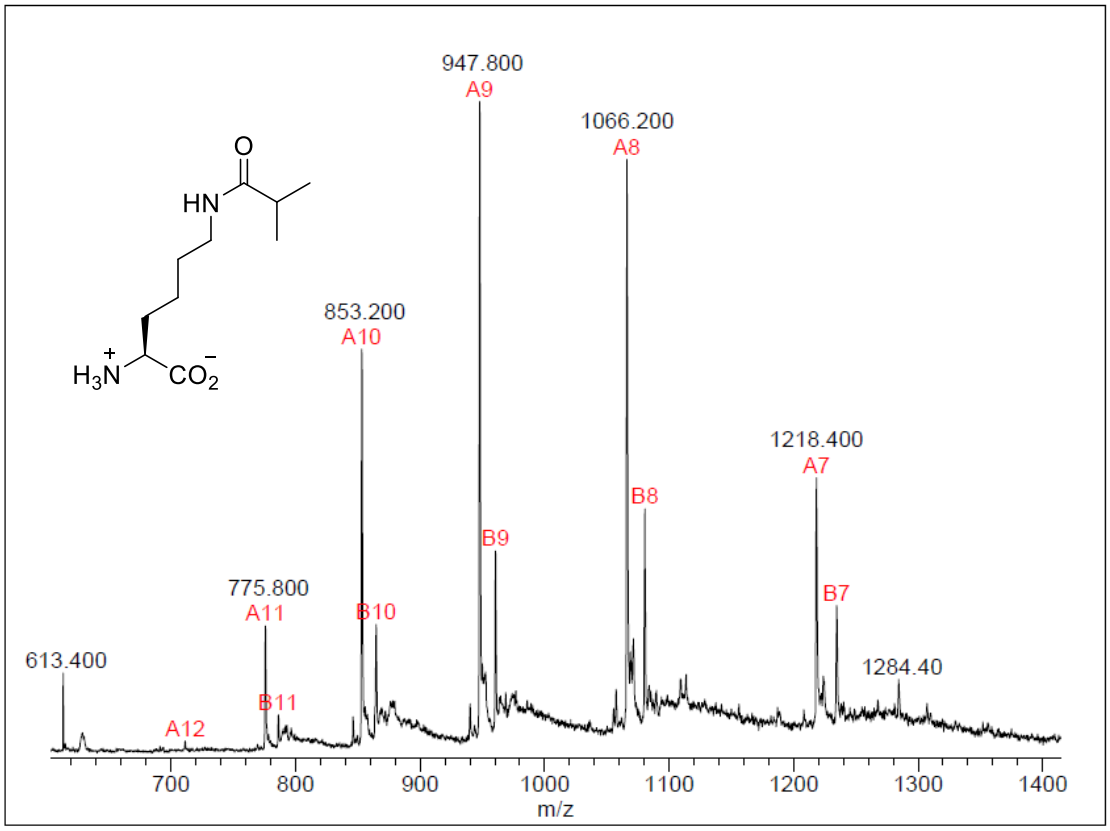


HRMS: M/Z+ calcd for $C_{13}H_{14}N_4O_4$ [M+H]+: 291.1088; found: 291.1.

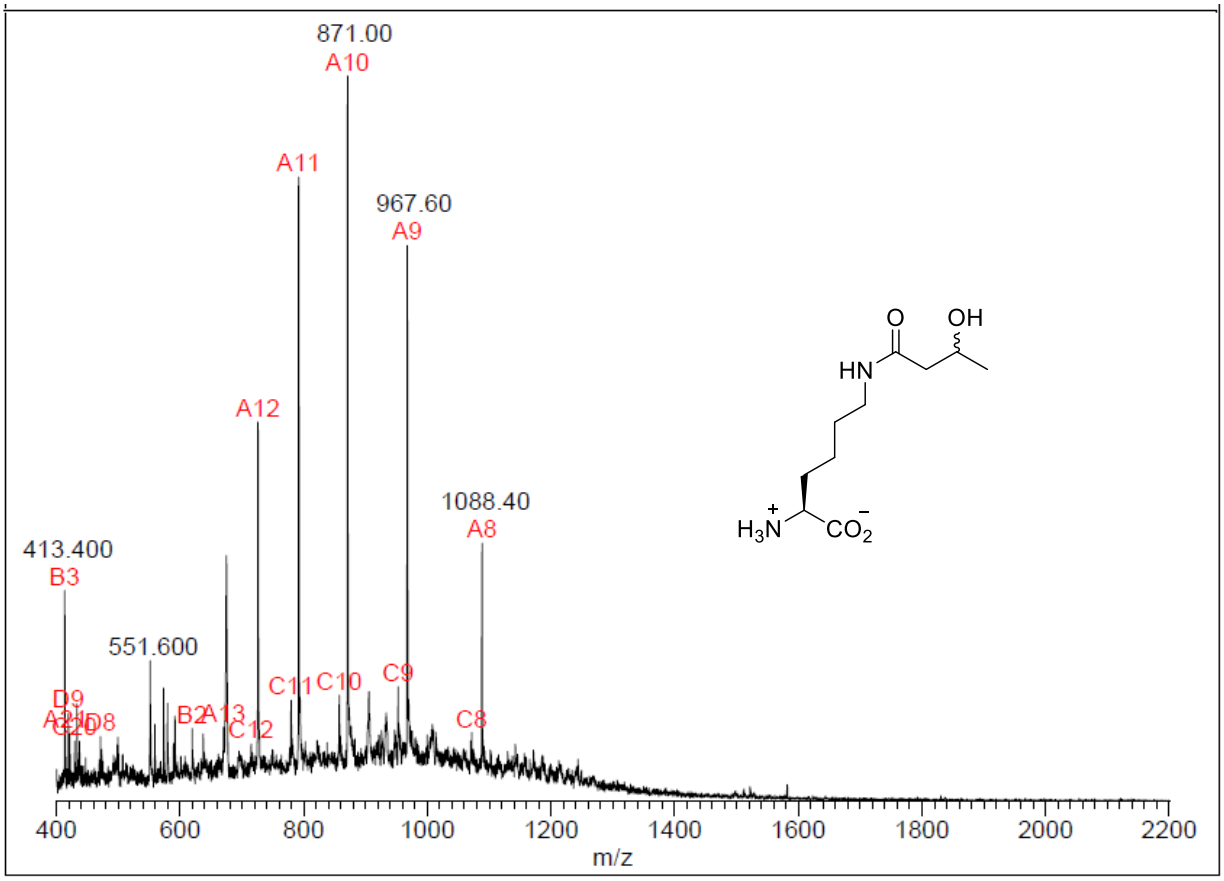
Appendix IV. Mass Spectra of UAA Incorporation into Ubiquitin

HPLC was run on an Agilent 1,100 using a Zorbax C18 150 mm x 0.3 mm column. MS was run on an Agilent 6340 ESI ion trap mass spectrometer.

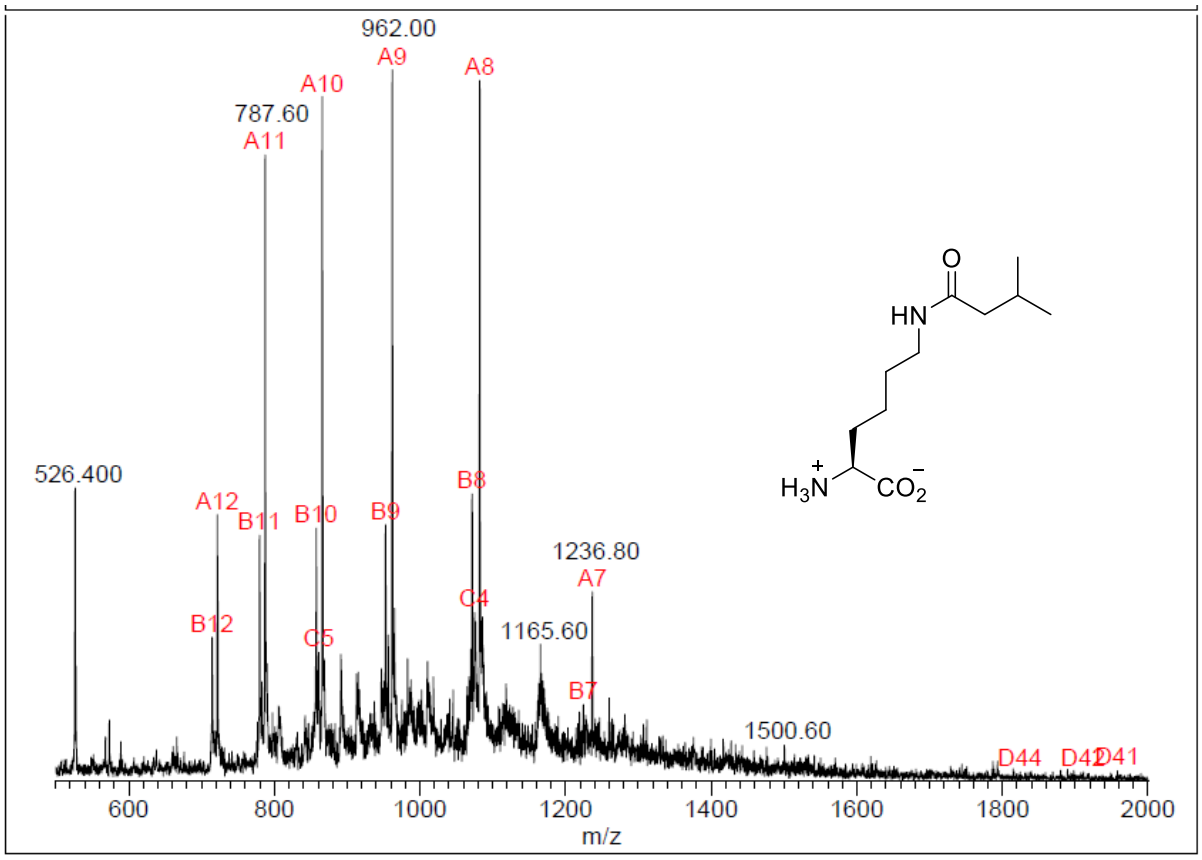
IV.I IBK



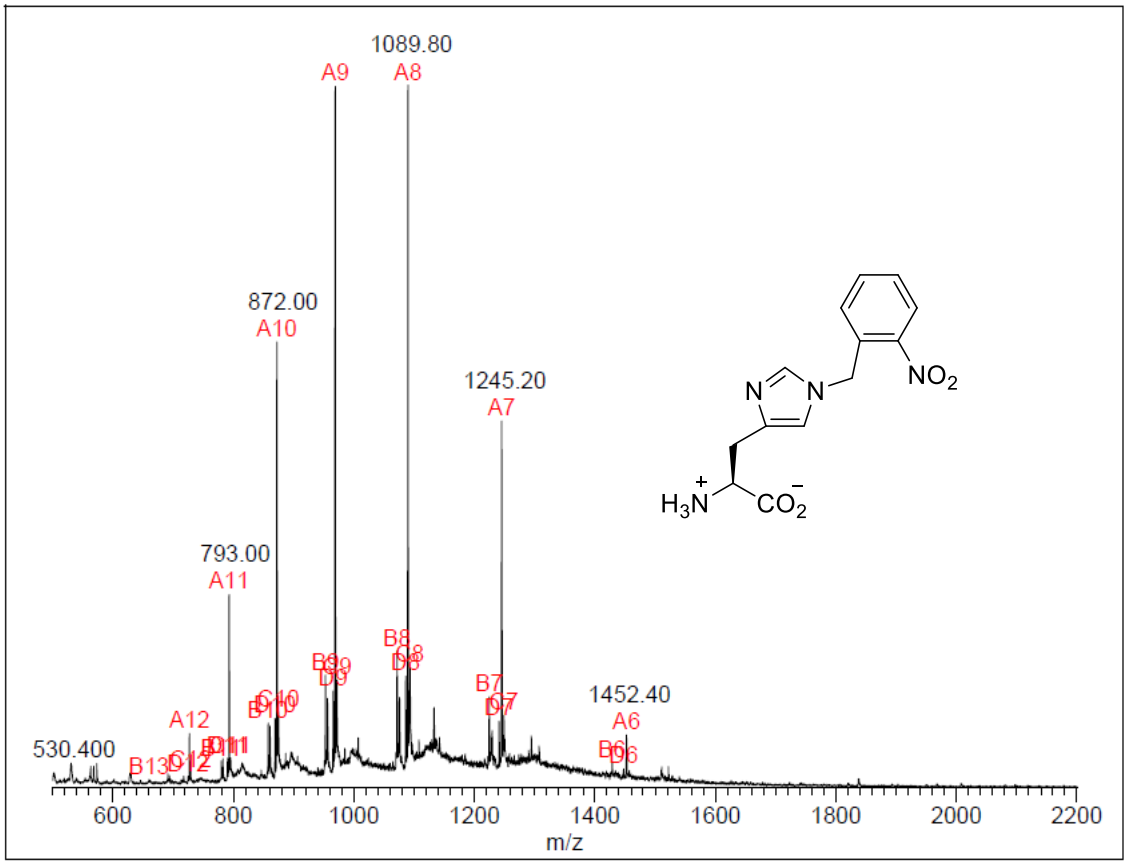
IV.II HBK


871.00
A10
A11
967.60
A9
A12
1088.40
A8
413.400
B3
551.600
C11
C10
C9
C8
B2
A13
C12
400
600
800
1000
1200
1400
1600
1800
2000
2200
m/z
O
OH
HN
H_3N^+
CO_2^-

IV.III IVK


962.00
A9
A8
A10
787.60
A11
526.400
A12
B11
B10
B9
B8
1236.80
A7
C4
B12
C5
1165.60
B7
1500.60
D44
D41
600
800
1000
1200
1400
1600
1800
2000
m/z
HN
O
H_3N^+
CO_2^-

IV.IV PCHis


1089.80
A9
A8
872.00
A10
1245.20
A7
793.00
A11
1452.40
A6
A12
530.400
B13
600
800
1000
1200
1400
1600
1800
2000
2200
m/z
N
N
NO2
H3N
CO2

Appendix V. PylRS Derived Mutants

PylRS Derived Mutants														
		Mutations											Genome	Additional unique mutations
		241	267	270	271	274	311	313	349	366	382	386	Barkeri (Mb)	
#	Synthetases	276	302	305	306	309	346	348	384	401	417	421	Mazei (Mm)	
1	WT	M	A	L	Y	L	N	C	Y	V	W	G	Mm	
2	BocLys	M	A	L	Y	L	N	C	F	V	W	G	Mm	
3	CbzLys	M	A	L	A	L	N	C	F	V	W	G	Mm	
4	OnbLys	M	A	L	M	A	N	A	F	V	W	G	Mm	
5	Interm	M	A	L	M	A	N	C	F	V	W	G	Mm	
6	WT+	M	A	L	Y	L	A	A	Y	V	W	G	Mm	
7	BocLys+	M	A	L	Y	L	A	A	F	V	W	G	Mm	
8	CbzLys+	M	A	L	A	L	A	A	F	V	W	G	Mm	
9	OnbLys+	M	A	L	M	A	A	A	F	V	W	G	Mm	
10	Interm+	M	A	L	M	A	A	A	F	V	W	G	Mm	
11	EV16-5	M	A	L	M	A	A	A	F	V	W	G	Mb	
12	EV17	M	A	L	M	A	N	A	F	V	W	G	Mb	
13	EV20	F	S	L	C	M	N	C	Y	V	W	G	Mb	
14	PCC1	M	A	L	Y	L	M	Q	Y	V	N	G	Mb	
15	PCC2	M	A	L	Y	L	Q	A	Y	M	W	G	Mb	
16	PylHRS	M	A	I	F	G	N	F	F	M	W	G	Mb	
17	ThzKRS	M	S	I	F	G	N	V	F	M	W	G	Mb	M315F, D344G
18	**IVKRS**	**M**	**A**	**L**	**Y**	**L**	**N**	**T**	**Y**	**V**	**W**	**G**	**Mb/Mm**	
19	**HBKRS**	**M**	**A**	**L**	**M**	**A**	**N**	**A**	**F**	**V**	**W**	**G**	**Mb**	
20	**PCHisRS**	**M**	**A**	**L**	**M**	**A**	**G**	**V**	**F**	**K**	**W**	**G**	**Mb**	

References

1. Verdin, E. & Ott, M. 50 years of protein acetylation: From gene regulation to epigenetics, metabolism and beyond. *Nature Reviews Molecular Cell Biology* **16,** 258–264 (2015).
2. Schaper, K., Etinski, M. & Fleig, T. Theoretical investigation of the excited states of 2-nitrobenzyl and 4,5-methylendioxy-2-nitrobenzyl caging groups. *Photochem. Photobiol.* **85,** 1075–1081 (2009).
3. Young, T. S. & Schultz, P. G. Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *Journal of Biological Chemistry* **285,** 11039–11044 (2010).
4. Burke, S. A., Lo, S. L. & Krzycki, J. A. Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. *J. Bacteriol.* (1998).
5. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20,** 1044–1048 (2002).
6. Wang, L. & Schultz, P. G. Expanding the genetic code. *Angewandte Chemie - International Edition* **44,** 34–66 (2004).
7. Ambrogelly, A., Palioura, S. & Söll, D. Natural expansion of the genetic code. *Nature Chemical Biology* **3,** 29–35 (2007).
8. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. *Science (80-.).* **296,** 1459–1462 (2002).
9. Zhao, M. *et al.* Desorption electrospray tandem MS (DESI-MSMS) analysis of methyl centralite and ethyl centralite as gunshot residues on skin and other surfaces. *Journal of Forensic Sciences* **53,** 807–811 (2008).
10. Nguyen, D. P. *et al.* Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* **136,** 2240–2243 (2014).
11. Li, F. *et al.* Expanding the genetic code for photoclick chemistry in E. coli, Mammalian Cells, and A. thaliana. *Angew. Chemie - Int. Ed.* **52,** 9700–9704 (2013).
12. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4,** 232–234 (2008).
13. Kim, C. H., Kang, M., Kim, H. J., Chatterjee, A. & Schultz, P. G. Site-specific incorporation of μ-N-crotonyllysine into histones. *Angew. Chemie - Int. Ed.* **51,** 7246–7249 (2012).
14. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).
15. Lang, K. *et al.* Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat. Chem.* **4,** 298–304 (2012).
16. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11020–4 (2002).
17. Hancock, S. M., Uprety, R., Deiters, A. & Chin, J. W. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J. Am. Chem. Soc.* **132,** 14819–14824 (2010).
18. Mukai, T. *et al.* Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **371,** 818–822 (2008).
19. Gautier, A. *et al.* Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132,** 4086–4088 (2010).

20. Bianco, A., Townsley, F. M., Greiss, S., Lang, K. & Chin, J. W. Expanding the genetic code of Drosophila melanogaster. *Nat. Chem. Biol.* **8,** 748–750 (2012).
21. Ai, H., Shen, W., Sagi, A., Chen, P. R. & Schultz, P. G. Probing Protein-Protein Interactions with a Genetically Encoded Photo-crosslinking Amino Acid. *ChemBioChem* **12,** 1854–1857 (2011).
22. Tippmann, E. M., Liu, W., Summerer, D., Mack, A. V. & Schultz, P. G. A Genetically Encoded Diazirine Photocrosslinker inEscherichia coli. *ChemBioChem* **8,** 2210–2214 (2007).
23. Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science (80-.).* **301,** 964–967 (2003).
24. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3,** 769–772 (2007).
25. Xie, J., Supekova, L. & Schultz, P. G. A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. *ACS Chem. Biol.* **2,** 474–478 (2007).
26. Chuh, K. N., Batt, A. R. & Pratt, M. R. Chemical Methods for Encoding and Decoding of Posttranslational Modifications. *Cell Chemical Biology* **23,** 86–107 (2016).
27. Ai, H. W., Lee, J. W. & Schultz, P. G. A method to site-specifically introduce methyllysine into proteins in E. coli. *Chem. Commun.* **46,** 5506–5508 (2010).
28. Wang, Y. S. *et al.* The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of l-phenylalanine and its derivatives. *Mol. Biosyst.* **7,** 714–717 (2011).
29. Davis, L. & Chin, J. W. Designer proteins: Applications of genetic code expansion in cell biology. *Nature Reviews Molecular Cell Biology* **13,** 168–182 (2012).
30. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126,** 14306–14307 (2004).
31. Gautier, A., Deiters, A. & Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J. Am. Chem. Soc.* **133,** 2124–2127 (2011).
32. Lang, K. & Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chem. Biol.* **9,** 16–20 (2014).
33. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149–2154 (1963).
34. Jackson, D. Y. *et al.* A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science (80-.).* **266,** 243–247 (1994).
35. Schnölzer, M. & Kent, S. B. H. Constructing proteins by dovetailing unprotected synthetic peptides: Backbone-engineered HIV protease. *Science (80-.).* **256,** 221–225 (1992).
36. CAMARERO, J. A., COTTON, G. J., ADEVA, A. & MUIR, T. W. Chemical ligation of unprotected peptides directly from a solid support. *J. Pept. Res.* **51,** 303–316 (2010).
37. Ishida, H., Qi, Z., Sokabe, M., Donowaki, K. & Inoue, Y. Molecular design and synthesis of artificial ion channels based on cyclic peptides containing unnatural amino acids. *J. Org. Chem.* **66,** 2978–2989 (2001).
38. Hackenberger, C. P. R. & Schwarzer, D. Chemoselective ligation and modification strategies for peptides and proteins. *Angewandte Chemie - International Edition* **47,** 10030–10074 (2008).
39. Sun, J., Chen, M., Xu, J. & Luo, J. Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. *J. Mol. Evol.* **61,** 437–444 (2005).

40. Hao, B. *et al.* A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science (80-.).* **296,** 1462–1466 (2002).
41. Burke, B., Lipman, R. S. A., Shiba, K., Musier-Forsyth, K. & Hou, Y. M. Divergent Adaptation of tRNA Recognition by Methanococcus jannaschii Prolyl-tRNA Synthetase. *J. Biol. Chem.* **276,** 20286–20291 (2001).
42. Ishii, R., Nureki, O., Yanagisawa, T., Fukunaga, R. & Yokoyama, S. Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina mazei. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62,** 1031–1033 (2006).
43. Harper, A. D., Bailey, C. B., Edwards, A. D., Detelich, J. F. & Keatinge-Clay, A. T. Preparative, in Vitro Biocatalysis of Triketide Lactone Chiral Building Blocks. *ChemBioChem* **13,** 2200–2203 (2012).
44. Piasecki, S. K. *et al.* Employing modular polyketide synthase ketoreductases as biocatalysts in the preparative chemoenzymatic syntheses of diketide chiral building blocks. *Chem. Biol.* **18,** 1331–1340 (2011).
45. Popp, B. V. & Ball, Z. T. Structure-selective modification of aromatic side chains with dirhodium metallopeptide catalysts. *J. Am. Chem. Soc.* **132,** 6660–6662 (2010).
46. Suzuki, T. *et al.* Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. *Nat. Chem. Biol.* **13,** 1261–1266 (2017).
47. Ambrogelly, A. *et al.* Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci.* **104,** 3141–3146 (2007).
48. Polycarpo, C. R. *et al.* Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580,** 6695–6700 (2006).
49. Yanagisawa, T. *et al.* Crystallographic Studies on Multiple Conformational States of Active-site Loops in Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* **378,** 634–652 (2008).
50. Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci.* **104,** 11268–11273 (2007).
51. Neumann, H., Slusarczyk, A. L. & Chin, J. W. De Novo generation of mutually orthogonal aminoacyl-TRNA synthetase/ tRNA pairs. *J. Am. Chem. Soc.* **132,** 2142–2144 (2010).
52. Kobayashi, T., Yanagisawa, T., Sakamoto, K. & Yokoyama, S. Recognition of Non-α-amino Substrates by Pyrrolysyl-tRNA Synthetase. *J. Mol. Biol.* **385,** 1352–1360 (2009).
53. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
54. L'Hernault, S. W. Chlamydomonas alpha-tubulin is posttranslationally modified in the flagella during flagellar assembly. *J. Cell Biol.* **97,** 258–263 (2004).
55. Wellen, K. E. *et al.* ATP-citrate lyase links cellular metabolism to histone acetylation. *Science (80-.).* **324,** 1076–1080 (2009).
56. Wang, X., Wei, X., Pang, Q. & Yi, F. Histone deacetylases and their inhibitors: molecular mechanisms and therapeutic implications in diabetes mellitus. *Acta Pharm. Sin. B* **2,** 387–395 (2012).
57. Kaypee, S. *et al.* Aberrant lysine acetylation in tumorigenesis: Implications in the development of therapeutics. *Pharmacology and Therapeutics* **162,** 98–119 (2016).
58. Guarente, L. Sirtuins, aging, and metabolism. *Cold Spring Harb. Symp. Quant. Biol.* **76,** 81–90 (2011).
59. Selvi, B. R. & Kundu, T. K. Reversible acetylation of chromatin: Implication in regulation

of gene expression, disease and therapeutics. *Biotechnology Journal* **4,** 375–390 (2009).
60. ALLFREY, V. G., FAULKNER, R. & MIRSKY, A. E. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **51,** 786–94 (1964).
61. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science (80-.).* **311,** 844–847 (2006).
62. Neumann, H. *et al.* A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Mol. Cell* **36,** 153–163 (2009).
63. Kuo, M. H. & Allis, C. D. Roles of histone acetyltransferases and deacetylases in gene regulation. *BioEssays* **20,** 615–626 (1998).
64. Tessarz, P. & Kouzarides, T. Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell Biol.* **15,** 703–708 (2014).
65. Xu, F., Zhang, Q., Zhang, K., Xie, W. & Grunstein, M. Sir2 Deacetylates Histone H3 Lysine 56 to Regulate Telomeric Heterochromatin Structure in Yeast. *Mol. Cell* **27,** 890–900 (2007).
66. Shahbazian, M. D. & Grunstein, M. Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu. Rev. Biochem.* **76,** 75–100 (2007).
67. Smith, S. & Stillman, B. Stepwise assembly of chromatin during DNA replication in vitro. *EMBO J.* **10,** 971–80 (1991).
68. Du, J. *et al.* Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase. *Science (80-.).* **334,** 806–809 (2011).
69. Zhang, Z. *et al.* Identification of lysine succinylation as a new post-translational modification. *Nat. Chem. Biol.* **7,** 58–63 (2011).
70. Frizzell, N., Lima, M. & Baynes, J. W. Succination of proteins in diabetes. *Free Radical Research* **45,** 101–109 (2011).
71. Kim, S. C. *et al.* Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* **23,** 607–18 (2006).
72. Garrity, J., Gardner, J. G., Hawse, W., Wolberger, C. & Escalante-Semerena, J. C. N-lysine propionylation controls the activity of propionyl-CoA synthetase. *J. Biol. Chem.* **282,** 30239–30245 (2007).
73. Fritz, K. S., Green, M. F., Petersen, D. R. & Hirschey, M. D. Ethanol Metabolism Modifies Hepatic Protein Acylation in Mice. *PLoS One* **8,** (2013).
74. Chen, Y. *et al.* Lysine Propionylation and Butyrylation Are Novel Post-translational Modifications in Histones. *Mol. Cell. Proteomics* **6,** 812–819 (2007).
75. Peng, C. *et al.* The First Identification of Lysine Malonylation Substrates and Its Regulatory Enzyme. *Mol. Cell. Proteomics* **10,** M111.012658 (2011).
76. Tan, M. *et al.* Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab.* **19,** 605–617 (2014).
77. Jiang, H. *et al.* SIRT6 regulates TNF-α secretion through hydrolysis of long-chain fatty acyl lysine. *Nature* **496,** 110–113 (2013).
78. Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146,** 1016–1028 (2011).
79. Baeza, J., Smallegan, M. J. & Denu, J. M. Mechanisms and Dynamics of Protein Acetylation in Mitochondria. *Trends in Biochemical Sciences* **41,** 231–244 (2016).
80. Gattner, M. J., Vrabel, M. & Carell, T. Synthesis of ε-N-propionyl-, ε-N-butyryl-, and ε-

N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem. Commun.* **49,** 379–381 (2013).
81. Paik, W. K., Pearson, D., Lee, H. W. & Kim, S. Nonenzymatic acetylation of histones with acetyl-CoA. *BBA Sect. Nucleic Acids Protein Synth.* **213,** 513–522 (1970).
82. Wagner, G. R. & Hirschey, M. D. Nonenzymatic Protein Acylation as a Carbon Stress Regulated by Sirtuin Deacylases. *Molecular Cell* **54,** 5–16 (2014).
83. Baeza, J., Smallegan, M. J. & Denu, J. M. Site-specific reactivity of nonenzymatic lysine acetylation. *ACS Chem. Biol.* **10,** 122–128 (2015).
84. Ruan, H. Bin, Singh, J. P., Li, M. D., Wu, J. & Yang, X. Cracking the O-GlcNAc code in metabolism. *Trends in Endocrinology and Metabolism* **24,** 301–309 (2013).
85. Laffel, L. Ketone bodies: A review of physiology, pathophysiology and application of monitoring to diabetes. *Diabetes/Metabolism Research and Reviews* **15,** 412–426 (1999).
86. Xie, Z. *et al.* Metabolic Regulation of Gene Expression by Histone Lysine β-Hydroxybutyrylation. *Mol. Cell* **62,** 194–206 (2016).
87. Yanagisawa, T. *et al.* Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chem. Biol.* **15,** 1187–1197 (2008).
88. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.* **260,** 649–663 (1996).
89. Chen, P. R. *et al.* A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chemie - Int. Ed.* **48,** 4052–4055 (2009).
90. Adams, S. R. & Tsien, R. Y. Controlling Cell Chemistry with Caged Compounds. *Annu. Rev. Physiol.* **55,** 755–784 (2003).
91. Givens, R. S., Weber, J. F. W., Jung, A. H. & Park, C. H. New photoprotecting groups: Desyl and p-hydroxyphenacyl phosphate and carboxylate esters. *Methods Enzymol.* **291,** 1–29 (1998).
92. Zhu, Y., Pavlos, C. M., Toscano, J. P. & Dore, T. M. 8-Bromo-7-hydroxyquinoline as a photoremovable protecting group for physiological use: Mechanism and scope. *J. Am. Chem. Soc.* **128,** 4267–4276 (2006).
93. Furuta, T. & Noguchi, K. Controlling cellular systems with Bhc-caged compounds. *TrAC - Trends Anal. Chem.* **23,** 511–519 (2004).
94. Momotake, A., Lindegger, N., Niggli, E., Barsotti, R. J. & Ellis-Davies, G. C. R. The nitrodibenzofuran chromophore: A new caging group for ultra-efficient photolysis in living cells. *Nat. Methods* **3,** 35–40 (2006).
95. Nakayama, K., Heise, I., Görner, H. & Gärtner, W. Peptide release upon photoconversion of 2-nitrobenzyl compounds into nitroso derivatives. *Photochem. Photobiol.* **87,** 1031–1035 (2011).
96. Goeldner, M. & Givens, R. S. *Dynamic Studies in Biology: Phototriggers, Photoswitches and Caged Biomolecules*. *Dynamic Studies in Biology: Phototriggers, Photoswitches and Caged Biomolecules* (2005). doi:10.1002/3527605592
97. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chemie - Int. Ed.* **45,** 2728–2731 (2006).
98. Martínez, A. Evidence for a functionally important histidine residue in human tyrosine hydroxylase. *Amino Acids* **9,** 285–292 (1995).
99. Uchida, K. Histidine and lysine as targets of oxidative modification. *Amino Acids* **25,** 249–257 (2003).

100. Remko, M. & Rode, B. M. Effect of metal ions (Li+, Na+, K+, Mg2+, Ca2+, Ni2+, Cu2+, and Zn 2+) and water coordination on the structure of glycine and zwitterionic glycine. *J. Phys. Chem. A* **110,** 1960–1967 (2006).
101. Li, F., Fitz, D., Fraser, D. G. & Rode, B. M. Catalytic effects of histidine enantiomers and glycine on the formation of dileucine and dimethionine in the salt-induced peptide formation reaction. *Amino Acids* **38,** 287–294 (2010).
102. Mrozek, A., Karolak-Wojciechowska, J. & Kieć-Kononowicz, K. Five-membered heterocycles. Part III. Aromaticity of 1,3-imidazole in 5+n hetero-bicyclic molecules. *J. Mol. Struct.* **655,** 397–403 (2003).
103. Priyakumar, U. D., Punnagai, M., Mohan, G. P. K. & Sastry, G. N. A computational study of cation-π interactions in polycyclic systems: Exploring the dependence on the curvature and electronic factors. *Tetrahedron* **60,** 3037–3043 (2004).
104. Reddy, A. S. & Sastry, G. N. Aromatic Motifs of Naturally Occurring Amino Acids : A Theoretical Study †. *J. Phys. Chem. A* 8893–8903 (2005).
105. Engerer, L. K. & Hanusa, T. P. Geometric effects in olefinic cation-π interactions with alkali metals: A computational study. *J. Org. Chem.* **76,** 42–49 (2011).
106. Vijay, D. & Sastry, G. N. Exploring the size dependence of cyclic and acyclic π-systems on cation-π binding. *Phys. Chem. Chem. Phys.* **10,** 582–590 (2008).
107. Matsumura, H. *et al.* Novel cation-π interaction revealed by crystal structure of thermoalkalophilic lipase. *Proteins Struct. Funct. Genet.* **70,** 592–598 (2008).
108. Reddy, A. S., Zipse, H. & Sastry, G. N. Cation-π interactions of bare and coordinatively saturated metal ions: Contrasting structural and energetic characteristics. *J. Phys. Chem. B* **111,** 11546–11553 (2007).
109. Schottel, B. L., Chifotides, H. T. & Dunbar, K. R. Anion-π interactions. *Chemical Society Reviews* **37,** 68–83 (2008).
110. Hughes, R. M. & Waters, M. L. Effects of lysine acetylation in a β-hairpin peptide: Comparison of an amide-π and a cation-π interaction. *J. Am. Chem. Soc.* **128,** 13586–13591 (2006).
111. Meyer, E. A., Castellano, R. K. & Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angewandte Chemie - International Edition* **42,** 1210–1250 (2003).
112. Kang, S. O., Hossain, M. A. & Bowman-James, K. Influence of dimensionality and charge on anion binding in amide-based macrocyclic receptors. *Coordination Chemistry Reviews* **250,** 3038–3052 (2006).
113. Dong, J., Callahan, K. L., Borotto, N. B. & Vachet, R. W. Identifying Zn-bound histidine residues in metalloproteins using hydrogen-deuterium exchange mass spectrometry. *Anal. Chem.* **86,** 766–773 (2014).
114. Lusic, H. Control of Biological Processes with Unnatural Nucleosides, Amino Acids, and Small Molecule Probes. 429 (2009).

Vita

William Daniel Kinney was born on February 10, 1990 in Phoenixville, Pennsylvania. He graduated from Owen J. Roberts High School, Pottstown, Pennsylvania in 2008. He received his Bachelor of Science in Chemistry from James Madison University, Harrisonburg, Virginia in 2012.