Resolving the Limitations of Genetic Code Expansion Platforms

Katherine Taylor Grasso

A dissertation submitted to the Department of Chemistry

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

March 2021

RESOLVING THE LIMITATIONS OF GENETIC CODE EXPANSION PLATFORMS

by

Katherine Taylor Grasso

Advisor: Professor Abhishek Chatterjee, Ph.D.

Abstract

Over the past twenty years, the site-specific incorporation of unnatural amino acids (UAAs) into a target protein through genetic code expansion (GCE) has emerged as one of the foremost technologies to selectively modify proteins in their native cellular context. This technology relies on engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that are orthogonal to the host cells' endogenous aaRS/tRNA pairs. Traditionally, scientists look towards evolutionarily distant domains of life to identify orthogonal aaRS/tRNA pairs that can be further engineered for GCE applications in the host system. For example, bacterial aaRS/tRNA pairs are used for GCE in eukaryotes. The directed evolution of orthogonal aaRS/tRNA pairs for eukaryotic GCE has been less fortuitous due to the cumbersome nature of established yeast-based selection platforms. Recently, our lab circumvented this platform-based limitation by developing "altered translational machinery" (ATM) *Escherichia coli* strains that enabled the directed evolution of bacterial aaRS/tRNA pairs for eukaryotic GCE applications. In the ATM-tyrosyl (ATMY) *E. coli* strain, reintroduction of the *E. coli* tyrosyl-tRNA ($tRNA^{Ec Tyr}_{CUA}$) as a nonsense suppressor led to cross-reactivity with the endogenous *E. coli* glutaminyl-tRNA synthetase (*Ec*GlnRS), restricting the activity range of aaRSs that could be selected, ultimately diminishing the scope of incorporable UAAs. To recover the dynamic range of this platform, cross-reactivity of the $tRNA^{Ec Tyr}_{CUA}$ was

eliminated through directed evolution of the tRNA acceptor stem. This new, orthogonal tRNA revealed weak mutant aaRSs whose suppression efficiencies were boosted through additional rounds of directed evolution. Improved aaRS mutants exhibited higher solubility, thermal stability, and suppression efficiency than their predecessor.

While the newly engineered, orthogonal $tRNA^{Ec\mathrm{Tyr}}_{\mathrm{CUA}}$ gave access to novel aaRS/tRNA pairs for eukaryotic GCE, some notable UAAs were still missing that could be incorporated with the archaeal *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/tRNA pair in bacteria. Following a systematic investigation into the discrepancy between the *E. coli* tyrosyl-tRNA synthetase (*Ec*TyrRS)/tRNA and *Mj*TyrRS/tRNA pairs, we found that it can be partially attributed to the low structural robustness of the *Ec*TyrRS. This limitation was overcome by rationally designing chimeric TyrRSs composed of *Ec*TyrRS and a structural homologue from the thermophilic bacterium *Geobacillus stearothermophilus*. The chimeric scaffolds demonstrated enhanced stability, activity, and resilience to destabilizing active site mutations, offering a potentially more attractive scaffold for GCE.

Acknowledgements

First, I would like to thank my family. They have supported me throughout my entire life and I am so grateful for all of the sacrifices they have made to help me achieve this goal. Special thanks to my parents, Fred and Kris Grasso, for showing me the meaning of hard work and perseverance. Thank you to my sisters, Stephanie Montgomery and Ally Grasso, for providing me shoulders to lean on when times were tough. Thank you to my grandma, Diane Tolson, for always making me feel loved and supported. I love all of you.

I would like to thank my advisor, Professor Abhishek Chatterjee. Five years ago, I emailed Abhishek asking if he would hire me as a lab tech. To my surprise, he responded suggesting that I should apply to BC's chemistry Ph.D. program and that if I had any questions that I could always talk to him over the phone. This one email exchange completely changed my future. I wandered down a career path that I had never before considered, and learned so many valuable life lessons along the way. So, thank you again, Abhishek. Not just for guiding me as a scientist over these past five years, but for believing in my potential and encouraging me to pursue this degree.

Thank you to my lab members, you are a great group of people. I feel lucky to have been immersed in a culture of hard work and intellectual curiosity over these past five years. You all have pushed me to fulfill my potential as both a scientist and person. I would especially like to thank Megan Yeo for being an amazing advisee and friend. I wish you great success in your future career as a scientist, I know that you will achieve amazing things. I would like to thank Christen Hillenbrand and Chris Latour for being a great support system in the lab and helping out with my experiments. I would like to thank Dr. Partha Sarathi Addy for mentoring me my first year. While the project didn't work out, our friendship certainly did. You are such a kind soul and I am grateful to have had the opportunity to work with you. I would like to thank Elise Ficaretta for being an

amazing bay-mate, team ATM member, and always volunteering to help out when I am in a crunch. I would like to thank the rest of the team ATM members past and present for helping push forward the BEST project in the lab, but I may be biased. Additionally, I would like to thank Lily Jewel and Rachel Huang for taking time out of their busy schedules to review this dissertation.

Thank you to all of my friends. In particular, Yejin Chang, Paul Hicks, Mike Kelly, Rachel Gadzicki (Kelly), and Ashley Hunt. All of you have provided me with such a strong sense of community and support over my time in graduate school. My favorite memories from these past five years were made during our Barcelona brunches, bar crawls, wine and paint nights, road trips, and visits to NYC. I am so happy and lucky to have all of you in my life.

Thank you to my past advisors, Professor Rudi Fasan and Professor Larry Hudson. Professor Fasan, thank you for introducing me to the world of chemical research. The opportunity to work in your lab during undergrad catalyzed my desire to pursue my Ph.D. in chemistry. Professor Hudson, thank you for challenging me to think for myself. Your courses always pushed me to creatively evaluate the world around me. They were by far the most impactful part of my undergraduate experience.

Thank you to the Boston College Office of Technology Transfer and Licensing staff for providing me with an unpaid internship over these past few years. This opportunity set me up for success in my future career by giving me a glimpse into the world of patent law and helping me land my first job. I would especially like to thank the women that I worked with, Dr. Jill Edgar and Dr. Mo-Ying Hsieh.

Thank you to all of the office staff in Merkert who keep the building running. This past year has been especially challenging, and everyone in Merkert appreciates all of the hard work you all have put in to make this place operable.

Finally, I would like to thank Adam. I could not have achieved this without you. My greatest accomplishment over these past five years was meeting you. I love you.

Table of Contents

List of Figures

Chapter 1 Figures

Chapter 2 Figures

Chapter 3 Figures

Chapter 4 Figures

Chapter 5 Figures

Chapter 6 Figures

List of Tables

Chapter 2 Tables

Chapter 3 Tables

Chapter 4 Tables

List of Abbreviations

Standard one-letter and three-letter abbreviations are used for the 20 natural amino acids.

4CDz	4-carboxybenzenediazonium
4MADz	4-methylalcoholdiazonium
4MDz	4-methoxydiazonium
5-FOA	5-fluorootic acid
5-HTP	5-hydroxytryptophan
aaRS	aminoacyl-tRNA synthetase
ABPP	activity-based protein profiling
AcK	acetyllysine
amp	ampicillin
ATM	altered translational machinery
ATMW	altered translational machinery tryptophan
ATMY	altered translational machinery tyrosine
bp	base-pair
cAA	canonical amino acid
CAT	chloramphenicol acetyl-transferase
carb	carbenicillin
$CDCl_3$	deuterated chloroform
CETSA	cell-free extract thermal shift assay
CFU	colony forming unit
Ch2TyrRS	chimera 2 tyrosyl-tRNA synthetase
Ch6TyrRS	chimera 6 tyrosyl-tRNA synthetase
chlor	chloramphenicol
CoA	acetyl coenzyme A
CRACR	chemoselective rapid azo-coupling reaction
Cs_2CO_3	cesium carbonate
CuAAC	copper-catalyzed azide alkyne cycloaddition
DCM	dichloromethane
DMEM	Dulbecco's Modified Eagle Medium
DMF	Dimethylformamide
DNA	deoxyribonucleic acid
dtyrS	deletion of tyrosyl-tRNA synthetase
*Ec*GlnRS	*E. coli* glutaminyl-tRNA synthetase
*Ec*TrpRS	*E. coli* tryptophanyl-tRNA synthetase
*Ec*Tyr-OMeYRS	*E. coli* o-methyl-L-tyrosyl-tRNA synthetase mutant
EcTyr-OMeYRS	*E. coli* tyrosyl-derived OMeY-selective tRNA synthetase
EcTyr-pBPARS	*E. coli* tyrosyl-derived pBPA-selective tRNA-synthetase
*Ec*TyrRS	*E. coli* tyrosyl-tRNA synthetase
EF-Tu	elongation factor Tu
EGFP-39-TAG	enhanced green fluorescent protein with an amber stop codon at amino acid position 39

EGFP	enhanced green fluorescent protein
em	emission
ePCR	error-prone polymerase chain reaction
EtOAc	ethyl acetate
ex	excitation
FACS	fluorescence activated cell sorting
FBS	fetal bovine serum
GCE	genetic code expansion
GFPmut3	green fluorescent protein mut3
GFPuv	green fluorescent protein uv
g	gram
*Gs*Tyr-pBPARS	*G. stearothermophilus* tyrosyl-derived pBPA-selective tRNA synthetase
GsTyrRS	*G. stearothermophilus* tyrosyl-tRNA synthetase
H_2O	water
HEK293T	human embryonic kidney 293 cells with a temperature-sensitive allele of the SV40 T antigen
IA	iodoacetamide alkyne
IPTG	isopropyl β-D-1-thiogalactopyranoside
kan	kanamycin
L	liter
LB	lysogeny broth
M	molar
*Mb*PylRS	*M. barkeri* pyrrolysyl-tRNA synthetase
MeOH	methanol
mg	milligram
$MgSO_4$	magnesium sulfate
min	minute
*Mj*TyrRS	*M. jannaschii* tyrosyl-tRNA synthetase
*Mj*Y	*M. jannaschii* tyrosyl
mL	milliliter
mM	millimolar
*Mm*PylRS	*M. mazei* pyrrolysl-tRNA synthetase
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NaCl	sodium chloride
NaH	sodium hydride
NaN_3	sodium azide
$NaNO_2$	sodium nitrite
NCL	native-chemical ligation
nm	nanometer
nt	nucleotide
pAcrF	*p*-acrylamido-(S)-phenylalanine
pAEY	*p*-azidoethyl-tyrosine
pAlkAcF	*p*-alkynylacetamido-phenylalanine
pAzAcF	*p*-azidoacetamido-phenylalanine
pAzPrAmF	*p*-azidopropanamido-phenylalanine

pBPA	*p*-benzoyl-L-phenylalanine
PBP	penicillin-binding protein
PBS	phosphate buffered saline solution
pCAcF	*p*-chloroacetamido-phenylalanine
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEI MAX	polyethylenimine Max
pFAcF	*p*-fluoroacetamido-phenylalanine
PIPE	polymerase incomplete primer extension
PTM	post-translational modification
RNA	ribonucleic acid
rpm	revolutions per minute
*Sc*TrpRS	*S. cerevisiae* tryptophanyl-tRNA synthetase
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sfGFP-151-TAG	superfolder green fluorescent protein with an amber stop codon at amino acid position 151
sfGFP	superfolder green fluorescent protein
sfGFPwt	superfolder green fluorescent protein wild type
SPAAC	strain-promoted azide alkyne cycloaddition
spec	spectinomycin
SPPS	solid-phase peptide synthesis
sTyr	sulfotyrosine
TAG	amber stop codon
tet	tetracycline
TEV	*Tobacco etch virus*
THF	Tetrahydrofuran
TLC	thin layer chromatograpy
TOP-ABPP	tandem-orthogonal proteolysis activity-based protein profiling
TPST	tyrosylprotein sulfotransferase
tRNA	transfer ribonucleic acid
TyrRS	tyrosyl-tRNA synthetase
UAA	unnatural amino acid
V	volts
µg	microgram
µL	microliter

Chapter 1

Introduction

1.1 Genetic code expansion

Proteins are the workhorses of the cell, facilitating chemical transformations, enabling the transport of molecules across cell membranes, and regulating cell-signaling.[1] Due to proteins' integral role in cellular biology their aberrant production and regulation are usually the cruxes of a disease.[2-6] Over the past two decades, the ability to identify new proteins has increased exponentially through scientific advances in the field of proteomics.[7, 8] Unfortunately, our ability to discern the relationship between a protein's structure and its function has not kept pace. Tremendous efforts have been put forth to fill this gap in knowledge because of the key roles that proteins play in cellular biology. Thus far, developed technologies have focused on the chemical modification of naturally expressed proteins,[9] the synthesis of full-length proteins,[10, 11] and the co-translational incorporation of chemically unique unnatural amino acids (UAAs) into a target protein through genetic code expansion (GCE).[7, 8] Over the past two decades, GCE has emerged as the foremost technology for the selective modification of proteins because of its specificity and ability to modify proteins of any size. The repertoire of UAAs currently accessible with GCE gives access to unique biorthogonal handles that can be site-specifically modified following installation,[12, 13] post-translational modification (PTM) mimics that enable studies of different protein interactions in the cell,[14, 15] and photoaffinity probes that covalently capture different transient protein interactions.[16, 17] This dissertation focuses on GCE advancements made in two major areas: (i) the selection platforms used for aminoacyl-tRNA synthetase (aaRS)/tRNA pair engineering for mammalian GCE and (ii) how to best approach choosing a parental aaRS scaffold for engineering.

1.1.1 Protein translation

Across all three domains of life, 64 degenerate codons (Figure 1.1) are used to encode the 20 canonical amino acids (Figure 1.2) and three stop codons are used to terminate protein synthesis.[1, 7, 8] While canonical amino acid derivatives have been identified in natural proteins, all but two (selenocysteine[18] and pyrrolysine[19]) are due to post-translational modifications (PTMs). This lack of chemical diversity is especially astounding considering the broad range of roles that proteins play in cell biology.[1]

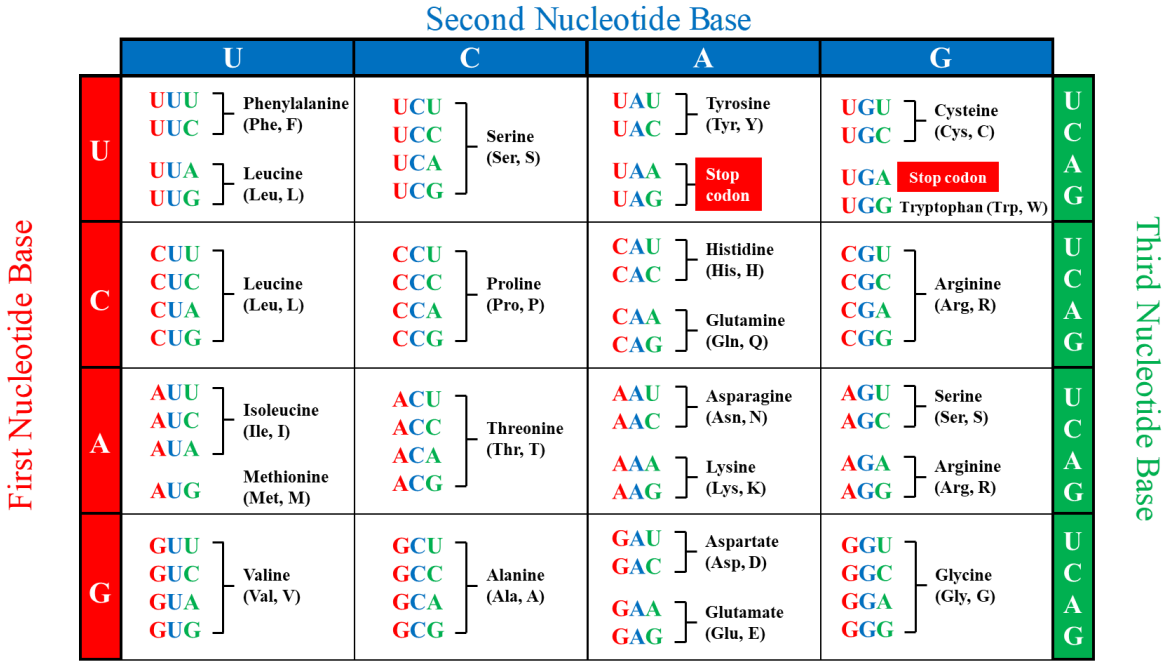


Figure 1.1. Amino acid mRNA codon chart.

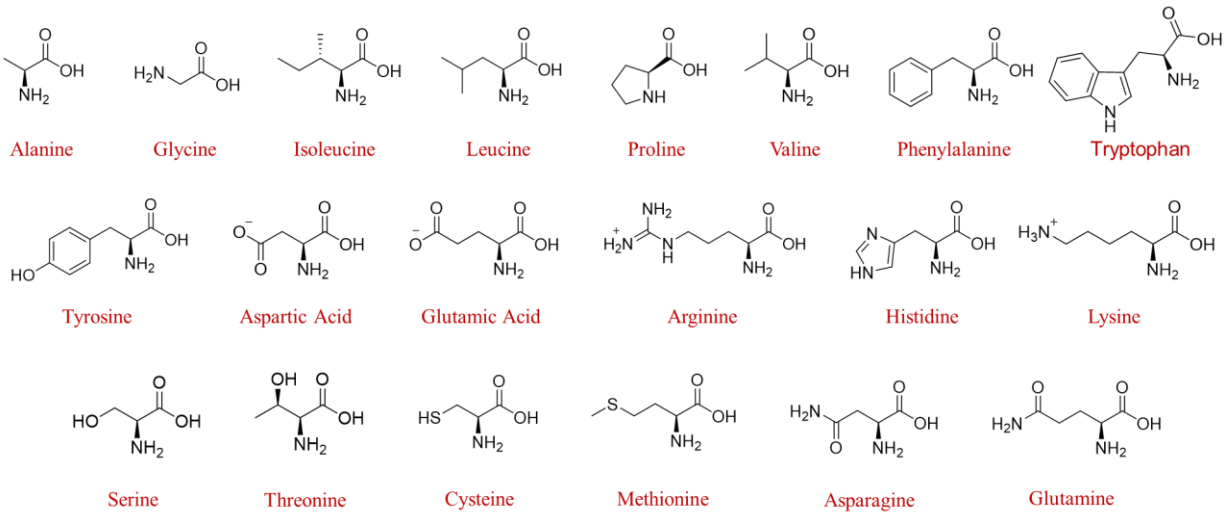


Figure 1.2. The structures of the twenty canonical amino acids.[20]

Individual amino acids are polymerized into proteins during a process called protein translation. During protein translation an aminoacyl-tRNA synthetase (aaRS) first charges a tRNA with the appropriate amino acid (Figure 1.3).[21] Then, the acylated tRNA shuttles the amino acid to the ribosomal machinery where the tRNA anticodon interfaces with its corresponding codon on the mRNA, polymerizing the amino acid into the growing peptide chain.[21] These full-length proteins can then be further modified in the cell through the covalent addition of other chemical groups or small protein modifiers called PTMs (Figure 1.4).[22]

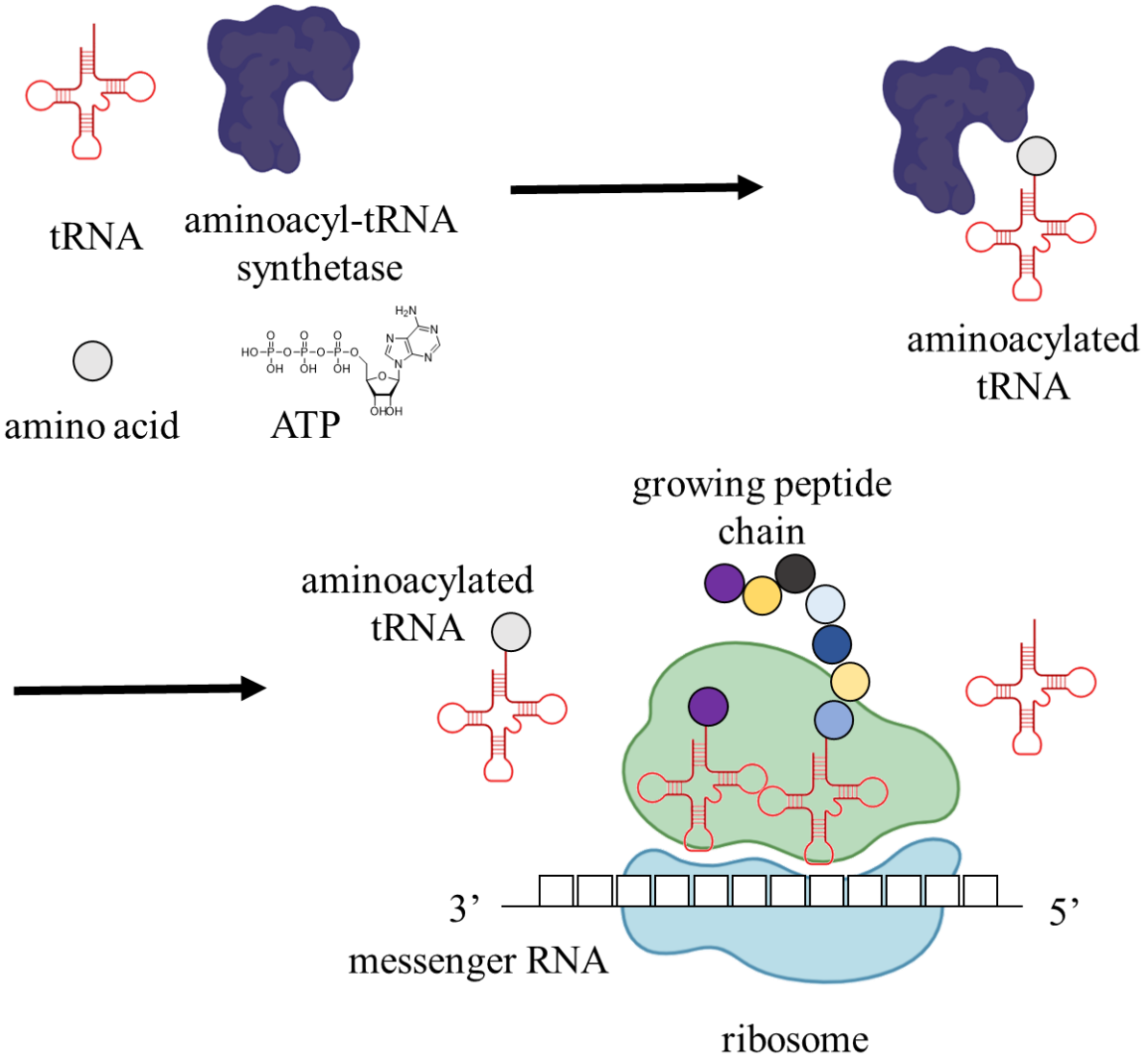


Figure 1.3. Protein translation.[21] Step 1: In the aminoacyl-tRNA synthetase active site an amino acid is coupled to AMP and aminoacyl-AMP is formed followed by the loss of two phosphate groups from the initial ATP structure (top left). Step 2: The uncharged tRNA binds to the aminoacyl-tRNA synthetase and the aminoacyl-tRNA synthetase transfers the amino acid from aminoacyl-AMP to the tRNA (top right). Step 3: The aminoacylated-tRNA is shuttled to the ribosome where it can covalently link the amino acid to the growing peptide chain (bottom).

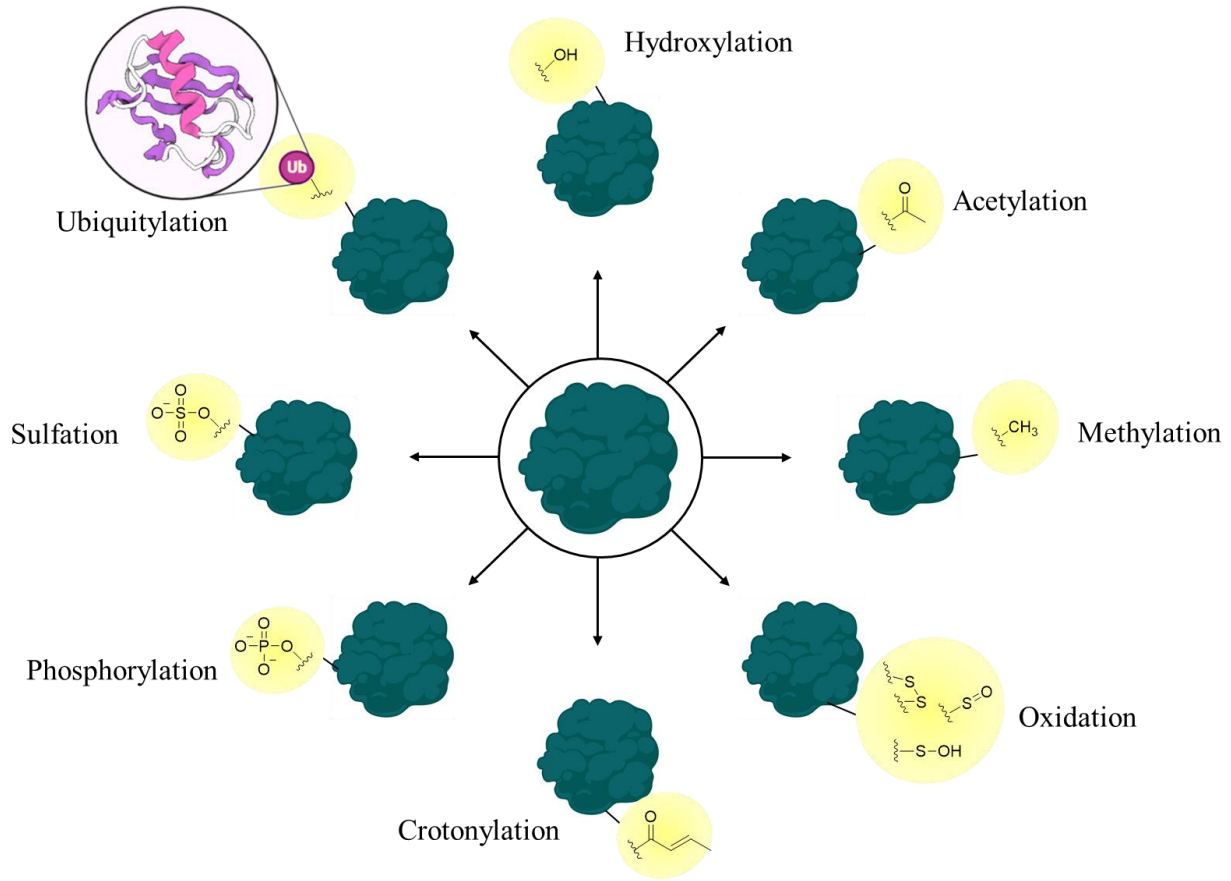


Figure 1.4. Common post-translational modifications.[22] The installation of different PTMs provide alternative methods to regulate cellular biology. For example, transient PTMs (e.g., phosphorylation) can regulate enzymatic activity. Other modifications (e.g., ubiquitylation) can target proteins for degradation.

1.1.2 Strategies to study protein structure/function relationships

Many strategies have emerged over the past 60 years to study, alter, or even improve upon the chemical activity of proteins. One of the first employed methods was the chemical modification of full-length, folded proteins, termed chemical mutagenesis.[9, 23] However, our ability to mimic the precision observed in nature through a chemical mutagenesis approach is limited due to the lack of unique chemical handles afforded by the canonical amino acids. In fact, most developed chemical mutagenesis methods to date rely on the reactivity of just a few nucleophilic amino acid residues (Figure 1.5).[9, 23] Selective targeting of a canonical amino acid generally proves fruitless as they are commonplace within the protein structure.

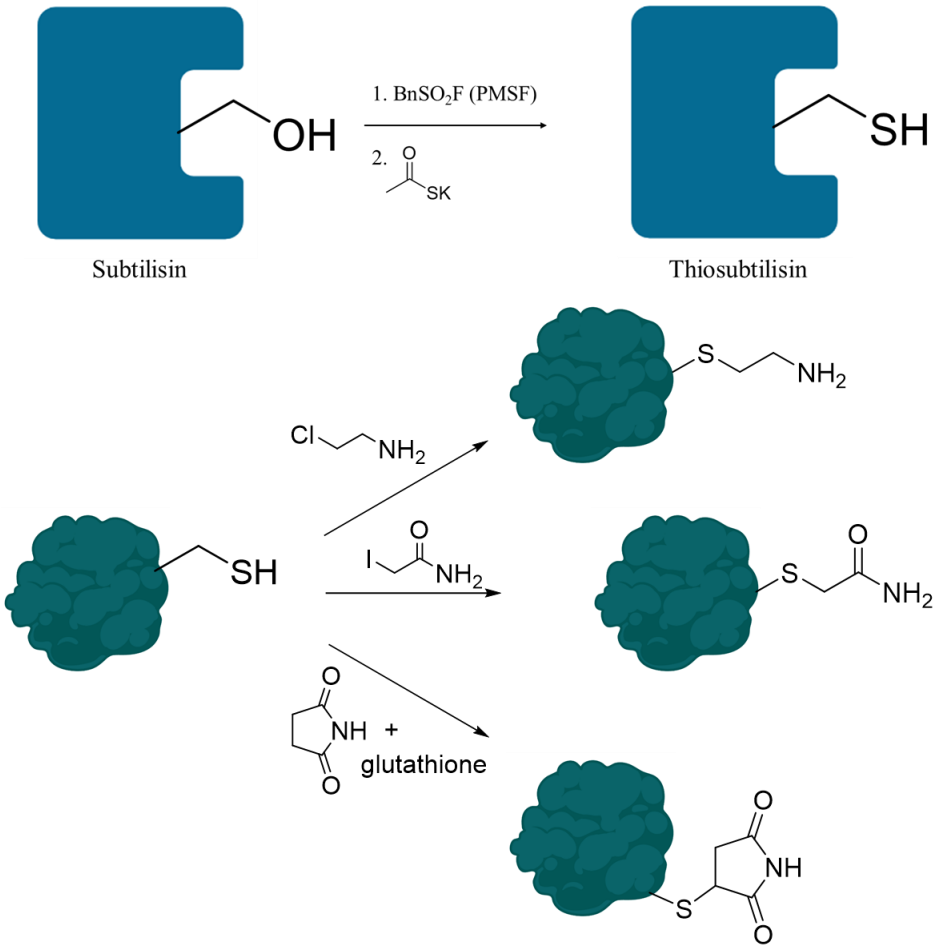


Figure 1.5. Chemical modification of full-length proteins. In 1966, both Koshland and Bender reported on the first point mutation of an enzyme.[9, 23, 24, 25] They had independently found chemical methods to selectively modify the subtilisin active site serine to a cysteine (top). This seminal discovery has been attributed to kick-starting the field of protein chemical mutagenesis. Since this discovery, other chemical modifications of the canonical amino acids have been identified such as cysteine $S_{protein}$-C bonds (bottom), cysteine $S_{protein}$-S bonds, and lysine $N_{protein}$-C bonds, among others.[9]

An alternate approach created to obtain residue-level selectivity in protein modification was through the chemical synthesis of native proteins. A few different tactics have been adapted for the chemical synthesis of proteins including solution-based synthetic chemistry,[11] solid-phase peptide synthesis (SPPS),[11] and native-chemical ligation (NCL)[10, 11] (Figure 1.6). However, all of these synthetic methods are limited by either the high-cost of production, time-consuming

protocols for synthesizing proteins, limited solubility of intermediates, a lack of high-resolution purification methods, an inability to make peptides larger than 50 amino acid residues, or site-modification requirements.[10, 11]

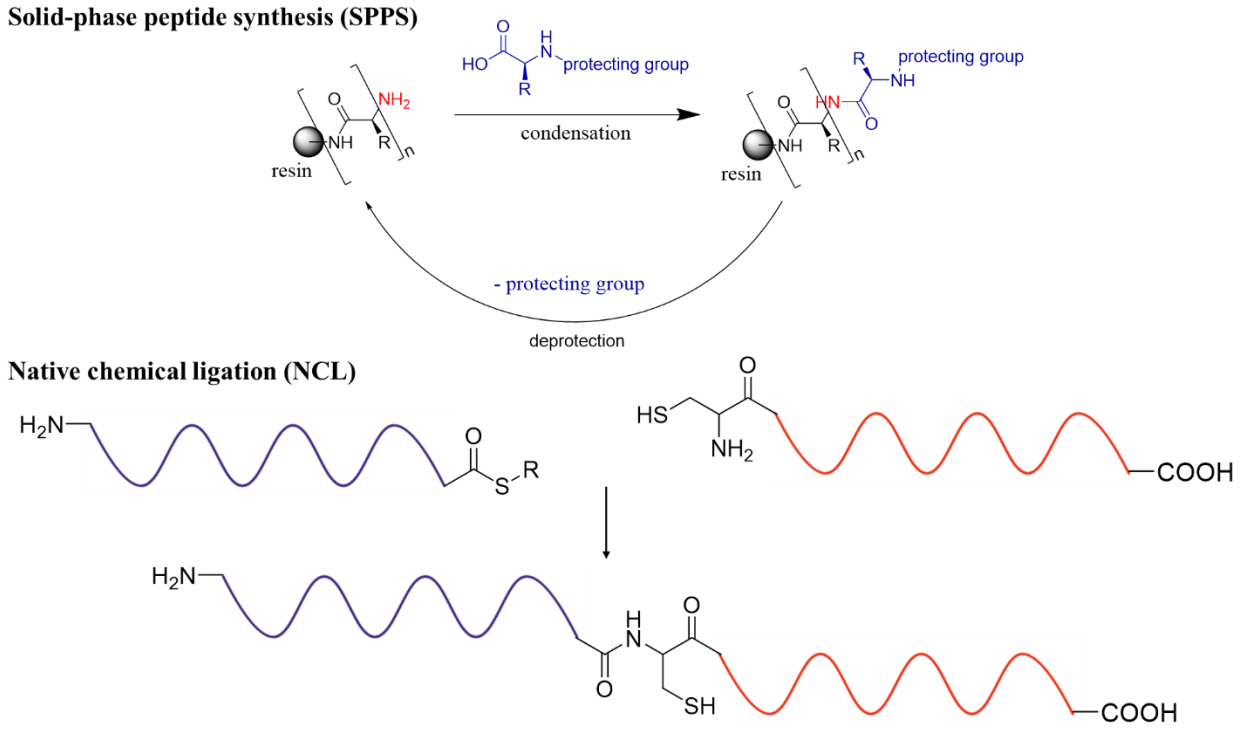


Figure 1.6. Approaches for chemically synthesizing full-length proteins.[10, 11] SPPS relies on the C-terminal attachment of an amino acid to a solid resin that is subsequently condensed with iterative rounds of protected amino acids (top).[11] In NCL, unprotected peptide segments are joined together through a thiol/thioester exchange, followed by a nucleophilic rearrangement, that results in the polypeptide product that is linked by a native peptide bond.[10]

1.1.3 Principles of GCE

Twenty years ago, genetic code expansion (GCE) was introduced as a new methodology to selectively modify a target protein. This technology harnesses the power of protein translation to introduce unnatural amino acids (UAAs) site-specifically into a protein of interest providing access to novel chemical modalities such as new enzymatic functions, imaging probes, bioconjugation handles, and PTM mimics.[7, 8, 26] Three components are required for this technology to function: (i) an orthogonal aaRS/tRNA pair that can be engineered for UAA incorporation, (ii) a blank codon that can be assigned for the site-specific insertion of the UAA, and (iii) efficient transportation of the UAA into the cell (Figure 1.7).[7, 8, 26] Traditionally, we look to evolutionarily distant organisms from alternate domains of life to find an orthogonal aaRS/tRNA pair.[7, 8, 26] This strategy is based on early reports that *in vitro* cross-species aminoacylation is often low.[27] For GCE to be successful, it is imperative that the orthogonal aaRS does not acylate any of the endogenous tRNA, and that it only charges the orthogonal tRNA with the UAA (Figure 1.7).[7, 8, 26] Additionally, the orthogonal tRNA should not be acylated by any of the endogenous aaRSs, should only be charged with the UAA, and should decode the blank codon assigned to it (Figure 1.7)[7, 8, 26] The UAA also cannot be a substrate for the endogenous aaRS/tRNA pair.[7, 8, 26] Any issues of cross-reactivity can result in toxicity problems for the host-organism from the non-specific insertion of the UAA throughout the proteome and the non-specific insertion of canonical amino acids into the protein of interest.[7, 8, 26]

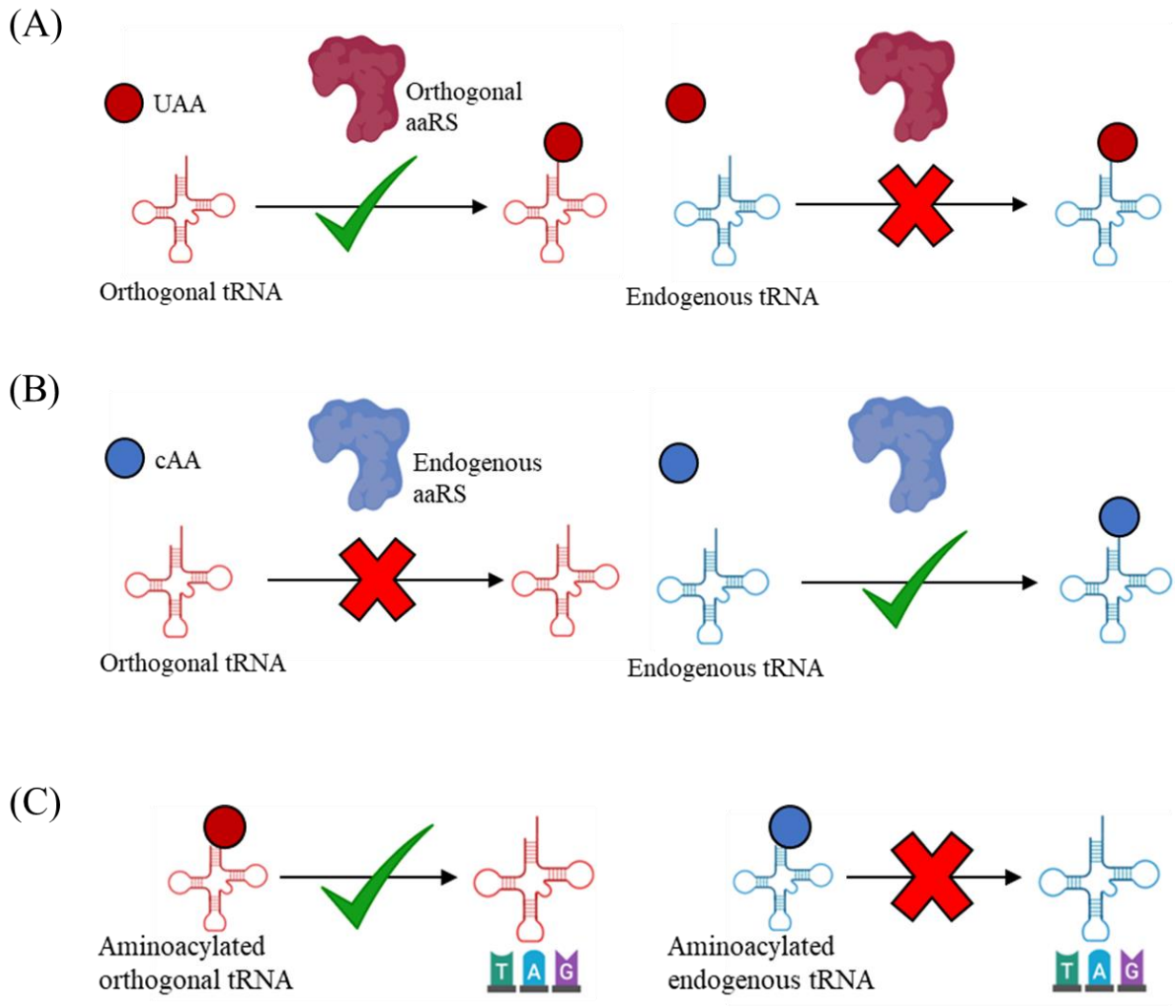


Figure 1.7. aaRS/tRNA pair orthogonality requirements for GCE.[7, 8, 26] (1) The orthogonal aaRS can only charge the orthogonal tRNA with the UAA. (2) The endogenous aaRS cannot charge the orthogonal tRNA with a canonical amino acid (cAA). (3) Only the charged orthogonal tRNA can recognize the reassigned codon (typically the amber stop codon—TAG).

Once an orthogonal aaRS/tRNA pair is identified, a simple workflow can be used to modify the substrate specificity of the aaRS for the UAA of choice through directed evolution. First, the orthogonal aaRS/tRNA pair is imported into the host organism that will be used for GCE.[7, 8, 26] Second, the anticodon of the orthogonal tRNA can be reassigned to a blank codon that isn't used for any of the twenty canonical amino acids.[7, 8, 26] Typically, this is the amber stop codon (TAG),

but other nonsense codons and quadruplet codons have been used.[7, 8, 16, 28-33] Third, the key active site residues in the aaRS are randomly mutagenized followed by a subsequent selection scheme to enrich for the desired activity (UAA incorporation).[7, 8, 26] Generally, discontinuous selection schemes have relied on both positive and negative selective pressures that enrich for active and orthogonal aaRS mutants, respectively.[7, 8, 26] This approach can be used for any genetically tractable organism, making this technology extremely powerful and easily translatable. In fact, over the past twenty years GCE has emerged as the most attractive technique for the selective modification of proteins.

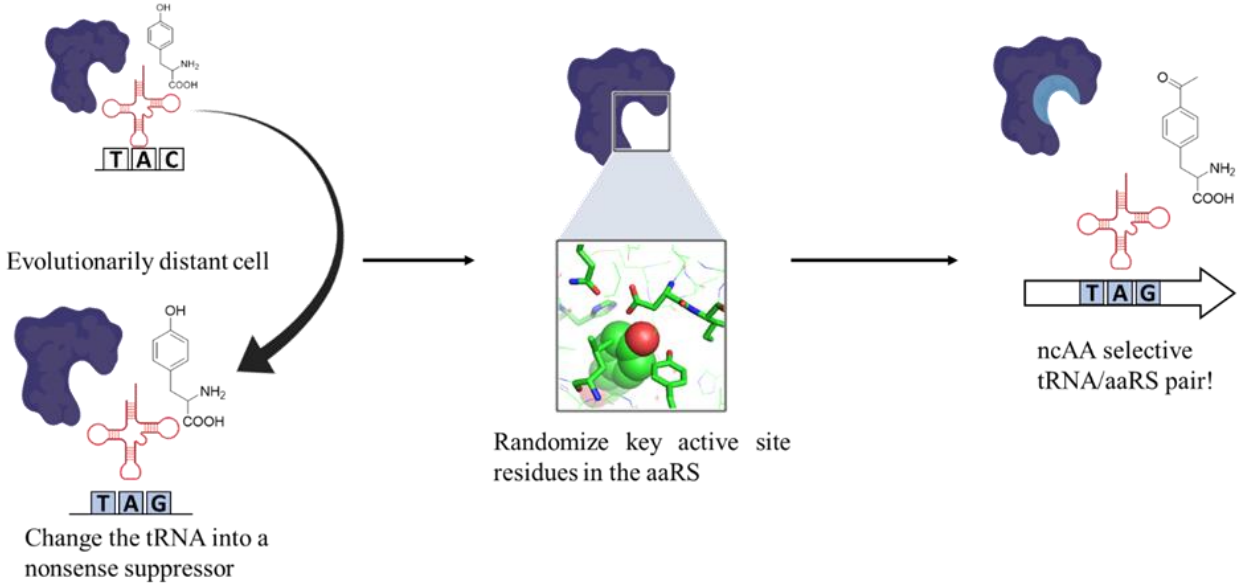


Figure 1.8. Engineering aaRS/tRNA pairs for UAA incorporation.[7, 8, 26] The paradigm for incorporating UAAs first relies on importing an aaRS that won't cross-react with the host cells endogenous tRNAs and synthetases (left). These aaRS/tRNA pairs are usually imported from an evolutionarily distant organism. Then, the anticodon of the tRNA can be switched to a codon that is not assigned to any of the twenty canonical amino acids (left). Key residues for amino acid recognition are randomly mutagenized (center). A selective pressure is applied to identify aaRSs that only charge the desired UAA (right).

1.1.4 GCE in bacteria

Escherichia coli was presented as an ideal host organism to initially test out the theory of GCE because its translational machinery had been extensively studied and the genetic

manipulation of *E. coli* is efficient and relatively straightforward.[7, 8, 26] The first aaRS/tRNA pair identified for GCE use in *E. coli* was the tyrosyl pair from the methanogenic archaeon *Methanococcus jannaschii* (*Mj*TyrRS/tRNA).[33] The *Mj*TyrRS/tRNA pair was chosen because of the lack of an editing domain on the *Mj*TyrRS, making it incapable of removing the charged UAA from the tRNA, and because the tRNA anticodon was not a main recognition element for the *Mj*TyrRS, so it could be more easily reassigned.[33] Initial UAAs incorporated with *Mj*TyrRS/tRNA pair were substituted at the *para* position of the phenyl ring, but the substrate specificity was not limited to these types of structural derivatives and soon *meta* and *ortho derivatives* were incorporated.[8] Since then, the pyrrolysyl pairs from methanogenic *Methanosarcina mazei* and *Methanosarcina barkerii* have been added to the repertoire of commonly used aaRS/tRNA pairs for bacterial GCE (Figure 1.9).[8] Unlike most aaRS/tRNA pairs that have codon-specificities, the *M. mazei* and *M. barkerii* naturally evolved as amber stop codon (TAG) -suppressors.[19, 34a, 34b] Therefore, no tRNA codon reassignment is necessary.

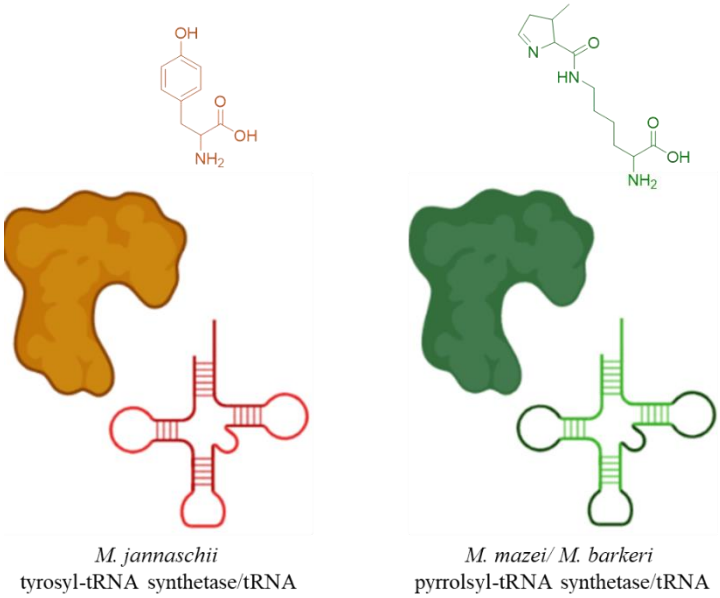


Figure 1.9. Commonly used aaRS/tRNA pairs for GCE in bacteria.[7, 8] The two most frequently used aaRS/tRNA pairs for bacterial GCE are predominantly based on the *Mj*TyrRS/tRNA pair and the pyrrolysyl-tRNA synthetase/tRNA pairs.

In most *E. coli* applications, reassignment of the amber stop codon (TAG) for UAA incorporation is used because it is the most infrequent among the three stop codons in *E. coli*.[8] Additionally, amber suppressor tRNAs have been identified or engineered in *E. coli* to read through the amber codon with little to no growth penalty.[8] While the use of amber-suppression has been more than successful, expanding the number of usable blank codons for GCE gives access to unique applications for pull-down experiments,[35] working towards the development of a fully orthogonal genetic code in the cell,[36] probing protein function in a cellular context, amongst other applications. In fact, work demonstrated by our lab[31, 35, 37] and numerous other groups[38-41] has provided insight into the unique biochemical space that multi-site incorporation can enable researchers to explore.

The facile engineering strategies of orthogonal aaRS/tRNA pairs in *E. coli* has given us access to a whole host of novel chemical space.[7, 8, 26] These engineering platforms benefit from *E. coli*'s fast growth-rate, *E. coli*'s ability to maintain and propagate DNA, and the ease of isolating the DNA encoding an engineered aaRS displaying desired activities from *E. coli*.[8] The commonly used selective pressures and selection platforms for the directed evolution of bacterial aaRS/tRNA pairs will be discussed in more detail in chapter 3 of this dissertation. Historically, these platforms have an antibiotic-based positive selection and a toxic barnase gene-based negative selection that enrich for UAA-active aaRSs.[8, 33] A newer, alternate platform termed phage-assisted continuous evolution (PACE) has been used to engineer orthogonal aaRSs by linking the activity of a biomolecule to the survival of bacteriophages infecting a pool of *E. coli* that is continuously being turned over.[42, 43] This selection strategy enables the directed evolution of multiple generations in a matter of hours. Both of these approaches have been highly successful, resulting in the incorporation of > 150 UAAs in bacteria. A comprehensive list of all of the UAAs that have been

incorporated in *E. coli* through 2015 has been compiled in a review by Ben Davis.[8] Since 2015, a few other notable UAAs have been incorporated such as the PTMs phosphoserine[15] and phosphotyrosine.[14] To depict the broad chemical space, provided through bacterial GCE, some of the UAA structures and functional applications have been included below (Figure 1.10).[8] Further developing new aaRS/tRNA pairs, beyond the *Mj*TyrRS and *Mm*PylRS/*Mb*PylRS, for bacterial GCE will give access to new chemical structures that are currently inaccessible with the structural limitations of the *Mj*TyrRS and *Mm*PylRS/*Mb*PylRS active sites.

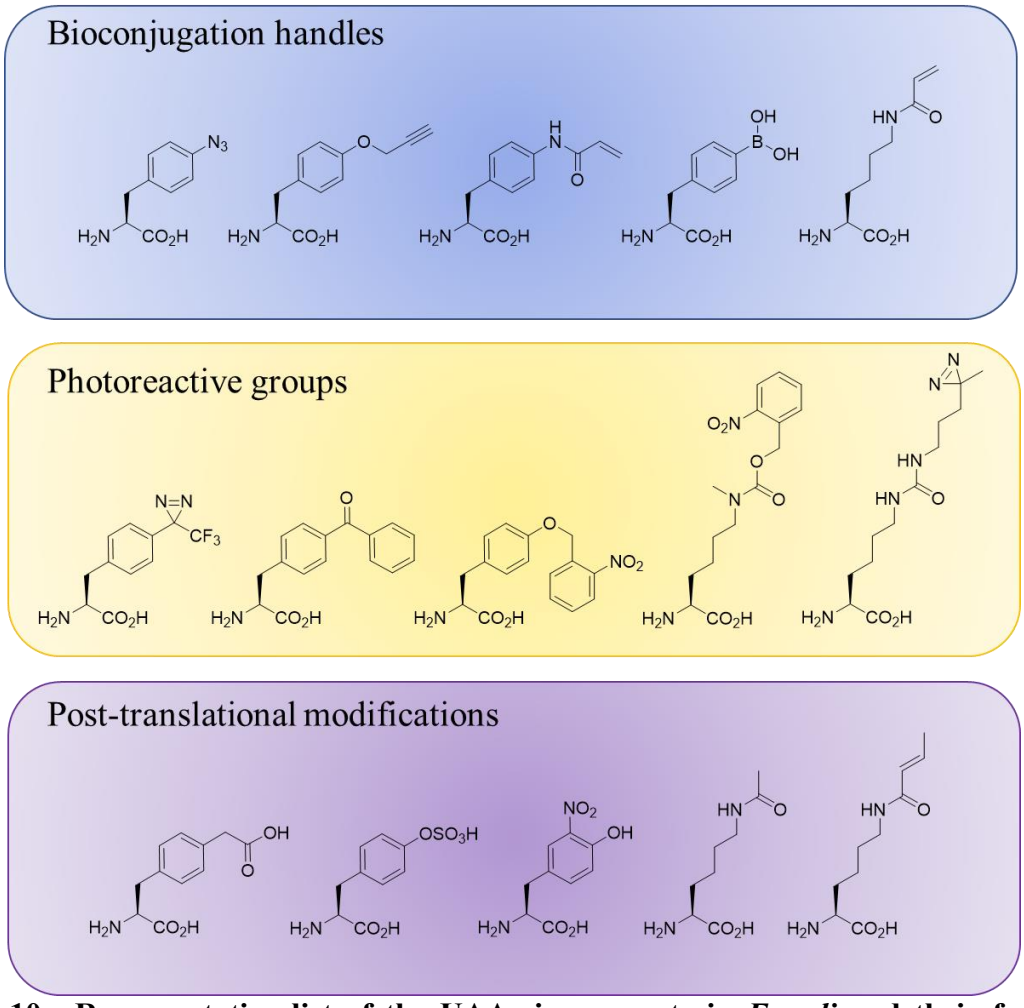


Figure 1.10. Representative list of the UAAs incorporate in *E. coli* and their functional applications.[8] This is only a subset of the >150 UAAs incorporated in *E. coli* to-date.

1.1.5 GCE in eukaryotes

While the development of the *Mj*TyrRS/tRNA pair significantly expanded the chemical space available in *E. coli*, it's orthogonality is not transferrable to eukaryotes. Therefore, alternative aaRS/tRNA pairs needed to be scrutinized for orthogonality for eukaryotic GCE applications. By turning to the established paradigm of finding orthogonal aaRS/tRNA pairs in evolutionarily distant domains of life, the lab of Paul Schimmel identified that the *E. coli* tyrosyl-tRNA synthetase (*Ec*TyrRS)/tRNA pair was orthogonal in the model organism *Saccharomyces cerevisiae.*[44] Subsequent work by Kensaku Sakamoto[45] and Jason Chin[16] demonstrated that the *Ec*TyrRS/tRNA pair could function as an amber suppressor for the incorporation of UAAs in eukaryotes. Since then, three other aaRS/tRNA pairs have been commonly used for eukaryotic GCE: the *M. mazei/M. barkeri* pyrryolysl-tRNA synthetase/tRNA pair, the *E. coli* leucyl-tRNA synthetase/tRNA pair, and the *E. coli* tryptophanyl-tRNA synthetase/tRNA pair (Figure 1.11).[7, 8, 26]

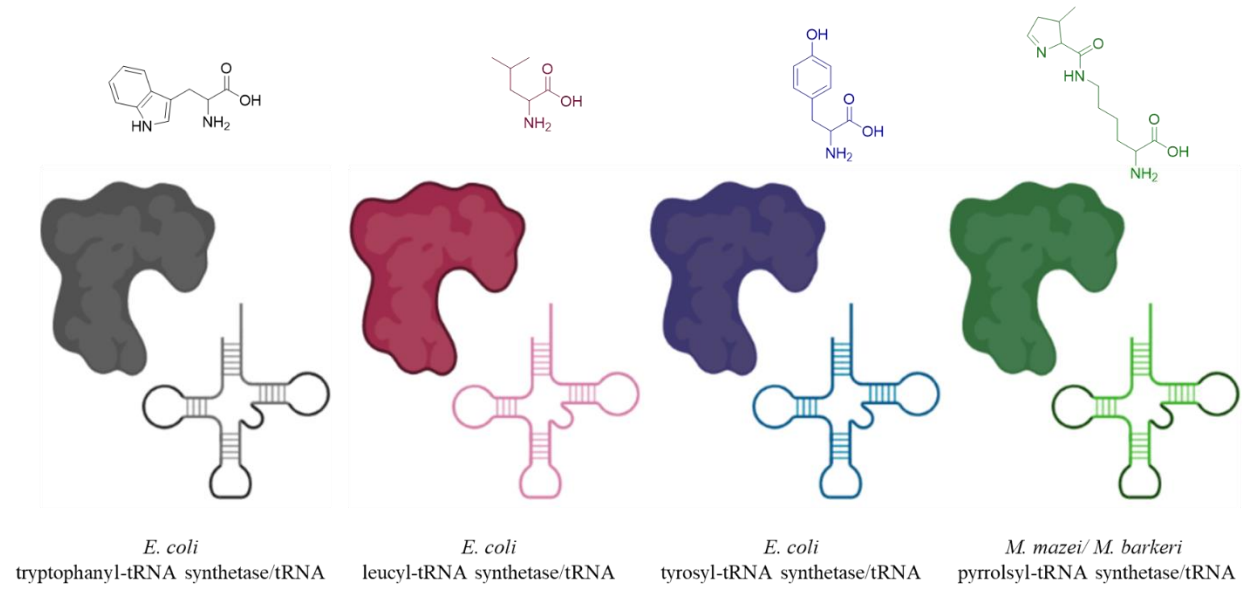


Figure 1.11. The four aaRS/tRNA pairs predominantly used for GCE in eukaryotes.[8]

Directed evolution of the *E. coli* aaRS/tRNA pairs for GCE in eukaryotes has been generally carried out in *S. cerevisae*.[7, 8, 16, 26] Unlike higher eukaryotes, *S. cerevisiae* is amenable to the same directed evolution strategies that are used in *E. coli*. Single mutant library members can be efficiently delivered through transformation of the DNA into the cell and subsequently isolated for characterization.[16] The standard dual-sieve selection scheme for *S. cerevisiae* was developed by Jason Chin and relies on the transcriptional activation of the GAL4-responsive *HIS3*, *URA3*, and *lacZ* reporter genes (Figure 1.12).[16] This strategy has enabled the successful incorporation of a narrow range of UAAs for eukaryotic GCE.[7, 8, 26] The *S. cerevisae* platform's main limitations are the organism's low transformation efficiency and slow propagation time.

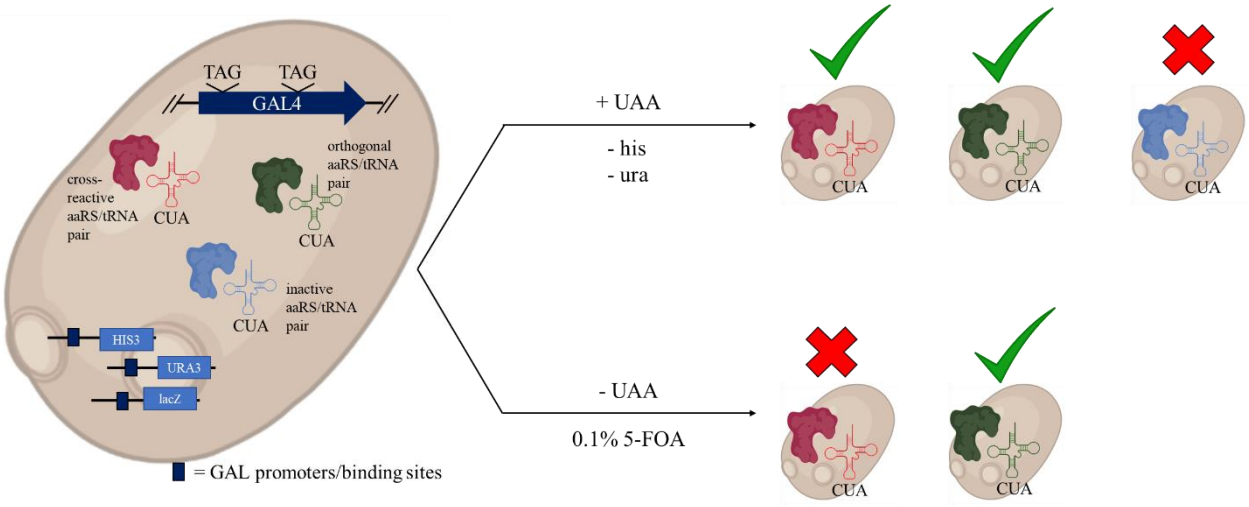


Figure 1.12. Directed evolution platform in *S. cerevisiae*.[16] *S. cerevisiae* are co-transformed with a GAL4 reporter plasmid and a plasmid containing a randomized aaRS library. Any active aaRS library members will cause full length expression of the GAL4 transcriptional regulator protein. In the positive selection (top), *S. cerevisiae* is plated in the presence of the UAA and the absence of ura and his. If the GAL4 gene has been successfully suppressed then the cell should be able to biosynthesize these essential nutrients (*HIS3* and *URA3* genes). In the negative selection (bottom), yeast is plated in the presence of 0.1% 5-fluoorotic acid (5-FOA), but the absence of UAA. Cells that express *URA3*, because of *GAL4* suppression by canonical amino acids, convert 5-FOA to a toxic product, killing the cell and eliminating cross-reactive aaRS variants.

Typically, archaebacterial aaRS/tRNA pairs are cross-reactive with their eukaryotic counterparts due to the evolutionary conservation of key identity elements of their aaRS/tRNA

pairs.[8] However, the *M. mazei* and *M. barkeri* pyrrolysl-tRNA synthetase (*Mm*PylRS/*Mb*PylRS)/tRNA pairs are orthogonal to eukaryotic aaRS/tRNA pairs due to their naturally evolved cognate recognition of the amber stop codon, amongst other unique structural features.[19, 34] As mentioned previously, both of these methanogenic archaebacterial aaRS/tRNA pairs are also orthogonal in bacteria.[8] This enables our ability to first engineer the *Mm*PylRS/*Mb*PylRS/tRNA pairs in the more facile *E. coli* selection platform and subsequently use the engineered *Mm*PylRS/*Mb*PylRS for GCE applications in eukaryotes. In fact, this approach has become so favored by the community that over the past ten years, eukaryotic GCE has been dominated by the *Mm*PylRS/*Mb*PylRS/tRNA pairs (Figure 1.13).[7, 8, 26, 46] Such heavy reliance on the active site of one-type of aaRS/tRNA pair greatly confines the incorporable UAA structural adaptations available for eukaryotic GCE, limiting the power of this technology.

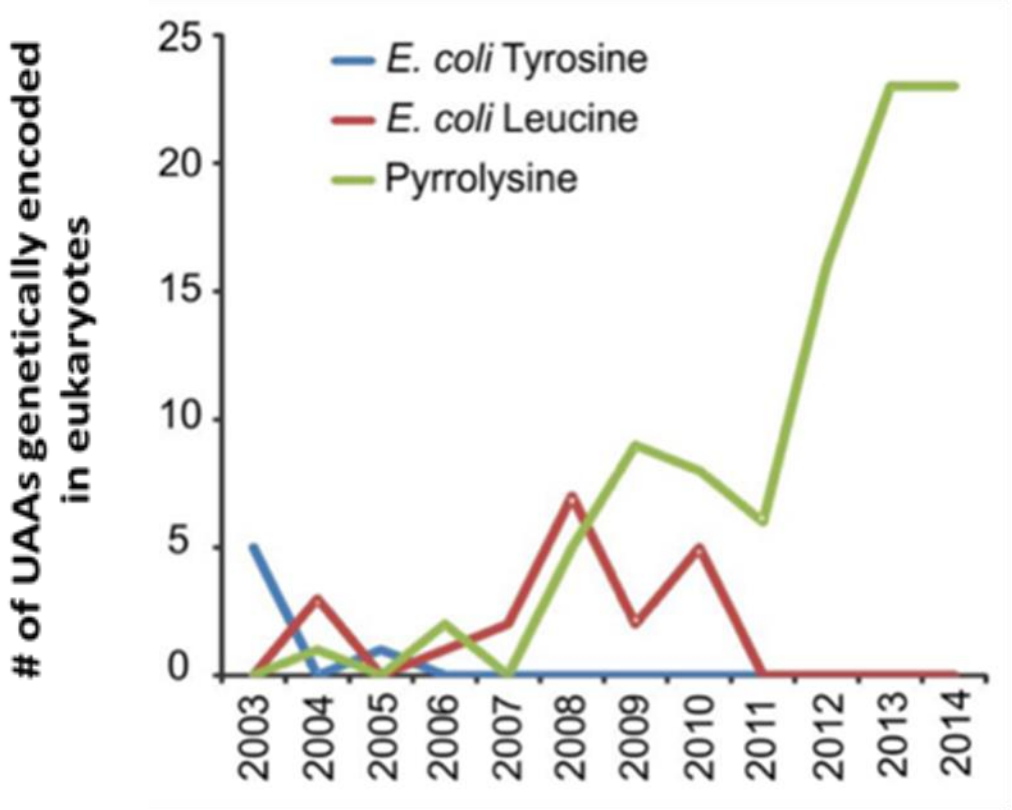


Figure 1.13. aaRS/tRNA pairs used for GCE in eukaryotes between 2003-2015. [7, 8, 26, 46]

1.2 Altered Translational Machinery *E. coli* strains

1.2.1 Limitations of current GCE platforms

The two *in vivo*-selection systems (*E. coli* and *S. cerevisiae*) developed for the generation of UAA-specific aaRS variants have jointly led to the successful incorporation of >160 UAAs through GCE.[7, 8, 26, 46] Yet, a limited set of these UAAs are available for eukaryotic incorporation because of the cumbersome nature of the *S. cerevisiae* platform relative to the *E. coli* platform. Furthermore, the majority of incorporated UAAs for eukaryotic GCE are accomplished with the *Mm*PylRS/*Mb*PylRS. This results in structural limitations of the UAAs because they have to be based around the same few core motifs of the pyrrolysine structure. Unless new orthogonal aaRS/tRNA pairs are added to the repertoire, many structures and functionalities cannot be incorporated. For instance, there are many naturally occurring PTMs, or close structural mimics, that cannot be incorporated in eukaryotes through GCE because no orthogonal aaRS/tRNA pair exists. One noticeably missing example is phosphotyrosine, which has been successfully incorporated in *E. coli*.[14] Tyrosine phosphorylation is an important PTM that plays key roles in the mechanisms used to control a wide variety of biological processes,[14, 47-51] but how they specifically function remains poorly understood. The ability to site-specifically incorporate phosphotyrosine, amongst other PTMs, will help give us insight into their physiological roles by studying them in their native context. Efforts towards overcoming this limitation for eukaryotic GCE will be discussed in more detail in chapters 2-4 of this dissertation.

Another challenge faced in the field of GCE has been creating streamlined platforms to improve the suppression efficiency of UAA-specific aaRS variants. Current selection platforms rely heavily on qualitative measurements of aaRS suppression efficiency (e.g., cell survival), making it difficult to discern between the dynamic range in activity of different aaRS variants.[52]

Little effort has been made towards transitioning these platforms to quantitative measurements.[52-54] In this dissertation, I disclose efforts made to improve the suppression efficiencies of engineered aaRS variants through two separate approaches: (i) the development of quantitative selection platforms and (ii) through choosing an optimal parental aaRS prior to directed evolution.

1.2.2 Engineering of ATM E. coli strains

Dominance of the *Mm*PylRS/*Mb*PylRS pairs for eukaryotic GCE has limited the structural diversity of incorporated UAAs. As mentioned before, this impairs our ability to harness the power of this technology for understanding the physiological roles of proteins in their native context. The slanted preference for the *Mm*PylRS/*Mb*PylRS pairs in eukaryotic GCE is due to their natural orthogonality in both prokaryotes and eukaryotes,[34, 53] enabling them to be initially engineered in the much more facile *E. coli* system to genetically encode new UAAs. These UAA-selective aaRS variants can be subsequently used in any type of eukaryotic application.

Our lab sought to overcome the inherent limitations of having two separate platforms for bacterial and eukaryotic GCE by engineering *E. coli* strains that could serve as host organisms for the directed evolution of any aaRS/tRNA pair, regardless of its domain of origin. This strategy relied on the functional replacement of an endogenous aaRS-tRNA pair of *E. coli* with an archaeal or eukaryotic counterpart. Work towards generating strains of a similar nature was accomplished previously, but these strains frequently experienced growth defects making them unamenable to GCE platforms.[55] Our novel approach focused on first optimizing the performance of the substituted aaRS/tRNA pair in *E. coli* to alleviate any possible growth penalties. Previously engineered archaeal/eukaryotic aaRS/tRNA pairs were targeted as candidates for functional replacement.[29, 33] Once that had been accomplished, the replaced *E. coli* aaRS/tRNA pair could be reintroduced as an amber suppressor and engineered for UAA incorporation (Figure 1.14). This

strategy was successfully applied to create two separate "altered translational machinery" (ATM) *E. coli* strains for the engineering of *Ec*TyrRS and *E. coli* tryptophanyl-tRNA synthetase (*Ec*TrpRS).[56, 57]

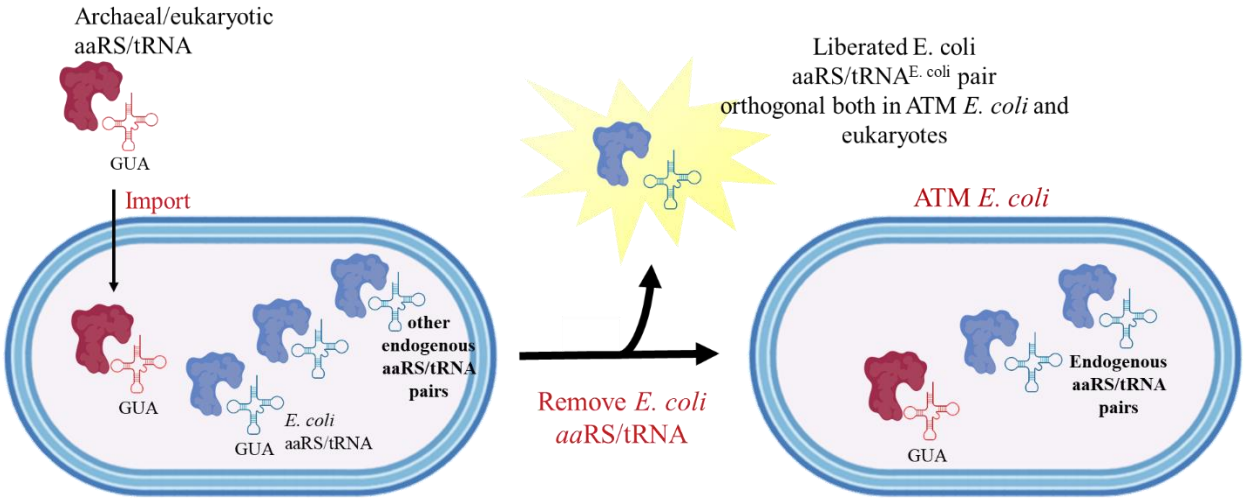


Figure 1.14. General strategy to create ATM *E. coli* strains. [56, 57] An evolutionarily distant aaRS/tRNA pair is imported into *E. coli* so that it can functionally replace one of the aaRS/tRNA pairs. This "liberated" pair can now be reintroduced into the strain as a nonsense suppressor.

1.2.3 Limitations of ATMY E. coli

Even though we engineered these ATM *E. coli* strains with minimal growth penalty, reintroduction of the tRNA$^{E.coli}$ as an amber suppressor, in both ATM-tyrosine (ATMY) and ATM-tryptophan (ATMW), lead to cross-reactivity with the endogenous *E. coli* glutaminyl-tRNA synthetase (*Ec*GlnRS)/tRNA pair. [56, 57] By altering the tRNA$^{E. coli}$ to an amber suppressor (TAG), we were inherently increasing its cross-reactivity with *Ec*GlnRS because this aaRS selectively recognizes the central U residue of its cognate tRNA$^{EcGkn}_{CUG}$.[58, 59] This hypothesis was confirmed by MS analysis and with a chloramphenicol resistance-based activity assay in both ATMW and ATMY with their amber suppressor tRNAs (Figure 1.15). In ATMW *E. coli*, orthogonality of the tRNA^{EcTrp} was easily recovered by switching the tRNA to an opal suppressor (UGA).[56] However, this same solution could not be used for the ATMY *E. coli* strain because reassigning the

tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ as an opal suppressor would result in loss of recognition by *Ec*TyrRS.[57] Instead, we chose to overexpress the native tRNA$^{Ec\text{Gln}}$ to mitigate the observed cross-reactivity to a level sufficient enough for the directed evolution of highly active *Ec*TyrRS/tRNA pairs.[57] While we were able to achieve a "quick fix" for engineering *Ec*TyrRS/tRNA pairs, moving forward a more sustainable option needs to be created that will eliminate all cross-reactivity with the endogenous *Ec*GlnRS.

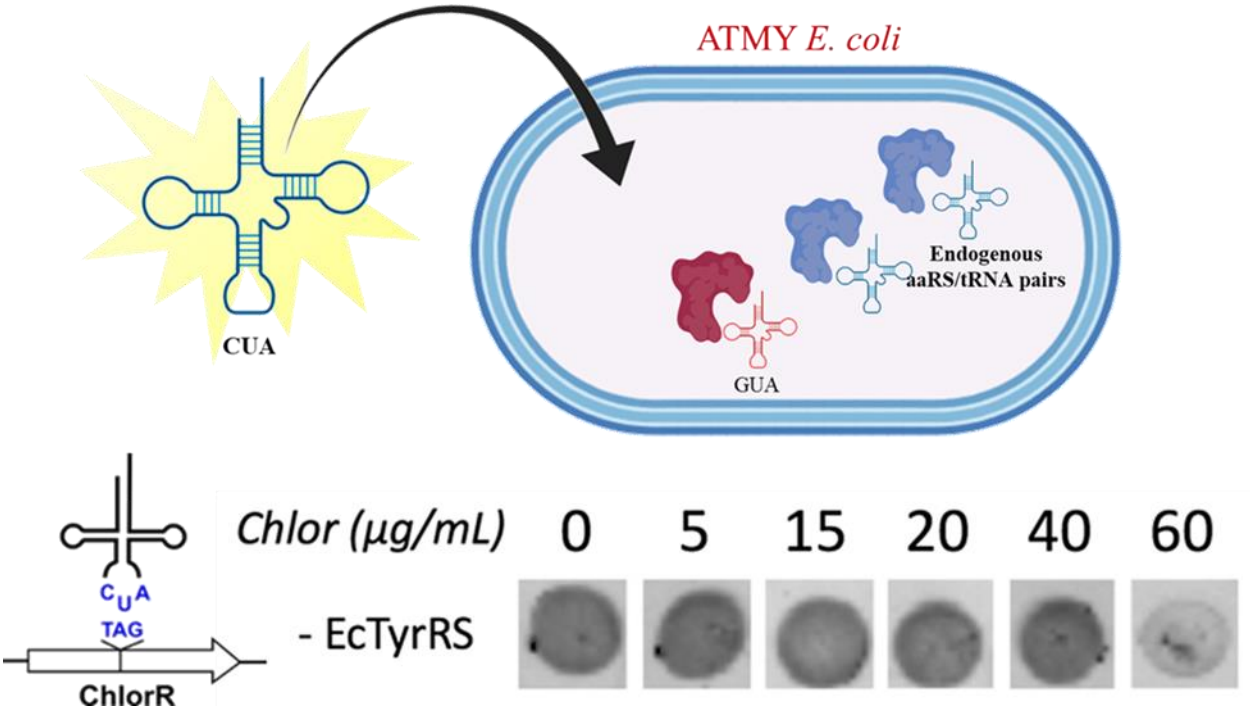


Figure 1.15. Chloramphenicol resistance-based assay to evaluate the cross-reactivity of *Ec*GlnRS with tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$.[57] Each spot is an inoculated cell culture on varying concentrations of chloramphenicol. The cells are only able to grow up if the amber stop codon tRNA gets charged with an amino acid. In the absence of its synthetase, the amber stop codon tRNA is being charged at a high enough level that the cells are able to grow up to 60 μg/mL of chloramphenicol. This is too high of a concentration of antibiotic to work with for engineering synthetases, ideally, we would want background suppression level less than or equal to 10 μg/mL of chloramphenicol.

1.3 References

1. Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darnell, J. Molecular Cell Biology. 4th edition. New York: W. H. Freeman (2000).
2. Della Corte, C.M., Viscardi, G., Di Liello, R., Fasano, M., Martinellia, E., Troiana, T., Ciardiello, F., Morgillo, F. Role and targeting of anaplastic lymphoma kinase in cancer. Mol Cancer 17, 30 (2018).
3. Guccione, E., Richard, S. The regulation, functions and clinical relevance of arginine methylation. Nat Rev Mol Cell Biol 20, 642 (2019).
4. Michalak, E.M., Burr, M.L., Bannister, A.J. et al. The roles of DNA, RNA and histone methylation in ageing and cancer. Nat Rev Mol Cell Biol 20, 573 (2019).
5. Rape, M. Ubiquitylation at the crossroads of development and disease. Nat Rev Mol Cell Biol 19, 59 (2018).
6. Thompson, A. A Targeted Agent for Sickle Cell Disease—Changing the Protein but Not the Gene. N Engl J Med 381, 579 (2019).
7. Chin, J.W. Expanding and Reprogramming the Genetic Code of Cells and Animals. Annu Rev Biochem 83, 379 (2014).
8. Dumas, A., Lercher, L., Spicer, C.D., Davis, B.G. Designing logical codon reassignment - Expanding the chemistry in biology. Chem Sci. 6(1), 50 (2015).
9. Spicer, C., Davis, B. Selective chemical protein modification. Nat Commun 5, 4740 (2014).
10. Dawson, P.E., Kent, S.B.H. Synthesis of Native Proteins by Chemical Ligation. Annu Rev Biochem. 69, 923 (2000).
11. Kent, S.B. Chemical synthesis of peptides and proteins. Annu Rev Biochem. 57, 957 (1988)
12. Furman, J. L., Kang, M., Choi, S., Cao, Y., Wold, E. D., Sun, S. B., Smider, V. V., Schultz, P. G., Kim, C. H. A genetically encoded aza-Michael acceptor for covalent cross-linking of protein-receptor complexes. J Am Chem Soc. 136(23), 8411 (2014).
13. Supekova, L., Zambaldo, C., Choi, S., Lim, R., Luo, X., Kazane, S.A., Young, T.S., Schultz, P.G. The genetic incorporation of *p*-azidomethyl-L-phenylalanine into proteins in yeast. Bioorganic & Medicinal Chemistry Letters 28(9), 1570 (2018).
14. Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., Liu, T., Lyu, X., Du, J., Xuan, W., Yao, A., Reed, S. A., Kang, M., Zhang, Y., Guo, H., Huang, C., Yang, P. Y., Wilson, I. A., Schultz, P. G., Wang, F. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. Nat Chem Biol. 13(8), 845 (2017).
15. Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S. M., Huguenin-Dezot, N., Muqit, M. M., Fry, A. M., Bayliss, R., Chin, J. W. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat Chem Biol 11(7), 496 (2015).
16. Chin, J.W., Cropp, T.A., Anderson, C., Mukherji, M., Schultz, P.G. An expanded eukaryotic genetic code. Science 301, 964 (2003).
17. 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 *Escherichia coli*. PNAS 99 (17). 11020 (2002).
18. Copeland, P. Making sense of nonsense: the evolution of selencysteine usage in proteins. Genome Biol 6, 221 (2005).

19. Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-tRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. Biochim. Biophys. Acta 1844 (6), 1059 (2014).
20. Tan, K. A., Rajagopal, I., Tan, T. Structure & Function - Amino Acids. https://chem.libretexts.org/@go/page/7809 (2019).
21. Cooper, G.M. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates (2000).
22. Prabakaran, S., Lippens, G., Steen, H., Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med. 4 (6), 565 (2012).
23. Chalker, J.M., Davis, B.G. Chemical mutagenesis: selective post-expression interconversion of protein amino acid residues. Curr Opin Chem Biol 14(6), 781 (2010).
24. Neet, K.E., Koshland, D.E. Jr. The conversion of serine at the active site of subtilisin to cysteine: a "chemical mutation". Proc Natl Acad Sci U S A. 56(5), 1606 (1966).
25. Polgar, L., Bender, M.L. A New Enzyme Containing a Synthetically Formed Active Site. Thiol-Subtilisin[1]. J. Am. Chem. Soc. 88(13), 3153 (1966).
26. Wang, L. Engineering the Genetic Code in Cells and Animals: Biological Considerations and Impacts. Acc Chem Res. 50(11), 2767 (2017).
27. Kwok, Y., Wong, J.T.F., Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes. Canadian Journal of Biochemistry 58(3), 213 (1980).
28. Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S., Schultz, P.G. An expanded genetic code with a functional quadruplet codon. PNAS 101(20), 7566 (2004).
29. Chatterjee, A., Xiao, A., Yang, P.Y., Soundararajan, G., Schultz, P.G. A Tryptophanyl-tRNA Synthetase/tRNA pair for Unnatural Amino Acid Mutagenesis in E. coli. Angew Chem Int Ed Engl. 52(19), 5106 (2013).
30. Chatterjee, A., Xiao, H., Schultz, P.G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. Proc Natl Acad Sci U S A 109, 14841 (2012).
31. Italia, J.S., Addy, P.S., Erickson, S.B., Peeler, J.C., Weerapana, E., Chatterjee, A. Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. J. Am. Chem. Soc. 141(15), 6204 (2019).
32. Wang, F. Robbins, S., Guo, J., Shen, W., Schultz, P.G. Genetic Incorporation of Unnatural Amino Acids into Proteins in *Mycobacterium tuberculosis*. PLoS One 5(e9354), (2010).
33. Wang, L., Brock, A., Herberich, B., Schultz, P.G. Expanding the Genetic Code of *Escherichia coli*. Science 292, 498 (2001).
34. (a) Kavran, J.M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., Steitz, T.A. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proc Natl Acad Sci U S A. 104(27), 11268 (2007). (b) Suzuki, T., Miller, C., Guo, L. T., Ho, J., Bryson, D. I., Wang, Y. S., Liu, D. R., & Söll, D. Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase. Nat Chem Biol. 13(12), 1261 (2017).
35. Zheng, Y., Gilgenast, M.J., Hauc, S., Chatterjee, A. Capturing post-translational modification-triggered protein-protein interactions using dual noncanonical amino acid mutagenesis. ACS Chemical Biology 13(5), 1137 (2018).
36. Liu, C.C., Jewett, M.C., Chin, J.W., Voigt, C.A. Toward and Orthogonal Central Dogma. Nat Chem Biol. 14(2), 103 (2018).

37. Zheng, Y., Mukherjee, R., Chin, M.A., Igo, P., Gilgenast, M.J., Chatterjee, A. Expanding the scope of single and double noncanonical amino acid mutagenesis in mammalian cells using orthogonal polyspecific leucyl-tRNA synthetases. Biochemistry 57(4), 441 (2018).
38. Chatterjee, A., Sun, S.B., Furman, J.L., Xiao, H., Schultz, P.G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*. Biochemistry 52(10), 1828 (2013).
39. Dunkelmann, D.L., Willis, J.C.W., Beattie, A.T., Chin, J.W. Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. Nat. Chem. 12, 535 (2020).
40. Wan, W., Huang, Y., Wang, Z., Russell, W.K., Pai, P.J., Russell, D.H., Liu, W.R. A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in *Escherichia coli*. 122, 3279 (2010).
41. Wu, B., Wang, Z., Huang, Y., Lu, W.R. Catalyst-Free and Site-Specific One-Pot Dual Labeling of a Protein Directed by Two Genetically Incorporated Noncanonical Amino Acids. Chembiochem 13(10), 1405 (2012).
42. Bryson, D.I., Fan, C., Guo, L.T., Miller, C., Söll, D., Liu, D.R. Continuous directed evolution of aminoacyl-tRNA synthetases. Nat Chem Biol. 13(13), 1253 (2017).
43. Esvelt, K., Carlson, J., Liu, D. A system for the continuous directed evolution of biomolecules. Nature 472, 499 (2011).
44. Edwards H., Schimmel P. A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase. *Mol. Cell. Biol.* 10, 1633 (1990).
45. Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I., & Yokoyama, S. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic acids res. 30*(21), 4692 (2002).
46. Liu, C. C., Schultz, P. G. Adding New Chemistries to the Genetic Code. Annu. Rev. Biochem. 79, 413 (2010).
47. Alonso A, et al. Protein tyrosine phosphatases in the human genome. Cell 117, 699 (2004).
48. Eckhart, W., Hutchinson, M.A., Hunter, T. An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. Cell. 18, 925 (1975).
49. Hunter, T., Cooper, J.A. Protein-tyrosine kinases. Annu Rev Biochem. 54, 897 (1985).
50. Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. The protein kinase complement of the human genome. Science 298, 1912 (2002).
51. Pawson, T. Specificity in signal transduction: From phosphotyrosine-SH2 domain interactions to complex cellular systems. Cell 116, 191 (2004).
52. Potts, K.A., Stieglitz, J.T., Lei, M., Van Deventer, J.A. Reporter system architecture affects measurements of noncanonical amino acid incorporation efficiency and fidelity. Mol. Syst. Des. Eng. 5, 573 (2020).
53. Hohl, A., Karan, R., Akal, A. Renn, D., Liu, X., Ghorpade, S., Groll, M., Rueping, M., Eppinger, J. Engineering a Polyspecific Pyrrolysyl-tRNA Synthetase by a High Throughput FACS Screen. Sci Rep 9, 11971 (2019).
54. Owens, A.E., Grasso, K.T., Ziegler, C.A., Fasan, R. Two-Tier Screening Platform for Directed Evolution of Aminoacyl-tRNA Synthetases with Enhanced Stop Codon Suppression Efficiency. ChemBioChem 18(12), 1109 (2017).

55. Iraha, F., Oki, K., Kobayashi, T., Ohno, S., Yokogawa, T., Nishikawa, K., Yokoyama, S., Sakamoto, K. Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion. *Nucleic Acids Res. 38*(11), 3682 (2010).
56. Italia, J.S., Addy, P.S., Wrobel, C.J.J., Crawford, L.A., Lajoie, M.J., Zheng, Y., Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol. 13, 446 (2017).
57. Italia, J.S., Latour, C., Wrobel, C.J.J., Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. Cell Chemical Biology 25(10), 1304 (2018).
58. Jahn, M., Rogers, M.J., Söll, D. Anticodon and acceptor stem nucleotides in tRNA(Gln) are major recognition elements for E. coli glutaminyl-tRNA synthetase. Nature 352, 258 (1991).
59. M.J., Adachi, T., Inokuchi, H., Söll, D. Switching tRNA(Gln) identity from glutamine to tryptophan. Proc. Natl. Acad. Sci. USA 89, 3463 (1992).

Chapter 2

Evolution of an orthogonal *E. coli* tyrosyl suppressor tRNA

2.1 Introduction

The principles of GCE are described in detail in Chapter 1. This chapter focuses on the aforementioned ATMY platforms, complications which arose from cross-reactivity, and how that was overcome.

2.1.1 Reintroduction of $tRNA^{EcTyr}_{CUA}$ results in cross-reactivity with EcGlnRS

The development of ATM *E. coli* strains enabled the successful incorporation of a broad range of previously inaccessible UAAs in mammalian cells (Figure 2.1).[1, 2] Pivotal to this technology was the identification of heterologous aaRS/tRNA pairs that could functionally replace the endogenous aaRS/tRNA pair.[3-5] This was achieved for both ATMW and ATMY *E. coli* strains by utilizing eukaryotic or archaeal aaRS/tRNA pairs that had been previously engineered for GCE in *E. coli*.[1, 2] In both of these ATM strains, the reintroduction of the respective *E. coli* $tRNA_{CUA}$ resulted in cross-reactivity with the endogenous *Ec*GlnRS (Figure 2.2).[1, 2] The tRNA anticodon is a key recognition element for the *Ec*GlnRS to identify its $tRNA^{EcGln}$.[6, 7] Specifically, the *Ec*GlnRS hones in on the central "U" of the anticodon loop.[6, 7] By changing the *E. coli* tryptophanyl and tyrosyl tRNA anticodons to "CUA," they become at risk to cross-react with the *Ec*GlnRS. In the ATMW system, the cross-reactivity of the $tRNA^{EcGln}$ was easily remedied by reassigning the suppressor tRNA from the amber stop codon "CUA" to the opal stop codon "UCA."[1] However, reassignment of the anticodon to an opal suppressor was not a translatable solution to the ATMY system because changing the central nucleotide in the anticodon loop of the $tRNA^{EcTyr}$ to a "C" would render the tRNA inactive.[2] A "quick fix" for the ATMY system was achieved by overexpressing the wild-type $tRNA^{EcGln}$.[2] While a host of UAAs were incorporated in the ATMY system, a few notable exceptions were missing. For example, Chin *et al.* previously engineered

the *Ec*TyrRS in *S. cerevisiae* to incorporate the photoaffinity probe *p*-benzoyl-L-phenylalanine (pBPA), but this same *Ec*TyrRS mutant could not be engineered with our ATMY selection-system due to cross-reactivity between the $tRNA^{EcTyr}_{CUA}$ and *Ec*GlnRS (Figure 2.2).[8] Further engineering of the $tRNA^{EcTyr}_{CUA}$ was required to restore orthogonality and gain access to the full range of UAAs available for GCE in ATMY.

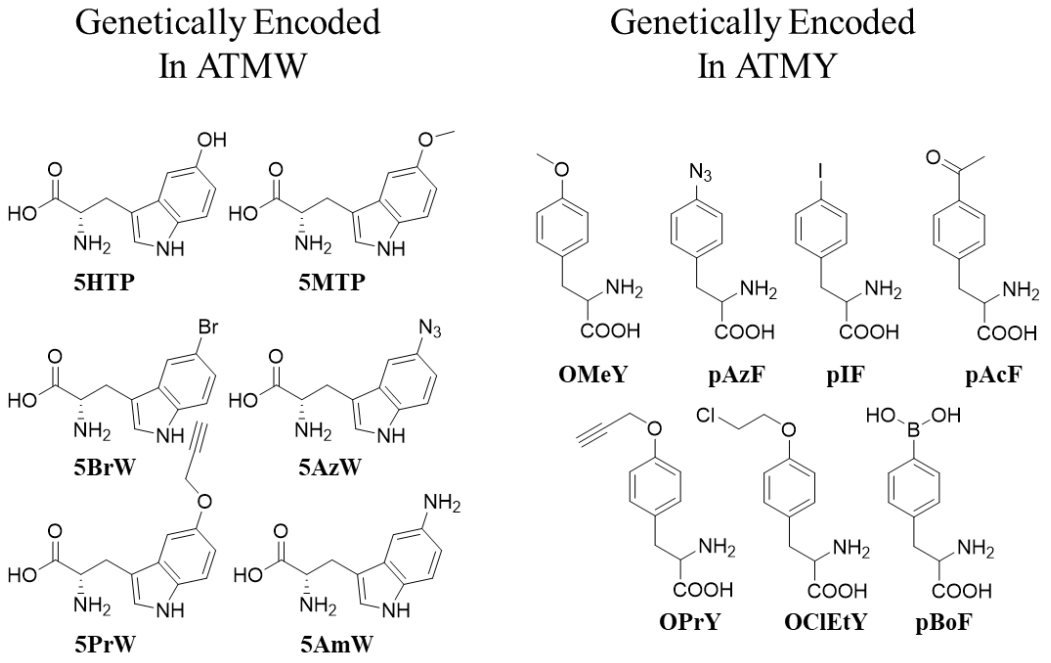


Figure 2.1. Structures of the UAAs incorporated with the newly developed ATM *E. coli* strains.[1, 2] All of the UAAs must be derived from the structure of the cognate amino acid their aaRS naturally charges.

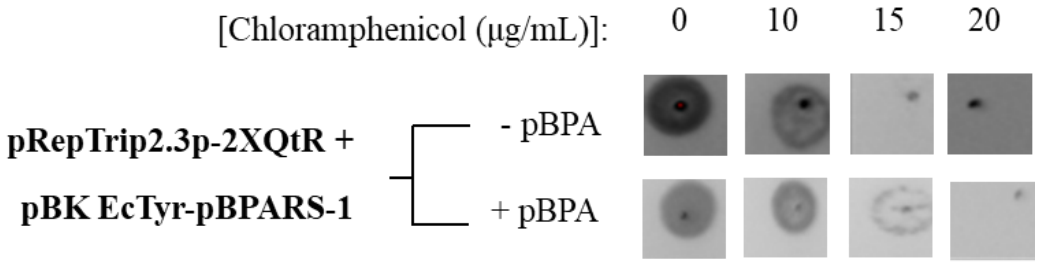


Figure 2.2. Cross-reactivity of the $tRNA^{EcTyr}_{CUA}$ with *Ec*GlnRS. A chloramphenicol acetyl-transferase (CAT)-TAG reporter assay shows that the cross-reactivity of the $tRNA^{EcTyr}_{CUA}$ with the *Ec*GlnRS masks the suppression activity of weakly active aaRSs, such as the *Ec*TyrRS mutant that charges pBPA. The overexpression of the $tRNA^{EcGln}$ in the pRep reporter plasmid does not attenuate this cross-reactivity enough for it to function as a viable selection platform for *Ec*TyrRS directed evolution.

2.1.2 *Structure and role of tRNA^{EcTyr} in the cell*

In the cell, tRNA functions as an adaptor molecule which translates the genetic code into proteins during protein synthesis.[9] Across all kingdoms of life, most tRNAs fold into a cloverleaf-shaped secondary structure consisting of several stem-loop motifs.[10] These tRNA structures play critical roles in regulating tRNA acylation through interactions with the tRNA's endogenous aaRS structures. To engineer the tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ for improved orthogonality, these key interactions between the tRNA$^{Ec\text{Tyr}}$ and *Ec*TyrRS must be upheld. For example, bacterial TyrRSs contain an extra C-terminal domain that is stabilized in between the elbow of the variable loop arm and the

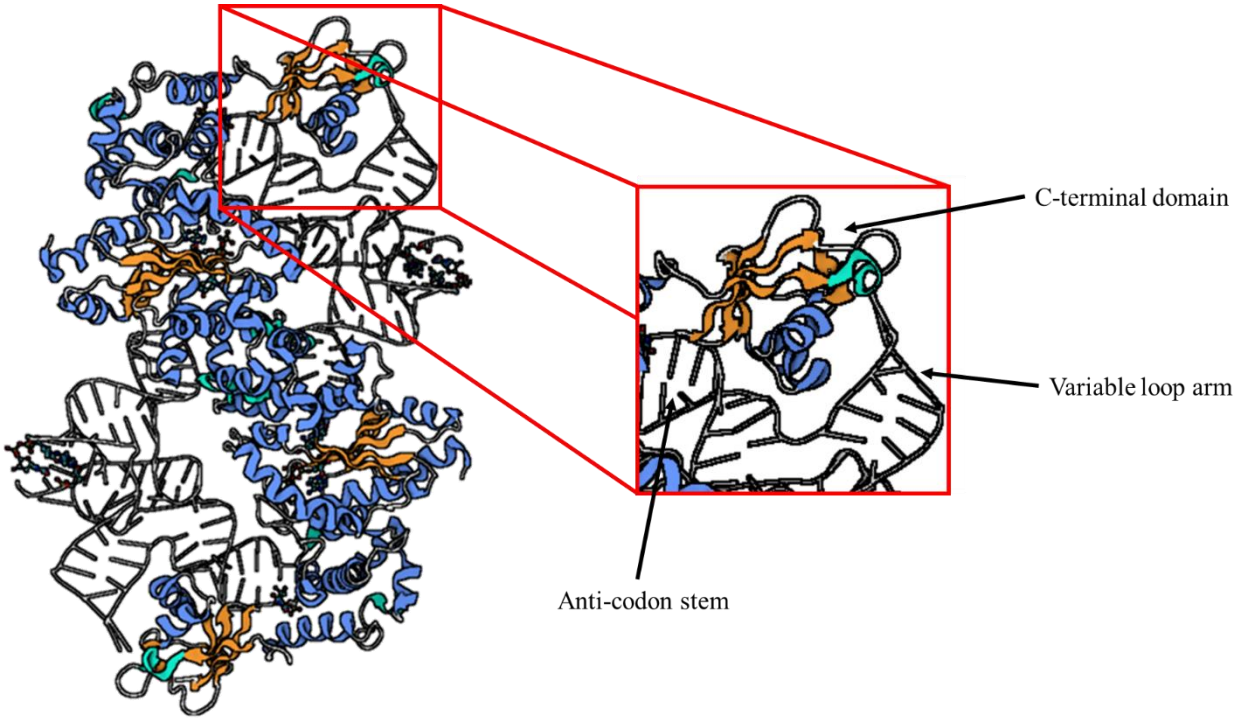


Figure 2.3. The crystal structure (PDB: 1H3E) of tyrosyl-tRNA synthetase from *Thermus thermophilus* complexed with its cognate tRNA$^{Tyr}_{GUA}$, ATP, and tyrosinol.[12] This was the first time that a bacterial TyrRS C-terminal domain had been successfully crystallized, which was difficult due to its highly flexible nature.

anti-codon stem during tRNA^{Tyr}-acylation (Figure 2.3).[11, 12] This unique binding pattern, in bacterial TyrRSs, aids in the recognition of the bacterial tRNA^{Tyr}.[11-13] Eukaryotic and archaeal TyrRSs have evolved either a different or no C-terminal domain, eliminating the possibility of

cross-reacting with bacterial TyrRSs. Other known key tRNA^{Tyr} recognition elements are the first base pair of the cognate acceptor stem (G1-C72 in prokaryotes/mitochondria, C1-G72 in eukaryotes/archaea) and the anticodon bases C34 and Ψ35 (Figure 2.4).[11-13] As mentioned previously, we aimed to preserve the elements listed above when engineering the tRNA$^{Ec\text{Tyr}}$ to maintain fidelity with the *Ec*TyrRS.

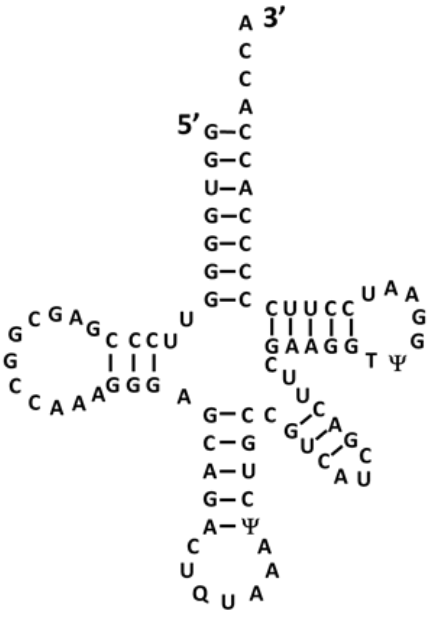


Figure 2.4. The sequence and secondary structure of the cognate tRNA$^{Ec\text{Tyr}}$.

2.1.3 Engineering orthogonal tRNAs

Two approaches can be used for engineering orthogonal tRNAs for GCE applications. The first, more tedious route is to exclusively use rational design based on prior knowledge of the tRNA recognition elements and host transcriptional/translational processing machinery components.[14, 15] One notable example of this approach was when the Schultz group took a *S. cerevisiae*-derived tryptophanyl tRNA (tRNA$^{Sc\text{Trp}}$) and introduced it as a nonsense suppressor into bacteria.[16] Reassignment of the tRNA$^{Sc\text{Trp}}$ as a nonsense suppressor led to cross-reactivity with the *E. coli* endogenous lysl-tRNA synthetase.[16] However, modification of the G:C content of the anticodon

stem, and ultimately the flexibility, recovered orthogonality of the *S. cerevisiae* tryptophanyl-tRNA synthetase (*Sc*TrpRS)/ $tRNA^{ScTrp}_{CUA}$ pair in bacteria.[16]

The more powerful route to engineer orthogonal tRNAs is through directed evolution. In bacteria, it has been repeatedly proven that randomized tRNA libraries, which target key structural identity elements of the tRNA, can be selected for enhanced tRNA suppression efficiencies and orthogonality.[17-20] The main regions that have successfully generated orthogonal tRNA variants include the T-stem, anticodon stem, and the variable loops, all of which are critical for EF-Tu binding or tertiary structural integrity.[17-20] The pool of randomized tRNA is designed based on the conserved loop regions and identity elements of the tRNA in question to maintain both structural integrity and recognition with its cognate aaRS. Since tRNA-directed evolution using large libraries has generally been inaccessible in mammalian systems, the development of orthogonal-tRNAs has been largely limited to the less favorable, rational design approach for mammalian applications.

Once a tRNA library is built, a robust selection methodology must be employed to sieve out the active, orthogonal tRNA library members. A viable platform requires an observable output signal tied to the suppressor tRNAs activity and the ability to subsequently isolate individual mutant library members following selection. Based on these premises, we hypothesized that recovery of the orthogonality of the $tRNA^{EcTyr}_{CUA}$ would be feasible through directed evolution in ATMY *E. coli*. This would be achieved by first rationally designing a $tRNA^{EcTyr}_{CUA}$ library that maintains key recognition elements of the *Ec*TyrRS,[11-13] but is randomized in other of the $tRNA^{EcTyr}_{CUA}$ regions that could be cross-reacting with the *Ec*GlnRS. Then, a well-established antibiotic-based screening platform[1-3] would be used to screen the $tRNA^{EcTyr}_{CUA}$ library.

Ultimately, we anticipated that an orthogonal tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ would be created through directed evolution for future use in synthetase engineering applications.

2.2 Results and discussion

2.2.1 Directed evolution of an orthogonal tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ *in ATMY5*

The feasibility of evolving suppressor tRNAs for novel functions has been shown on numerous occasions.[18, 19, 21, 22] Additionally, key *Ec*TyrRS recognition elements- such as Ade73 and the base-pair Gua1:Cyt72- of the tRNA$^{Ec\text{Tyr}}$ have been well characterized through x-ray crystallography.[11-13] With this knowledge in hand, we designed a scheme to engineer a more orthogonal tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ in ATMY5 *E. coli* (EcNR1GT pUltraBR-MjY dtyrS tryTV::tRNA$^{\text{EcTyr}}_{\text{Inact}}$ tyrU::tRNA$^{\text{EcTyr}}_{\text{Inact}}$ lambda::tolc) (Figure 2.5).[2] First, a tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ library was generated (theoretical diversity 1.05 x 10^6, > 99% coverage) in which five base pairs of the acceptor stem (2-6, 67-71) were randomized without consideration for base-pairing interactions. Enrichment of the tRNA library for active hits was conducted with an established chloramphenicol-acetyl transferase-TAG (CAT-TAG) expression assay in the presence of an *Ec*TyrRS mutant previously engineered to incorporate o-methyl-L-tyrosine (*Ec*Tyr-OMeYRS) (Figure 2.6).[2] The *Ec*Tyr-OMeYRS was chosen due to its high suppression efficiency of the tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ with OMeY.

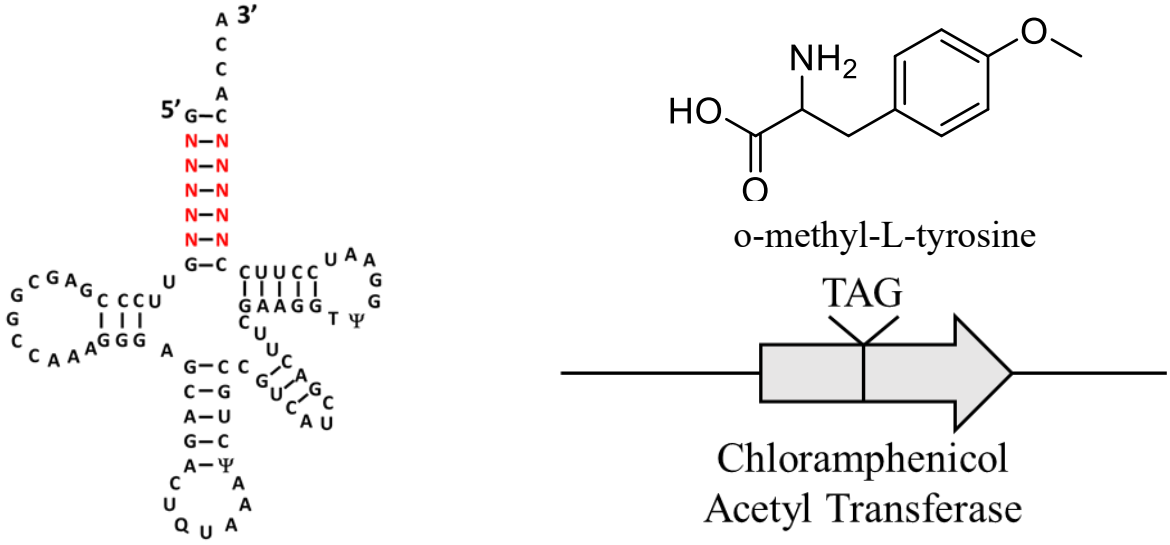
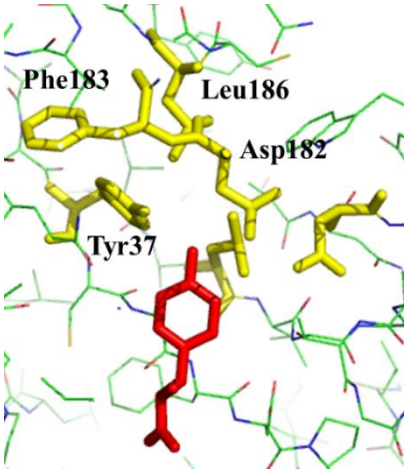


Figure 2.5. Strategy to evolve the tRNA$^{EcTyr}_{CUA}$. The acceptor stem of the tRNA$^{EcTyr}_{CUA}$ was randomized through directed evolution. Positive selection conditions in the presence of 1 mM OMeY led to the suppression of the CAT-TAG reporter and, ultimately antibiotic resistance.



	Tyr37	Asp182	Phe183	Leu186
*Ec*Tyr-OMeYRS	Val	Ser	Met	Leu

Figure 2.6. The *Ec*TyrRS active site and the active site mutations of *Ec*Tyr-OMeYRS.

Elimination of cross-reactive tRNA library members was accomplished with the TAG-inactivated toxic barnase gene, accompanied by a subsequent positive selection on the tRNA library.[2, 3] Individual library member activities were screened in the presence of *Ec*Tyr-OMeYRS with a CAT-TAG assay to identify tRNAs that afforded orthogonality. Only one tRNA sequence

(tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1) (Figure 2.7) conferred a significant improvement in orthogonality relative to the wildtype tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$, and survived up to 30 μg/mL of chloramphenicol (Figure 2.8).

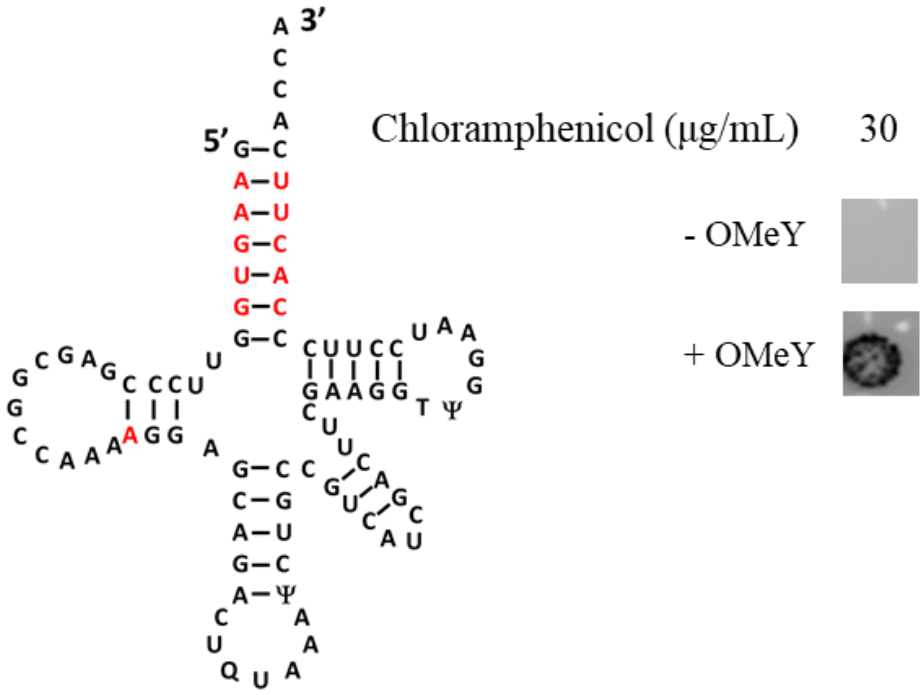


Figure 2.7. The mutant tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1 sequence. The only identified tRNA library member with validated orthogonality at 30 μg/mL of chloramphenicol.

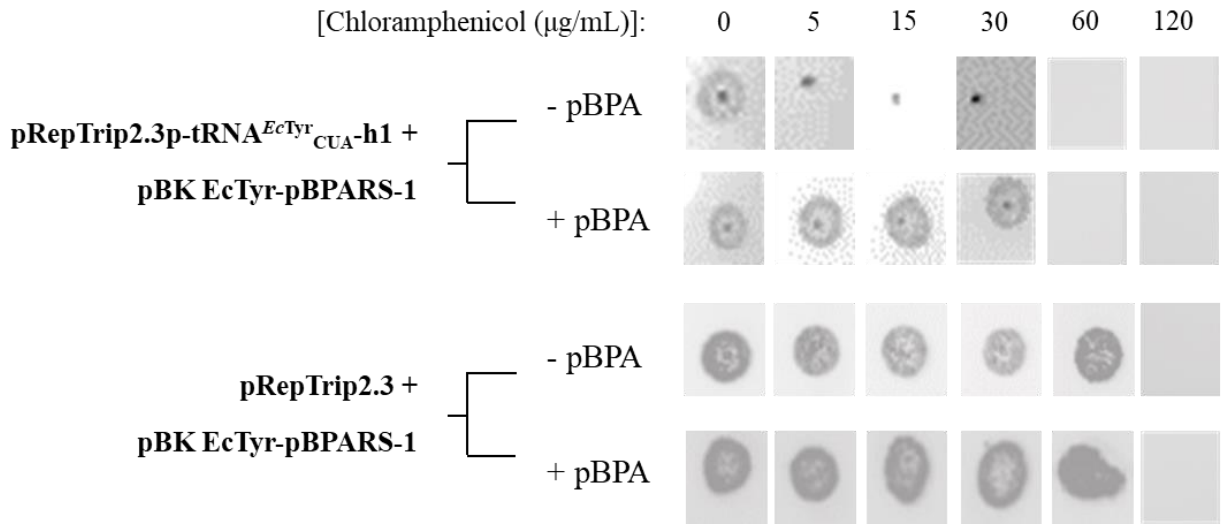


Figure 2.8. Characterization of tRNA$^{EcTyr}_{CUA}$-h1 orthogonality. A chloramphenicol gradient was used to assess the orthogonality of the mutant tRNA$^{EcTyr}_{CUA}$-h1 relative to its progenitor. In the absence of any tRNA^{EcGln}, the tRNA$^{EcTyr}_{CUA}$ + pBPA-selective *Ec*TyrRS mutant leads to cross-reactive growth up to 60 μg/mL of chloramphenicol in both the presence and absence of pBPA (bottom two rows). Cross-reactivity is completely eliminated with the new tRNA$^{EcTyr}_{CUA}$-h1 + pBPA-selective *Ec*TyrRS mutant (top two rows).

The mutant tRNA$^{EcTyr}_{CUA}$-h1 contained a G24A point mutation in the D-stem. This point mutation eliminates base-pairing interactions in the D-stem, most likely leading to tRNA structural integrity issues. To gain insight into how the G24A point mutation in the D-stem affected the tRNA activity, this point mutation was introduced into the progenitor tRNA$^{EcTyr}_{CUA}$. Additionally, the tRNA$^{EcTyr}_{CUA}$-h1 A24 point mutation was reverted to G24. The tRNAs' suppression efficiency was then measured with an sfGFP-151-TAG reporter in ATMY5 *E. coli* (Figure 2.9). It was observed that the introduction of the A24 mutation in the D-stem leads to decreased activity in both the tRNA$^{EcTyr}_{CUA}$-h1 and progenitor tRNA$^{EcTyr}_{CUA}$. Further studies would need to be carried out to fully understand how this mutation fundamentally affects the tRNA structure.

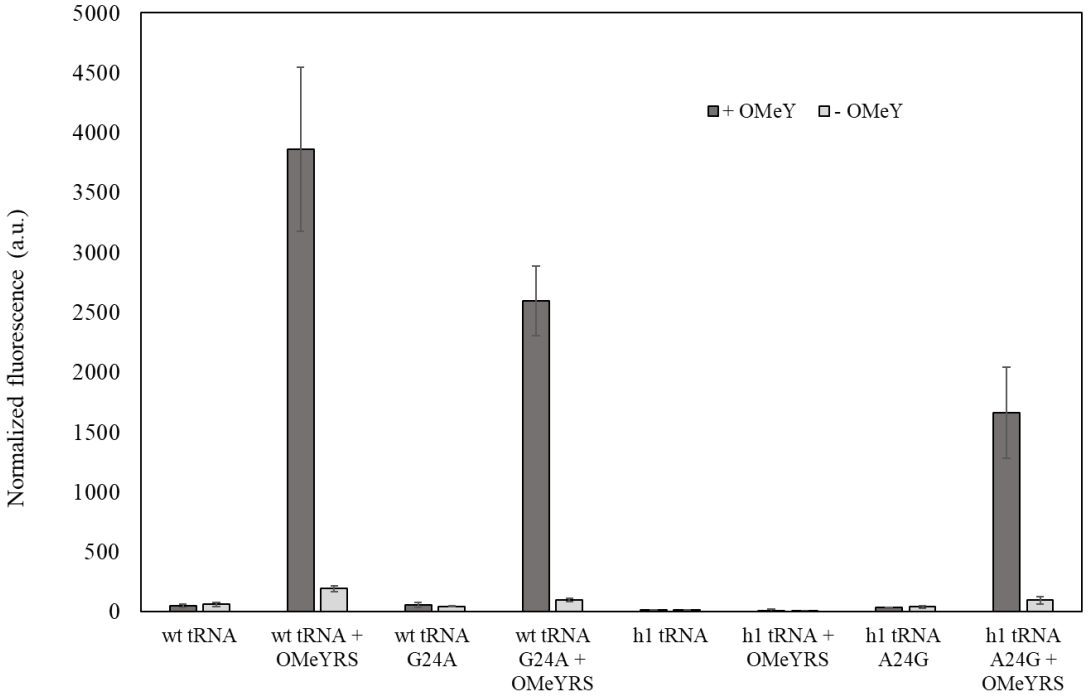


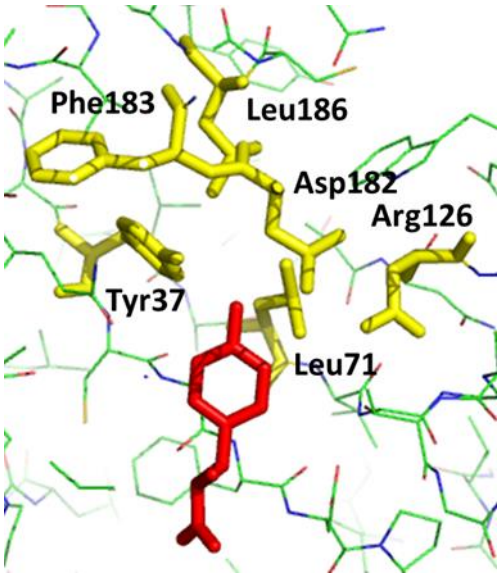
Figure 2.9. Characterization of the tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1 D-stem point mutation. The tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1 (h1 tRNA) and tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ (wt tRNA) were co-transformed with the *Ec*Tyr-OMeYRS into ATMY5. The suppression efficiencies of the tRNA were measured as the normalized fluorescence output of full-length sfGFP in the presence and absence of OMeY.

2.2.2 Engineering of pBPA-selective EcTyrRS mutants

To demonstrate the value of having an orthogonal tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ available for directed evolution in ATMY *E. coli*, we sought to engineer alternative UAA-selective *Ec*TyrRSs that could be used to investigate protein-protein interaction (PPI) networks. PPIs are critical components in numerous biological networks. However, their apparent functions are difficult to decipher due to their transient nature.[23] To overcome this limitation, further methodology must be developed to capture these interactions in their native context. Currently, most methods rely on affinity purification of these transient interactions, which can break down during the purification process

since they are not linked through covalent bonds.[24, 25] This hurdle can be overcome through the introduction of an unbreakable covalent bond, creating a more accurate picture of these protein interaction networks.

Benzophenone photophores provide an attractive strategy to capture protein-protein interactions *in vivo* due to their stability, ability to be manipulated in ambient light, activation at 350-360 nm (which is gentler on proteins), and preferential reactivity with unreactive C-H bonds.[26] An amino acid analog, pBPA has been successfully incorporated through GCE in both bacteria and eukaryotes.[3, 8] However, previous selections with the ATMY3 selection platform yielded no mutant *Ec*TyrRSs that were capable of incorporating this UAA.[2] Ultimately, incorporation was not possible due to high levels of background cross-reactivity between the $tRNA^{EcTyr}_{CUA}$ and *Ec*GlnRS (Figure 2.2). With the new, orthogonal $tRNA^{EcTyr}_{CUA}$-h1, we hypothesized that the engineering of these weaker, currently inaccessible mutant *Ec*TyrRSs with our ATMY5 platform could be achieved. To accomplish this feat, an *Ec*TyrRS library was constructed based on the available crystal structure (PDB: 1X8X), and six key amino acid residues (Tyr37, Leu71, Asn126, Asp182, Phe183, Leu186) in the active site of *Ec*TyrRS were randomized by site-saturation mutagenesis (theoretical diversity 1.06×10^7, > 99% coverage) (Figure 2.10).[27] Additionally, a C-terminal residue (Asp265) was mutated to Arg to enhance TAG-codon recognition.[28] The *Ec*TyrRS library was subjected to a two-tier selection scheme in ATMY5. First, a positive selection (+ 1 mM pBPA) was used to enrich mutants that charged pBPA. A negative selection (- pBPA) was used to remove *Ec*TyrRS variants that charged canonical amino acids. After one round of selection, the mutant library exhibited a significant pBPA-dependent survival, suggesting pBPA-selective *Ec*TyrRS mutants' successful enrichment.



Y37	L71	N126	D182	F183	L186	D265
FLIMVSTAYHCG	NBT	NSPTACGDH	NST	NNK	NNK	R

Figure 2.10. Active site residues of *Ec*TyrRS. The active site was randomized through site-directed mutagenesis.

Individual mutant library member activities were characterized with a CAT-TAG assay (+ 1 mM pBPA), identifying two unique *Ec*TyrRS sequences (Table 2.1). Tyr37 goes from a large, nonpolar amino acid to a small, nonpolar amino acid (Gly/Ala). Leu71 went to a smaller, nonpolar amino acid (Val). Asn126 remained the same. Asp182 is a key residue in the active site that forms a hydrogen bonding network with the phenol -OH in tyrosine.[11] The Asp182 residue was expected to drastically change in order to accommodate the benzophenone structure. In fact, Asp182 changed from a polar, acidic amino acid to smaller amino acid (Ser/Ala). Phe183 remained a large, nonpolar, aromatic amino acid (Phe/Tyr). Leu186 remained a nonpolar amino acid (Ala/Tyr).

Clones	Tyr37	Leu71	Asn126	Asp182	Phe183	Leu186
pBPARS-1*	Gly	Leu	Asn	Gly	Phe	Ala
pBPARS-7	Gly	Val	Asn	Ser	Tyr	Ala
pBPARS-8	Ala	Val	Asn	Ala	Phe	Tyr

*engineered by Chin *et al.*

Table 2.1. The active site mutations for the two newly identified mutants, pBPARS-7 and pBPARS-8. Both mutants had exhibited active sites with residues similar to the previously identified pBPARS-1 active site.

Suppression efficiencies of the three mutants were characterized in ATMY4 (EcNR1GT pUltraBR-MjY dtyrS tryTV::tRNA$^{EcTyr}_{CUA}$ tyrU::tRNA$^{EcTyr}_{CUA}$ lambda::tolc) with an sfGFP-151-TAG reporter (Figure 2.11).[2] Successful incorporation of pBPA into sfGFP-151-TAG was confirmed by fluorescence (Figure 2.11), and all three mutants were observed to have comparable activities. The engineered pBPA-selective *Ec*TyrRS mutants were cloned into the previously described mammalian expression plasmid pB1U to characterize suppression efficiencies in HEK293T cells (Figure 2.11).[2] Expression efficiencies of EGFP-39-pBPA was confirmed through fluorescence imaging (Figure 2.12).

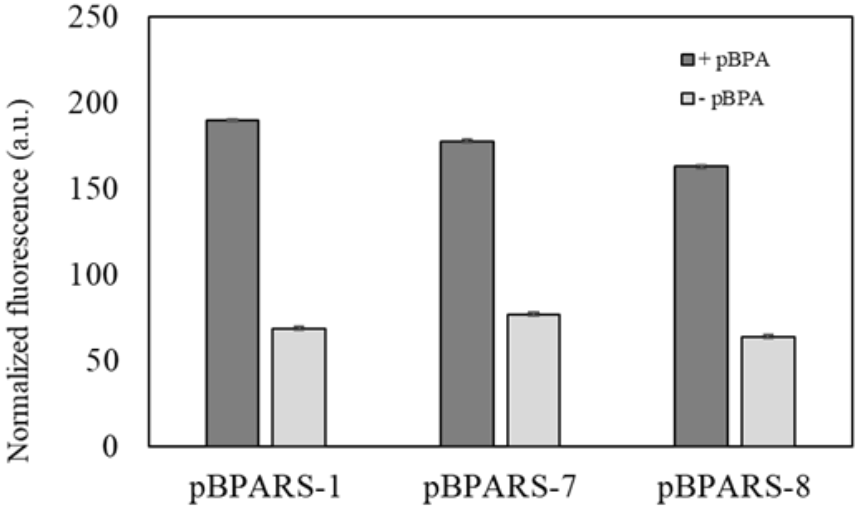


Figure 2.11. Suppression efficiencies of the engineered pBPA-selective mutants. The suppression efficiencies were characterized by sfGFP-151-TAG reporter fluorescence in ATMY4. The two new mutants exhibited similar activity to the previously identified pBPARS-1.

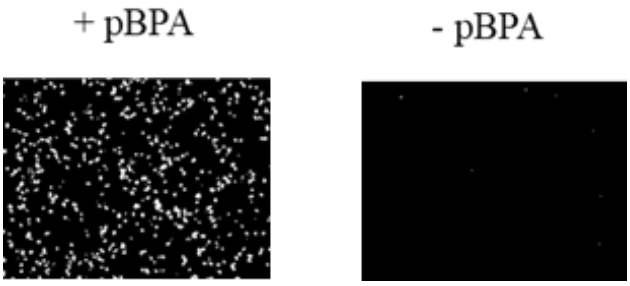


Figure 2.12. Fluorescence microscopy images of HEK293T cells with pBPARS-7 (+/- 1 mM pBPA). These images depict pBPA-dependent EGFP-39-TAG expression.

2.3 Conclusions

In summary, we have engineered an orthogonal tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ for GCE in ATMY5 *E. coli*. The mutant tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1 had an A:U rich acceptor stem and an erroneous G24A mismatch point mutation in the D-stem. It was observed that the introduction of A24 into the D-stem of the tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1 led to lower suppression efficiencies of both the tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1 and its progenitor. Further investigation is required to fully understand the mechanisms behind this observation.

By rescuing the tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$ orthogonality, we demonstrated that weakly active *Ec*TyrRS mutants are accessible with the "universal" ATMY5 platform, furthering its utility. While we have gained access to a broader dynamic range of *Ec*TyrRS mutants, these weakly active variants have little value for mammalian cell applications. However, the ATMY5 platform can now be adapted to improve the weakly active mutants.

2.3.1 Ongoing and future directions

This new ATMY5 platform will be used to evolve novel bacterial TyrRS/tRNA pairs to incorporate a broader range of UAAs in eukaryotes. There are still PTMs, bioconjugation, handles, and other important chemical modalities that have evaded us (e.g., phosphotyrosine[29]). It will also

be adapted for enhancing *Ec*TyrRS mutant suppression efficiencies through directed evolution, among other applications.

2.4 Acknowledgements

Dr. James Italia trained me, helped design the library cloning strategies, helped design the selection scheme, and gave me access to all of his plasmids/primers. Chris Latour helped me work through the tRNA library cloning protocol and assisted in electrocompetent cell preparation. Megan Yeo assisted in the characterization of the $tRNA^{EcTyr}_{CUA}$-h1 D-stem point mutation.

2.5 Experimental procedures

All cloning and plasmid propagation was done in DH10B *E. coli* cells. Restriction enzymes, Phusion HS II High-Fidelity DNA polymerase, and IPTG were obtained from Fisher. T4 DNA ligase was obtained from Enzymatics. DNA extraction and PCR clean up were conducted with Macherey-Nagel Binding Buffer NTI and Epoch mini spin columns from Thermo Fisher Scientific. Media components were obtained from Fisher Scientific.

2.5.1 Strains, cell lines

ATMY5 and ATMY4 *E. coli* cells were obtained from the Chatterjee lab glycerol stocks. These strains were originally engineered by a previous student in the Chatterjee lab, Dr. James Italia. A detailed description of how ATMY5 and ATMY4 *E. coli* were engineered has been published in Dr. Italia's Cell Chemical Biology manuscript.[2]

HEK293T cells (ATCC) were maintained at 37 ºC and 5% CO_2 in DMEM-high glucose (HyClone) supplemented with penicillin/streptomycin (Hyclone, final concentration of 100 U/L penicillin and 100 μg/mL streptomycin) and 10% fetal bovine serum (Corning).

2.5.2 Plasmids

All primer and plasmid sequences can be found in the Appendix.

pRepTrip2.3-EcOMeYRS was constructed by first PCR amplifying EcOMeYRS-VSML out of pBK EcOMeYRS-VSML[2] with the primers BKrep-SpeI-F and BKrep-BglII-R. The PCR product was digested with SpeI/BglII and inserted into the pRep vector backbone.

pRepTrip2.3p-EcYtR-h1 was created by PCR amplifying EcYtR-h1 from pBK EcYtR-h1 with BKrep-SpeI-F and BKrep-BglII-R. The PCR product was digested with SpeI/BglII and inserted into the pRep vector backbone.

pNeg-EcYtR-h1-barnase-2X TAG was created by PCR amplifying EcYtR-h1 from pBK EcYtR-h1 with NEGrep-SphI-F and BKrep-NcoI-R. The PCR product was digested with SphI/NcoI and inserted into the pNeg backbone that was digested with the same restriction enzymes.

pNeg-NoYtR was propagated from a former student's stocks.

The pEvol-sfGFP151-TAG reporter construct was propagated from Dr. James Italia's plasmid stocks by transforming it into DH10B cells, inoculating a culture, and mini-prepping the plasmid DNA.

All pB1U-aaRS plasmids were generated by PCR amplifying the mutant EcTyrRS from its pBK plasmid with primers EcYRS-NheI-F and EcYRS-XhoI-R, followed by digestion with NheI/XhoI. The digested product was then inserted into the pB1U vector backbone.

2.5.3 Unnatural amino acids

The o-methyl-l-tyrosine used in the present experiments was obtained from Fisher Scientific (catalog number AAH6309606).

para-benzoyl-l-phenylalanine was purchased from Chem-Impex International (catalog number 05110).

2.5.4 Assessment of aaRS-tRNA activity using a chloramphenicol reporter

Inoculated 5 mL LB media cultures of ATMY5 harboring pRepTrip2.3-EcYtR-h1 and pBK aaRS were grown overnight. Following overnight growth, the ATMY5 LB media cultures were diluted to an OD600 of 0.03 then 3 µl of this dilution was inoculated on LB agar plates supplemented with kanamycin, spectinomycin, tetracycline, carbenicillin, 0.02% arabinose, varying chloramphenicol concentrations, and +/- 1 mM of UAA. Growth on the plates was analyzed 24 hours following inoculation.

2.5.5 Assessment of aaRS-tRNA activity using a sfGFP-151 reporter

The pEvol-sfGFP-151-TAG reporter and pBK aaRS plasmids were co-transformed into ATMY4 for expression.[2] A 5 mL overnight culture was inoculated with a single colony from the transformation plates. Following overnight growth, 200 µL of the overnight culture was used to inoculate a 20 mL LB media culture supplemented with spectinomycin, kanamycin, and chloramphenicol. This culture was then grown at 37 ºC with shaking (250 rpm) to a final OD600 of 0.6 then induced with IPTG (1 mM) and incubated for 16 hours at 30 ºC with shaking (250 rpm). The cultures were then spun down (4,500 g, 10 min, 4 ºC), the LB media was removed, and the pelleted cells were resuspended in 1X PBS buffer. The resuspended cultures were diluted with

1 X PBS (15 µL culture added to 135 µL 1 X PBS) and fluorescence readings were collected in a 96 well-plate using a SpectraMAX M5 (Molecular Devices) (ex = 488 nm and em = 510 nm). Mean of two independent experiments were reported, and error bars represent standard deviation.

2.5.6 Construction of the pBK tRNA$^{EcTyr}_{CUA}$ library

Site-directed mutagenesis was used to randomize five base-pairs of the acceptor stem of the EcTyr-tRNA to create the library (theoretical diversity 1.05 x 10^6). Two PCR products of the EcYtR-tRNA were amplified with Phusion HSII with the following primers: pBKseqtF, mutiR, mutiF, and JI MCS sqR. The 5' and 3' PCR products were joined together by primerless overlap extension. The full-length PCR product was gel purified (1% agarose gel, 150 V) followed by amplification with the terminal primers pBKseqtF and JI MCS sqR. The PCR amplified insert was digested with BamHI/NcoI and ligated by T4 DNA Ligase into pBK vector digested with the same restriction enzymes. The ligation mixture was ethanol precipitated with Yeast-tRNA (Ambion) and transformed into DH10B electrocompetent cells. The library was covered using > 10^7 distinct CFU.

2.5.7 Selection for an orthogonal tRNA$^{EcTyr}_{CUA}$

ATMY5 was co-transformed with the pBK tRNA$^{EcTyr}_{CUA}$ library and the positive selection reporter plasmid pRepTrip2.3-EcOMeYRS. This reporter construct contains a glnS-promoted *EcTyr*-OMeYRS, a CAT reporter mutagenized at Q98 to a stop codon (TAG), an arabinose inducible T7 RNA polymerase mutagenized at positions 8 and 114 to stop codons (TAG), and t7 promoted wild-type GFPuv. Successful suppression of the TAG codons by active tRNA$^{EcTyr}_{CUA}$ library members leads to full-length expression of the CAT-reporter conferring chloramphenicol resistance. Additionally, suppression of the T7 RNA polymerase drives expression of GFPuv. 2.3 x 10^7 CFU

were plated on LB + 1x Spec/1x Tet/1x Kan/0.5x Amp/0.02% arabinose + chloramphenicol (30 and 50 μg/mL) in the presence of 1 mM OMeY for 24 hours at 37 ºC.

Following the positive selection, ATMY5 was co-transformed with the pBK $tRNA^{EcTyr}_{CUA}$ library and the negative reporter plasmid pNeg-noEcYtR. This reporter construct harbors a toxic barnase gene with stop codons at amino acids 1 and 90 under the control of an arabinose inducible promoter. 1.8 x 10^7 CFU were plated on LB + 1x Amp/Kan/Spec/0.2% arabinose in the absence of UAA for 12 hours at 37 º C.

2.5.8 Construction of the pBK EcTyrRS library

Overlap extension was used to introduce the D265R mutation into the previously constructed EcYRS-lib1 (theoretical diversity 1.06 x 10^7).[2] Two PCR products of the EcYRS-lib1 were amplified with Phusion HSII with the following primers: BKrep-SpeI-F, EcYRS-D265R-R, EcYRS-D265R-F, and BkrepBglII-R. The two products were joined together by primerless overlap extension followed by agarose gel purification (1% agarose gel, 150 V). The purified product was amplified with the terminal primers BKrep-SpeI-F and BKrep-BglII-R and digested with NdeI/PstI. This digested product was ligated by T4 DNA ligase into pBK vector digested with the same restriction enzymes. The ligation mixture was ethanol precipitated with Yeast-tRNA (Ambion) and transformed into DH10B electrocompetent cells. The library was covered using > 10^8 distinct CFU.

2.5.9 Assessment of aaRS-tRNA activity using a EGFP39 reporter

HEK293T cells were cultured in Dulbeco's modified Eagle's medium (high glucose DMEM) supplemented with 10% fetal bovine serum (FBS) and penicillin-streptomycin (0.5x) at 37 ºC in

the presence of 5 % CO_2. The cells were seeded at a density of 600,000 cells per 12-well plate 24 hours prior to transfection. Co-transfection with pAcBac1-EGFP-39-TAG and pB1U aaRS was carried out when the cells were ~70% confluent. For the co-transfection, PEI (Sigma) and DNA were mixed at a ratio of 4 μL PEI (1 mg/mL) to 1 μg of total DNA (500 ng of each plasmid) in DMEM. This PEI/DNA mixture was incubated for 10 minutes and then added to each well (100 μL per well). Then, UAAs were added to each well (1 mM final concentration). Fluorescence images were taken 48 hours after transfection using a Zeiss Axio Observer fluorescence microscope.

2.6 References

1. Italia, J.S., Addy, P.S., Wrobel, C.J.J., Crawford, L.A., Lajoie, M.J., Zheng, Y., Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol. 13, 446 (2017).
2. Italia, J.S., Latour, C., Wrobel, C.J.J., Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. Cell Chemical Biology 25(10), 1304 (2018).
3. Wang, L., Brock, A., Herberich, B., Schultz, P.G. Expanding the Genetic Code of *Escherichia coli*. Science 292, 498 (2001).
4. Chatterjee, A., Xiao, A., Yang, P.Y., Soundararajan, G., Schultz, P.G. A Tryptophanyl-tRNA Synthetase/tRNA pair for Unnatural Amino Acid Mutagenesis in E. coli. Angew Chem Int Ed Engl. 52(19), 5106 (2013).
5. Iraha, F., Oki, K., Kobayashi, T., Ohno, S., Yokogawa, T., Nishikawa, K., Yokoyama, S., Sakamoto, K. Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion. Nucleic Acids Res. 38(11), 3682 (2010).
6. Jahn, M., Rogers, M.J. & Söll, D. Anticodon and acceptor stem nucleotides in tRNA(Gln) are major recognition elements for *E. coli* glutaminyl-tRNA synthetase. Nature 352, 258–260 (1991).
7. Rogers, M.J., Adachi, T., Inokuchi, H. & Söll, D. Switching tRNA(Gln) identity from glutamine to tryptophan. Proc. Natl. Acad. Sci. USA 89, 3463–3467 (1992).
8. Chin, J.W., Cropp, T.A., Anderson, C., Mukherji, M., Schultz, P.G. An expanded eukaryotic genetic code. Science 301, 964 (2003).
9. Cooper, G.M. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates (2000).
10. Giegé, R., Jüheng, F., Pütz, J., Stadler, P., Sauter, C., Florentz, C. Structure of transfer RNAs: similarity and variability. WIREs RNA 3, 37 (2012).
11. Bedouelle, H. Tyrosyl-tRNA Synthetases. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience (2000-2013).
12. Yaremchuk A, Kriklivyi I, Tukalo M, Cusack S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. EMBO J. 21(14), 3829 (2002).
13. Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. Nat. Struct. Biol. 10(6), 425 (2003).
14. Javahishvili, T., Manibusan, A., Srinagesh, S., Lee, D., Ensari, S., Shimazu, M., Schultz, P.G. Role of tRNA Orthogonality in an Expanded Genetic Code. ACS Chem. Biol. 9 (4), 874 (2014).
15. Serfling, R., Lorenz, C., Etzel, M., Schicht, G., B ttke, T., M rl, M., and Coin, I. Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in mammalian cells. Nucleic Acids Res. 46, 1 (2018).
16. Hughes RA, Ellington AD. Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA. Nucleic Acids Res. 38(19), 6813 (2010).
17. Baldridge, K.C., Jora, M., Maranhao, A.C., Quick, M.W., Addepalli, B., Brodbelt, J.S., Ellington, A.D., Limbach, P.A., Contreras, L.M. Directed Evolution of Heterologous

tRNAs Leads to Reduced Dependence on Post-transcriptional Modifications. ACS Synth. Biol. 7 (5), 1315 (2018).
18. Chatterjee, A., Xiao, H., Schultz, P.G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. PNAS 109 (37), 14841 (2012).
19. Guo, J., Melançon, C.E., Lee, H.S., Groff, D., Schultz, P.G. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. Angew. Chemie Int. Ed. 48, 9148 (2009).
20. Fan, C., Xiong, H., Reynolds, N.M., Söll, D. Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. Nucleic Acids Res. 43, 1 (2015).
21. Chatterjee, A., Lajoie, M.J., Xiao, H., Church, G.M., Schultz, P.G. A bacterial strain with a unique quadruplet codon specifying non-native amino acids, ChemBioChem 15, 1782 (2014).
22. Maranho, A.C.; Ellington, A.D. Evolving orthogonal suppressor tRNAs to incorporate modified amino acids, ACS Synth. Biol. 6 (1), 108 (2017).
23. Prabakaran, S., Lippens, G., Steen, H., Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med. 4 (6), 565 (2012).
24. Gavin, A.C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B, Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., Superti-Furga, G. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141 (2002).
25. Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, A., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulson, V., Sorenson, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W.V., Figeys, D., Tyers, M. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 415, 180 (2002).
26. Dormán, G., Prestwich, G.D. Benzophenone Photophores in Biochemistry. Biochemistry 33 (19), 5661 (1994).
27. Kobayashi, T., Takimura, T., Sekine, R., Kelly, V.P., Kamata, K., Sakamoto, K., Nishimura, S., Yokoyama, S. Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase. J. Mol. Biol. 346, 105 (2005).
28. Takimoto, J.K., Adams, K.L. Xiang, Z., Wang, L. Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. Mol. Biosyst. 5, 931 (2009).
29. Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., Liu, T., Lyu, X., Du, J., Xuan, W., Yao, A., Reed, S. A., Kang, M., Zhang, Y., Guo, H., Huang, C., Yang, P. Y., Wilson, I. A., Schultz, P. G., Wang, F. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. Nat Chem Biol. 13(8), 845 (2017).

Chapter 3

Development of a universal selection platform to improve aminoacyl-tRNA synthetase suppression efficiency

3.1 Introduction

3.1.1 Antibiotic-based selection platforms

To carry out directed evolution of aaRS/tRNA pairs, a strategy must be employed that selectively isolates the desired synthetase library members from a pool of mostly undesirable options. This is typically accomplished by linking synthetase activity to some type of phenotypic output that can be used to enrich desired library members.[1] For example, antibiotic resistance is frequently used as a selectable marker for the directed evolution of orthogonal aaRS/tRNA pairs in *E. coli*.[2-4] In this strategy, an amber stop codon is introduced at a permissible location in an enzymatic protein that confers antibiotic resistance when fully translated (Figure 3.1). It is an easily generalized approach in bacteria because it does not rely on the UAA's chemistry to observe a phenotypic output.[4] Antibiotic-based selections also provide tunability for selection stringency through modification of antibiotic concentrations.[4] Library members that survive this initial "positive" selection will have the ability to either charge a UAA or a canonical amino acid on its respective tRNA. To eliminate cross-reactive library members from the pool a "negative" selection must follow the positive selection. A negative selection would entail using a toxic gene that has an amber stop codon at a permissible location.[4] Full-length translation of this gene in the absence of the UAA, enabled only by cross-reactive synthetase mutants, would result in the death of their host cells.

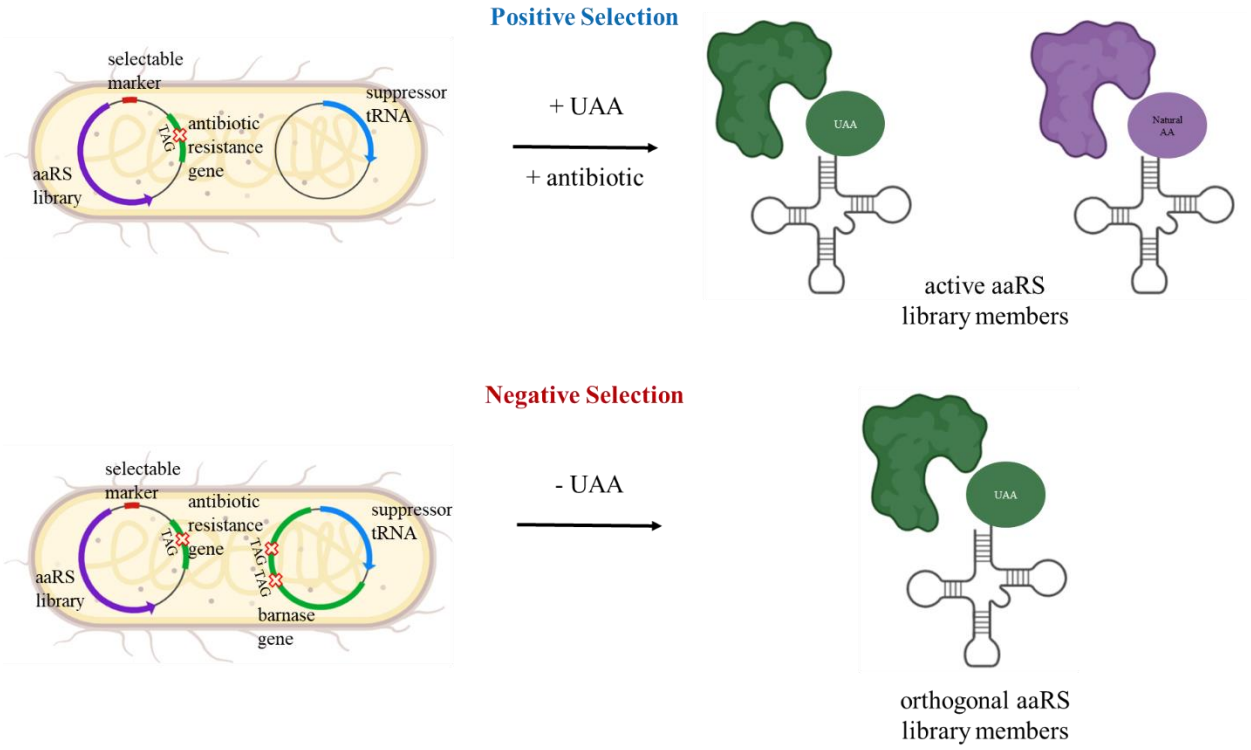


Figure 3.1. A general selection method for mutant aaRS/tRNA pairs that charge UAAs.[1-4] First, a positive selection is carried out in the presence of UAA that enriches for active library members (orthogonal and cross-reactive). Second, a negative selection is carried out in the absence of the UAA to eliminate any cross-reactive library members from the pool of mutants.

Initial antibiotic-based selection strategies utilized the suppression of β-lactamase to select for orthogonal aaRS/tRNA pairs in the presence of ampicillin.[4, 5] The amber nonsense mutation was introduced into the β-lactamase gene at position Ala-184 because it is not conserved among β-lactamases, so it was hypothesized that insertion of a UAA at this position would not alter the enzyme's activity.[6, 7] This positive selection has been coupled with the toxic barnase-based negative selection to enrich for numerous engineered orthogonal aaRS/tRNA pairs.[1] The main disadvantage to using a β-lactamase positive selection is that once the enzyme is translated, it is naturally secreted by the cell into the surrounding media (Figure 3.2).[8] This potentially helps neighboring cells survive the ampicillin selection that do not harbor plasmids with active synthetases.

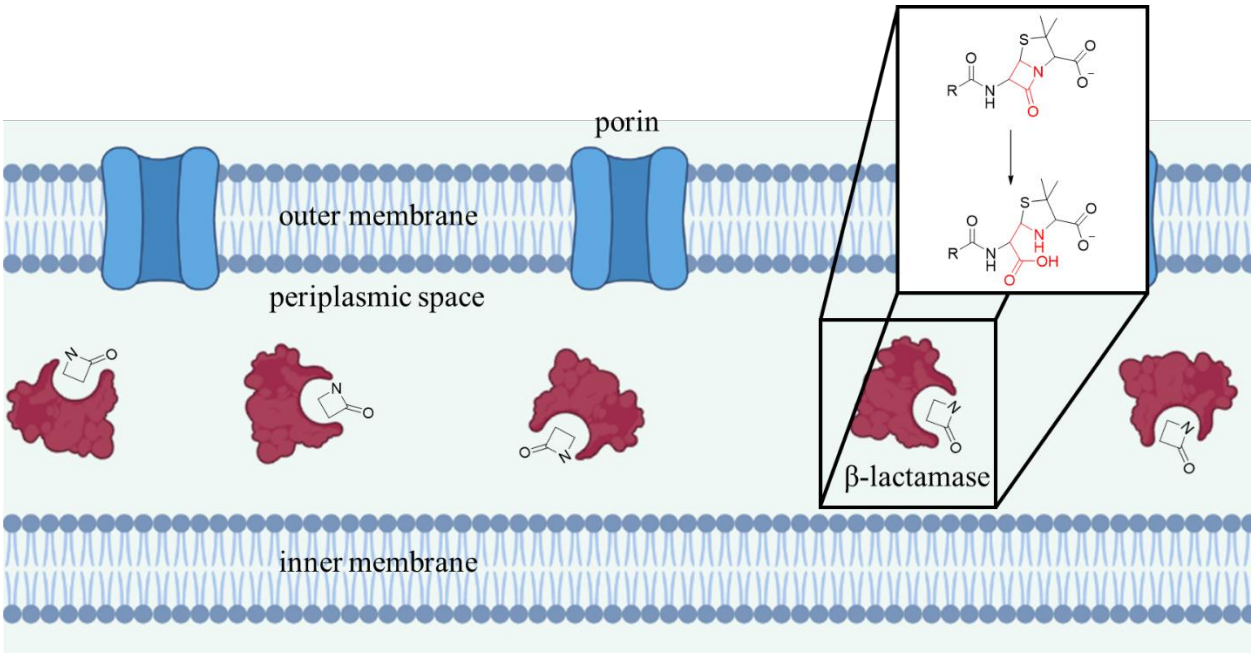


Figure 3.2. β-lactamase mechanism of action. In gram-negative bacteria, β-lactam antibiotics localize in the periplasmic space and bind to penicillin-binding proteins (PBPs) weakening the cell wall. β-lactamases are enzymes that break the β-lactam ring open of antibiotics (e.g., penicillins, cephalosporins, cephamycins, and carbapenems) through hydrolysis, as outlined in the figure above.[8-10] These enzymes are localized to the periplasmic space where these antibiotics reside, and can be secreted from high-level resistant bacteria (> 1,000 μg/mL ampicillin).[8]

Alternative pairs of antibiotics and antibiotic resistance genes have been explored for synthetase selection schemes. Over time, chloramphenicol and the chloramphenicol acetyltransferase (CAT) gene became widely considered as the optimal antibiotic-based selection platform because of the cytoplasmic localization of CAT and chloramphenicol's bacteriostatic properties, which minimize the weeding out of weakly active synthetase hits (Figure 3.3).[11] Currently, it is contentious in the field of directed evolution whether or not the maintenance of weak hits in a selection pool is advantageous.[12-14] However, the rationale of maintaining weakly active hits caused a shift in the community toward a reliance on CAT amber suppression platforms for the directed evolution of synthetases in bacteria.[1-3, 11] Initial studies comparing ampicillin and chloramphenicol platforms focused on the IC_{50} values rather than a correlation between cell-

survival at specific antibiotic concentrations and mutant synthetase suppression efficiencies.[11] We sought to more fully characterize the ability of ampicillin-based and chloramphenicol-based selection systems to differentiate between synthetase suppression efficiencies. In the future, this would enable our ability to easily discern strong mutants from weak ones.

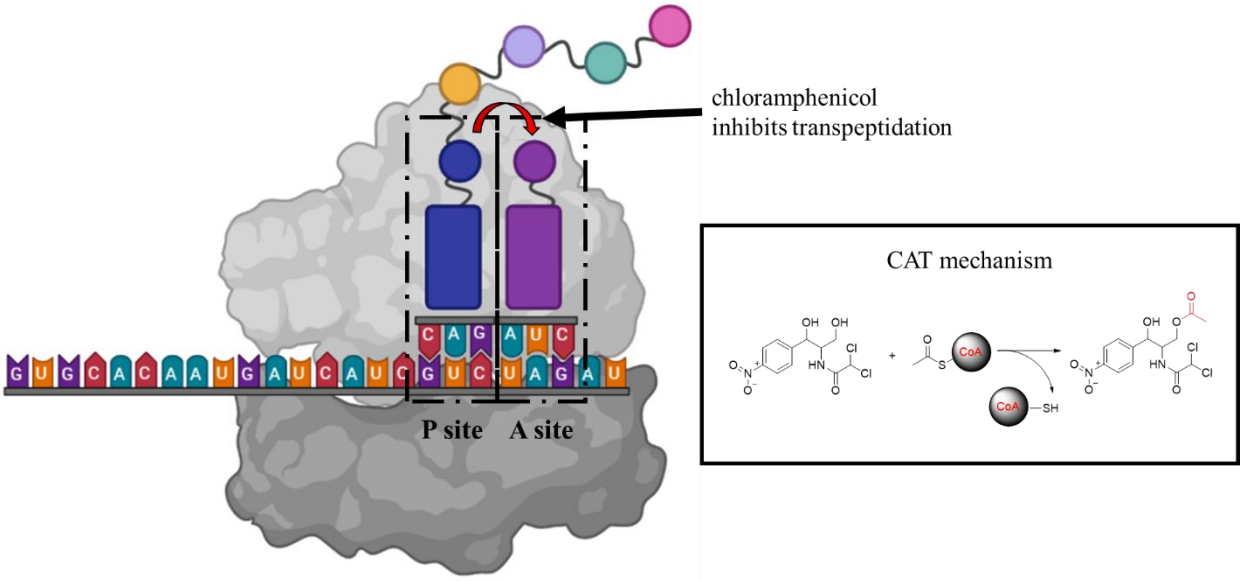


Figure 3.3. CAT mechanism of action. Chloramphenicol is a broad-spectrum antibiotic that inhibits protein synthesis by binding to the A-site of the 50S subunit of the bacterial ribosome.[15] The 3-hydroxyl and 1-hydroxyl of chloramphenicol form key hydrogen bonds with the ribosomal RNA-structure.[16-17] CAT disrupts the bonding of chloramphenicol to the ribosome by transferring an acetyl group from acetyl-coenzyme A (CoA) to the 3-hydroxyl group of chloramphenicol.[16-17]

3.1.2 Fluorescence-based selection platforms

E. coli is an ideal expression host for library screening platforms because it is possible to isolate library members within individual cells through transformation, guaranteeing that a cell's observed phenotype can be directly linked to the genotype of a single library member (Figure 3.1).[1-4] Individually screening each clone through antibiotic selection schemes is low-throughput ($< 10^4$ library members), time-consuming, and infrastructure intensive (Figure 3.1).[18] A low-throughput selection scheme is not ideal when approaching directed evolution strategies where structure-activity relationships of the target protein are poorly understood. This is because the

probability of identifying a desired mutant with the novel activity becomes even lower since so few library members can realistically be screened.

High-throughput selection schemes make it easier to identify library mutants with the desired phenotype when structure/function relationships are unknown.[18] These schemes typically rely on color, fluorescence, luminescence, or turbidity. Synthetases do not directly display any of these observable phenotypes. However, a synthetases' enzymatic activity can be tied to fluorescence in order to facilitate synthetase directed evolution through established schemes.[19-22] This is accomplished by using fluorescent genetic reporters such as green fluorescent protein (GFP)[20, 22] and β-galactosidase[19] that are mutated to contain amber suppression codons (Figure 3.4). Successful suppression of the amber codon in the fluorescent reporter leads to a fluorescent output signal that can be used to screen ~10^8 library members within 24 hours through fluorescence activated cell-sorting (FACS).[23, 24] Additionally, the use of fluorescent reporters removes the possibility of false positives resulting from survival mechanisms induced by antibiotics.[8] Current reported FACS screens for the directed evolution of synthetases require multiple rounds of positive and negative selection, resulting in a cumbersome selection process.[25] We sought to couple the quantitative power of FACS with the long-established antibiotic-based selection platform to minimize both of their shortcomings.

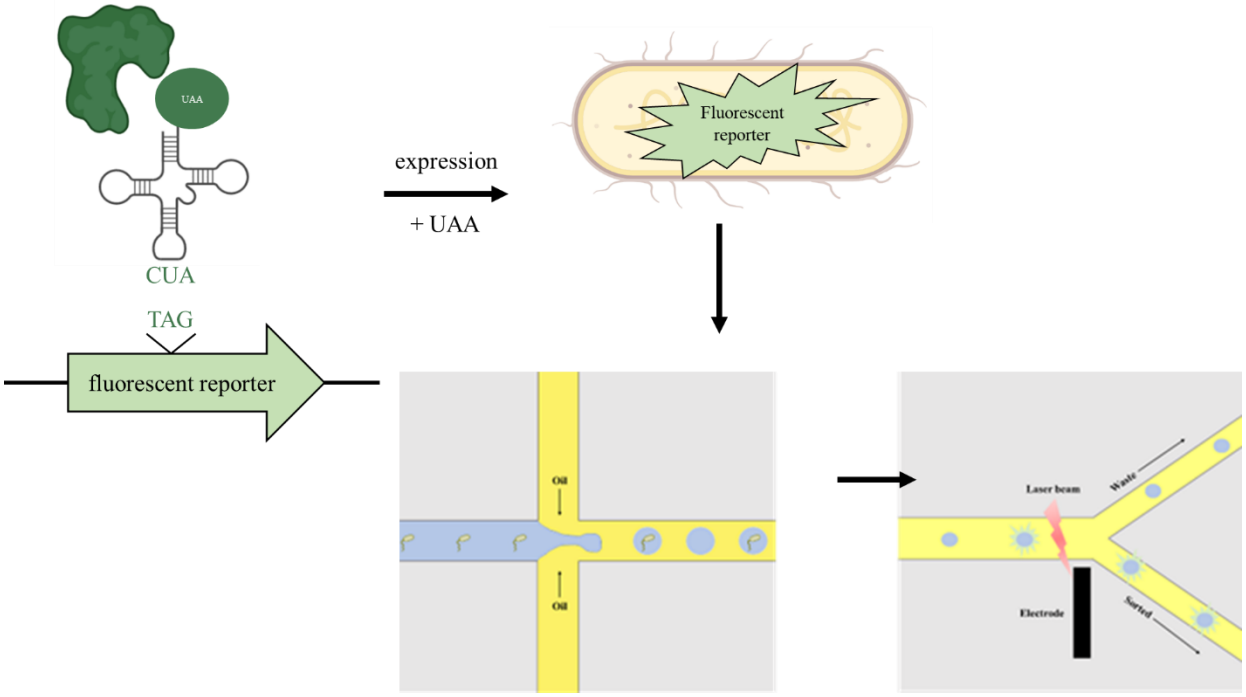


Figure 3.4. General fluorescent reporter-based selection scheme.[19-22] A fluorescent reporter encoded with an amber nonsense codon is suppressed in the presence of an active aaRS. Each individual cell with full-length fluorescent reporter protein is subsequently sorted through FACS. Droplets formed through the mixing of two immiscible phases sequester individual cells that can be sorted based on a fluorescent output signal.

3.1.3 Using directed evolution to improve synthetase activity

Identifying new aaRS/tRNA pairs for GCE is difficult due to the requirement for pairs that are (i) orthogonal with the host's endogenous aaRS/tRNA pairs as well as (ii) compatible with the host's transcriptional and translational machinery. This limits the accessible chemistries for UAA incorporation since only a few pairs have been successfully developed for GCE (Figure 3.5).[1] Additionally, not every engineered aaRS/tRNA pairs exhibits optimal suppression efficiencies, further limiting our toolbox of incorporable UAAs for relevant applications.

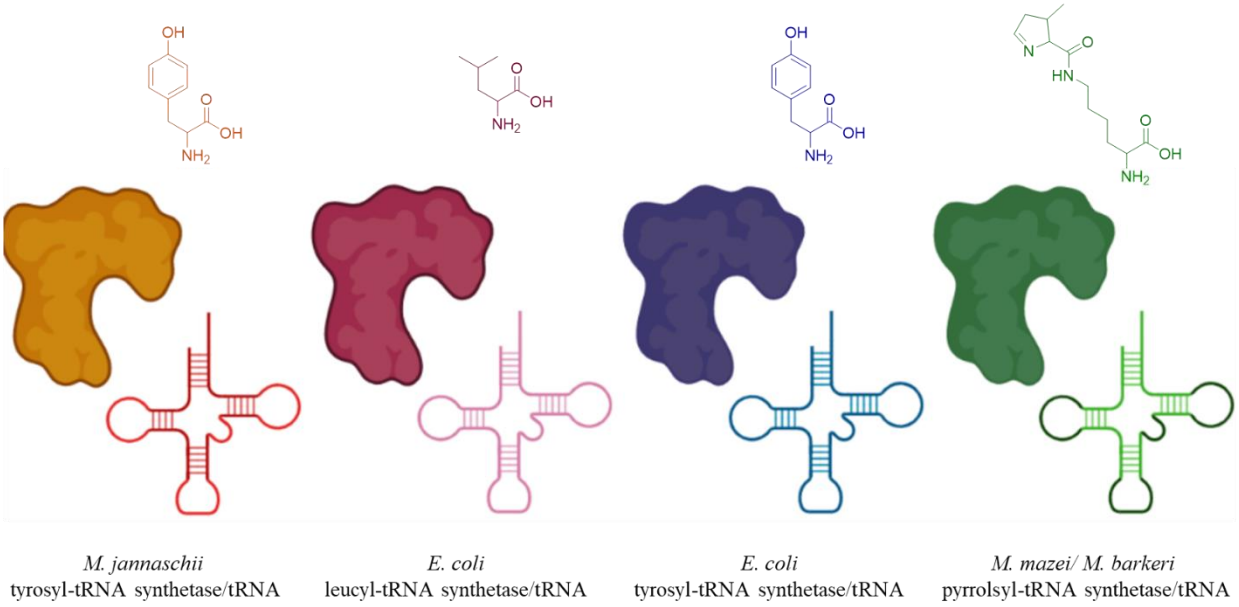


Figure 3.5. aaRS/tRNA pairs that have been engineered for genetic code expansion.[1] The four synthetase structures shown above account for ~90 % of the UAAs incorporated to-date. The structures of UAAs that can be incorporated are limited to derivations of these four synthetases' cognate amino acid structures, which are depicted above each aaRS/tRNA pair.

Suboptimal amber nonsense suppression can be partially attributed to the natural competition between the ribosomal incorporation of the desired UAA with termination of translation in response to the stop codon.[1] Efforts towards rectifying this limitation have mainly focused on optimizing aaRS/tRNA copy numbers through plasmid construction, engineering the nonsense suppressor tRNA, modification of the translational machinery (e.g. removal of EF-Tu), modification of sequence contexts, and genome-wide engineering of the host cell.[26-33] Contrastingly, little work has been done towards improving the enzymatic activity of the synthetase. The only recent developments for next-generation synthetase engineering have been limited to bacterial applications due to the difficulty of working with eukaryotic GCE platforms.[19] With the ATMY selection system, we believe that we can expand next-generation synthetase

engineering technology for eukaryotic applications, ultimately providing access to a broader range of chemical space.

3.2 Results and discussion

3.2.1 Construction of an ATMY selection platform to improve aaRS suppression efficiency

Successful engineering of the orthogonal tRNA$^{Ec\text{Tyr}}_{\text{CUA}}$-h1 enabled identification of *Ec*TyrRS mutants with low suppression efficiencies (*Ec*Tyr-pBPARS-7 and *Ec*Tyr-pBPARS-8). While these solutions provide access to unique chemistry in mammalian cells, they are suboptimal in activity for relevant applications (e.g., *in vivo* incorporation of multiple UAAs).[34, 35] Prior work by Arnold *et al*. showcases the power of altering an enzyme through error-prone PCR for novel chemical activity, then improved catalytic efficiency.[36, 37] Additionally, previous efforts have been made towards improving synthetase suppression efficiencies for GCE.[19, 38] Based on these precedents, we hypothesized that bacterial aaRSs could be improved through similar strategies for eukaryotic applications in ATMY *E. coli*.

The key to creating a successful selection platform for improving aaRS suppression efficiency is identifying an appropriate selectable marker whose phenotypic output can be quantifiable relative to synthetase activity. For an antibiotic-based platform, there must be a quantitative linear relationship between aaRS suppression efficiency and cell survival. Ampicillin- and chloramphenicol-based platforms have been pivotal in the engineering of novel orthogonal aaRS/tRNA pairs for GCE.[1-4, 7] However, no direct comparison has been made, in either of these platforms, to determine the linear relationship exhibited between antibiotic resistance and synthetase suppression efficiencies. We sought to determine which antibiotic provides a refined dynamic range to differentiate between weakly active and strongly active synthetases. This was

accomplished by screening both ampicillin and chloramphenicol resistance with two *Ec*Tyr-OMeYRSs that have an observed 6-fold difference in activity when evaluated with an sfGFP-151-TAG assay (Figure 3.6). Based on these antibiotic assays, we observed that ampicillin resistance depicts a more linear relationship between strong and weak aaRS activities than chloramphenicol resistance.

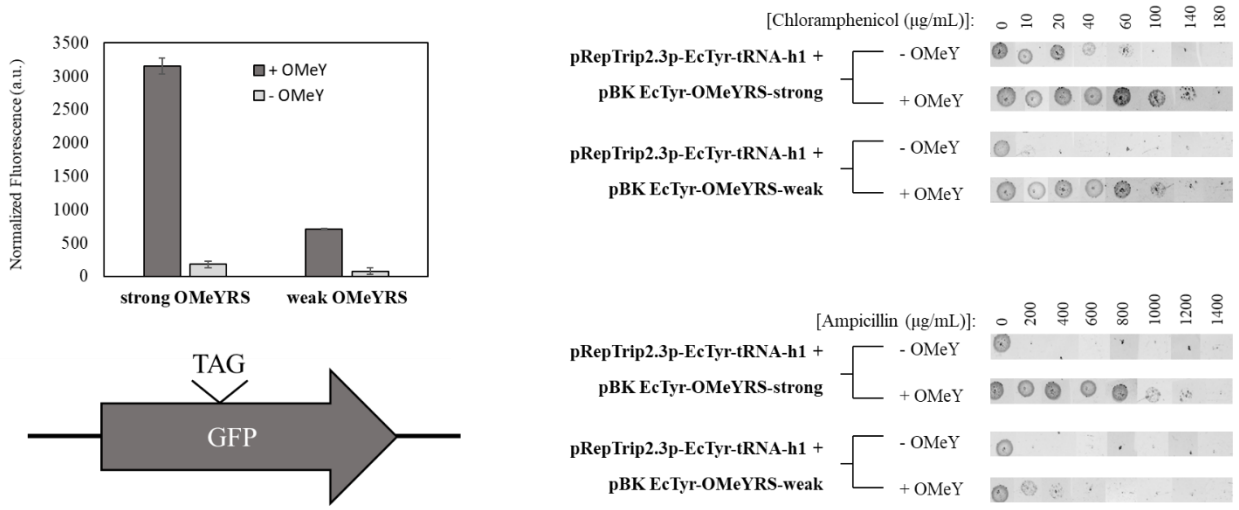


Figure 3.6. Characterization of antibiotic-based selection platforms. The *Ec*Tyr-OMeYRS strong and weak synthetase activities were characterized with a fluorescent reporter (left) where they exhibited a six-fold difference in activity. This was reflected by a six-fold difference in cell survival with the ampicillin-based selection (bottom right).

A purely ampicillin-based selection platform could be susceptible to cheaters due to the natural secretion of β-lactamase following its translation.[8, 11] To mitigate this issue, we decided to couple the antibiotic selection with a subsequent fluorescent-reporter screen. A selection construct was designed containing two fluorescent reporters (wild-type mCherry and sfGFP-151-TAG) (Figure 3.7). The wild-type mCherry protein fluorescence is used as an internal standard to control for protein expression variability between individual cells. The amber-suppressed sfGFP fluorescence is the marker used to quantify synthetase suppression efficiency.

Figure 3.7. The dual-fluorescent reporter construct. The sfGFP-151-TAG and mCherry WT were expressed under a T5/lac inducible promoter system. Full-length expression of sfGFP is contingent upon the successful incorporation of a UAA at the 151-Tyr amber suppressor codon location. The mCherry WT expression levels are contingent upon nascent expression levels of the individual cell. The selectable marker for this reporter construct is chloramphenicol (ChlorR).

Characterization of the dual-fluorescent reporter was carried out with the strong *Ec*Tyr-OMeYRS and the weak *Ec*Tyr-pBPARS. ATMY4 *E. coli* (EcNR1GT pUltraBR MjY dtyrS tyrTV::CUA tyrU::CUA lambda::tolC)[3] was co-transformed with the dual fluorescent reporter and one of the synthetases. Expression was induced in the presence of IPTG (1 mM) and UAA (1 mM). Following expression, it was observed that the ratio of sfGFP fluorescence to mCherry fluorescence exhibited an 8-fold difference between the strongly active *Ec*Tyr-OMeYRS and weakly active *Ec*Tyr-pBPARS (Figure 3.8). Consequently, it was concluded that the dynamic range of this reporter should be large enough to carry out a library selection.

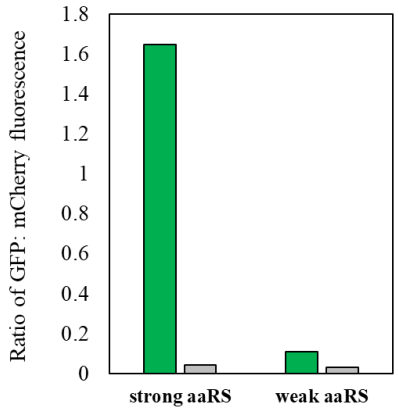


Figure 3.8. Ratio of sfGFP: mCherry fluorescence for a strong aaRS vs. weak aaRS in ATMY4. Normalization of the sfGFP fluorescence by the mCherry fluorescence accounts for the protein expression variability of individual cells.

To further validate that this construct could be used for aaRS library selections, a mock selection was carried out. ATMY4 cells harboring the dual fluorescent reporter and the *Ec*Tyr-OMeYRS, or the *Ec*Tyr-pBPARS, were cultured, grown, and induced for expression (+ 1 mM IPTG, + 1 mM UAA). Subsequently, the cells were mixed at a ratio of 100 highly active aaRS (*Ec*Tyr-OMeYRS) to 1 weakly active aaRS (*Ec*Tyr-pBPARS). The cells were sorted based on the ratio of their sfGFP to mCherry fluorescence (Figure 3.9) and collected (~ 6,000 cells). They were recovered in enriched LB Media and plated in the presence of the appropriate antibiotics (Spec/Kan/Chlor). Following enrichment, individual colonies were characterized by Sanger sequencing and fluorescence analysis (Figure 3.10), the majority of which turned out to be the stronger *Ec*Tyr-OMeYRS (80%). The successful enrichment of the strongly active *Ec*Tyr-OMeYRS validated our ability to use this fluorescent reporter for library selections.

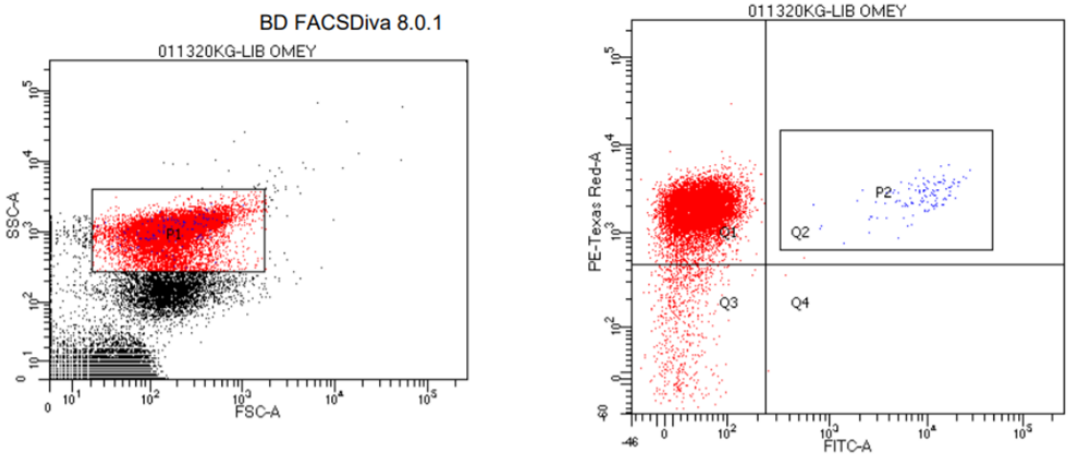


Figure 3.9. FACS of the *Ec*Tyr-OMEYRS + *Ec*Tyr-pBPARS mock library. The population of cells sorted were chosen based on their forward scatter and side scatter light ratio (left). The sorted cells were further scrutinized for their ratio of sfGFP (FITC-A): mCherry (PE-Texas Red-A) fluorescence. Parameters were outlined as quadrant 2 (Q2), resulting in the collection of only 1.1% of the total parent population (right).

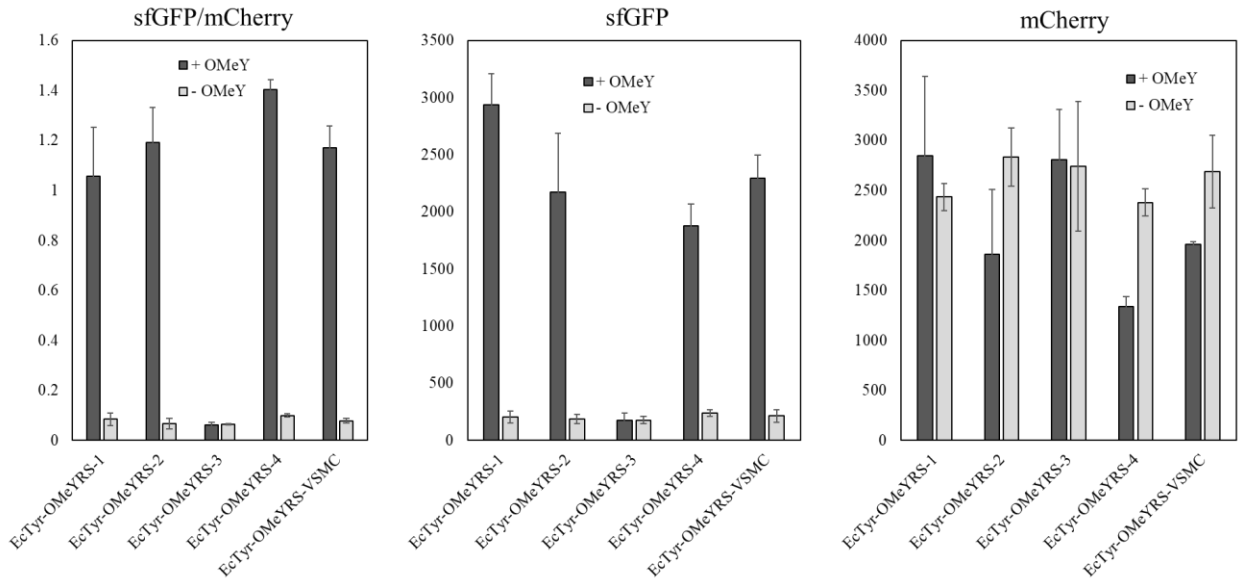


Figure 3.10. Fluorescence analysis of individual colonies following the mock selection. Fluorescence analysis validated that the majority of the cells (~80%) sieved through the mock selection were the expected strongly active *Ec*Tyr-OMeYRS. The colonies were characterized based on their ratios of mCherry fluorescence: OD600 (right), sfGFP fluorescence: OD600 (middle), and sfGFP fluorescence: mCherry fluorescence (left).

3.2.2 Improving the suppression efficiency of the EcTyr-pBPARS mutant

To demonstrate the utility of coupling the established ampicillin-based platform with our new dual-fluorescent reporter system for directed evolution, we engineered the *Ec*Tyr-pBPARS mutant (*Ec*Tyr-pBPARS-1) due to its potential value as an unquenchable photoaffinity probe in mammalian cells.[35, 39-41] Libraries were generated through error-prone PCR (ePCR)[42] amplification of the *Ec*Tyr-pBPARS-1 (average error rate of 2-4 nt/1275 bp) followed by polymerase incomplete primer extension (PIPE)[43] cloning into the pBK vector (~10^6 transformants). Following one round of ampicillin-based positive selection (+ 1mM pBPA) individual library members were characterized by a β-lactamase-TAG assay.[4] Several mutants exhibited a 2-fold improvement in activity. After sequencing analysis, it was determined that all of the screened mutants had a single-point mutation, reversion of Arg265 to Asp. The Asp265 mutant was used as the library template for the second round of mutagenesis, which produced no

improved hits. An additional round of ePCR mutagenesis followed by a third round of ampicillin-based selection (+ 1 mM pBPA) and the dual-fluorescent reporter screen yielded two unique mutants with ~5.5-fold improvement in activity in ATMY4 *E. coli* (Figure 3.11). Both hits had point mutations scattered throughout the aaRS (Table 3.1). Their ability to incorporate pBPA into sfGFP was validated by MS analysis and SDS-PAGE (Figure 3.12).

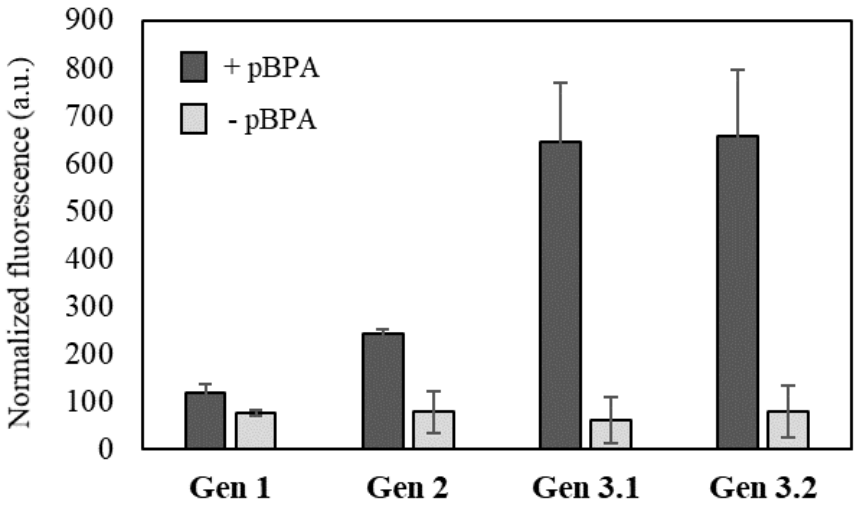


Figure 3.11. sfGFP fluorescence characterization of the *Ec*Tyr-pBPARS mutants. Over three generations a ~5.5-fold improvement in suppression efficiency is observed in bacteria.

Clones	Tyr37	Leu71	Asn126	Asp182	Phe183	Leu186	Additional mutations
Gen 1 pBPARS	Gly	Leu	Asn	Gly	Phe	Ala	Asp265Arg
Gen 2 pBPARS	Gly	Leu	Asn	Gly	Phe	Ala	N/A
Gen 3.1 pBPARS	Gly	Leu	Asn	Gly	Phe	Ala	Ile7Phe, Gly180Ser
Gen 3.2 pBPARS	Gly	Leu	Asn	Gly	Phe	Ala	Ser366Phe

Table 3.1. *Ec*Tyr-pBPARS mutant amino acid sequence modifications.

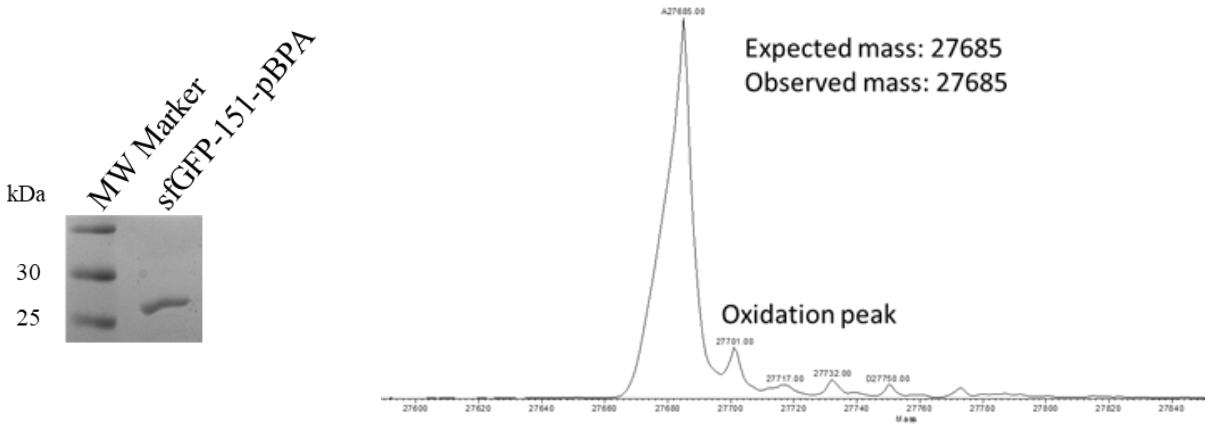


Figure 3.12. SDS-PAGE and MS analysis of sfGFP-151-pBPA.

The mutations for *Ec*Tyr-pBPARS-Gen 3.1 were individually reverted to identify if any of them conferred the improved activity alone, however, they appeared to be additive when compared with an sfGFP-151-TAG reporter fluorescence assay performed in ATMY4 (Figure 3.13). Furthermore, addition of the S366F point mutation from *Ec*Tyr-pBPARS-Gen 3.2 into *Ec*Tyr-pBPARS-Gen 3.1 resulted in a lower suppression efficiency of the sfGFP-151-TAG reporter in ATMY4 (Figure 3.13).

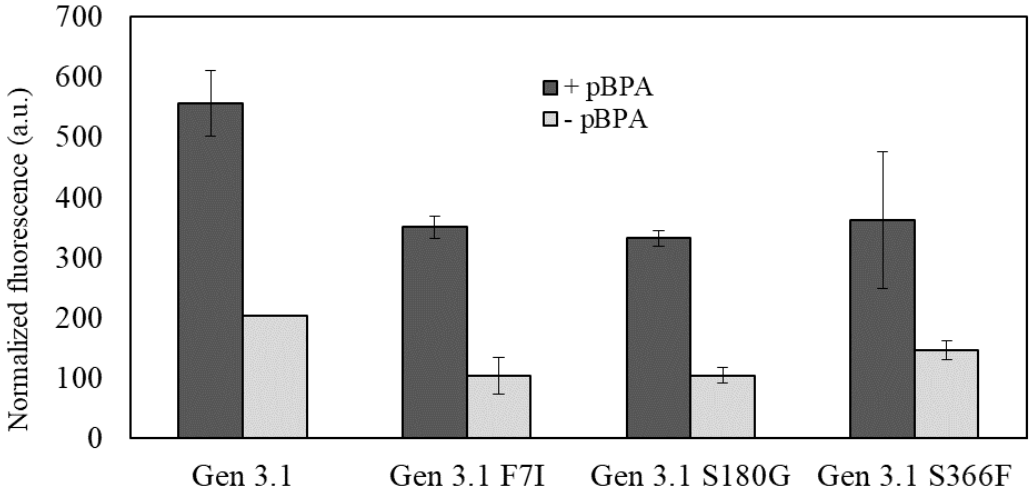


Figure 3.13. sfGFP fluorescence characterization of the *Ec*Tyr-pBPARS-Gen 3.1 reversion point mutations. The point mutations of the *Ec*Tyr-pBPARS-Gen 3.1 were reverted to the wild-type amino acid. Reversion of each point mutation resulted in reduced suppression efficiency relative to the parental aaRS (Gen 3.1), suggesting that the mutations were mutually necessary for enhanced activity. Addition of the S366F point mutation, into Gen 3.1, resulted in decreased fluorescence.

*Ec*Tyr-pBPARS- Gen 1 and *Ec*Tyr-pBPARS-Gen 3.1 were cloned into the mammalian expression plasmid pB1U for characterization in HEK293T cells. *Ec*Tyr-pBPARS-Gen3.1 and *Ec*Tyr-pBPARS-Gen 3.2 maintained their improved suppression efficiencies relative to *Ec*Tyr-pBPARS-Gen 1 and *Ec*Tyr-pBPARS-Gen 2 based on fluorescence imaging in HEK293T cells (Figure 3.14). SDS-PAGE and MS analysis were carried out to validate successful incorporation of pBPA into the EGFP-39-TAG reporter by *Ec*Tyr-pBPARS-Gen 3.1 (Figure 3.15).

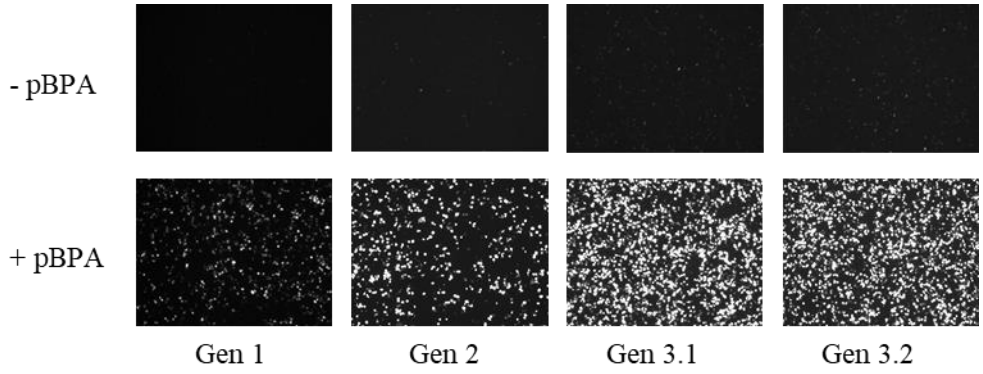


Figure 3.14. HEK293T fluorescence images. The improved *Ec*Tyr-pBPARS mutants (Gen 3.1 and Gen 3.2) maintain their enhanced activity in mammalian cells.

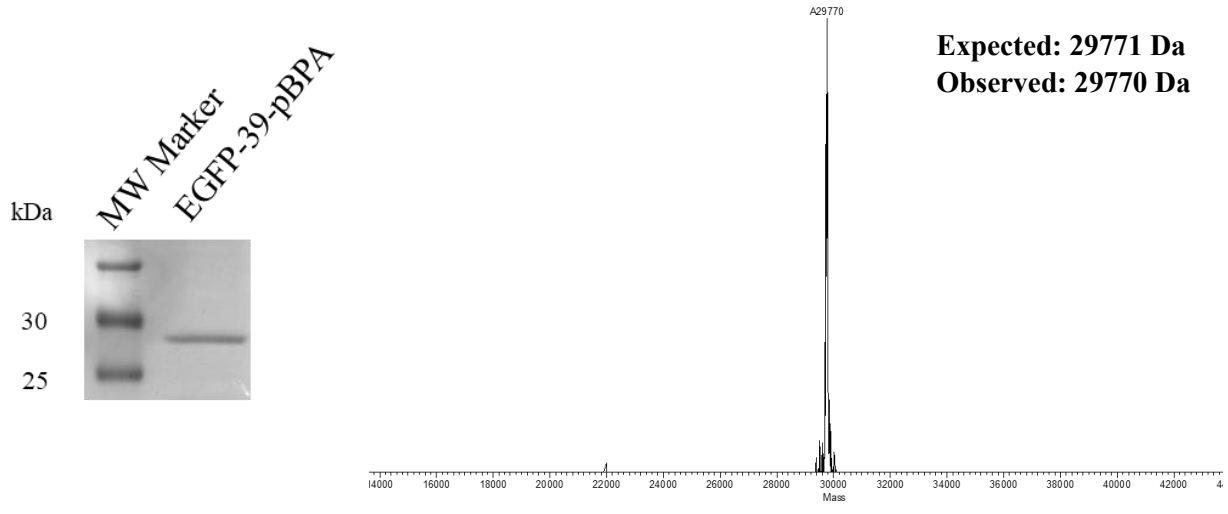


Figure 3.15. SDS-PAGE and MS analysis of EGFP-39-pBPA.

3.2.3 Double suppression with the improved EcTyr-pBPARS mutant

Post-translational modifications (PTMs) are used by the cell to expand the chemical space available to natural proteins beyond the twenty canonical amino acids.[44] Generally, the addition of PTMs is used to regulate different types of protein-protein interactions (PPIs) resulting in a cascade of downstream events (e.g., cell-signaling networks, DNA transcription). A host of PTMs have been identified as key players in these dynamic PPIs, such as bromodomain recognition of acetyllysine (AcK),[35] but their specific functional and physiological roles remain elusive. Identification of PPI roles remains a challenge for the scientific community due to the typically weak interactions between PPIs, making it difficult for them to survive pull-down techniques, and lack of methods to site-specifically produce homogenously PTM-modified protein at a large scale.

One solution for studying PTMs in their native context is to site-specifically incorporate multiple UAAs through GCE.[27, 30, 34, 35] By co-translationally incorporating a photoaffinity probe with a PTM, PPIs can be covalently captured and survive isolation techniques.[35] The *Methanococcus jannaschii*-derived tyrosyl pair (*Mj*Tyr) and *Methanosarchina*-derived pyrrolysyl pair (Pyl) have been successfully engineered for co-translational incorporation in *E. coli*.[1] In fact, this platform was previously adapted for the incorporation of a photoaffinity UAA probe and PTM mimic into a full-length protein to capture known bromodomain-histone binding partners.[35]

So far, multi-site UAA incorporation technology has been largely limited to *E. coli* expression systems.[27, 30, 34, 35] Ideally, to better understand how PTMs behave in mammalian cell biology these recombinant proteins would need to be expressed in mammalian cells. However, such applications have been limited due to the lack of accessible, orthogonal aaRS/tRNA pairs that can charge photoaffinity probes and PTM mimics in mammalian cells. We sought to demonstrate that the more active *Ec*Tyr-pBPARS-Gen 3.1 provides access to this unreachable technology in

eukaryotes. *Ec*Tyr-pBPARS-1 and *Ec*Tyr-pBPARS-Gen3.1 were cloned into the mammalian expression plasmid pAcBac1 for characterization in HEK293T cells. *Ec*Tyr-pBPARS-Gen3.1 and *Ec*Tyr-pBPARS-1 were co-transfected with pAcBac3-MbPylRS with $tRNA^{MbPyl}_{UCA}$ in HEK293T. Based on fluorescence output (Figure 3.16) and fluorescence imaging (Figure 3.17), double-suppression was only achieved with the improved *Ec*Tyr-pBPARS-Gen 3.1. The ability of our improved *Ec*Tyr-pBPARS-Gen 3.1 to carry out double-suppression applications demonstrated the utility of further engineering synthetases for enhanced suppression efficiency.

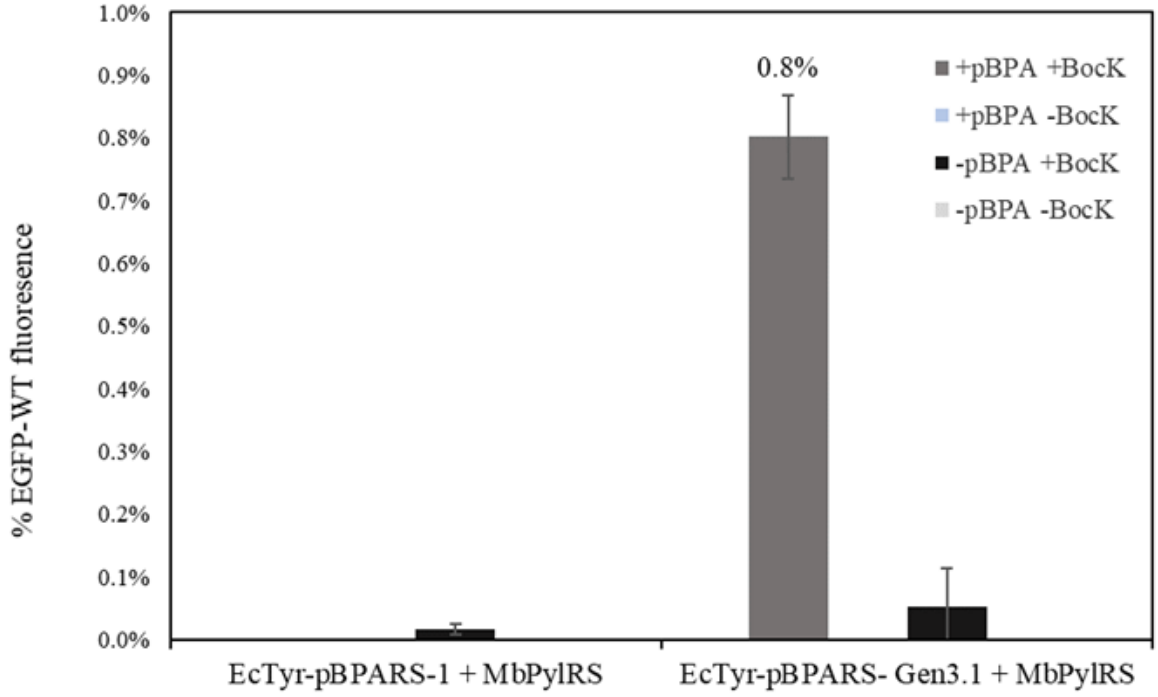


Figure 3.16. HEK293T fluorescence from *Ec*Tyr-pBPARS + *Mb*PylRS. EGFP fluorescence (0.8% of WT EGFP fluorescence) was only observed in the presence of both UAAs (+ pBPA and + BocK) as well as the Gen 3.1 aaRS.

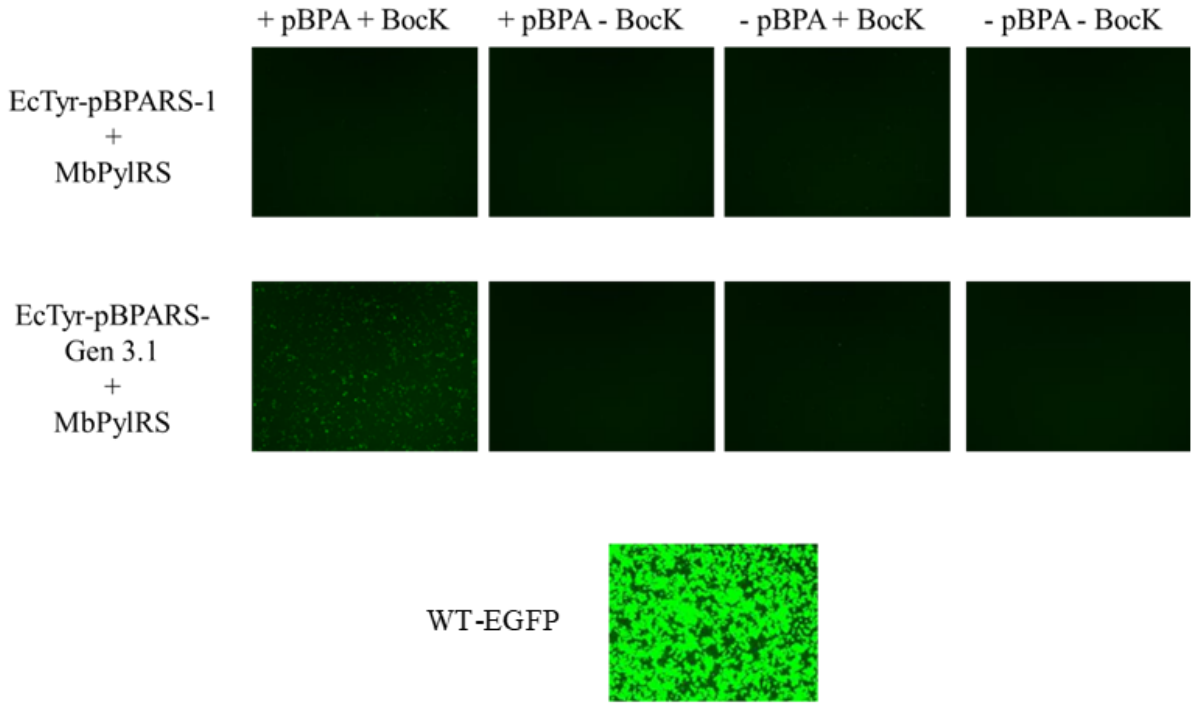


Figure 3.17. HEK293T cell images. No EGFP fluorescence is observed in the absence of the UAAs (pBPA and BocK). Only co-transfection of the Gen 3.1 aaRS and the *Mb*PylRS results in observable EGFP fluorescence.

3.3 Conclusions

In summary, we created the first selection platform that enhances engineered synthetases beyond the PylRS/tRNA pairs for GCE in eukaryotes. Further investigation into the dynamic range of established antibiotic-based selection platforms informed us that, contrary to popular belief, ampicillin-based selections were best suited for the directed evolution of synthetases in bacteria. Additionally, we were able to create a new dual-fluorescent reporter platform that quantitatively differentiates between weakly active and strongly active synthetases. By coupling the ampicillin-based platform with our newly created dual-fluorescent reporter system, we were able to improve the activity of a weakly active *Ec*Tyr-pBPARS ~5.5-fold in bacteria. This

enhanced activity was maintained in HEK293T cells and enabled access to double-suppression applications.

3.3.1 Ongoing and future directions

The work described here can be further applied to improve other weakly active aaRSs for novel mammalian applications. An undergraduate researcher, Megan Yeo, was pursuing this by carrying out this new selection scheme on an azide-specific *Ec*TyrRS. Double suppression applications are being furthered by another graduate student in the lab, Arianna Osgood. She will be working towards creating double suppression libraries, eliminating the need to laboriously rationally design the location of the two nonsense suppression positions.

3.4 Acknowledgements

Rachel Kelemen provided intellectual insight into the design of the dual-fluorescent reporter system. I was struggling to get this construct to work and her suggestions helped me overcome the hurdles I was facing in the initial weeks of designing this plasmid. Megan Yeo assisted with the PIPE cloning of the libraries. Arianna Osgood, one of the newest additions to the ATM subgroup, carried out the double suppression experiments demonstrating the utility of having an improve *Ec*Tyr-pBPARS mutant. Dr. Patrick Autissier assisted with the FACS sorting experiments.

3.5 Experimental procedures

Reference chapter 2's experimental procedures section for general materials and methods.

3.5.1 Strains, cell lines

Directed evolution of the EcTyr-pBPARS mutant occurred in the ATMY4 and ATMY5 *E. coli* strains. These strains were propagated from Dr. James Italia's glycerol stocks. Engineering of these strains is outlined in detail in his manuscript.[3]

All cloning and plasmid propagation were carried out in DH10B *E. coli* cells.

HEK293T cells (ATCC) were maintained at 37 ^{o}C and 5% CO_2 in DMEM-high glucose (HyClone) supplemented with penicillin/streptomycin (HyClone, final concentration of 100 U/L penicillin and 100 μg/mL streptomycin) and 10% fetal bovine serum (Corning).

3.5.2 Plasmids

All primer and plasmid sequences can be found in the Appendix.

pBK-pBPARS-Gen1 was propagated from a former student's DNA plasmid stocks.

The pBK-pBPARS-Gen3 point mutation reversion mutants were made through QuikChange mutagenesis (Agilent) with the following primers: pBPARS-F7I-F, pBPARS-S180G-F, and pBPARS-S366F-F.

pEvol-mCherry-sfGFP151-TAG was constructed by PCR amplification of the mCherry-wt from pET22b-mCherrywt with EcoNI-t7 promoter-F and EcoNI-lambda t0 term-R. The PCR product was then digested with EcoNI and inserted into the pEvol-sfGFP151-TAG.

pRepTrip2.3p-EcYtR-h1 was created by PCR amplification of EcYtR-h1 from pBK EcYtR-h1 with BKrep-SpeI-F and BKrep-BglII-R. The PCR product was digested with SpeI/BglII and inserted into the pRep vector backbone.

pB1U pBPARS-Gen1 and Gen3 were made by amplifying the pBPA-RS Gen 1 and Gen 3 from their respective pBK plasmids with EcYRS-NheI-F and EcYRS-XhoI-R. The amplified PCR products were subsequently digested with NdeI/XhoI and ligated into the pB1U vector backbone, which had been digested with the same restriction enzymes.

The pAcBac1 EGFP39-TAG construct was propagated from another student's plasmid stocks.

3.5.3 Unnatural amino acids

para-benzoyl-l-phenylalanine was purchased from Chem-Impex International (catalog number 05110).

3.5.4 EcTyr-pBPARS library generation through ePCR

pBK-EcTyr-pBPARS Gen 1 was used to ePCR[43] amplify out the EcTyr-pBPARS with the primers JI MCS sqR and pBKseqtF. TAQ polymerase was used to amplify the insert under the following conditions: 0.3 or 0.15 M $MnCl_2$, 0.4 mM dCTP, 0.4 mM dTTP, 0.08 mM dATP, 0.08 mM dGTP, and 25 ng of DNA. The amplified EcTyr-pBPARS library was inserted into the pBK vector through PIPE cloning.[43] To execute the PIPE cloning, the pBK vector was pre-digested with NdeI/NcoI, agarose gel purified (1%), and amplified with the following primers: pBK-backbone-F and pBK-backbone-R. The vector and insert PCR products were then mixed at equal volumes and co-transformed into DH10B cells generating 4.5 x 10^6 CFU.

3.5.5 Ampicillin selection of ePCR randomized pBPARS mutant library

The pBK EcYRS-pBPARS ePCR randomized library and the positive selection reporter plasmid pRepTrip2.3-pEcYtR-h1 were co-transformed into ATMY5. This pRepTrip reporter construct

contains a proK-promoted mutant $tRNA^{EcTyr}_{CUA}$, a β-lactamase reporter mutagenized at the third amino acid to a stop codon (TAG), a CAT reporter mutagenized at Q98 to a stop codon (TAG), an arabinose inducible T7 RNA polymerase mutagenized at amino acids 8 and 114 to stop codons (TAG), and T7-promoted wild-type GFPuv. The incorporation of pBPA at the stop codons by active EcYRS-pBPARS ePCR library members leads to full-length expression of the β-lactamase conferring ampicillin resistance. Additionally, suppression of the T7 RNA polymerase drives expression of GFPuv. 5.7 x 10^7 CFU were plated on LB + 0.5x Spec/0.5x Tet/0.5x Kan /0.02% arabinose + ampicillin (400, 600, 800 and 1000 μg/mL) + 3 μg/mL chloramphenicol in the presence of 1 mM pBPA for 24 hours at 37 ºC.

3.5.6 mCherry-sfGFP-151-TAG expression and FACS

The pBK EcYRS-pBPARS ePCR randomized library and the positive selection reporter plasmid pEvol mCherry sfGFP-151-TAG were co-transformed into ATMY4 and grown at 37 ºC overnight. The next day, the plates were harvested and used to inoculate LB media cultures at an OD600 of 0.03 and were grown with shaking to an OD600 of 0.4 (37 ºC, 250 rpm). The cultures were induced with 1 mM IPTG at 30 ºC with shaking (250 rpm) for 1.5 hours in the presence of 1 mM pBPA. Following induction, the cultures were spun down (4,000 g, 10 min) and resuspended in 1X PBS. The cells were diluted to ~1.0 x 10^6 cells/mL and submitted for FACS analysis. More than 1.0 x 10^6 cells were recovered in a 2 mL 2X concentration LB media culture and grown at 37 ºC for 4 hours with shaking (250 rpm). The recovered cell culture was then plated on LB agar media supplemented with antibiotics (1X Spec/Kan/Chlor) and grown overnight.

3.5.7 sfGFP151-TAG expression and purification

The pBK aaRS and pEvol sfGFP-151-TAG plasmids were co-transformed into ATMY4 *E. coli* cells. A single transformant was used to inoculate a 5 mL LB media overnight culture (1X Spec/Kan/Chlor). The next day, 200 µL of this overnight culture was used to inoculate a 20 mL LB media culture (1X Spec/Kan/Chlor) that was grown, with shaking, to an OD600 of 0.6 (37 ºC, 250 rpm). Once the culture reached the optimal OD600, it was induced with IPTG (final concentration 1 mM) in the presence of pBPA (final concentration 1 mM) for 18 hours with shaking (30 ºC, 250 rpm). The culture was then spun down, the LB media was removed, and the remaining cell pellet was resuspended in B-PER lysis buffer (Thermo Scientific), 1X Halt Protease Inhibitor Cocktail (Thermo Scientific), and 0.01% Pierce Universal Nuclease (Thermo Scientific). The resuspension was left at room temperature on a nutator for 30 min and then spun down (5,000 g, 10 min, 4 ºC). The clarified lysate was loaded onto a HisPur Ni-NTA resin (Thermo Scientific) column, and the sfGFP151-pBPA was purified following the manufacturer's protocol. Protein purity and mass were validated by SDS-PAGE and LC-MS.

3.5.8 EGFP-39-TAG expression and purification

The pAcBac1-EGFP-39-TAG and pB1U aaRS plasmids were co-transfected into HEK293T cells. One day before transfection, the HEK293T cells were seeded at a density of 8 x 10^6 cells per 10 cm dish. Once the cells reached ~90% confluence, they were transfected with a mixture of of PEI MAX (50 µL PEI MAX: 10 µg DNA), DNA, and DMEM. The transfection mixture incubated for 10 minutes prior to transfection. The transfection mixture was then added to the 10 cm dish, followed by pBPA (final concentration of 1 mM). After 48 hours of expression, cells were harvested, washed twice with 1X PBS buffer (spun at 5,000 g, 5 min, 4 ºC), and lysed with

CelLytic M lysis buffer (Sigma) supplemented with 1x Halt Protease Inhibitor and 0.01% Pierce Universal Nuclease. Following resuspension, the lysed cells were left to incubate at room temperature for 20 min. The cells were then spun down (spun at 5,000 g, 5 min, 4 ºC) and the EGFP-39-pBPA was purified from the clarified lysate with HisPur Ni-NTA resin following the manufacturer's protocol. Protein purity was characterized with both SDS-PAGE and LC-MS.

3.5.9 EGFP-2X-TAG expression and fluorescence analysis

The pAcBac1-*Ec*Tyr-pBPARS-Gen3-16X-YtR-TAG and pAcBac3-UbiC-MbPylRS-8XPytR-TGA-EGFP-39-TAG plasmids were co-transfected into HEK293T cells. One day before transfection, the HEK293T cells were seeded at a desity of 8 x 10^6 cells per 10 cm dish. Once the cells reached ~90% confluence, they were transfected with a mixture of of PEI MAX (50 µL PEI MAX: 10 µg DNA), DNA, and DMEM. The transfection mixture incubated for 10 minutes prior to transfection. The transfection mixture was then added to the 10 cm dish, followed by pBPA and BocK (1 mM of each). After 48 hours of expression, cells were harvested, washed twice with 1X PBS buffer (spun at 5,000 g, 5 min, 4 ºC), and fluorescence analysis was carried out on the cell suspension (ex = 488 nm, em = 520 nm).

3.6 References

1. Dumas, A., Lercher, L., Spicer, C.D., Davis, B.G. Designing logical codon reassignment - Expanding the chemistry in biology. Chem Sci. 6(1), 50 (2015).
2. Italia, J.S., Addy, P.S., Wrobel, C.J.J., Crawford, L.A., Lajoie, M.J., Zheng, Y., Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol. 13, 446 (2017).
3. Italia, J.S., Latour, C., Wrobel, C.J.J., Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. Cell Chemical Biology 25(10), 1304 (2018).
4. Wang, L., Brock, A., Herberich, B., Schultz, P.G. Expanding the Genetic Code of *Escherichia coli*. Science 292, 498 (2001).
5. Wang, L., Schultz, P.G. A general approach for the generation of orthogonal tRNAs. Chemistry & Biology 8, 883 (2001).
6. Huang, W., Petrosino, J., Hirsch, M., Shenkin, P.S., Palzkill, T. Amino acid sequence determinants of beta-lactamase structure and activity. J Mol Biol. 258 (4), 688 (1996).
7. Liu, D.R., Schultz, P.G. Progress toward the evolution of an organism with an expanded genetic code. Proc Natl Acad Sci U S A 96 (9), 4780 (1999).
8. Neu, H.C. Effect of β-Lactamase Location in *Escherichia coli* in Penicillin Synergy. Applied Microbiology 17 (6), 783 (1969).
9. Abraham, E.P., Chain, E. An Enzyme from Bacteria able to Destroy Penicillin. Nature 3713, 837 (1940).
10. Rice, L.B. Mechanisms of Resistance and Clinical Relevance of Resistance to β-Lactams, Glycopeptides, and Fluoroquinolones. Mayo Clin Proc. 87 (2), 198 (2012).
11. Pastrnak, M., Magliery, T.J., Schultz, P.G. A New Orthogonal Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for Evolving an Organism with an Expanded Genetic Code. Helvetica Chimica Acta 83, 2277 (2000).
12. Zheng, J., Guo, N., Wagner, A. Selection enhances protein evolvability by increasing mutational robustness and foldability. Science 370, eabb5962 (2020).
13. Bloom, J.D., Labthavikul, S.T., Otey, C.R., Arnold, F.H. Protein stability promotes evolvability. Proc. Natl. Acad. Sci U.S.A. 103, 5869 (2006).
14. Bershtein, S., Tawfik, D.C. Ohno's model revisited: Measuring the frequency of potentially adaptive mutations under various mutational drifts. Mol. Biol. Evol. 25, 2311 (2008).
15. Biswas, T., Houghton, J. L., Garneau-Tsodikova, S., Tsodikov, O. V. The structural basis for substrate versatility of chloramphenicol acetyltransferase CATI. Protein science: a publication of the Protein Society, 21(4), 520 (2012).
16. Dunkle, J.A., Xiong, L., Mankin, A.S., Cate, J.H. Structures of the *Escherichia coli* ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. Proc Natl Acad Sci USA. 107, 17152 (2010).
17. Bulkley, D., Innis, C.A., Blaha, G., Steitz, T.A. Revisiting the structures of several antibiotics bound to the bacterial ribosome. Proc Natl Acad Sci USA. 107, 17158 (2010).
18. Packer, M.S., Liu, D.R. Methods for the directed evolution of proteins. Nat Rev Genet 16, 379 (2015).
19. Owens, A.E., Grasso, K.T., Ziegler, C.A., Fasan, R. Two-Tier Screening Platform for Directed Evolution of Aminoacyl-tRNA Synthetases with Enhanced Stop Codon Suppression Efficiency. ChemBioChem 18(12), 1109 (2017).

20. 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 (10), 1044 (2002).
21. Link, A.J., Vink, M.K.S., Agard, N.J., Prescher, J.A., Bertozzi, C.R., Tirrell, D.A. Discovery of aminoacyl-tRNA synthetase activity through cell-surface display of noncanonical amino acids. Proc Natl Acad Sci 103 (27), 10180 (2006).
22. Kuhn, S.M., Rubini, M., Fuhrmann, M., Theobald, I., Skerra, A. Engineering of an Orthogonal Aminoacyl-tRNA Synthetase for Efficient Incorporation of the Non-natural Amino Acid O-Methyl-L-tyrosine using Fluorescence-based Bacterial Cell Sorting. J. Mol. Biol. 404 (1), 70 (2010).
23. Matula, K., Rivello, F., Huck, W.T.S. Single-Cell Analysis Using Droplet Microfluidics. Adv. Biosys. 4 (1), (2020).
24. Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.C., Marquez, M., Klibanov, A.M., Griffiths, A.D., Weitz, D.A. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. Proc Natl Acad Sci U S A. 107 (9), 4004 (2010).
25. Hohl, A., Karan, R., Akal, A., Renn, D., Liu, X., Ghorpade, S., Groll, M., Rueping, M., Eppinger, J. Engineering a Polyspecific Pyrrolysyl-tRNA Synthetase by a High Throughput FACS Screen. Sci Rep 9, 11971 (2019).
26. Young, T. S., Schultz, P. G. Beyond the canonical 20 amino acids: expanding the genetic lexicon. The Journal of biological chemistry, 285(15), 11039 (2010).
27. Chatterjee, A., Sun, S.B., Furman, J.L., Xiao, H., Schultz, P.G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*. Biochemistry 52(10), 1828 (2013).
28. Guo, J., Melançon, C.E., Lee, H.S., Groff, D., Schultz, P.G. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. Angew. Chemie Int. Ed. 48, 9148 (2009).
29. Mukai, T., Hayashi, A., Iraha, F., Sato, A., Ohtake, K., Yokoyama, S., Sakamoto, K. Codon reassignment in the Escherichia coli genetic code. Nucleic Acids Res. 38 (22), 8188 (2010).
30. Johnson D.B., Xu, J., Shen, Z., Takimoto, J.K., Schultz, M.D., Schmitz, R.J., Xiang, Z., Ecker, J.R., Briggs, S.P., Wang, L. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol. 7 (11), 779 (2011).
31. Amiram, M., Haimovich, A.D., Fan, C., Wang, Y.S., Aerni, H.R., Ntai, I., Moonan, D.W., Ma, N.J., Rovner, A.J., Hong, S.H., Kelleher, N.L., Goodman, A.L., Jewett, M.C., Söll, D., Rinehart, J., Isaacs, F.J. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. Nat Biotechnol. 33 (12), 1272 (2015).
32. Johnson, D.B.F., Wang, C., Xu, J., Schultz, M.D., Schmitz, R.J., Ecker, J.R., Wang, L. Release Factor One is Nonessential in *Escherichia coli*. ACS Chem. Biol. 7 (8), 1337 (2012).
33. Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A., Rohland, N., Schultz, P.G., Jacobson, J.M., Rinehart, J., Church, G.M., Isaacs, F.J. Genomically recoded organisms expand biological functions. Science 342 (6156), 357 (2013).
34. Italia, J.S., Addy, P.S., Erickson, S.B., Peeler, J.C., Weerapana, E., Chatterjee, A. Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. J Am Chem Soc. 141 (15), 6204 (2019).

35. Zheng, Y., Gilgenast, M.J., Hauc, S., Chatterjee, A. Capturing post-translational modification-triggered protein-protein interactions using dual noncanonical amino acid mutagenesis. ACS Chem Biol. 13 (5), 1137 (2018).
36. Dick, M., Sarai, N.S., Martynowycz, M.W., Gonen, T., Arnold, F.H. Tailoring tryptophan synthase TrpB for selective quaternary carbon bond formation. J. Am. Chem. Soc. 141 (50), 19817 (2019).
37. Bloom, J.D., Labthavikul, S.T., Otey, C.R., Arnold, F.H. Protein stability promotes evolvability. PNAS 103 (15), 5869 (2006).
38. Bryson, D.I., Fan, C., Guo, L.T., Miller, C., Söll, D., Liu, D.R. Continuous directed evolution of aminoacyl-tRNA synthetases. Nat Chem Biol. (12), 1253 (2017).
39. Chin, J.W., Cropp, T.A., Anderson, C., Mukherji, M., Zhang, Z., Schultz, P.G. An expanded eukaryotic genetic code. Science 301, 964 (2003).
40. 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 Escherichia coli. Proc. Natl. Acad. Sci. 99 (17), 11020 (2002).
41. Krishnamurthy, M., Dugan, A., Nwokoye, A., Fung, Y.H., Lancia, J.K., Majmudar, C.Y., Mapp, A.K. Caught in the act: covalent cross-linking captures activator-coactivator interactions in vivo. ACS Chem Biol. 6 (12), 1321 (2011).
42. McCullum EO, Williams BA, Zhang J, Chaput JC. Random mutagenesis by error-prone PCR. Methods Mol Biol. 634, 103 (2010).
43. Klock, H.E., Lesley, S.A. The polymerase incomplete primer extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. Methods Mol Biol. 498, 91 (2009).
44. Prabakaran, S., Lippens, G., Steen, H., Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. Wiley Interdiscip Rev Syst Biol Med. 4 (6), 565 (2012).

Chapter 4

Application of the ATMY5 selection system to evolve new UAA-selective *Ec*TyrRS mutants

4.1 Introduction

4.1.1 Discrepancies between the tyrosyl-derived UAA toolbox in eukaryotes vs. bacteria

A broad chemical range of UAAs have been incorporated in *E. coli* with the *M. jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/tRNA pair.[1-4] This includes UAAs that function as bioconjugation handles, photoaffinity probes, post-translational modification (PTM) mimics, etc.[1-4] While alternate aaRS/tRNA pairs have enabled access to these chemistries in eukaryotes, some of the structures (e.g., tyrosine PTM mimics) rely on the unique architecture of the TyrRS active site structure.[1-4] Unfortunately, the powerful *Mj*TyrRS/tRNA pair cannot be used for eukaryotic GCE due to its cross-reactivity with the host organism's aaRS/tRNA pairs. Instead, researchers rely on bacterial aaRS/tRNA pairs for eukaryotic GCE due to their naturally evolved orthogonality.[1-6] In fact, two decades ago, the *E. coli* tyrosyl-tRNA synthetase (*Ec*TyrRS)/tRNA pair was established for UAA incorporation in eukaryotes.[7] However, compared to the *Mj*TyrRS/tRNA pair, the *Ec*TyrRS/tRNA pair has been wildly unsuccessful (Figure 4.1) due to the challenges associated with the established directed evolution platforms used to engineer the *Ec*TyrRS/tRNA pair.[1-4] To broaden the range of genetically encoded UAAs with the *Ec*TyrRS/tRNA pair new selection paradigms must be established. If successful, the scope of GCE technology in eukaryotes will be greatly expanded furthering our ability to investigate the physiological roles of proteins.

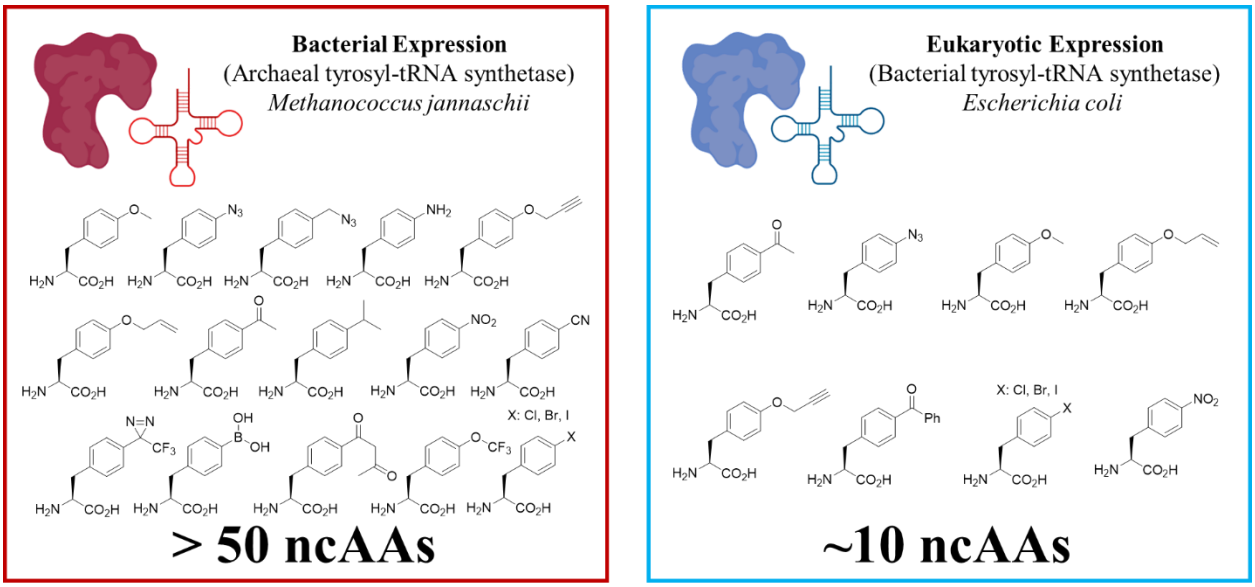


Figure 4.1. Tyrosyl-derived UAAs incorporated in eukaryotes and bacteria.[1-4]

4.1.2 Efforts towards engineering the EcTyrRS

We hypothesized that engineering the *Ec*TyrRS/tRNA pair was partially less fruitful, relative to the *Mj*TyrRS/tRNA pair, due to the cumbersome nature of the established *Ec*TyrRS directed evolution platform in yeast.[2, 6-9] This perceived limitation was overcome through the development of engineered *E. coli* strains in which the endogenous *Ec*TyrRS/tRNA pair was functionally replaced with its archaeal counterpart (ATMY strains).[6, 9] The engineered ATMY *E. coli* strains did not exhibit significant growth penalties and the "liberated" *Ec*TyrRS/tRNA pair was successfully reintroduced as an orthogonal TAG-codon suppressor.[6,9] Ultimately, we believed that the ATMY *E. coli* strains would provide access to a more easily manipulated selection platform for the engineering of *Ec*TyrRS to expand our UAA toolbox for eukaryotic GCE applications.

Although the ability to rapidly engineer the *Ec*TyrRS/tRNA pair using the facile ATMY *E. coli* directed evolution platform has provided access to several new UAAs,[9] in some instances, previously engineered mutants were not accessible (e.g., pBPA-selective *Ec*TyrRS[7]). Eventually, this observation was attributed to the cross-reactivity of the $tRNA^{EcTyr}_{CUA}$ with the endogenous *E. coli* glutaminyl-tRNA synthetase (*Ec*GlnRS) (Figure 4.2). We remedied the cross-reactivity of the $tRNA^{EcTyr}_{CUA}$ through directed evolution and subsequently used the orthogonal $tRNA^{EcTyr}_{CUA}$ to engineer the previously inaccessible pBPA-selective *Ec*TyrRS (chapter 2 of this dissertation). The suppression efficiency of this weakly active pBPA-selective *Ec*TyrRS was further enhanced by coupling our orthogonal ATMY5-based selection platform with a dual-fluorescent reporter system (Figure 4.3). Following these successes accomplished with our ATMY5 selection platform, we sought to further engineer the *Ec*TyrRS/tRNA pair for the incorporation of a broader range of UAAs.

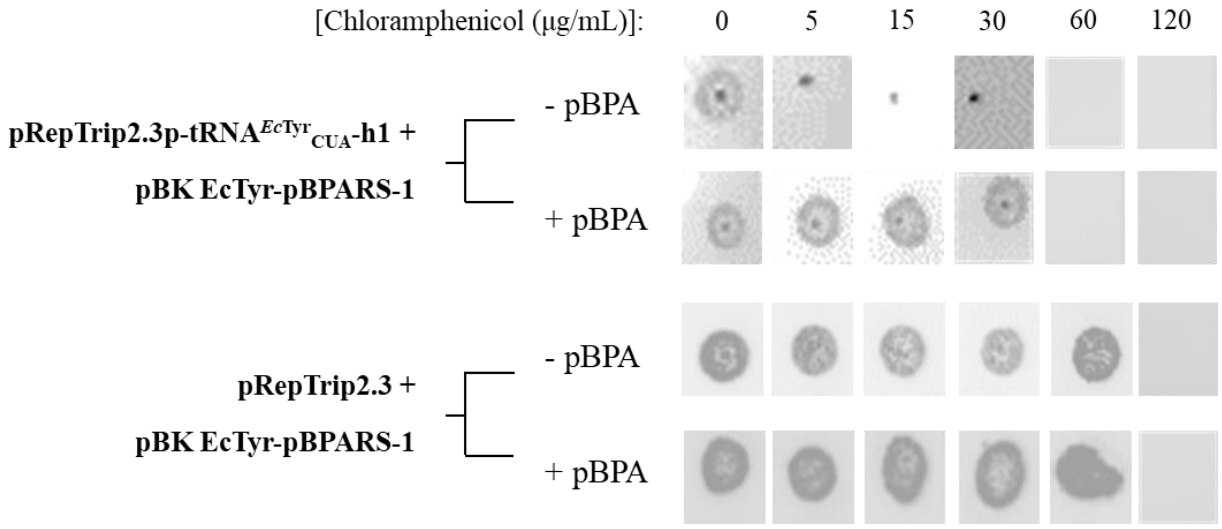


Figure 4.2. $tRNA^{EcTyr}_{CUA}$ cross-reactivity with the endogenous *Ec*GlnRS. A chloramphenicol acetyl-transferase (CAT)-TAG reporter assay shows that cross-reactivity of the $tRNA^{EcTyr}_{CUA}$ is eliminated once it was engineered (top). In the presence of a pBPA-selective *Ec*TyrRS (*Ec*Tyr-pBPARS-1) the orthogonal $tRNA^{EcTyr}_{CUA}$-h1 facilitates cell-survival only in the presence of pBPA (top). Contrastingly, a reporter plasmid containing the progenitor $tRNA^{EcTyr}_{CUA}$ has no observed difference in cell-survival in the presence and absence of pBPA (bottom).

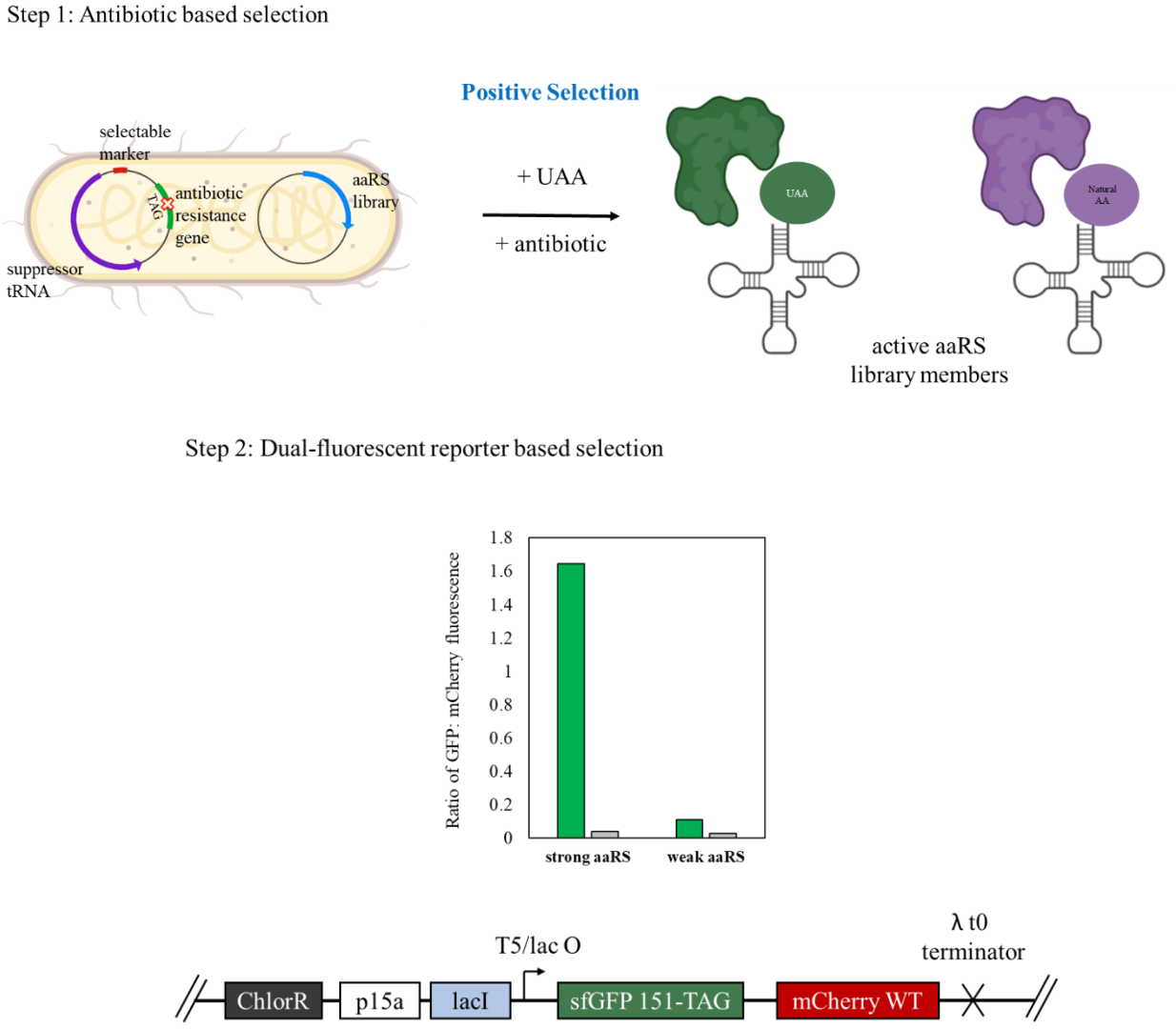


Figure 4.3. Directed evolution platform to improve aaRS suppression efficiency. First, an ampicillin-based selection is used to eliminate any inactive aaRS library members (Step 1). Second, a dual-fluorescent reporter construct is used to sort cells based on the ratio of full-length sfGFP production relative to mCherry production (Step 2).

4.2 Results and discussion

4.2.1 Engineering of pAEY-selective EcTyrRS mutants

The site-specific incorporation of azide-containing UAAs provides a powerful route to selectively click various tags (fluorescent probes, antibodies, etc.) onto proteins through copper-

catalyzed azide alkyne cycloaddition (CuAAC) and strain-promoted azide alkyne cycloaddition (SPAAC) (Figure 4.4).[10-12] Incorporation of *p*-azido-L-phenylalanine (pAzF) with the *Ec*TyrRS/tRNA$_{CUA}$ pair has enabled access to these chemistries, but at an expense. In the cellular environment the azide is quickly reduced to the amine, rendering the azide unusable. We hypothesized the installation of a UAA with an alkyl azide would experience diminished reduction of the incorporated azide by eliminating resonance of the triazo group with the aromatic ring.[11]

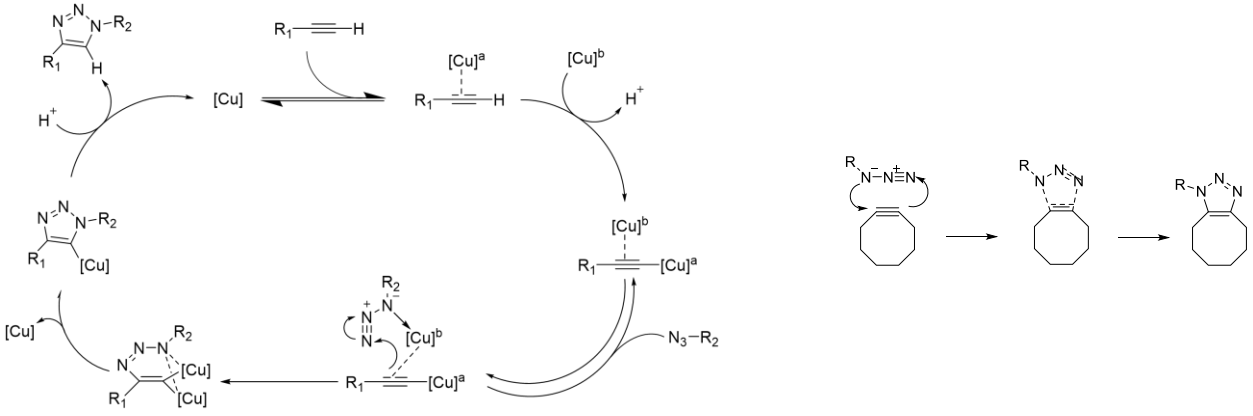
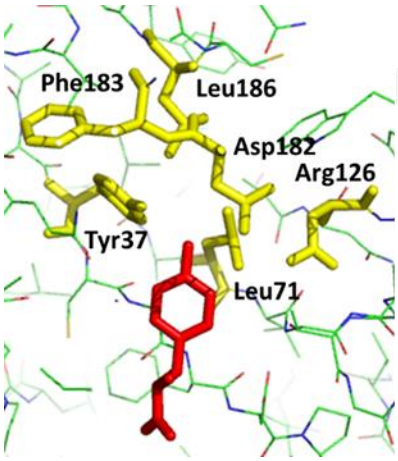


Figure 4.4. Azide-alkyne coupling reaction mechanisms.[13, 14] Copper-catalyzed azide alkyne cycloaddition is pictured left and strain-promoted azide alkyne cycloaddition is pictured right.

Based on our hypothesis outlined above, we synthesized *p*-azidoethyl-tyrosine (pAEY) (Figure 4.8) as an alkyl-azide UAA alternative to pAzF. To engineer an *Ec*TyrRS mutant that could incorporate pAEY an *Ec*TyrRS library[9] (Figure 4.5) (theoretical diversity 1.06 x 10^7, >99% coverage) was subjected to a round of positive selection (+1 mM pAEY) followed by TAG inactivated toxic barnase negative selection (- pAEY). Enrichment of *Ec*TyrRS mutants that charged pAEY was observed following this first round of selection. Characterization of individual library members identified three unique *Ec*TyrRS sequences that incorporated pAEY (Table 4.1). Tyr37 goes from a large, nonpolar amino acid to a small, nonpolar amino acid (Cys/Gly). Leu71 remained a nonpolar amino acid (Val/Ile/Cys). Asn126 was unaltered. Asp182 changed from a

large polar, acidic amino acid to small amino acid (Ser). Phe183 changed to a large, nonpolar amino acid (Met). Leu186 became a small, nonpolar amino acid (Gly, Ala, Cys).



Y37	L71	N126	D182	F183	L186	D265
FLIMVSTAYHCG	NBT	NSPTACGDH	NST	NNK	NNK	R

Figure 4.5. *Ec*TyrRS library active site and mutations.

Clones	Tyr37	Leu71	Asn126	Asp182	Phe183	Leu186	Asp265
pAEYRS-6	Cys	Cys	Asn	Ser	Met	Ala	Arg
pAEYRS-11	Gly	Val	Asn	Ser	Met	Cys	Arg
pAEYRS-16*	Cys	Ile	Asn	Ser	Met	Gly	Arg

*contains an additional point mutation D165N

Table 4.1. *Ec*TyrRS pAEY-selective mutants active site mutations.

Suppression efficiencies of the three mutants were characterized in ATMY4[9] with an sfGFP-151-TAG reporter. Successful incorporation of pAEY into sfGFP-151-TAG was confirmed by fluorescence analysis (Figure 4.6). The engineered pAEY-selective *Ec*TyrRS mutants were cloned into the mammalian expression plasmid pB1U to characterize suppression efficiencies in HEK293T cells. Expression efficiencies of EGFP-39-pAEY was confirmed by fluorescence imaging (Figure 4.7).

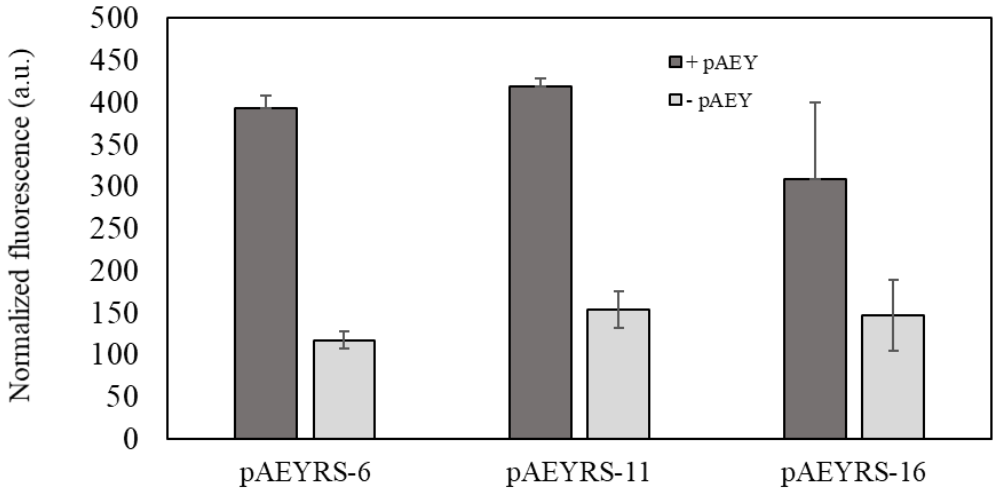


Figure 4.6. sfGFP-151-pAEY fluorescence in ATMY4. Fluorescence signal of full-length sfGFP was measured in the presence and absence of pAEY for the three pAEY-selective mutants.

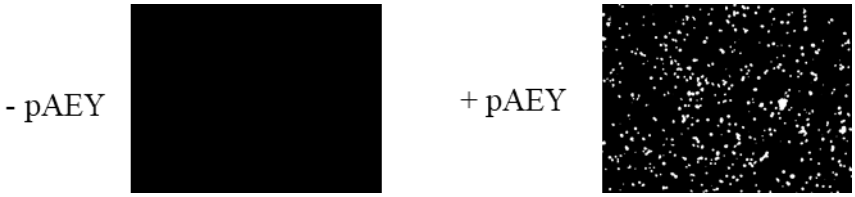


Figure 4.7. EGFP-39-pAEY fluorescence images in HEK293T. HEK293T cells were co-transfected with a pB1U plasmid harboring pAEYRS-11 and a pAcBac1 plasmid containing the EGFP reporter. Fluorescent images were taken in the presence and absence of pAEY.

4.2.2 Creation of a polyspecific EcTyrRS mutant

The selective coupling of electrophilic small-molecule probes with naturally occurring nucleophilic residues of enzymes has been extensively exploited for activity-based proteomic profiling[15, 16] and medicinal chemistry applications.[17-19] However, this application has not extended greatly into the development of protein-based inhibitors, etc. due to the technical difficulties of site-specific modification of the desired protein-agonist. GCE circumvents this limitation by evolving an orthogonal aaRS/tRNA pair for the incorporation of electrophilic UAAs that selectively react with nucleophilic amino acids in a target protein. In fact, a variety of electrophilic UAAs have already been genetically incorporated in bacteria, such as, vinyl sulfonamides, fluoroketones, alkyl and aryl ketones, and acrylamides.[1-4, 20] However, this technology has been

generally restricted from mammalian cell applications due to the difficulty of evolving orthogonal aaRS/tRNA pairs for electrophilic UAA incorporation in mammalian cells.

With our new ATMY5 directed evolution platform, we sought to identify a mutant *Ec*TyrRS that was capable of incorporating an electrophilic UAA. The aforementioned *Ec*TyrRS library (theoretical diversity 1.06 x 10^7, >99% coverage) and antibiotic-based selection strategy was used to identify mutants that charged *p*-acrylamido-(S)-phenylalanine (pAcrF) (Figure 4.8). One unique mutant was identified by DNA sequencing through these selections. The Tyr37 residue changed from a large nonpolar amino acid to a small nonpolar amino acid (Gly). The Leu71 remained a hydrophobic amino acid (Val). The Asp182 became Cys. The Phe183 remained an aromatic amino acid (Tyr). The Leu186 was modified to a smaller amino acid (Cys) (Table 4.2).

Clones	Tyr37	Leu71	Asn126	Asp182	Phe183	Leu186	Asp265
pAcrFRS-9	Gly	Val	Asn	Cys	Tyr	Cys	Arg

Table 4.2. pAcrF-selective *Ec*TyrRS active site mutations.

Due to similarities between the active site structures of the pAEY-selective *Ec*TyrRS mutant active sites (Table 4.1) and the pAcrF- selective *Ec*TyrRS mutant active site (Table 4.2), we evaluated the ability of *Ec*Tyr-pAcrFRS-9 to charge pAEY. The polyspecificity of the pAcrF-selective *Ec*TyrRS mutant was further evaluated for a range of other structurally similar UAAs. Ultimately, this mutant *Ec*TyrRS was capable of charging numerous UAAs containing unique biorthogonal handles and photoaffinity probes at high levels (Figure 4.8, Table 4.5, Table 4.6). An sfGFP-151-TAG reporter fluorescence assay was used to characterize suppression efficiencies of the pAcrF-*Ec*TyrRS mutant in ATMY4 (Figure 4.9). Fluorescence and MS analysis were used to validate the successful incorporation of all of the UAAs into sfGFP-151-TAG (Figure 4.10-18).

The engineered pAcrF-selective *Ec*TyrRS mutant was cloned into the mammalian expression plasmid pB1U to characterize suppression efficiencies in HEK293T cells. Expression efficiencies of EGFP-39-UAA was confirmed by SDS-PAGE and MS analysis (Figure 4.19-26). Notably, the alkyl-azide UAAs experienced minimal to no reduction once incorporated in EGFP in HEK293T cells (Table 4.6), validating our earlier hypothesis. The incorporation of pCAcF in HEK293T cells was not achieved due to observed toxicity of the UAA (Figure 4.27).

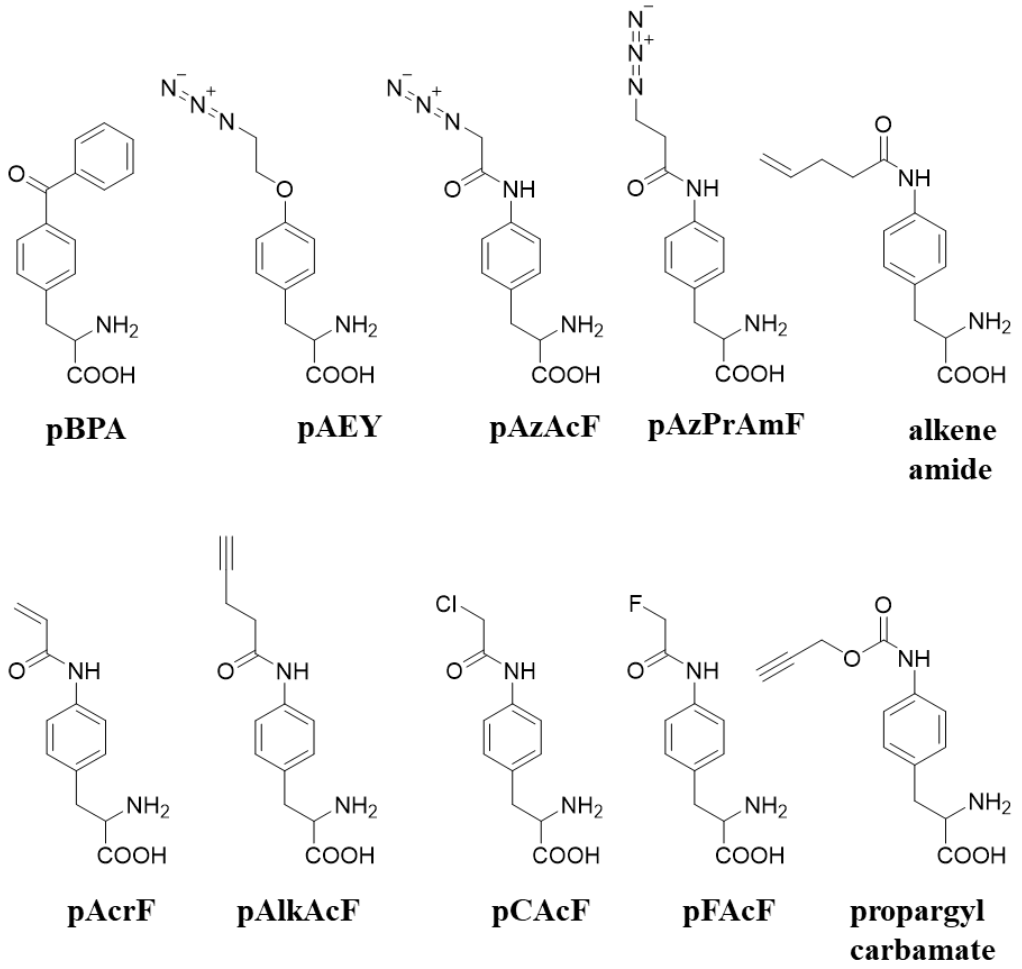


Figure 4.8. UAAs incorporated by *Ec*Tyr-pAcrFRS.

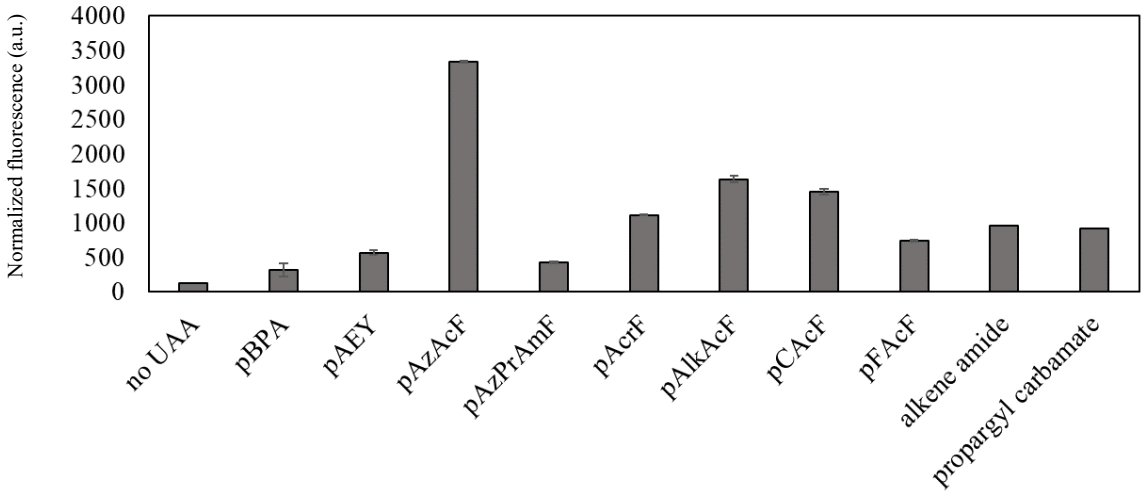


Figure 4.9. sfGFP-151-UAA fluorescence analysis in ATMY4. Fluorescence signal of full-length sfGFP was measured in the presence and absence of each UAA with the *Ec*Tyr-pAcrFRS-9.

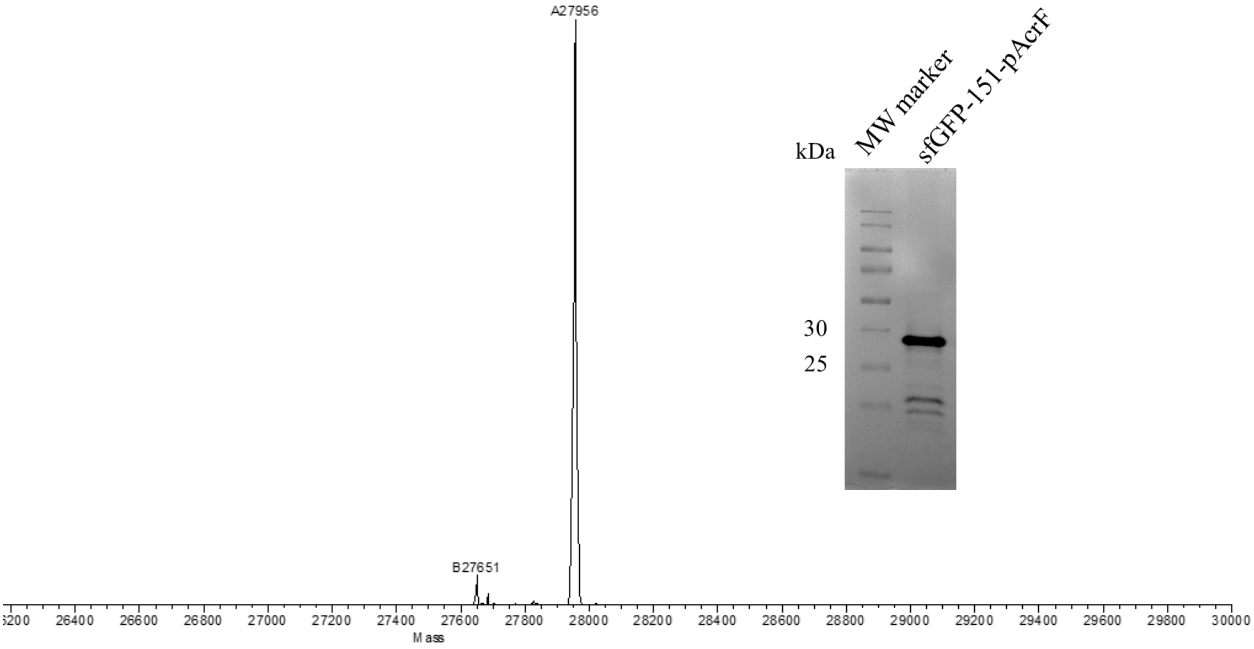


Figure 4.10. sfGFP-151-pAcrF MS analysis and SDS-PAGE.

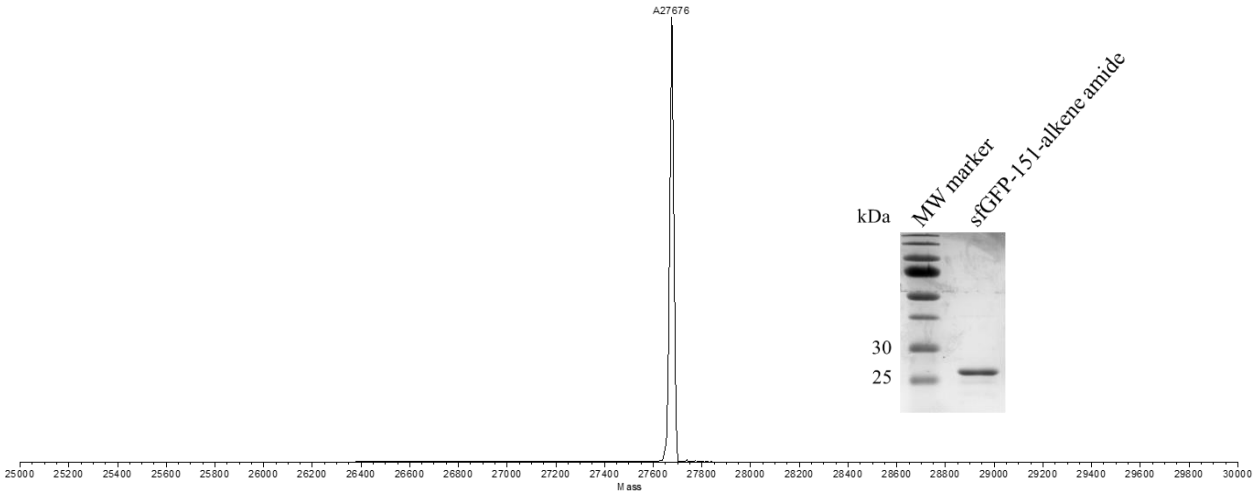


Figure 4.11. sfGFP-151-alkene amide MS analysis and SDS-PAGE.

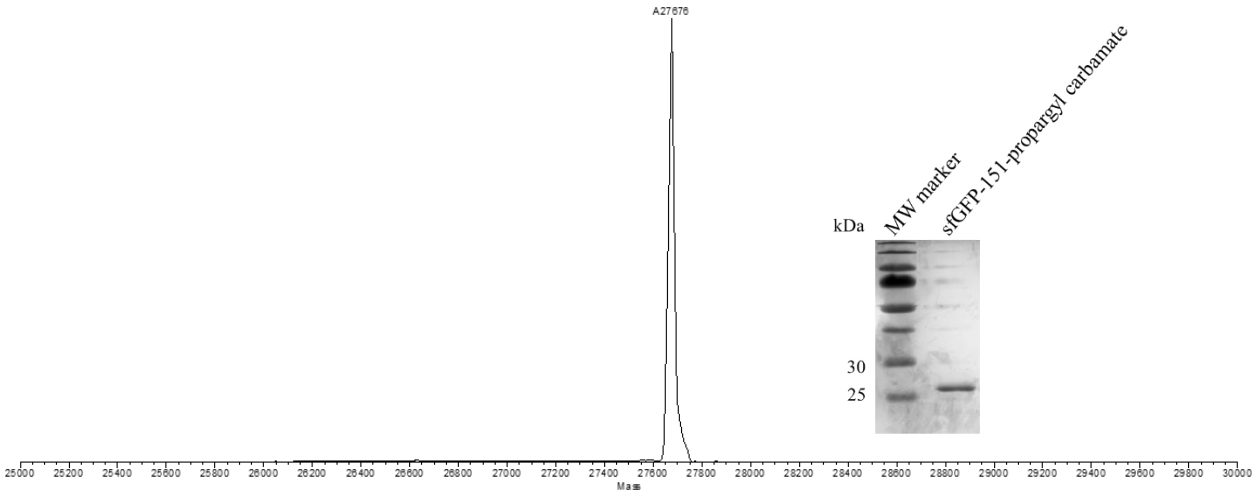


Figure 4.12. sfGFP-151-propargyl carbamate MS analysis and SDS-PAGE.

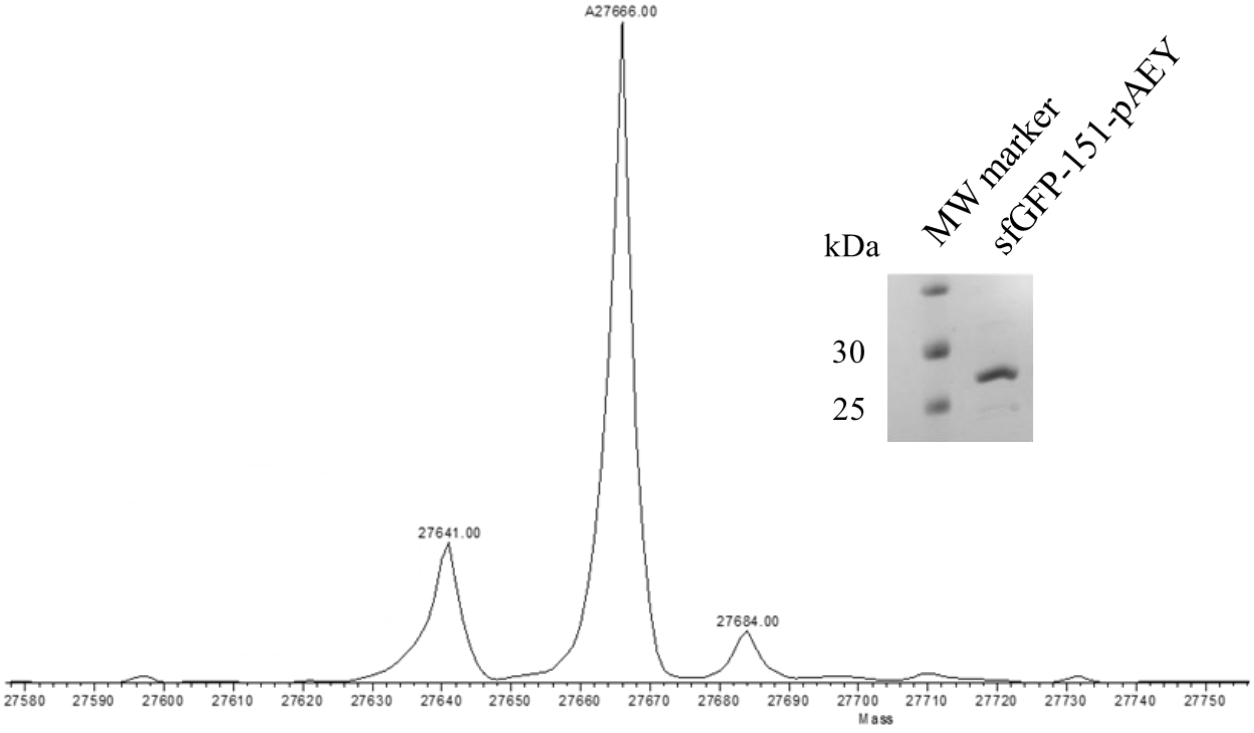


Figure 4.13. sfGFP-151-pAEY MS analysis and SDS-PAGE.

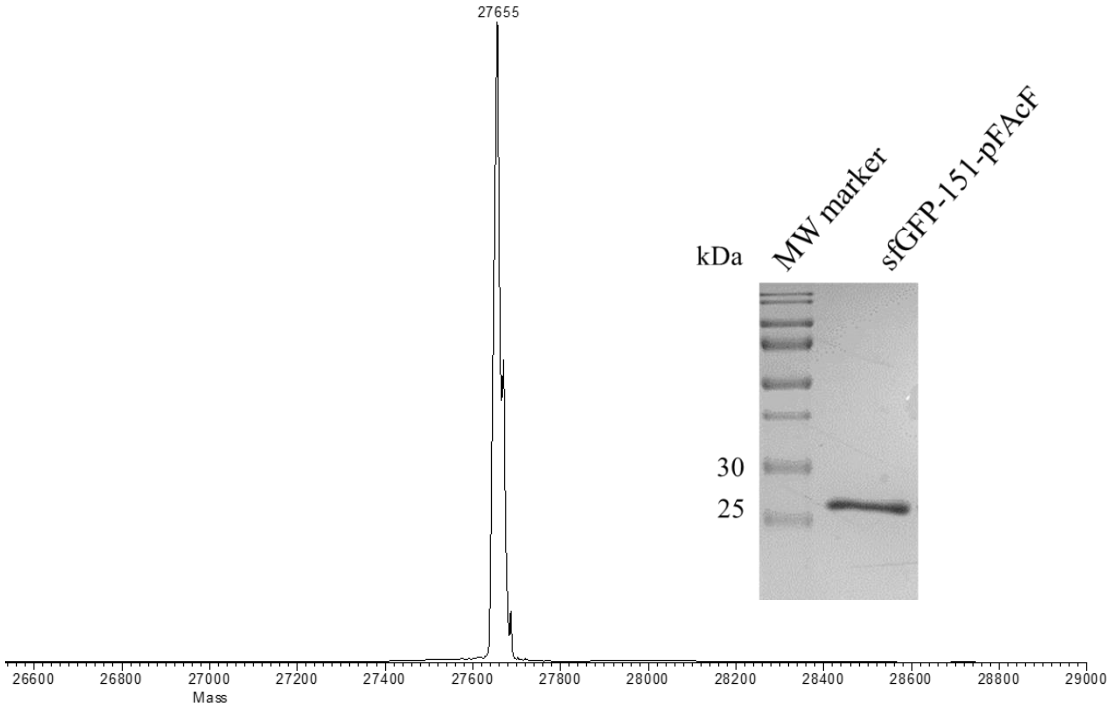


Figure 4.14. sfGFP-151-pFAcF MS analysis and SDS-PAGE.

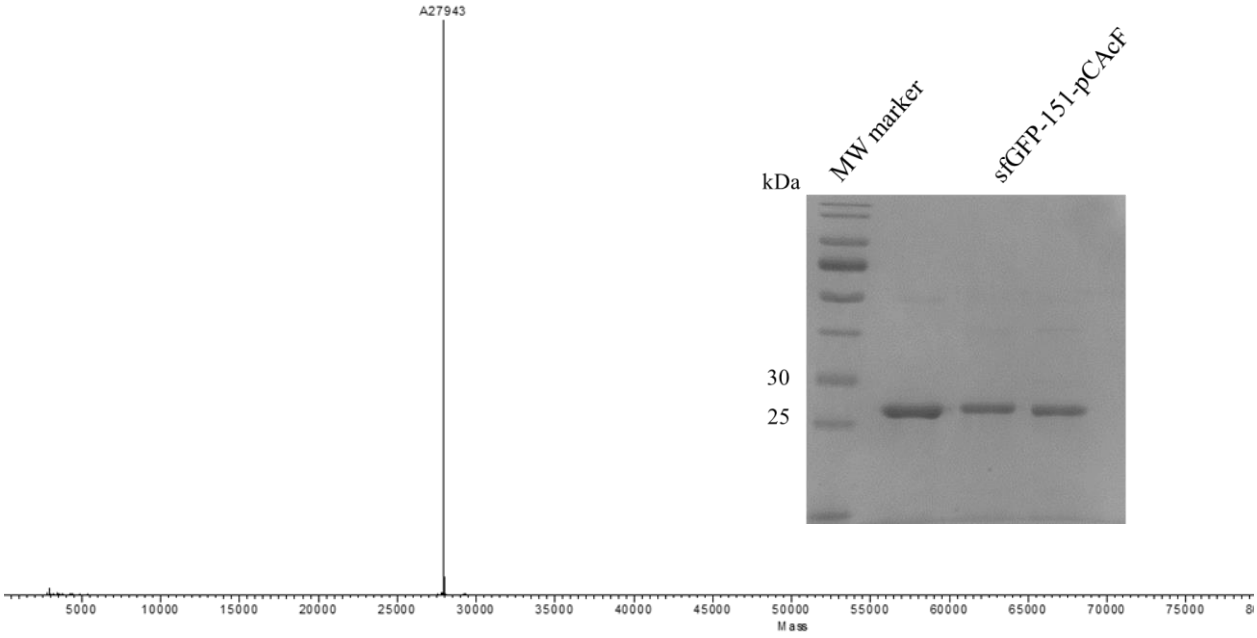


Figure 4.15. sfGFP-151-pCAcF-MS analysis and SDS-PAGE.

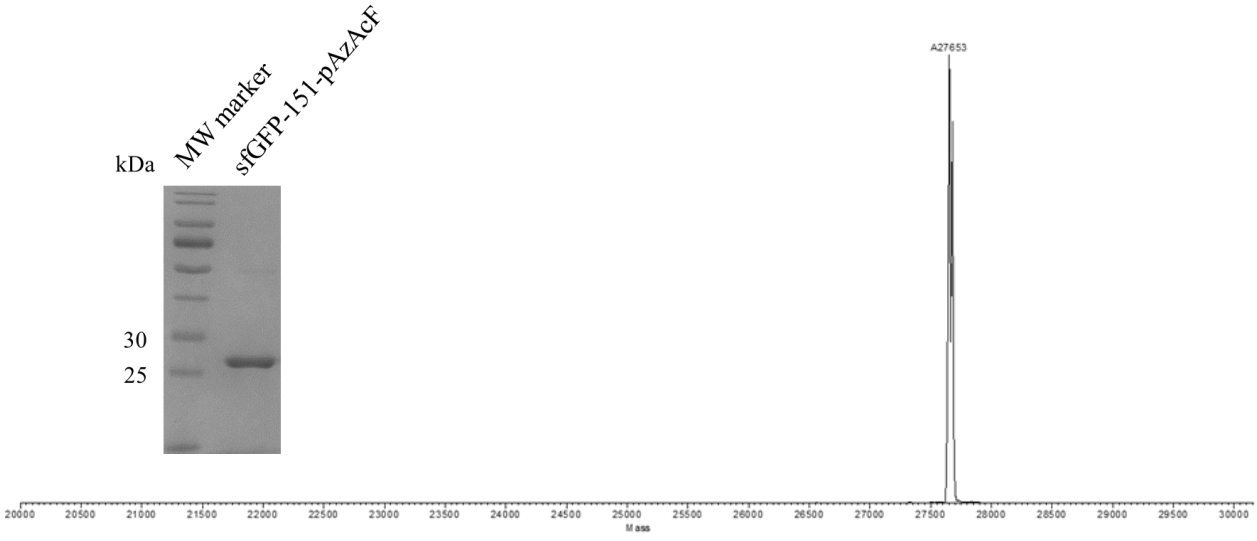


Figure 4.16. sfGFP-151-pAzAcF MS analysis and SDS-PAGE.

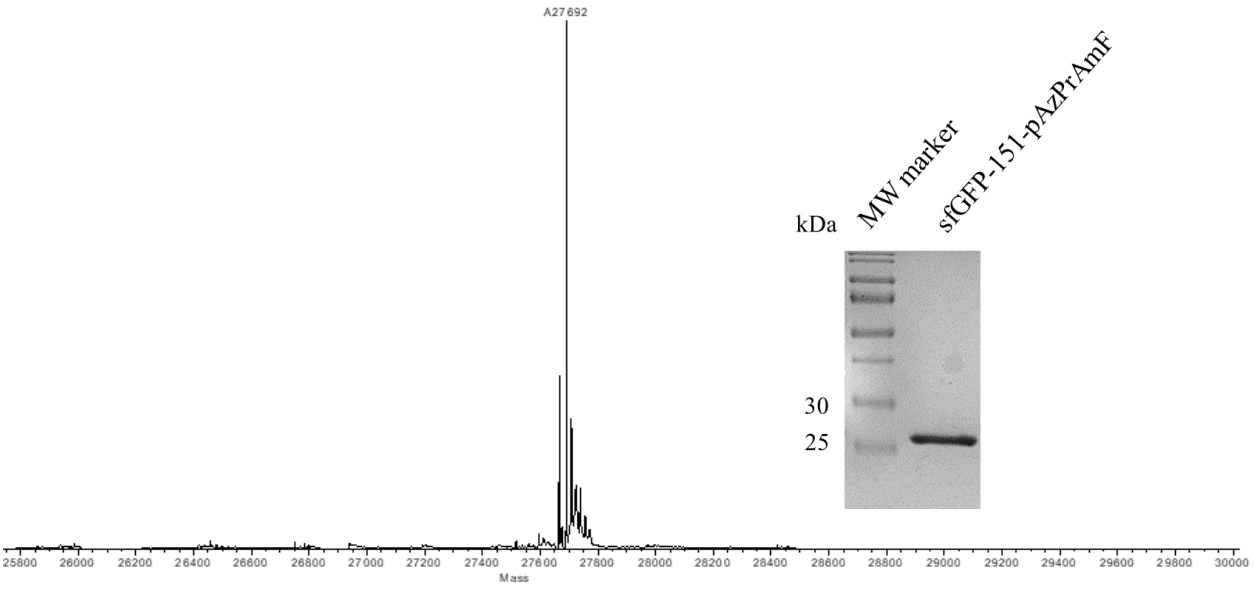


Figure 4.17. sfGFP-151-pAzPrAmF MS analysis and SDS-PAGE.

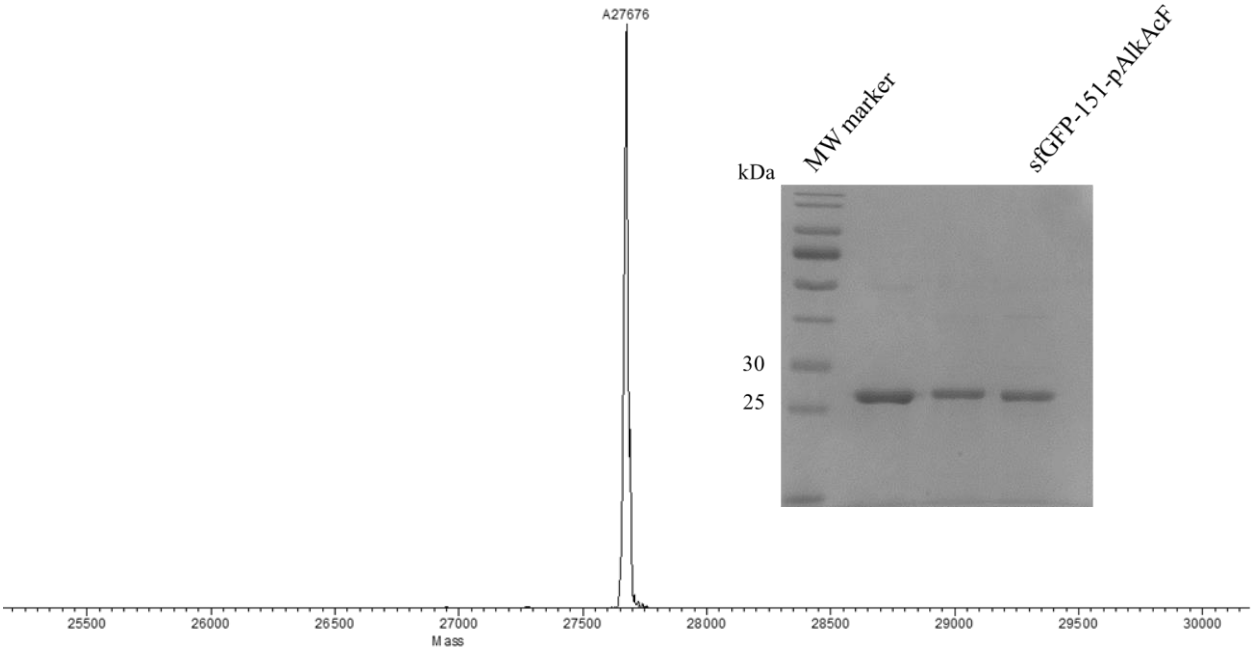


Figure 4.18. sfGFP-151-pAlkAcF MS analysis and SDS-PAGE.

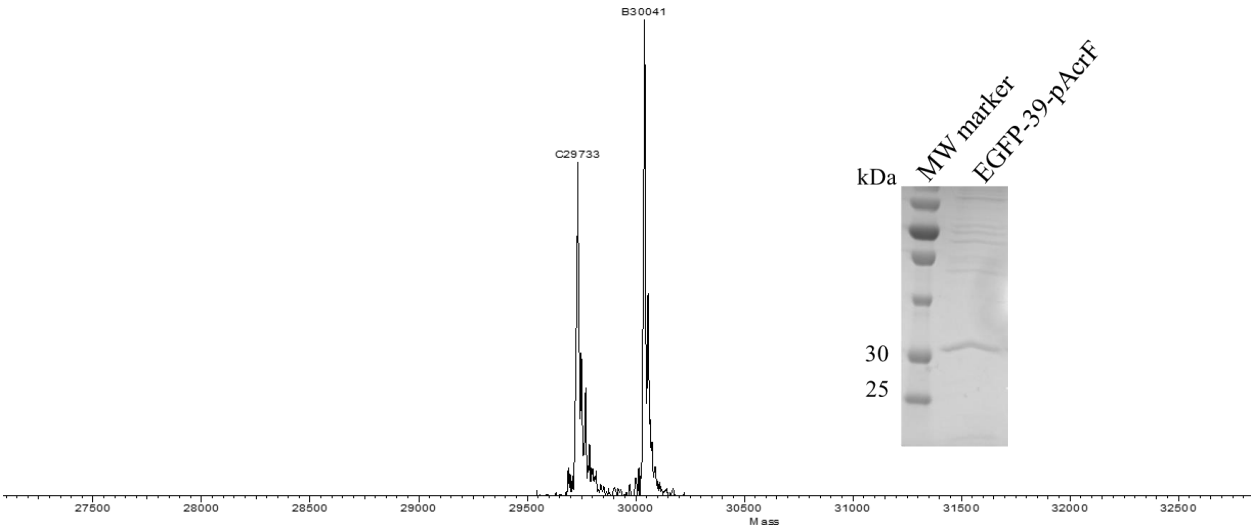


Figure 4.19. EGFP-39-pAcrF MS analysis and SDS-PAGE.

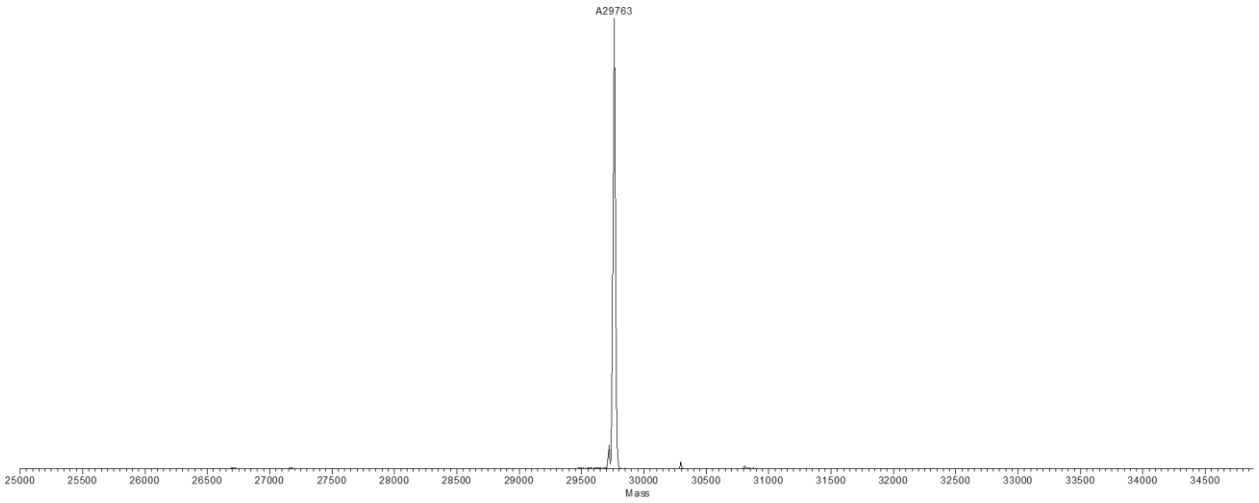


Figure 4.20. EGFP-39-alkene amide MS analysis.

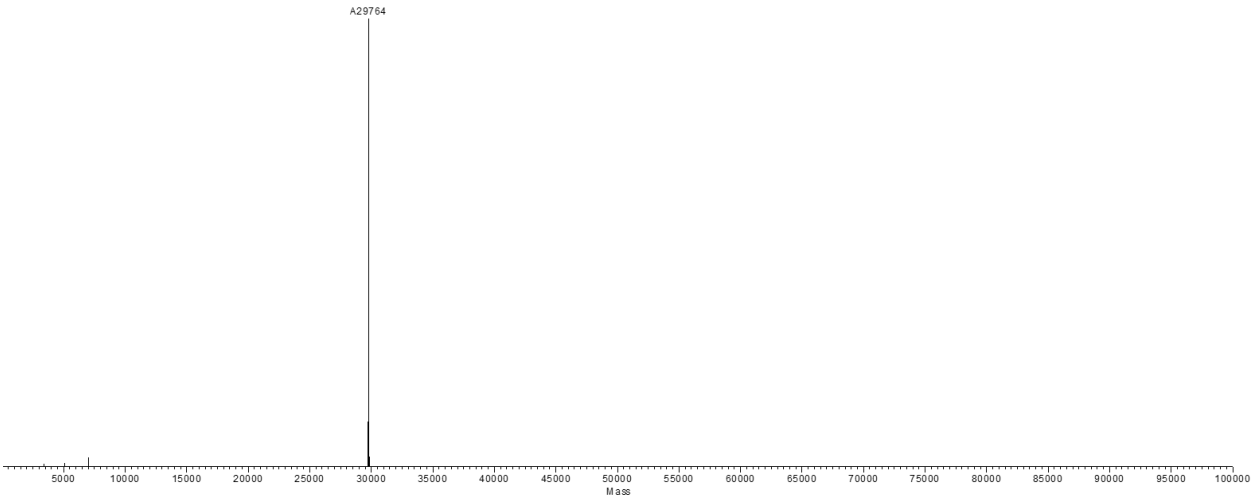


Figure 4.21. EGFP-39-propargyl carbamate MS analysis.

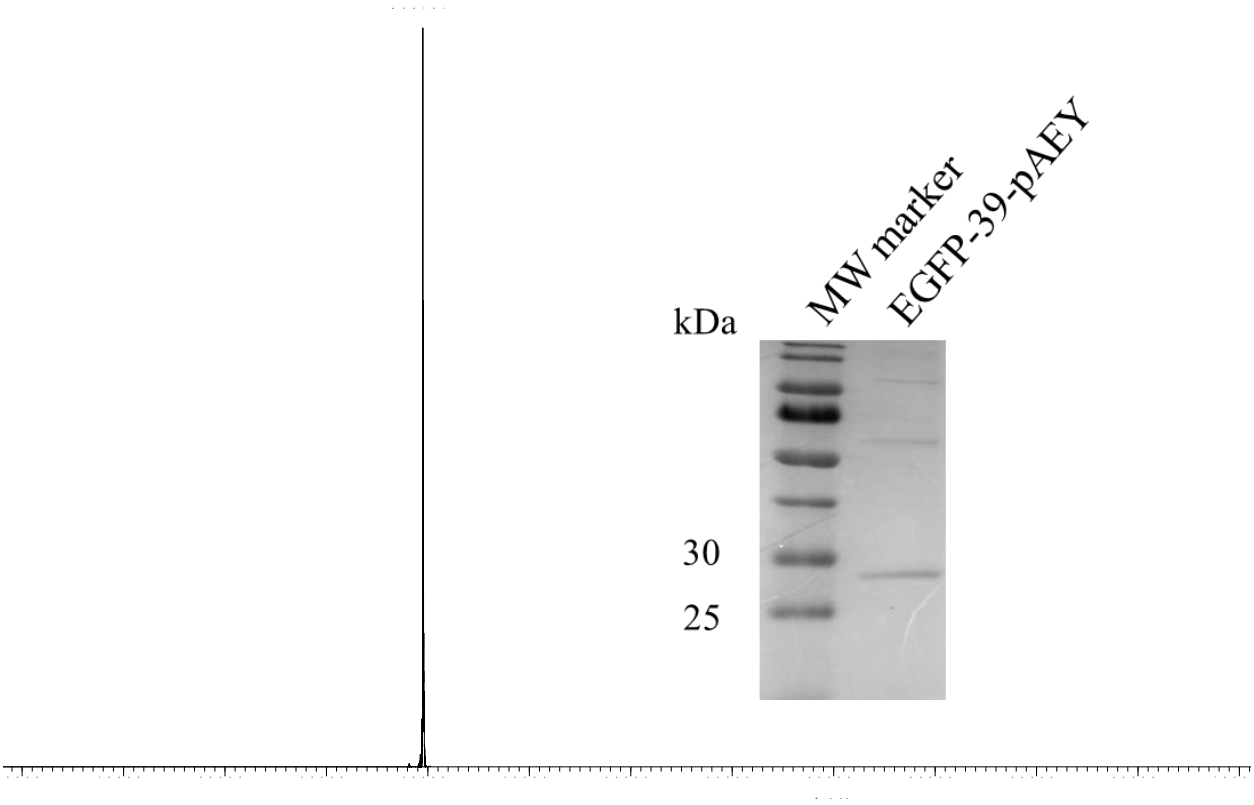


Figure 4.22. EGFP-39-pAEY MS analysis and SDS-PAGE.

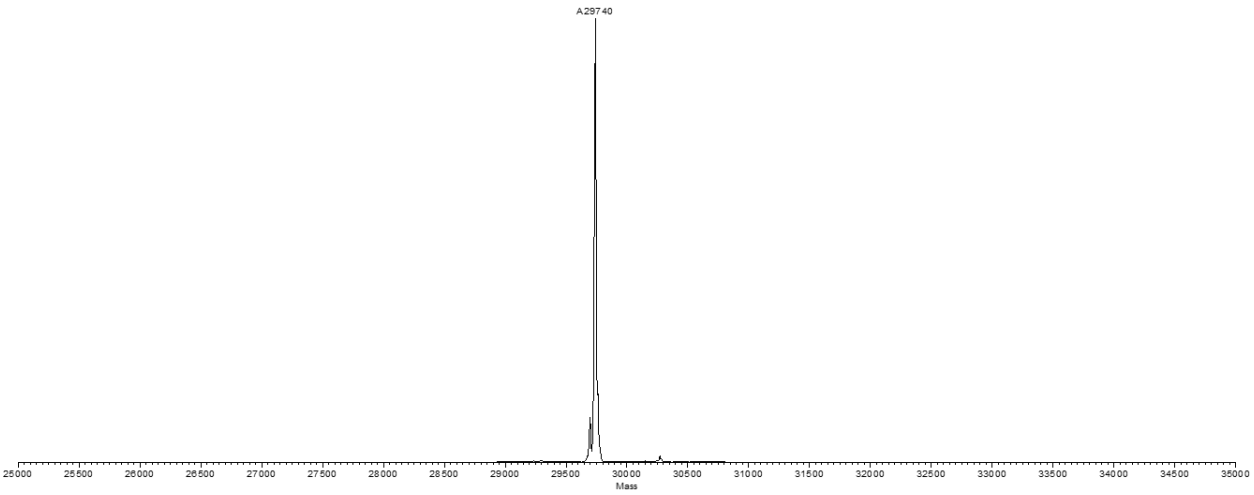


Figure 4.23. EGFP-39-pFAcF-MS analysis and SDS-PAGE.

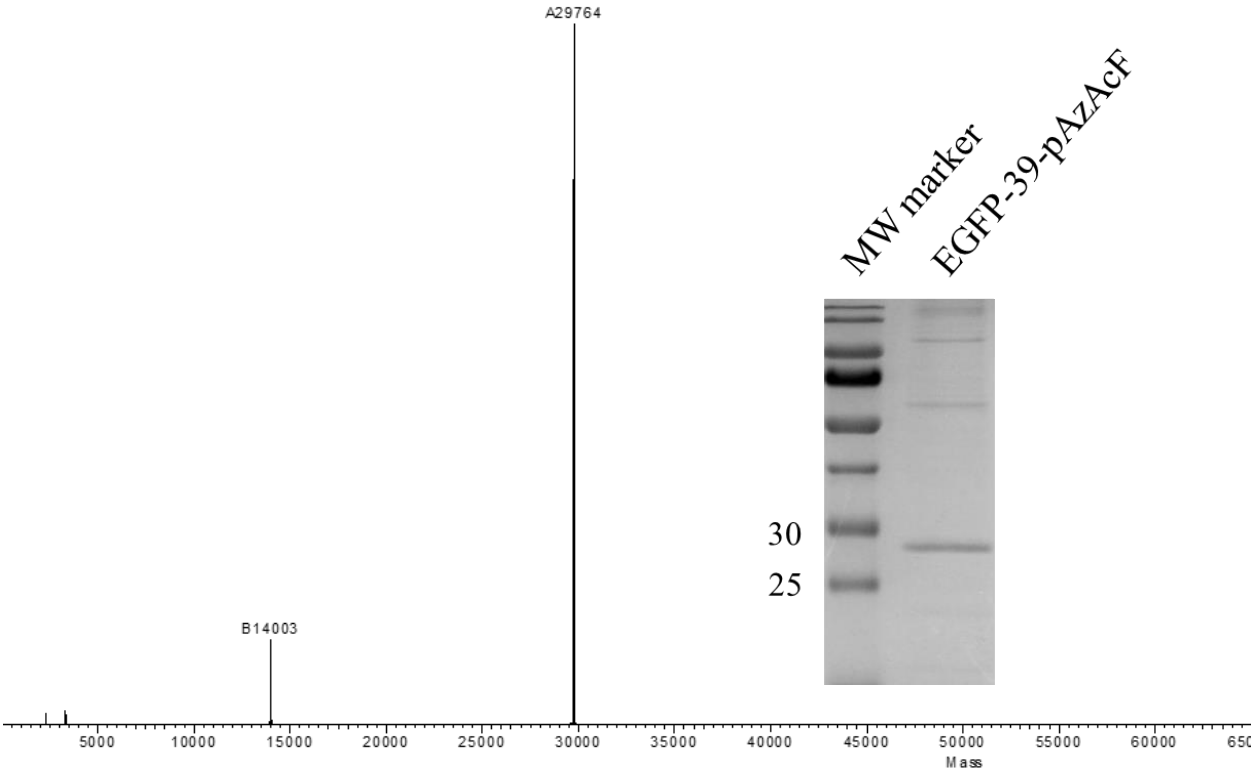


Figure 4.24. EGFP-39-pAzAcF MS analysis and SDS-PAGE.

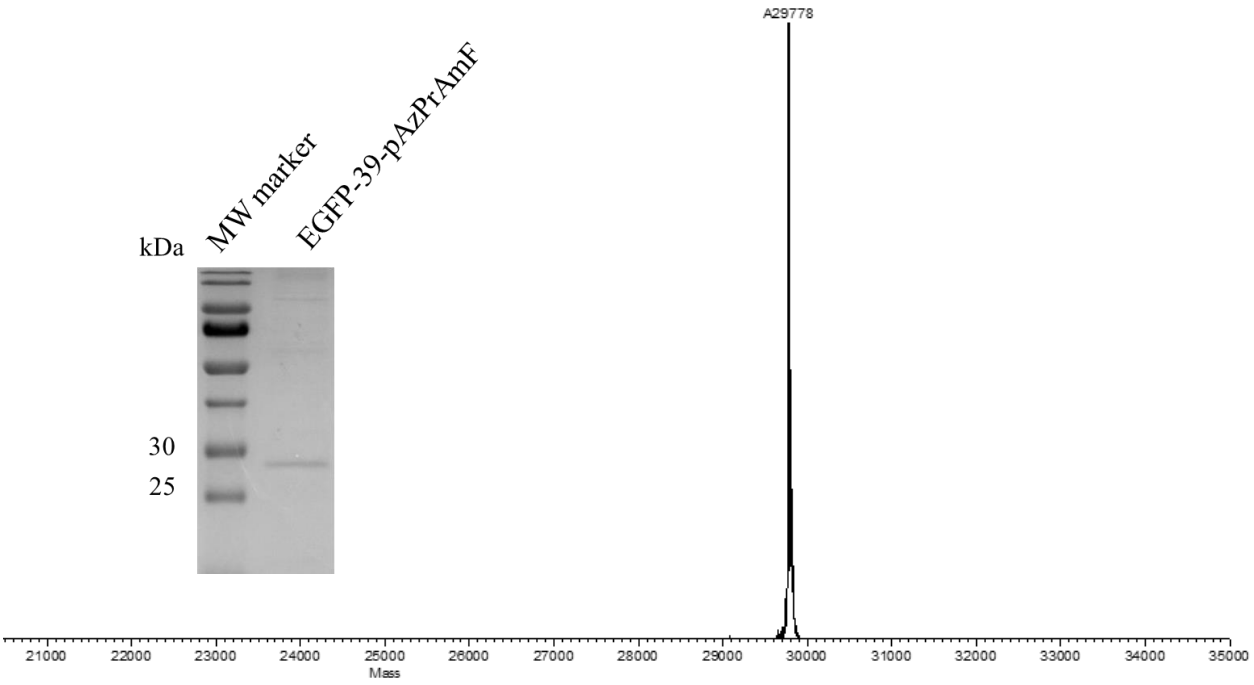


Figure 4.25. EGFP-39-pAzPrAmF MS analysis and SDS-PAGE.

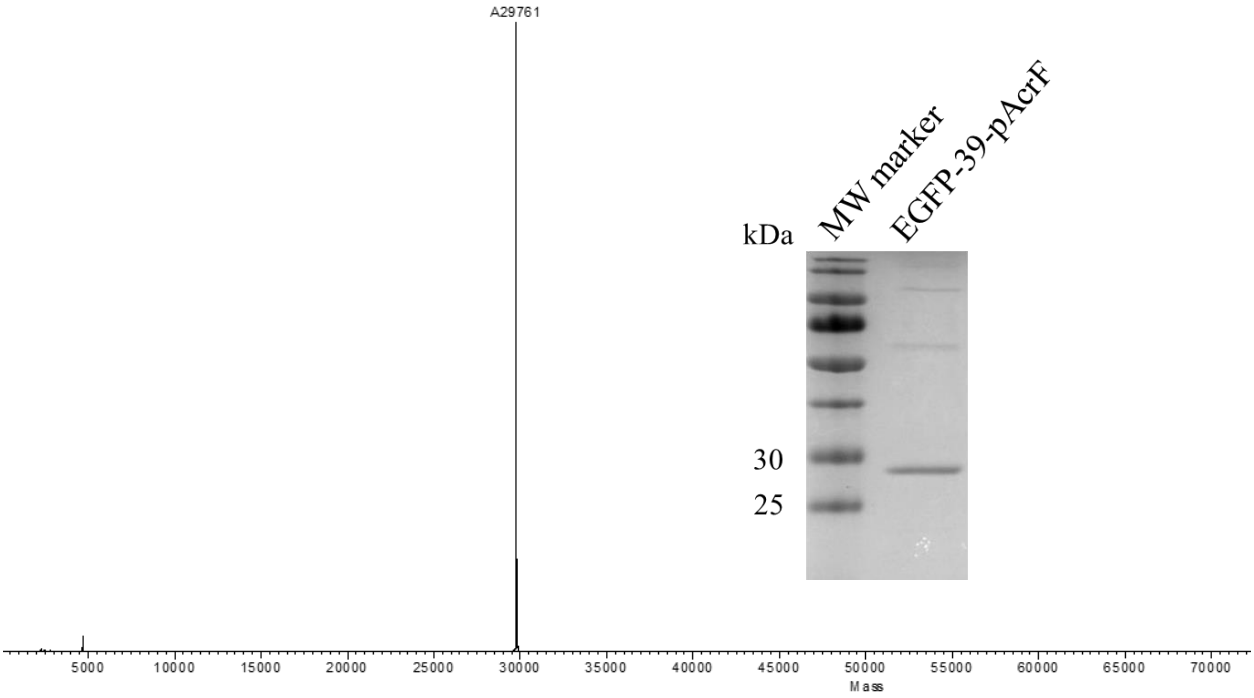


Figure 4.26. EGFP-39-pAlkAcF MS analysis and SDS-PAGE.

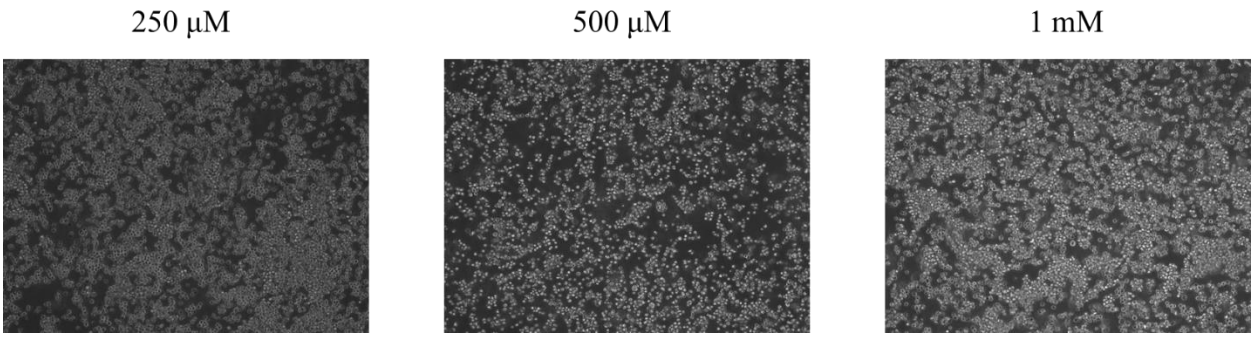


Figure 4.27. Toxicity of pCAcF in HEK293T cells. HEK293T cells were co-transfected with the EGFP fluorescent reporter and AcrFRS-9 in the presence of varying concentrations of pCAcF (top). Significant levels of cell-death were observed at every pCAcF concentration.

UAA	aaRS	Yield (mg/L)	Expected mass (Da)	Observed mass (Da)
pAEY	pAcrFRS	1.3	27665	27666, 27641
Alkene amide	pAcrFRS	6	27674	27676
Propargyl carb	pAcrFRS	7.8	27674	27676
pAcrF	pAcrFRS	16	27650	27956
pFAcF	pAcrFRS	5.3	27656	27655
pCAcF	pAcrFRS	8	27673	27943
pAzAcF	pAcrFRS	12	27680	27679, 27653
pAzPrAmF	pAcrFRS	1.4	27693	27692, 27666
pAlkAcF	pAcrFRS	6	27677	27676
sTyr	sTyr-RS-1	3.4	27678	27677
WT (Tyr)	N/A	49	27598	27597

Table 4.3. sfGFP reporter protein expression yields.

UAA	aaRS	Yield (μg/10^7 cells)	Expected mass (Da)	Observed mass (Da)
pAEY	pAcrFRS	54	29751	29751
Alkene amide	pAcrFRS	129	29760	29763
Propargyl carb	pAcrFRS	123	29760	29764
pAcrF	pAcrFRS	118	29736	29734, 30041
pFAcF	pAcrFRS	76	29740	29740
pCAcF	pAcrFRS	N/A	-	-
pAzAcF	pAcrFRS	148	29766	29764
pAzPrAmF	pAcrFRS	88	29779	29778
pAlkAcF	pAcrFRS	77	29763	29761
sTyr	sTyr-RS-1	150	29762	29761
WT (Tyr)	N/A	293	29684	29681

Table 4.4. EGFP reporter protein expression yields.

4.2.3 Engineering of sTyr-selective mutants

The first recorded observation of tyrosine *O*-sulfation was in bovine fibrinopeptide B by Bettelheim in 1954.[21] However, it wasn't until the 1980s that the scientific community began to appreciate the importance of this modification. This was due to the work of Huttner, who established that a variety of naturally occurring proteins in mammalian cells and tissues contain this PTM.[22] Huttner and Lee further identified the enzymes responsible for the installation of the sulfate group called tyrosylprotein sulfotransferases (TPSTs).[23] TPSTs form the tyrosine O^4-sulfate ester by catalyzing the transfer of a sulfate group from 3'-phosphoadenosine 5'-phosphosulfate to the hydroxyl group of peptidyl tyrosine residues (Figure 4.28).[21] These enzymes reside in the *trans*-Golgi network and the proteins they modify play important roles in inflammation, hemostasis, immunity, etc.[21, 24]

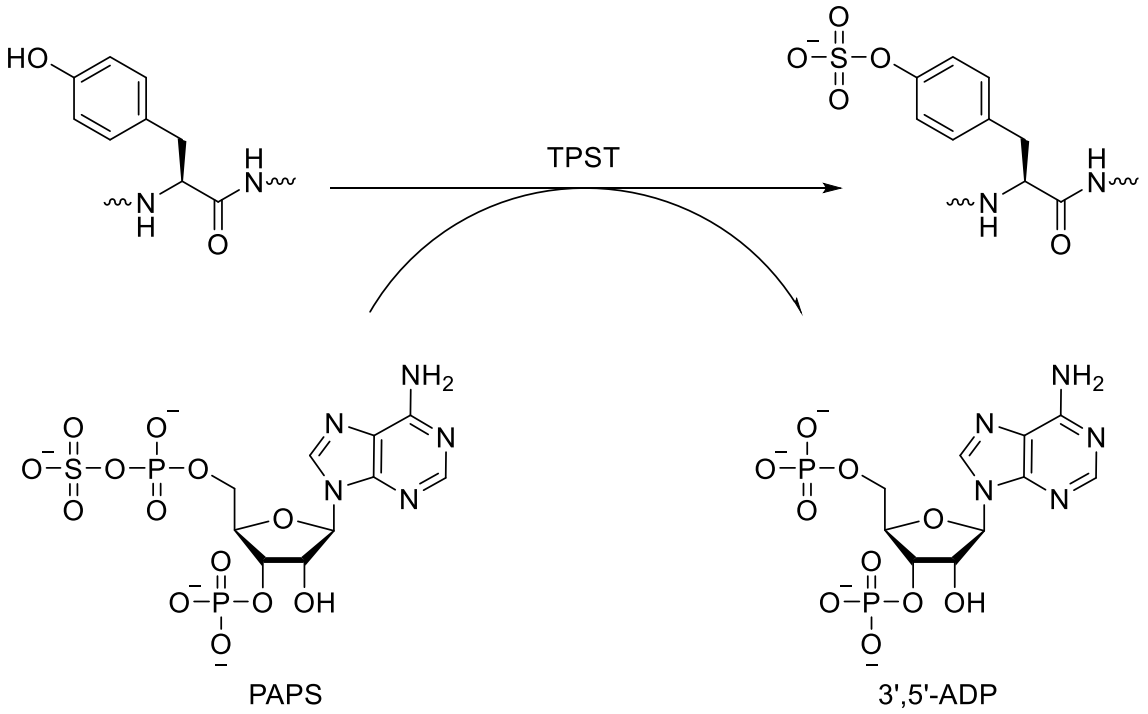


Figure 4.28. Natural sulfation of tyrosine residues by TPSTs in cells.[21]

Sulfo-tyrosine (sTyr) presented as an ideal UAA to validate the dynamic range of the ATMY5-based selection system for numerous reasons. Firstly, while tyrosine *O*—sulfation has been identified in a broad range of eukaryotic proteins (~1% of the total proteome) the role that this modification plays in the functions of these proteins is poorly understood.[25, 26] Three requirements must be met in order to identify the roles of specific tyrosine *O*-sulfation residues: (i) the modified protein must be identified and isolated at high enough concentrations to study, (ii) the individual residues that are sulfated must be characterized based on their degree of sulfation, and (iii) dynamic studies must be carried out to understand the functional roles of the sulfated residues. One method that achieves all three of these requirements is GCE. Secondly, sTyr was an ideal UAA for validating the dynamic range of the ATMY5-based selection system because our lab has previously engineered highly active mutant *Ec*TyrRSs that incorporate sTyr, enabling the study of sulfated proteins in eukaryotes (Table 4.3).[25] In summary, the physiological importance of sTyr and our prior capability of engineering *Ec*TyrRS mutants that could incorporate this UAA provided the impetus to use it as a control system for testing the ATMY5 platform.

Hit	Leu71	Asp182	Phe183	Leu186	Asp265
sTyr-VGL	Val	Gly	Tyr	Leu	Arg
sTyr-VGM	Val	Gly	Phe	Met	Arg

Table 4.5. Active site mutations of previously engineered sTyr-selective *Ec*TyrRS mutants.[25]

To engineer an *Ec*TyrRS mutant that charged sTyr the previously engineered *Ec*TyrRS-D265R library (theoretical diversity 1.06 x 10^7, > 99% coverage) (Figure 4.5) was subjected to a double-sieve selection scheme in the presence and absence of sTyr (1 mM). Following this first round of selection, enrichment of *Ec*TyrRS mutants that charged sTyr was observed (Figure 4.29). Characterization of individual library members identified four unique *Ec*TyrRS sequences that incorporated sTyr (Table 4.4). Tyr37 remained unchanged. Leu71 remained a nonpolar amino acid (Val). Asp182 changed from a large, polar amino acid to a small amino acid (Gly). Phe183 either remained the same or changed to Tyr. Leu186 remained a hydrophobic amino acid (Ile/Leu/Val/Met).

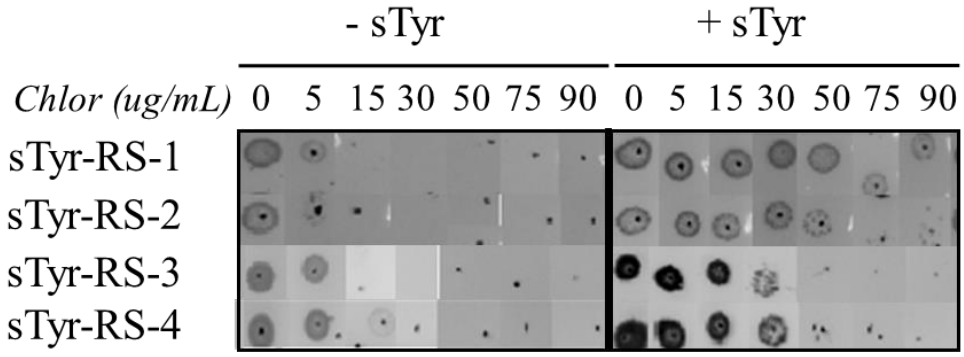


Figure 4.29. Characterization of sTyr-selective *Ec*TyrRS mutants with a CAT-TAG assay.

Hit	Leu71	Asp182	Phe183	Leu186	Asp265
sTyr-RS-1	Val	Gly	Tyr	Ile	Asp
sTyr-RS-2	Val	Gly	Tyr	Leu	Asp
sTyr-RS-3	Val	Gly	Tyr	Leu	Arg
sTyr-RS-4	Val	Gly	Phe	Val	Arg

Table 4.6. Active site-mutations of the sTyr-selective *Ec*TyrRS mutants engineered in ATMY5.

Suppression efficiencies for the four mutants were characterized in ATMY4 with an sfGFP-151-TAG reporter (Figure 4.30). Successful incorporation of sTyr into sfGFP-151-TAG was confirmed by SDS-PAGE and MS analysis (Figures 4.31). The newly identified hits demonstrated similar suppression efficiencies to the previously engineered sTyr-VGM (Figure 4.30). The engineered sTyr-selective mutants were cloned into the pB1U expression vector to characterize suppression efficiencies in HEK293T cells. Expression efficiencies of EGFP-39-sTyr was confirmed by fluorescence (Figure 4.32), SDS-PAGE, and MS analysis (Figure 4.33). Notably, the sTyr-selective mutants engineered with the ATMY5 selection platform all exhibited lower cross-reactivity in HEK293T than the previously engineered sTyr-VGM and sTyr-VGL variants (Figure 4.32).

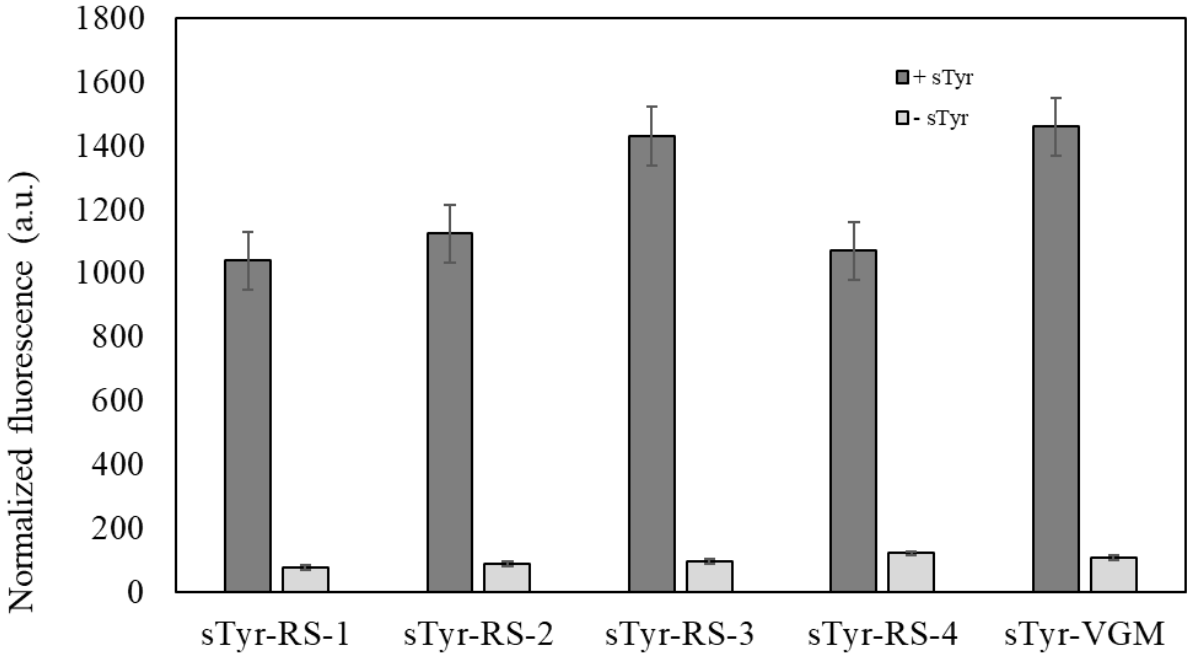


Figure 4.30. sfGFP-151-sTyr fluorescence analysis in ATMY4.

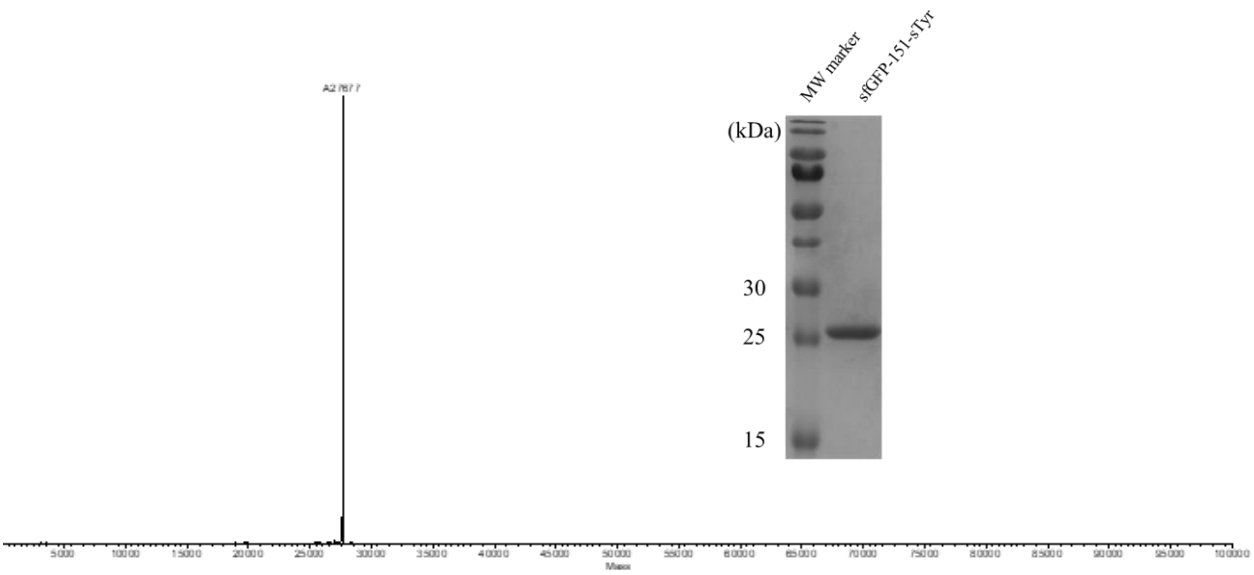


Figure 4.31. sfGFP-151-sTyr MS analysis and SDS-PAGE.

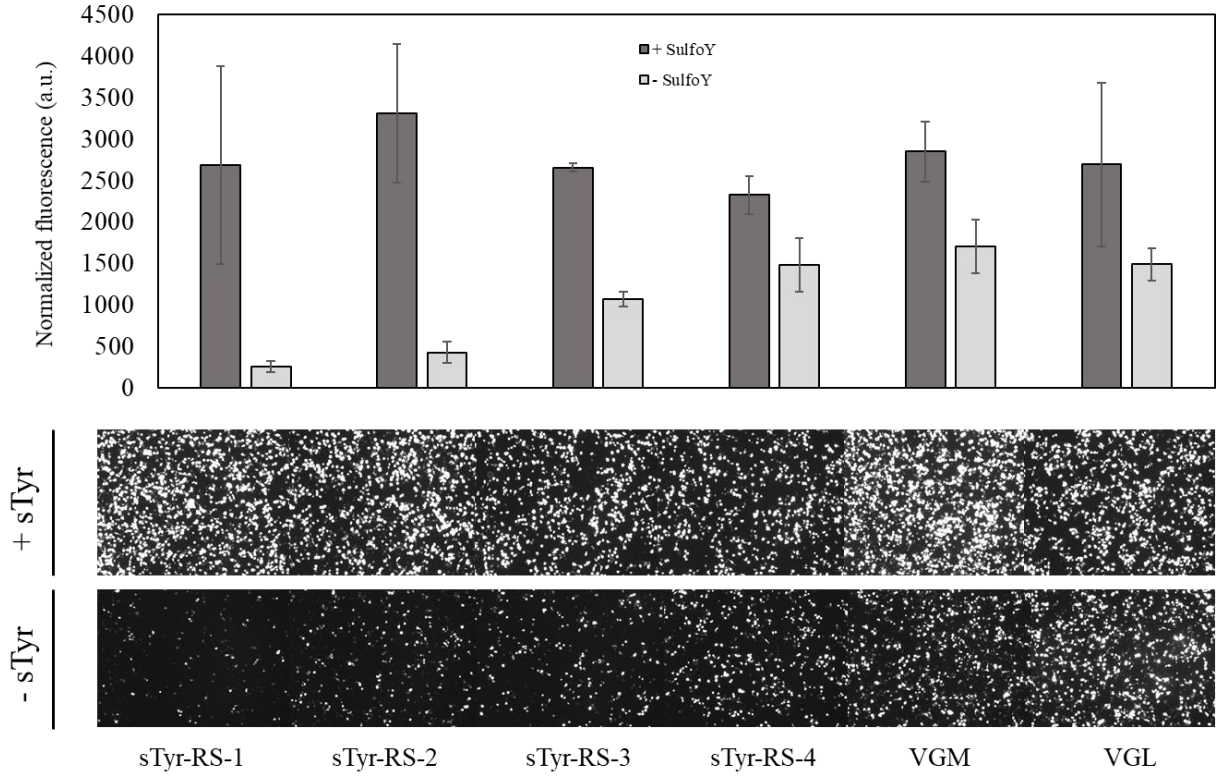


Figure 4.32. EGFP-39-sTyr fluorescence analysis in HEK293T.

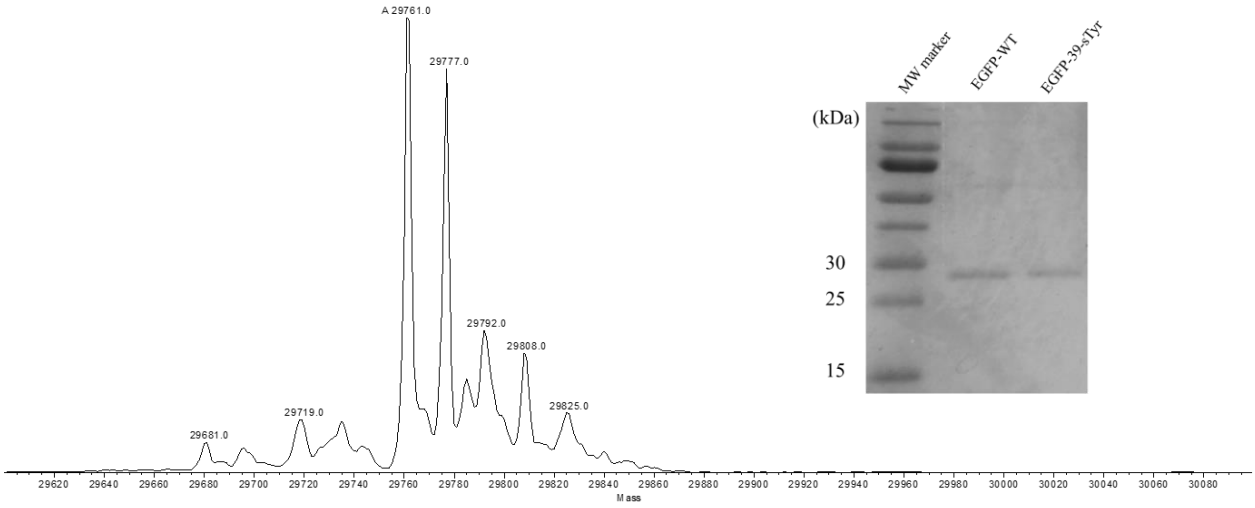


Figure 4.33. EGFP-39-sTyr MS analysis and SDS-PAGE.

4.3 Conclusions

Further engineering of the *Ec*TyrRS provides access to structurally unique UAAs in eukaryotic GCE. By adapting the ATMY5 selection platform, we created multiple novel mutant *Ec*TyrRSs that could incorporate a range of bioconjugation handles, PTM mimics, and photoaffinity probes (Table 4.5, Table 4.6). While we were able to gain access to all of this exciting new chemistry, we still could not engineer the *Ec*TyrRS for some notable UAAs (e.g., phosphotyrosine). Alternate strategies beyond solely modifying the selection platform must be considered to overcome this limitation.

4.3.1 Ongoing and future directions

The ability to incorporate azide-containing UAA derivatives with the *Ec*TyrRS has enabled the AAV project members to work towards engineering the $tRNA^{EcTyr}_{CUA}$ in mammalian cells with an established directed evolution scheme originally developed by Dr. Rachel Kelemen.[27] Additionally, other weakly active *Ec*TyrRS mutants (e.g., *Ec*Tyr-pAEYRS) can be further

engineered for improved suppression efficiencies with our two-tier antibiotic/fluorescent reporter selection scheme.

4.4 Acknowledgements

Megan Yeo, one of my former undergraduate mentees, played a large role in the engineering of the *Ec*Tyr-pAEYRS. This polyspecific hero of an aaRS wouldn't exist without her. Christen Hillenbrand, another former undergraduate mentee, helped out with innumerable day-to-day tasks like making competent cells, implementing cloning strategies, and sfGFP expression. Finally, Dr. Soumya J. Roy synthesized almost all of the UAAs mentioned in this chapter and assisted with GFP expressions, purifications, and characterizations.

4.5 Experimental procedures

The general materials used for experiments described in this chapter were the same as those outlined in chapter 2's experimental procedures.

4.5.1 Strains, cell lines

ATMY4 and ATMY5 *E. coli* cells were obtained from the Chatterjee lab glycerol stocks. This cell line was originally engineered by Dr. James Italia. A detailed description of how they were engineered can be found in his manuscript.[9]

4.5.2 Plasmids

All primer and plasmid sequences can be found in the Appendix.

The creation of the following plasmids is outlined in the experimental procedures section of chapter 2: pRepTrip2.3p-EcYtR-h1, pNeg-NoYtR, and all pB1U-aaRS plasmids.

The pEvol-sfGFP151-TAG reporter construct was propagated from Dr. James Italia's plasmid stocks by transforming it into DH10B cells, inoculating a culture, and mini-prepping the plasmid DNA.

4.5.3 Unnatural amino acids

The o-sulfo-L-tyrosine used in the present experiments was purchased from Bachem AG- (Lot no. 1071296).

para-benzoyl-l-phenylalanine was purchased from Chem-Impex International (catalog number 05110).

The following amino acids were synthesized by Dr. Soumya J. Roy: LCA, pAzAcF, pAzPrAmF, alkene amide, pAcrF, pAlkAcF, pCAcF, pFAcF, and propargyl carbamate.

4.5.4 Construction of the pBK EcTyrRS library

A detailed discussion of the *Ec*TyrRS-D265R library construction can be found in the experimental protocols of chapter 2.

4.5.5 Selection for synthetases charging UAAs

The antibiotic selection platform is described in the experimental protocols section of chapter 2.

4.5.6 Assessment of aaRS-tRNA activity using a chloramphenicol reporter

The protocol for the CAT-TAG assay is described in the experimental protocols section of chapter 2.

4.5.7 Assessment of aaRS-tRNA activity using a sfGFP-151-TAG reporter

All fluorescence reporter assays used in the present experiments mirrored those performed in the experimental protocols section of chapter 2.

4.5.8 sfGFP-151-UAA expression and purification

Protein expressions and purifications completed in the present experiments were homologous those performed in the experimental protocols section of chapter 2.

4.5.9 EGFP-39-UAA expression and fluorescence analysis

All fluorescence reporter assays used in the present experiments mirrored those performed in the experimental protocols section of chapter 2.

4.5.10 EGFP39-UAA expression and purification

Protein expressions and purifications completed in the present experiments were homologous those performed in the experimental protocols section of chapter

4.6 References

1. Dumas, A., Lercher, L., Spicer, C.D., Davis, B.G. Designing logical codon reassignment - Expanding the chemistry in biology. Chem Sci. 6(1), 50 (2015).
2. Chin, J. W. Expanding and reprogramming the genetic code. Nature 550, 53 (2017).
3. Dumas, A., Lercher, L., Spicer, C. D., Davis, B. G. Designing logical codon reassignment−Expanding the chemistry in biology. Chemical Science 6, 50 (2015).
4. Young, D. D., Schultz, P. G. Playing with the molecules of life. ACS Chem. Biol. 13, 854−870 (2018).
5. Vargas-Rodriguez, O., Sevostyanova, A., S ll, D., Crnkovic,́ A. Upgrading aminoacyl-tRNA synthetases for genetic code expansion. Curr. Opin. Chem. Biol. 46, 115 (2018).
6. Italia, J. S., Addy, P. S., Wrobel, C. J., Crawford, L. A., Lajoie, M. J., Zheng, Y., Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat. Chem. Biol. 13, 446 (2017).
7. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., Schultz, P. G. An expanded eukaryotic genetic code. Science 301, 964−967 (2003).
8. Italia, J. S., Zheng, Y., Kelemen, R. E., Erickson, S. B., Addy, P. S., and Chatterjee, A. Expanding the genetic code of mammalian cells. Biochem. Soc. Trans. 45, 555 (2017).
9. Italia, J. S., Latour, C., Wrobel, C. J., and Chatterjee, A. Resurrecting the bacterial tyrosyl-tRNA Synthetase/tRNA pair for expanding the genetic code of both E. coli and eukaryotes. Cell chemical biology 25, 1304 (2018).
10. (a) Kolb, H.C., Finn, M.G., Sharpless, K.B. Click Chemistry: diverse chemical function from a few good reactions. Angew. Chem. Int. Ed. Engl. 40, 2004 (2001). (b) Sletten, E.M., Bertozzi, C.R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. Angew. Chem. Int. Ed. 48, 38, 6974 (2009).
11. L'abbe, G. Decomposition and addition reactions of organic azides. Chem. Rev. 69 (3), 345 (1969).
12. Supekova, L., Zambaldo, C., Choi, S., Lim, R., Luo, X., Kazane, S.A., Young, T.S., Schultz, P.G. The genetic incorporation of *p*-azidomethyl-L-phenylalanine into proteins in yeast. Bioorg. Med. Chem. 28, 9, 1570 (2018).
13. Worrell, B.T., Malik, J.A., Fokin, V.V. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. Science 340 (6131), 457 (2013).
14. Jewett, J.C., Bertozzi, C.R. Cu-free click cycloaddition reactions in chemical biology. Chem Soc Rev. 39 (4), 1272 (2010).
15. Cravatt, B.F., Wright, A.T., Kozarich, J.W. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. Annu. Rev. Biochem. 77, 383 (2008)
16. Wang, S., Tian, Y., Wang, M., Wang, M., Sun, G., Sun, X. Advanced Activity-Based Protein Profiling Application Strategies for Drug Develpoment. Front. Pharmacol. 9, 353 (2018).
17. Hofer, T., Skeffington, L.R., Chapman, C.M., Rader, C. Molecularly defined antibody conjugation through a selenocysteine interface. Biochemistry 48 (50), 12047 (2009).
18. Miller, R.M., Paavilainen, V.O., Krishnan, S., Serafimova, I.M., Taunton, J. Electrophilic Fragment-Based Design of Reversible Covalent Kinase Inhibitors. J. Am. Chem. Soc. 135 (14), 5298 (2013).

19. Agarwal, P., Bertozzi, C.R. Site-Specific Antibody-Drug Conjugates: The Nexus of Bioorthogonal Chemistry, Protein Engineering, and Drug Development. Bioconjugate Chem. 26(2), 176 (2015).
20. Furman, J.L., Kang, M., Choi, S., Cao, Y., Wold, E.D., Sun, S.B., Smider, V.V., Schultz, P.G., Kim, C.H. A Genetically Encoded aza-Michael Acceptor for Covalent Cross-Linking of Protein-Receptor Complexes. J. Am. Chem. Soc. 136 (23), 8411 (2014).
21. Moore, K.L. The Biology and Enzymology of Protein Tyrosine O-Sulfation. J. Biol. Chem. 278 (27), 24243 (2003).
22. Huttner, W.B. Sulphation of tyrosine resides—a widespread modification of proteins. Nature 299, 273 (1982).
23. Lee, R.W., Huttner, W.B. Tyrosine-O-sulfated proteins of PC12 pheochromocytoma cells and their sulfation by a tyrosylprotein sulfotransferase. J. Biol. Chem. 258 (18), 11326 (1983).
24. Moore, K.L. Protein tyrosine sulfation: A critical posttranslational modification in plants and animals. Proc. Natl. Acad. Sci. U S A 106 (35), 14741 (2009).
25. Italia, J.S., Peeler, J.C., Hillenbrand, C.M., Latour, C., Weerapana, E., Chatterjee, A. Genetically encoded protein sulfation in mammalian cells. Nat Chem Biol. 16 (4), 379 (2020).
26. Stone, M.J., Chuang, S., Hou, X., Shoham, M., Zhu, J.Z. Tyrosine sulfation: an increasingly recognized post-translational modification of secreted proteins. New Biotechnology 25 (5), 299 (2009).
27. Kelemen, R.E., Jewel, D., Pasha, M., Cao, X., Chatterjee, A. Virus-assisted directed evolution of suppressor tRNAs in mammalian cells. (*Submitted*).

Chapter 5

Stability of the aminoacyl-tRNA synthetase affects its engineerability

A significant portion of the work described in this chapter has been published in:

Grasso, K.T., Yeo, M.J.R., Hillenbrand, C.M., Ficaretta, E.D., Italia, J.S., Huang, R.L., Chatterjee, A., Structural robustness affects the engineerability of aminoacyl-tRNA synthetases for genetic code expansion. *Biochemistry* 60 (7), 489 (**2021**).

5.1 Introduction

5.1.1 Brief synopsis of TyrRS/tRNA engineering for eukaryotic GCE applications

Selection of the *Ec*TyrRS/tRNA$^{Ec\text{Tyr}}$ pair for eukaryotic GCE applications was predominantly based on our extensive knowledge of this synthetase's structure and previous work demonstrating orthogonality of the tRNA$^{Ec\text{Tyr}}$ in *S. cerevisiae*.[1] Further investigation into the natural orthogonality of bacterial TyrRSs in eukaryotic organisms attributed a few key recognition elements in the tRNA acceptor stem to this phenomenon.[2-4] While the *Ec*TyrRS/tRNA$^{Ec\text{Tyr}}$ pair has been successfully engineered for eukaryotic GCE applications, the structural diversity of UAAs it can incorporate are lacking.[5-7] Initially, we attributed this to limitations of the *S. cerevisiae* platform that was historically used for engineering the *Ec*TyrRS.[8] However, we overcame this limitation through the creation of ATMY *E. coli* strains[9] and still could not identify *Ec*TyrRS variants for a host of UAAs (e.g., phosphotyrosine[10]). This led us to believe that alternative limitations remained undiscovered.

5.1.2 Effects of stability on the directed evolution of synthetases

For bacterial GCE, the *Mj*TyrRS/tRNA$^{Mj\text{Tyr}}$ pair has been used for the incorporation of tyrosyl-derived UAAs.[5-7] This pair has successfully encoded a broad range of useful UAAs in *E. coli* including those containing bioconjugation handles, photoaffinity probes, biophysical probes, models for natural post-translational modifications, etc.[5-7] The main justifications for choosing the *Mj*TyrRS/tRNA$^{Mj\text{Tyr}}$ pair were that its tRNA had different identity elements than the *E. coli* tRNA, the *Mj*TyrRS had a minimalist anticodon loop binding domain, and that the *Mj*TyrRS did not have an editing mechanism.[11] However, what if there was a fourth benefit to choosing this scaffold that had gone largely unnoticed by the GCE community?

In the field of directed evolution, it is a well-established paradigm that when a protein's structure is altered to obtain a new function, the stability of the resulting mutants will often be compromised.[12-15] Consequently, how extensively a protein can be engineered is often limited by how stable it is. We hypothesized that the *Mj*TyrRS/ $tRNA^{MjTyr}$ pair could be more successful due to its hidden benefit of being derived from a thermophilic archaeon. Proteins from thermophilic organisms are generally more stable up to higher temperatures which could be facilitating the structural stability of the *Mj*TyrRS variants. Contrastingly, the *Ec*TyrRS/$tRNA^{EcTyr}$ pair is from a mesophilic bacterium, which could be causing it to be a less stable scaffold, resulting in fewer viable, stable variants following directed evolution.

5.2 Results and discussion

5.2.1 Engineered TyrRS mutants exhibit lower activity and stability than the parental TyrRS

The newly optimized ATMY5 *E. coli* selection platform conferred the ability to engineer the *Ec*TyrRS/$tRNA^{EcTyr}_{CUA}$ pair in a more amenable host organism than yeast. Originally, we thought overcoming this singular, platform-based limitation would solve the issue of identifying all glaring, missing UAA-selective *Ec*TyrRSs. Unfortunately, we were sorely mistaken. While we now had the ability to rapidly engineer the *Ec*TyrRS/$tRNA^{EcTyr}_{CUA}$ pair for eukaryotic applications, there were still some notable UAAs missing and for some UAAs only weakly active mutants were accessible. For example, *Ec*TyrRS mutants engineered with our selection system for the incorporation of the photoaffinity probe *p*-benzoyl-L-phenylalanine (pBPA) exhibited significantly lower suppression efficiencies relative to the parental TyrRS (Figure 5.1). This observation was not merely the result of our selection system. In fact, one of the *Ec*Tyr-pBPARS

mutants had the same active site structure as the *Ec*Tyr-pBPARS previously engineered in yeast[8] (Figure 5.2). This observation indicated that the observed set of mutations optimally reconfigure the *Ec*TyrRS active site for charging pBPA.

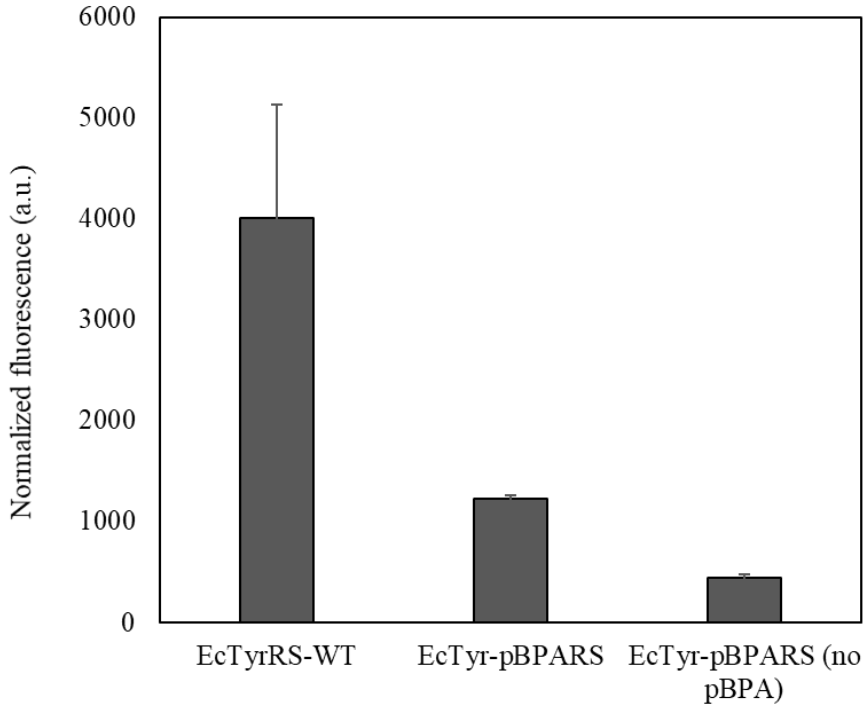
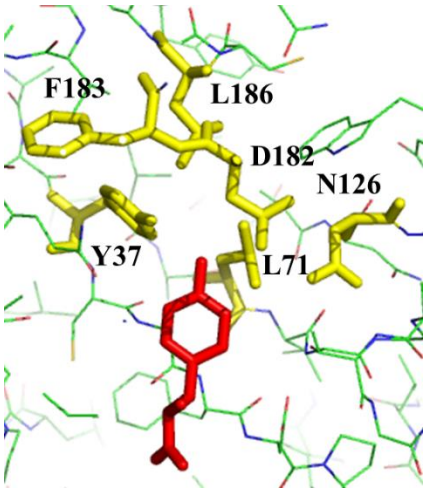


Figure 5.1. The activity of a representative *Ec*Tyr-pBPARS is significantly weaker than *Ec*TyrRS. Activity was evaluated using the sfGFP-151-TAG reporter, expressed in ATMY6 *E. coli*, by measuring the characteristic fluorescence of the full-length reporter resuspended in cells. For *Ec*Tyr-pBPARS, the expression was measured in the presence or absence of pBPA (1 mM).



Clones	Tyr37	Leu71	Asn126	Asp182	Phe183	Leu186
pBPARS-1*	Gly	Leu	Asn	Gly	Phe	Ala
pBPARS-7	Gly	Val	Asn	Ser	Tyr	Ala
pBPARS-8	Ala	Val	Asn	Ala	Phe	Tyr

Figure 5.2. *Ec*Tyr-pBPARS mutant active site mutations.

By developing a dual-sieve antibiotic and fluorescent reporter-based selection system we showed that we could improve the weakly active mutant *Ec*TyrRSs to functional suppression activity levels. However, it was unclear how the randomly distributed point mutations in the highly active mutants were improving the activity of the *Ec*Tyr-pBPARS. The mutations were randomly scattered throughout the aaRS structure (Figure 5.3) and no individual mutation seemed to cause an improvement in catalytic activity. In fact, the individual removal of each point mutation led to a drop in the overall suppression efficiency (Chapter 3, Figure 3.14). A systematic characterization was carried out and we identified that the *Ec*Tyr-pBPARS first-generation (Gen 1) mutant was largely insoluble in *E. coli*, whereas the *Ec*Tyr-pBPARS third-generation (Gen 3.1) mutant was largely soluble, potentially explaining the differences that we observed in activity. This was determined through western blot analysis of the soluble and insoluble fractions of cell-free extracts

of *E. coli* expressing polyhistidine-tagged *Ec*TyrRS-pBPA-Gen 1 and *Ec*TyrRS-pBPA-Gen 3.1, revealing that the former was nearly exclusively found in the insoluble fraction (Figure 5.4).

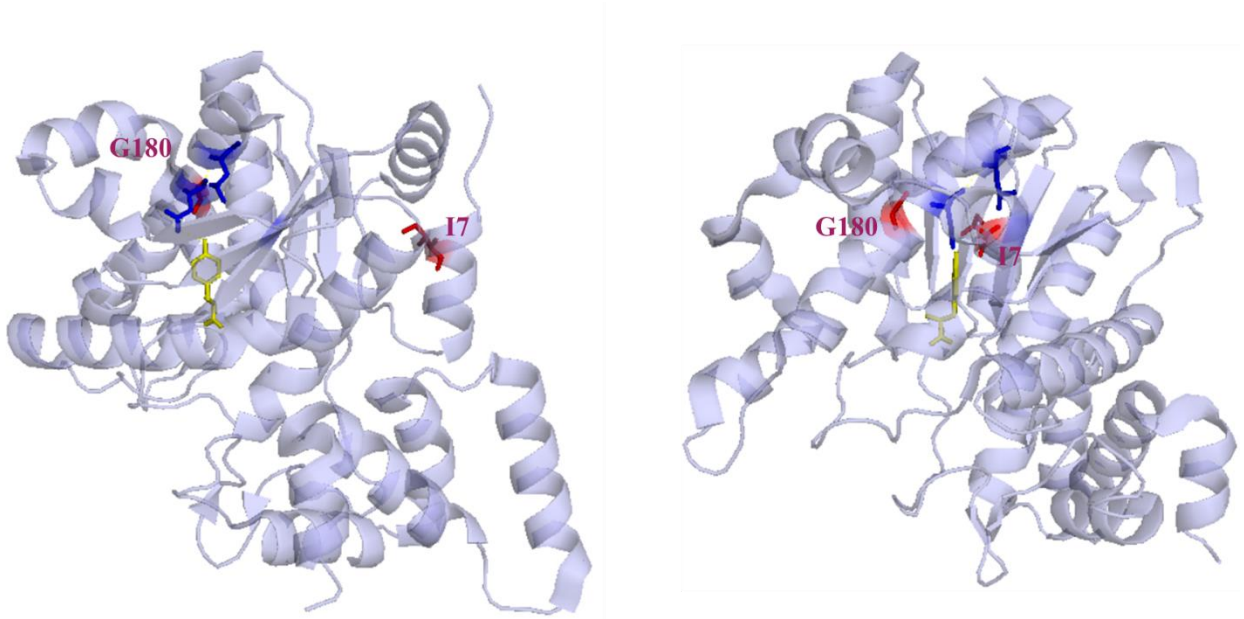


Figure 5.3. Point mutations identified in highly active *Ec*Tyr-pBPARS mutants.

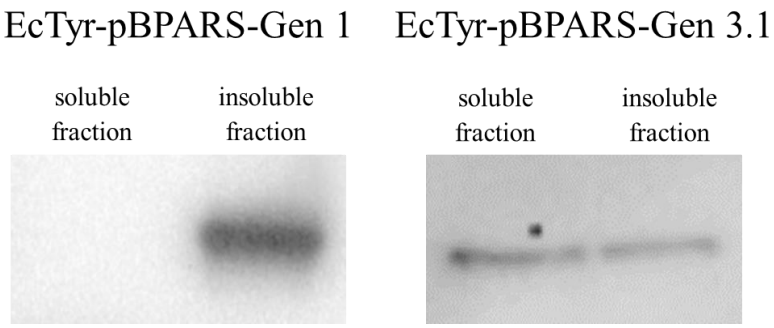


Figure 5.4. Western blot analysis of soluble and insoluble fractions of *E. coli* cell-free extracts expressing *Ec*TyrRS-pBPA-Gen 1 and *Ec*TyrRS-pBPA-Gen 3.1 reveals that the former is largely insoluble.

We observed that the *Ec*Tyr-pBPARS-Gen 3 exhibited improved solubility relative to its evolutionary starting point (Figure 5.4) and that none of its point mutations individually conferred enhanced catalytic activity (Chapter 3, Figure 3.13). It is also common knowledge that altered

protein structures can become unstable.[12-15] Based on this information, we hypothesized that these mutations enhanced the structural stability of *Ec*Tyr-pBPARS-Gen 3. To test this notion, we took advantage of a modified cellular thermal shift assay (CETSA).[16, 17] In this assay, a cell-free extract expressing a target protein is heated and the amount of protein remaining in the soluble fraction is subsequently tested by immunoblotting. The temperature range at which a protein is lost from the soluble fraction provides an estimate of its thermostability. An N-terminal hexa-histidine tag was appended to the proteins to facilitate their detection in a dot-blot assay using an anti-polyhistidine antibody. As anticipated, *Ec*Tyr-pBPARS-Gen 3 was more thermostable than *Ec*Tyr-pBPARS-Gen 1 (Figure 5.5). The Gen 3 mutant maintained solubility up to 50 ºC whereas the Gen 1 mutant was largely insoluble at room temperature (Figure 5.5). This supported our hypothesis that the Gen 3 mutant was more active partially due to enhanced stability/solubility characteristics.

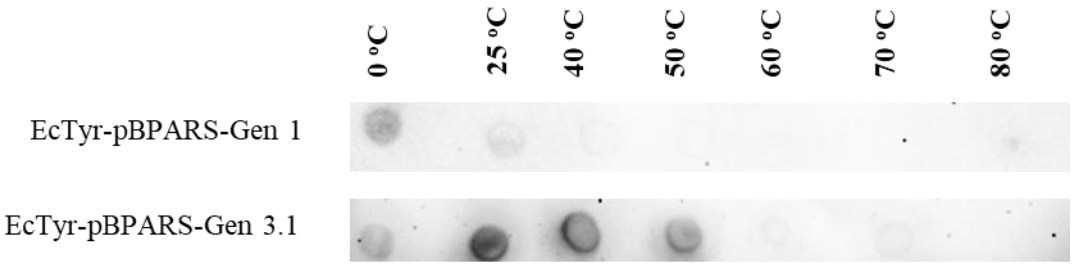
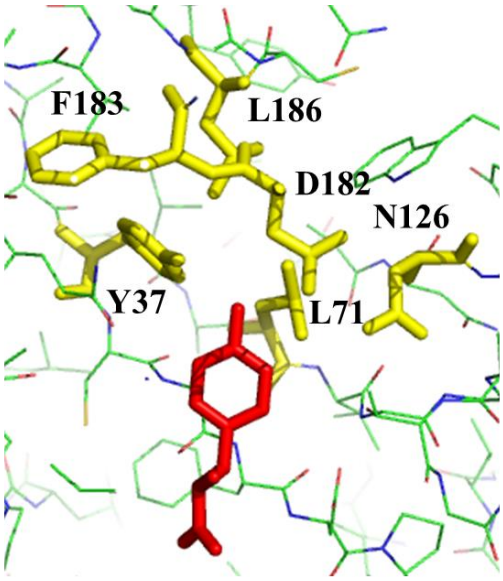


Figure 5.5. Thermal shifty assay of pBPA-selective *Ec*TyrRS mutants. Dot blot was performed (using an anti-polyhistidine antibody) on the soluble fraction of *E. coli* cell-free extracts expressing the indicated TyrRS variants (left), after incubation at the indicated temperature (top).

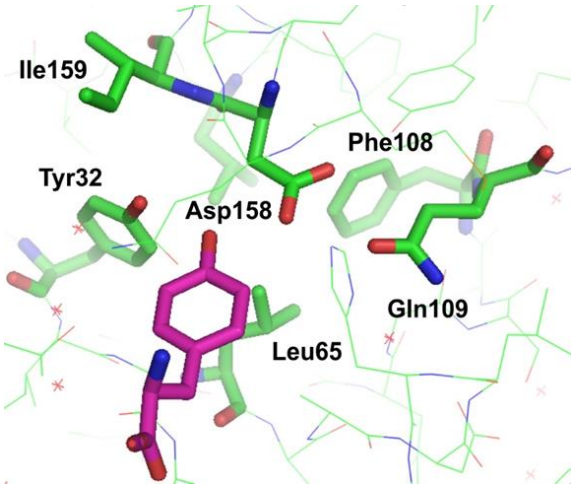
The observed differences between the *Ec*Tyr-pBPARS-Gen 3 and *Ec*Tyr-pBPARS-Gen 1 caused us to reevaluate the stability of the other *Ec*TyrRS mutants we have engineered. Additionally, we hypothesized that a more thermostable synthetase scaffold would result in more stable mutants. The *Mj*TyrRS presented as an ideal candidate for comparison to the *Ec*TyrRS due to its homologous structure and extensive history of being engineered for GCE applications.[5-7] We

chose to evaluate the *Ec*TyrRS, *Ec*Tyr-pBPARS, and a polyspecific *Ec*TyrRS mutant (*Ec*Tyr-PolyRS), which exhibits high activity (Figure 5.6). For comparison, we included *Mj*TyrRS, as well as two of its comparable engineered mutants: one selective for pBPA[18] (*Mj*Tyr-pBPARS) and another that exhibits UAA polyspecificity[19] (*Mj*Tyr-PolyRS) (Figure 5.7). All of these proteins contained an N-terminal hexa-histidine tag to facilitate their detection in a dot-blot assay using an anti-polyhistidine antibody. As anticipated, *Mj*TyrRS was found to be highly thermostable, maintaining solubility up to 80 ºC (Figure 5.8). In contrast, *Ec*TyrRS was much less stable and was lost from the soluble fraction between 50 and 60 ºC (Figure 5.8). These values are consistent with previously reported thermostability measurements.[20] All of the engineered mutants exhibited reduced stability relative to their wild-type counterparts. The *Mj*Tyr-pBPARS mutant was slightly less stable than its polyspecific counterpart, but both were soluble at physiological temperature. In contrast, for *Ec*TyrRS, only the polyspecific mutant was soluble at physiological temperature; the pBPA-selective mutant was not detected in the soluble fraction even at the lowest temperature tested (Figure 5.8). These observations support the hypothesis that the lower stability of the *Ec*TyrRS negatively impacts its engineerability. While some of its active site mutants, such as *Ec*Tyr-PolyRS, are adequately stable and active under physiological conditions, more destabilizing mutants such as *Ec*Tyr-pBPARS are not viable.



	Tyr37	Asp182	Phe183	Leu186
EcTyr-PolyRS	Val	Ser	Met	Ala

Figure 5.6. Active site mutations of *Ec*Tyr-PolyRS.



	Tyr32	Leu65	Glu107	Phe108	Gln109	Asp158	Ile159
MjTyr-pBPARS	Gly	Leu	Pro	Phe	Gln	Thr	Ser
MjTyr-PolyRS	Leu	Val	Glu	Trp	Met	Gly	Ala

Figure 5.7. Active site mutations of *Mj*TyrRS mutants.

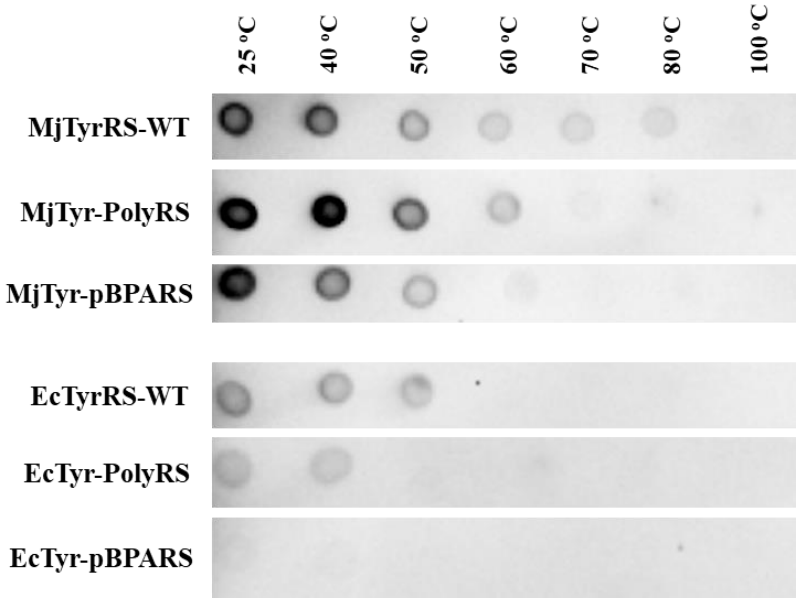


Figure 5.8. Thermal shift assay of various TyrRS variants. Dot blot was performed (using an anti-polyhistidine antibody) on the soluble fraction of *E. coli* cell-free extracts expressing the indicated TyrRS variants (left), after incubation at the indicated temperature (top).

5.2.2 Rationally designing chimeric TyrRS scaffolds with improved thermostability

To overcome the challenge of engineering stable UAA-selective *Ec*TyrRS variants, we sought to identify a better parental scaffold. One potential solution was the tyrosyl-tRNA synthetase from the thermophilic bacterium *Geobacillus stearothermophilus* (*Gs*TyrRS). Three main reasons distinguished the *Gs*TyrRS as an attractive candidate for engineering UAA-selective mutants: (i) the *Gs*TyrRS has been extensively characterized, (ii) the *Gs*TyrRS is more thermostable than the *Ec*TyrRS, and (iii) the *Gs*TyrRS is homologous to *Ec*TyrRS (Figure 5.9).[21-23] To evaluate its candidacy, we subjected the *Gs*TyrRS to both the CETSA and sfGFP-151-TAG assays. The CETSA assay reinforced that the *Gs*TyrRS was more thermostable than the *Ec*TyrRS (Figure 5.10). However, the *Gs*TyrRS displayed lower activity than the *Ec*TyrRS (Figure 5.11). It was not surprising that the *Gs*TyrRS was less active than *Ec*TyrRS because enzymes from thermophilic bacteria often exhibit weaker activity at lower temperatures.[24] We hypothesized that

we could marry the two TyrRSs into a chimeric structure, ultimately obtaining a balance between the stability of the *Gs*TyrRS and the activity of the *Ec*TyrRS. This hope was not unfounded, as previous work has demonstrated the utility of chimeric enzyme scaffolds for protein engineering.[25, 26] Guided by the work of Guez-Ivanier and Bedouelle[27], we engineered multiple chimeric scaffolds of the *Gs*TyrRS combined with the *Ec*TyrRS. Two of these scaffolds, termed chimera 2 tyrosyl-tRNA synthetase (Ch2TyrRS) and chimera 6 tyrosyl-tRNA synthetase (Ch6TyrRS) (Figure 5.12), demonstrated higher activity relative to their parental TyrRSs in ATMY *E. coli* (Figure 5.11). To assess whether these chimeric scaffolds were amenable to UAA incorporation we engineered in the pBPA-selective active site modifications and subsequently characterized them through the CETSA and sfGFP-151-TAG assays. Excitingly, both the Ch2 and Ch6 pBPA selective TyrRSs exhibited increased stability in the CETSA assay relative to both of the parental TyrRSs (Figure 5.13) in addition to increased activity (Figure 5.14). The only anomaly in this CETSA assay was that the pBPA-selective *Gs*TyrRS mutant was largely insoluble when expressed in *E. coli*. We are unsure why this occurred; however, it suggests that the chimeric TyrRSs are better suited for engineering new UAA-selective variants.

```
EcTyrRS-WT     1 MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALYCGFDPTADSLHLGHLVPLLCLKR
GsTyrRS-WT     1 M---DLLAELQWRGLVNQTTDEDGLRKLLNEERVTLYCGFDPTADSLHIGHLATILTMRR

EcTyrRS-WT    61 FQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWVDKIRKQVAPFLDFDCGENS
GsTyrRS-WT    58 FQQAGHRPIALVGGATGLIGDPSGKKSERTLNAKETVEAWSARIKEQLGRFLDFEADGNP

EcTyrRS-WT   121 AIAANNYDWFGNMNVLTFLRDIGKHFSVNQMINKEAVKQRLNREDQGISFTEFSYNLLQG
GsTyrRS-WT   118 AKIKNNYDWIGPLDVITFLRDVGKHFSVNYMMAKESVQSRI---ETGISFTEFSYMMLQA

EcTyrRS-WT   181 YDFACLNKQYGVVLQIGGSDQWGNITSGIDLTRRLH-QNQVFGLTVPLITKADGTKFGKT
GsTyrRS-WT   175 YDFLRLYETEGCRLQIGGSDQWGNITAGLELIRKTKGEARAFGLTIPLVTKADGTKFGKT

EcTyrRS-WT   240 EGGAVWLDPKKTSPYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRA
GsTyrRS-WT   235 ESGTIWLDKEKTSPYEFYQFWINTDDRDVIRYLKYFTFLSKEEIEALEQELREAPEKRAA

EcTyrRS-WT   300 QYVLAEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEMEKGADL
GsTyrRS-WT   295 QKTLAEEVTKLVHGEEALRQAIRISEALFSGDIANLTAAEIEQGFKDVPSFVHEGGDVPL

EcTyrRS-WT   360 MQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLFGRFTLLRRGKKNYC
GsTyrRS-WT   355 VELLVSAGISPSKRQAREDIQNGAIYVNGERLQDVGAILTAEHRLEGRFTVIRRGKKKYY

EcTyrRS-WT   420 LICWK
GsTyrRS-WT   415 LIRYA
```

Figure 5.9. Sequence alignment of the *Ec*TyrRS and *Gs*TyrRS.

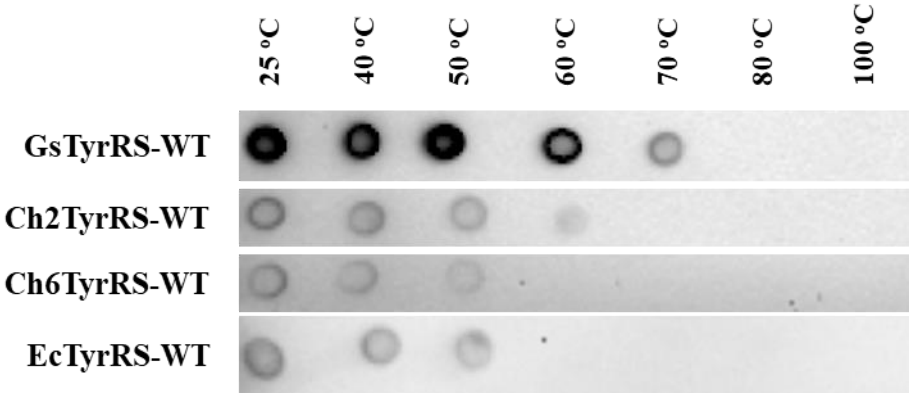


Figure 5.10. Thermal shift assay of the TyrRSs in *E. coli*. Dot blot was performed with an anti-polyhistidine antibody on the soluble fraction of *E. coli* cell-free extracts expressing the indicated TyrRS variants (left), after incubation at the indicated temperature (top).

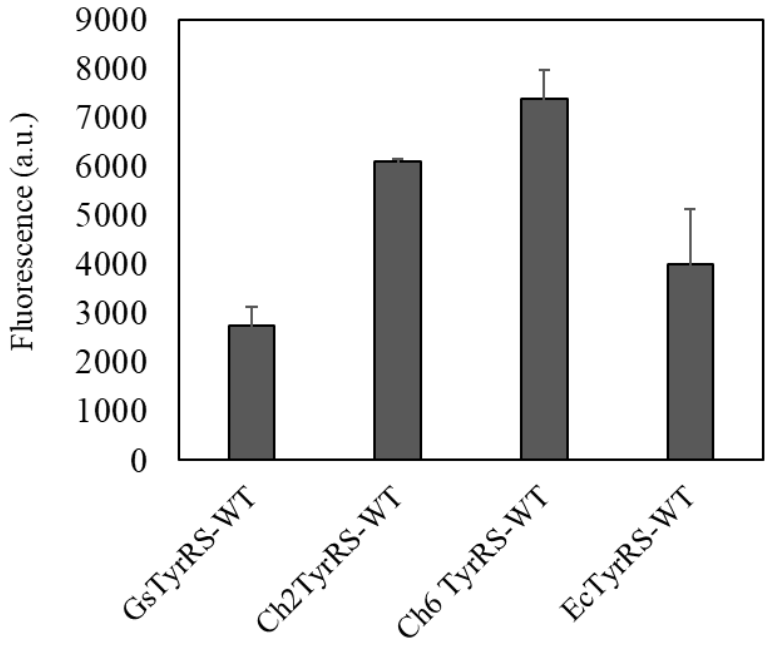


Figure 5.11. sfGFP-151-TAG assay of the TyrRSs in *E. coli*. The TyrRSs were expressed in *E. coli* in the presence of the suppressor tRNA, $tRNA^{EcTyr}_{CUA}$, and subsequent fluorescence of full-length sfGFP fluorescence was measured in resuspended cells.

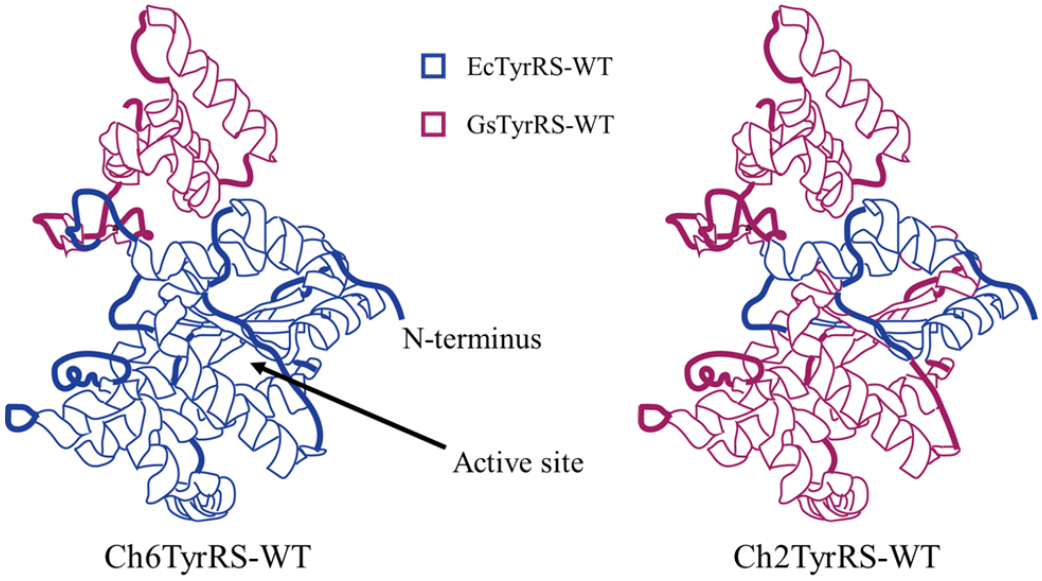


Figure 5.12. Bacterial TyrRS chimeras created from GsTyrRS and EcTyrRS. The *Ec*TyrRS crystal structure was used to highlight the progenitor sequences in the two chimeras.

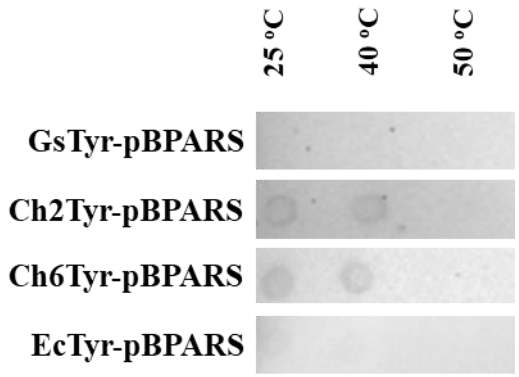


Figure 5.13. Thermal shift assay of the pBPA-selective TyrRSs in *E. coli*. Dot blot was performed with an anti-polyhistidine antibody on the soluble fraction of *E. coli* cell-free extracts expressing the indicated pBPA-selective TyrRS variants (left), after incubation at the indicated temperature (top).

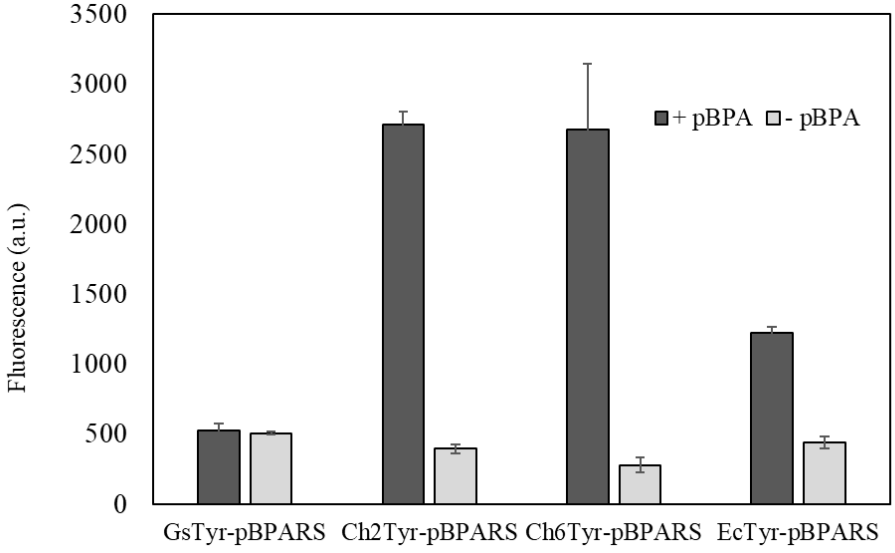


Figure 5.14. sfGFP-151-TAG assay of the pBPA-selective TyrRSs in *E. coli*. The pBPA-selective TyrRSs were expressed in *E. coli* in the presence of the suppressor tRNA, $tRNA^{EcTyr}_{CUA}$, and subsequent fluorescence of full-length sfGFP fluorescence was measured in resuspended cells.

Ultimately, our goal is to use these bacterial TyrRSs for eukaryotic GCE applications. To demonstrate the utility of these new chimeras for eukaryotic applications, we tested out their activities in HEK293T cells. First, we cloned the pBPA-selective *Ec*TyrRS, *Gs*TyrRS, Ch2TyrRS,

and Ch6TyrRS into a mammalian expression vector that has the aaRS under a UbiC promoter and contains 16 copies of the suppressor tRNA expression cassette. These plasmids were co-transfected with a pAcBac1 reporter plasmid containing an EGFP with an amber stop codon at position 39 (EGFP-39-TAG). Full-length expression of the EGFP could by monitored by fluorescence, in the presence and absence of 1 mM pBPA, and was used to evaluate the activity of the different TyrRS variants (Figure 5.15). Notably, both chimeras displayed much higher activity levels than their parental counterparts (Figure 5.16). To validate that the chimeras were selectively incorporating pBPA into the reporter protein, EGFP-39-pBPA was subsequently isolated using Ni-NTA chromatography and characterized by SDS-PAGE and MS analysis (Figure 5.17). One notable observation from the HEK293T expression data was that the pBPA-selective *Gs*TyrRS was highly active (Figure 5.16), contradicting the *E. coli* expression data (Figure 5.14). This observed difference between host organism expression could be attributed to the fact that eukaryotes have more sophisticated protein folding machinery, allowing them to better process unstable engineered proteins like pBPA-selective *Gs*TyrRS. This hypothesis was reinforced by a CETSA expression assay in HEK293T where it was observed that the *Gs*Tyr-pBPARS was expressed at higher levels than both chimeras, however, both chimeras displayed higher activities, underscoring their higher intrinsic activity (Figure 5.18).

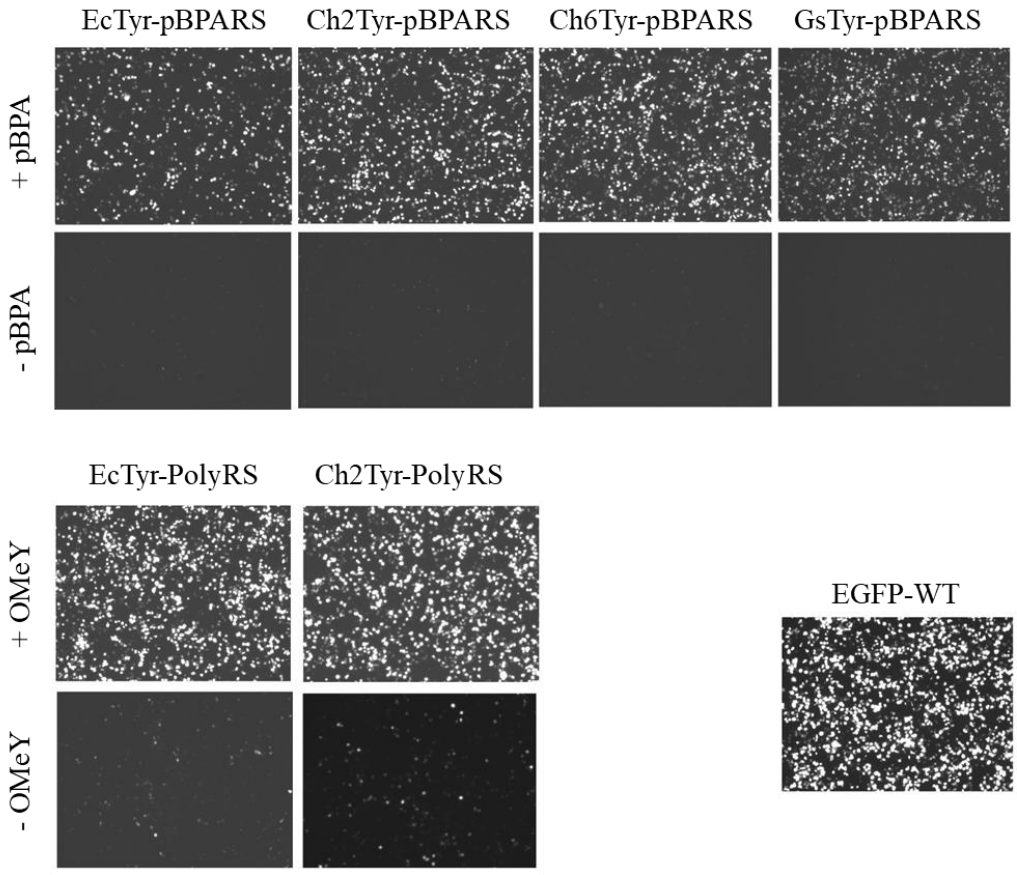


Figure 5.15. Fluorescence imaging of HEK293T cells expressing EGFP-39-TAG using various TyrRS/tRNA pairs (shown above) in the presence and absence of UAA.

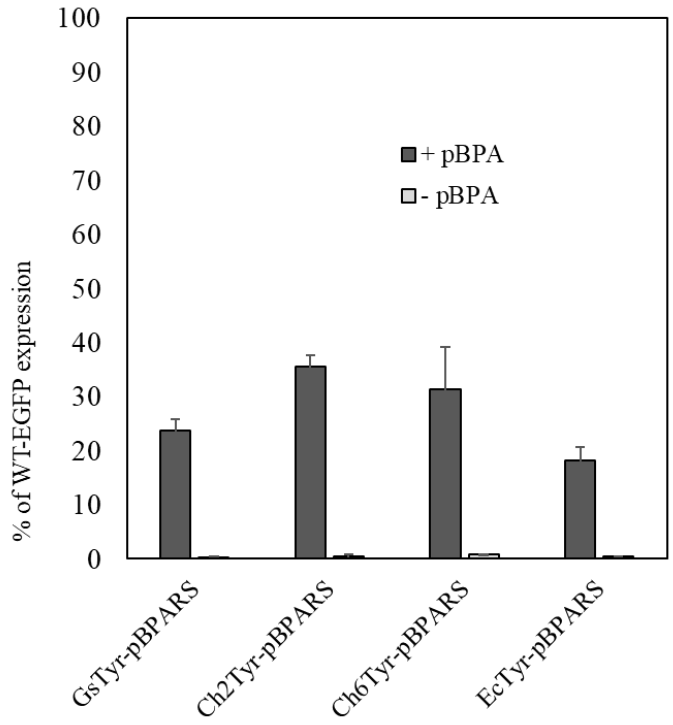


Figure 5.16. EGFP-39-TAG assay of the TyrRSs in *E. coli*. The TyrRSs were expressed in HEK293T cells in the presence of the suppressor tRNA and subsequent fluorescence of full-length EGFP fluorescence was measured in resuspended cells.

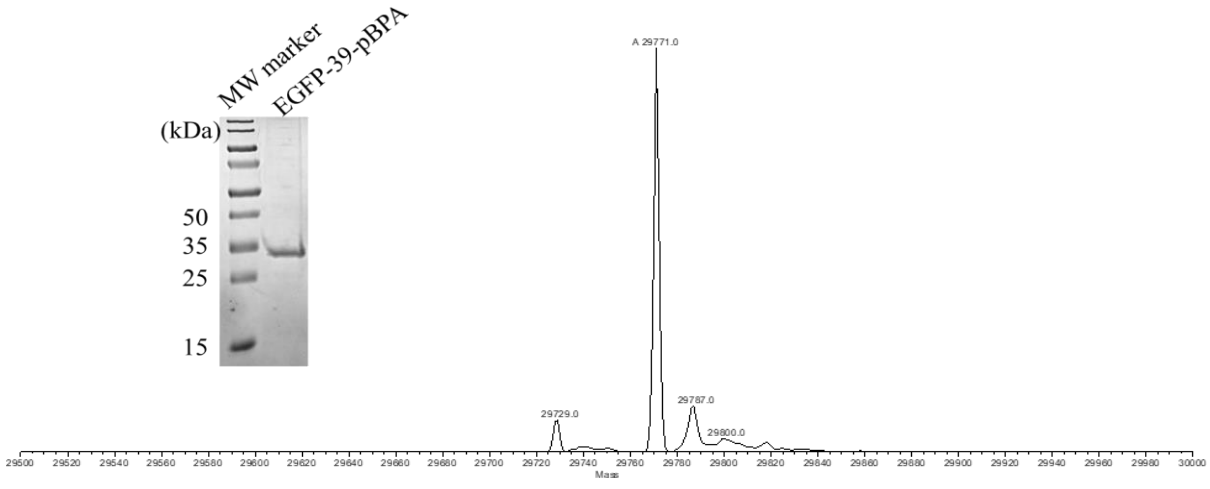


Figure 5.17. SDS-PAGE and MS analysis of EGFP-39-pBPA expressing in the presence of pBPA-selective Ch2TyrRS.

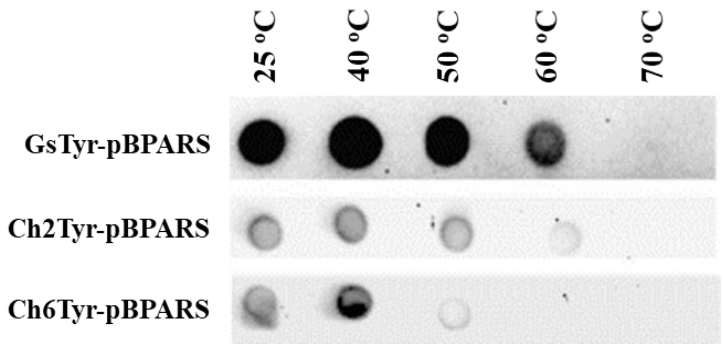


Figure 5.18. Thermal shift assay of the pBPA-selective TyrRSs in HEK293T cells. Dot blot was performed with an anti-polyhistidine antibody on the soluble fraction of HEK293T cell free extracts expressing the indicated pBPA-selective TyrRS variants (left), after incubation at the indicated temperature (top).

To validate the generalizability of using the chimeric TyrRS scaffolds for UAA incorporation we transferred the active site mutations of a polyspecific *Ec*TyrRS into the Ch2 structure. The activity of the Ch2 polyspecific variant was compared to its parental *Ec*TyrRS scaffold through an EGFP-39-TAG fluorescence assay in HEK293T cells (Figure 5.19). Both of these TyrRSs exhibited similar activity (Figure 5.19), demonstrating that the chimeric scaffold can

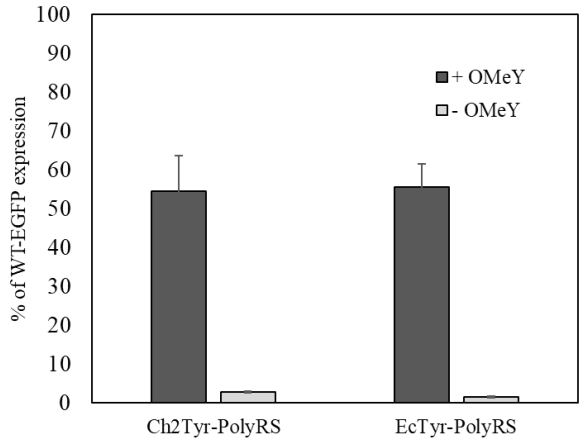


Figure 5.19. Activity of the polyspecific TyrRSs. Activity was measured in HEK293T cells using the expression of the EGFP-29-TAG reporter in the presence of OMeY. Characteristic fluorescence of the full-length EGFP-39-TAG reporter was measured in resuspended cells and normalized relative to EGFP-WT expression fluorescence.

re-capitulate the activity of well-behaved bacterially engineered *Ec*TyrRS mutants, while providing a better scaffold for those with suboptimal activity.

5.2.3 Development of a stability-based FACS selection

Our ability to rationally design chimeric bacterial TyrRSs generated more stable scaffolds for UAA-selective incorporation. While this strategy proved fruitful, rational design approaches for engineering aaRSs are limited based on our knowledge of aaRS structure-function relationships, minimizing the pool of candidates for chimera genesis. We sought to develop a directed evolution strategy that circumvents the necessity of prior knowledge of the aaRS structure. Instead, it would rely on the sexual recombination of two aaRS genes to create a library of random chimeric aaRS variants that can be sieved through stability and activity-based selections. We have already developed and extensively characterized multiple activity-based selection platforms (e.g., the dual-fluorescent reporter system developed in chapter 3 of this dissertation). Therefore, we only needed to develop and characterize a stability-based selection platform.

The creation of aaRS libraries through homologous and non-homologous recombination results in a majority of inactive variants.[28, 29] Realistically, larger library sizes ($>10^7$) would need to be generated and screened in order to identify an active variant. This consideration was central to the design of our stability-based selection reporter. We would need to achieve high-throughput screening capabilities if we desired to engineer any active aaRS chimeras through recombination. This could be achieved if we incorporated a fluorescent-reporter to enable fluorescence activated cell-sorting (FACS) as part of our platform.

Guided by previous work,[30, 31] we explored the use of a monomeric GFP variant (GFPmut3) fusion assay. In our construct the GFPmut3 would be fused to the C-terminus of the aaRS (Figure 5.20). Upon expression, well-folded aaRS permits folding of the C-terminal GFPmut3 and

subsequent chromophore formation. Contrastingly, if the N-terminal aaRS is folded improperly then the GFPmut3 will not fold and no fluorescence can be detected. To test the dynamic range of our construct, fluorescence analysis was carried out in ATMY5 *E. coli* with TyrRSs of varying stabilities (Figure 5.21). As anticipated, the *Ec*TyrRS native structure was the most stable, followed by the polyspecific variant, and the pBPA variants. To validate the utility of this construct for the directed evolution of TyrRSs we carried out a mock selection with known unstable and stable *Ec*TyrRS variants, pBPA-selective and polyspecific respectively. A mock library was generated (1000 pBPA-selective *Ec*TyrRS: 1 polyspecific *Ec*TyrRS) and subjected to FACS (Figure 5.22). Following one round of enrichment, individual mock library members were characterized by fluorescence analysis and Sanger sequencing revealing that predominantly the stable, polyspecific *Ec*TyrRS variant was enriched (Figure 5.23). This demonstrated the feasibility of selecting more stable chimeric bacterial aaRSs with this platform.

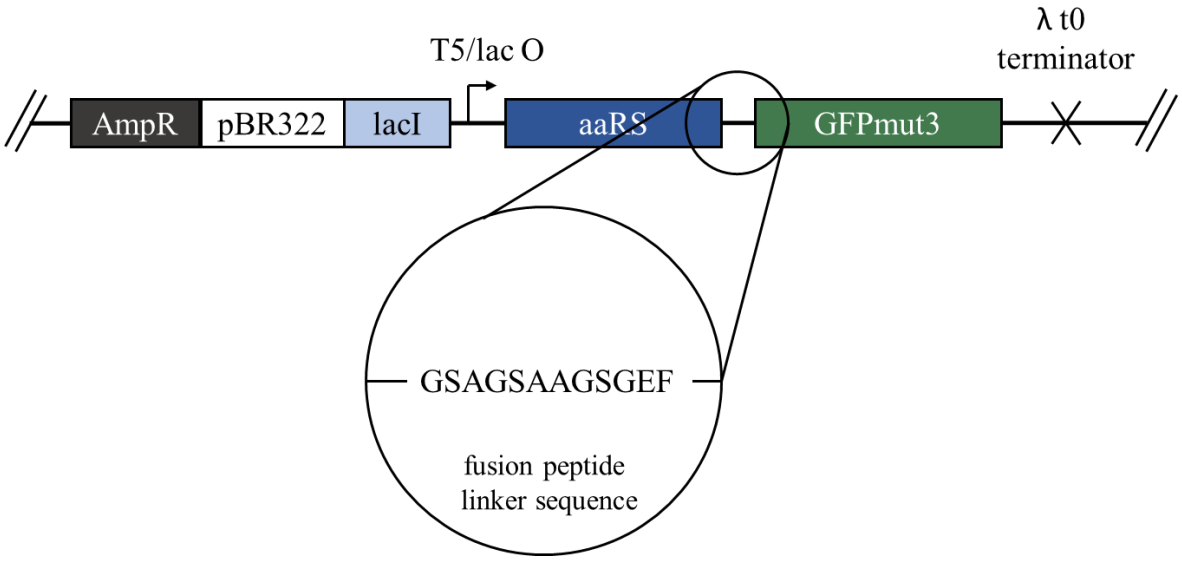


Figure 5.20. pET22b aaRS-GFPmut3 fusion construct. The N-terminal aaRS is linked to the C-terminal GFPmut3 with a previously reported linker sequence.[31]

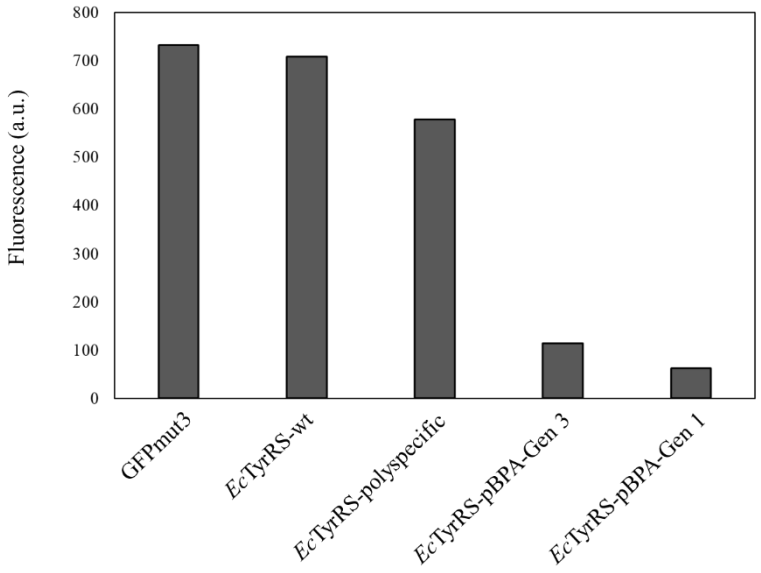


Figure 5.21. Fluorescence analysis of the *Ec*TyrRS-GFPmut3 fusion constructs. Expression of the aaRS was carried out in ATMY5 and monitored based on the formation of the GFPmut3 chromophore.

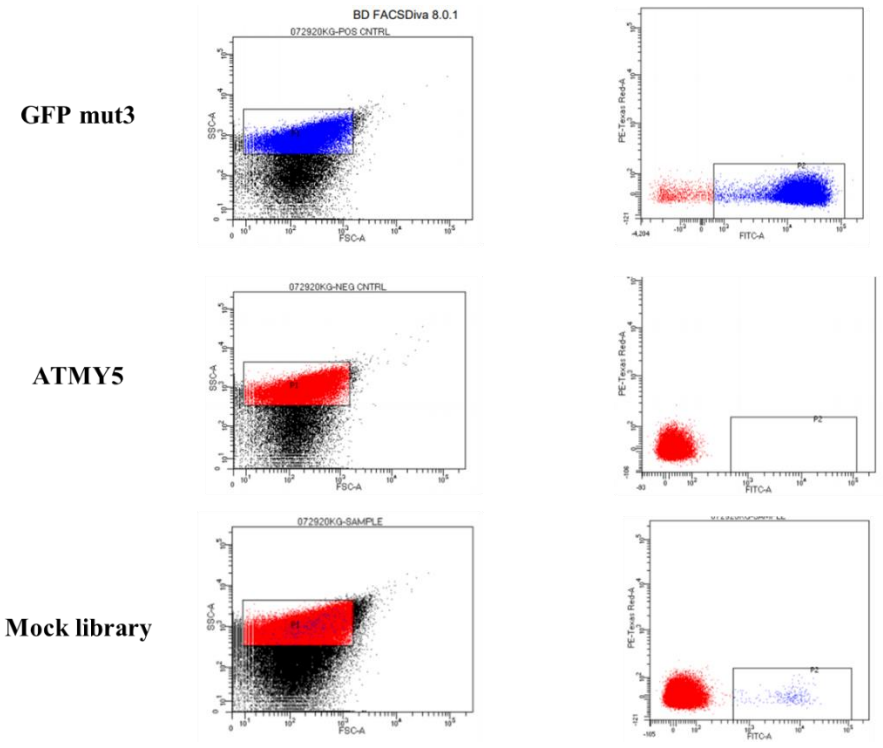


Figure 5.22. FACS selection of the mock library. The population of cells sorted were chosen based on their forward scatter and side scatter light ratio (left column). The sorted cells were further scrutinized for GFP fluorescence (right column). Parameters were outlined as quadrant P2,

resulting in the collection of < 1% of the total parent population. GFP mut 3 (top row) and cell only controls (middle row) in addition to the mock library (bottom row).

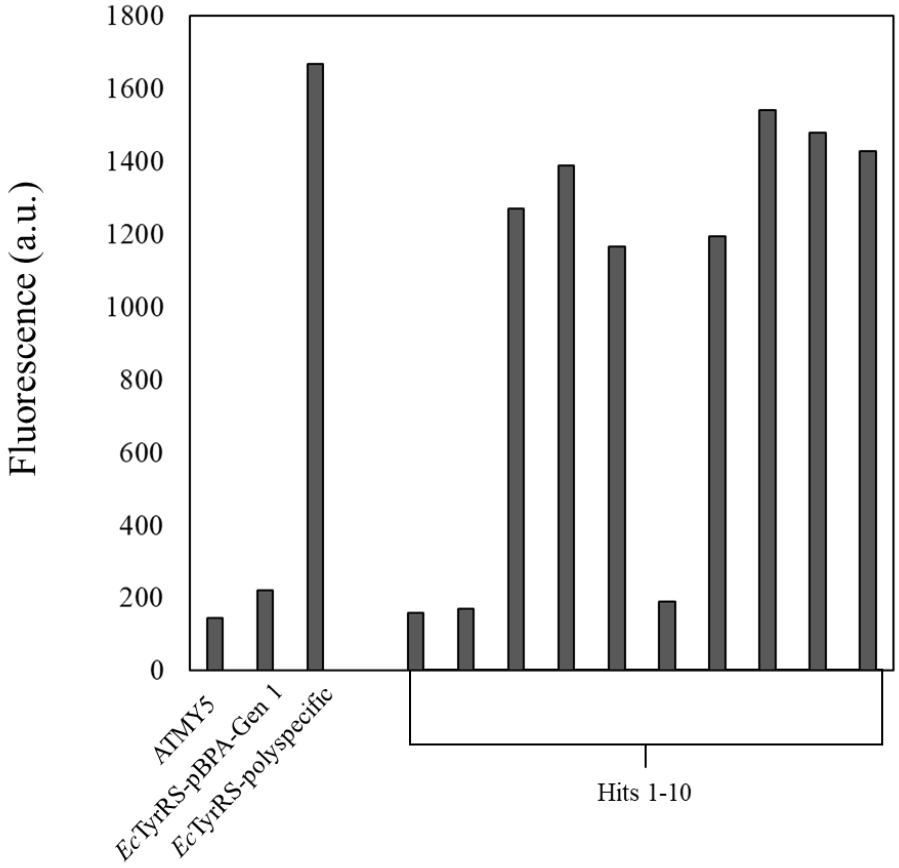


Figure 5.23. Fluorescence analysis of enriched *Ec*TyrRS variants following FACS selection. Expression of the aaRS was carried out in ATMY5 and monitored based on the formation of the GFPmut3 chromophore.

5.3 Conclusions

In summary, we show that the structural robustness of an aaRS affects its engineerability for GCE. We provide a solution for creating optimal parental TyrRS scaffolds for directed evolution by engineering chimeras composed of homologous TyrRSs from mesophilic and thermophilic bacteria. Furthermore, we describe a platform for the directed evolution of new chimeric scaffolds so that we are not limited by rational design. The directed evolution of these

new chimeric TyrRSs with our optimized ATMY platform should give access to new UAA-selective variants.

5.3.1 Ongoing and future directions

Moving forward, the stability fluorescent reporter construct can be used to engineer new chimeric aaRS scaffolds. One of our newest graduate students, Chintan Soni, will be working on the development of identifying the optimal TyrRS scaffold for GCE applications. Additionally, this stability construct can be further applied to demonstrate how the parental aaRS scaffold affects the total number of aaRS variants possible. I am currently working towards the creation of three homologous libraries in TyrRS scaffolds of varying stability (*Ec*TyrRS, *Gs*TyRS, Ch2TyrRS). These three libraries will be subjected to a stability selection followed by next-generation sequencing analysis of the library input and output. Comparison of the library input and output would provide a snapshot of the variability in the number and types of mutants accessible with each scaffold.

5.4 Acknowledgements

Christen Hillenbrand kickstarted this project when she struggled to express different mutant synthetases for an activity assay she was conducting. Megan Yeo helped conduct fluorescence assays. Chintan Soni is the current graduate student being trained to carry this project forward. Finally, the work of Hugues Bedouelle inspired the key facet for this project. His meticulous work investigating *Ec*TyrRS structural stability elements made this project possible.

5.5 Experimental procedures

General materials and methods are described in the experimental procedures section of chapter 2.

5.5.1 Strains, cell lines

ATMY6 *E. coli* cells were obtained from the Chatterjee lab glycerol stocks. This cell line was engineered by a current student in the Chatterjee lab, Elise Ficaretta. This strain is derived from the C321 strain developed by the Church lab.[32] The same strategy used to engineer prior ATMY strains was applied to create this one.

HEK293T cells (ATCC) were maintained at 37 ^{o}C and 5% CO_2 in DMEM-high glucose (HyClone) supplemented with penicillin/streptomycin (Hyclone, final concentration of 100 U/L penicillin and 100 μg/mL streptomycin) and 10% fetal bovine serum (Corning).

5.5.2 Plasmids

All primer and plasmid sequences can be found in the appendix.

The pBK *G. stearothermophilus* tyrosyl-RS was created by amplifying the gene from a gBlock, provided by IDT, with the terminal primers GsYRS NdeI-F and GsYRS NcoI-R. The amplified insert was then digested with NdeI/NcoI and inserted into the pBK vector backbone, which had been digested with the same restriction enzymes. The mutant *G. stearothermophilus* tyrosyl-RS mutants were then generated via standard site-directed mutagenesis of the appropriate active site residues with the following primers: GeobacYRS-Y34G-R, GeobacYRS-D176G-R, GeobacYRS-GGFL-L180A-R, GeobacYRS-D176G-F, and GeobacYRS-GGFL-L180A-F. The pBK *E. coli*

tyrosyl-RS wild type and mutants were previously reported and propagated from a former student's stocks, Dr. James Italia.[8, 9]

The pBK chimeric H2 and H6 tyrosyl-RS' were constructed through overlap amplification of fragments of the *E. coli* tyrosyl-RS N-terminus and the *G. stearothermophilus* C-terminus. The PCR fragments were made with the following primers: EcYRS NdeI-F, H2 EcYRS-iR, H6 EcYRS-iR, H2 GsYRS-iF, H6 GsYRS-iF, and GsYRS NcoI-R. The overlap amplification of these fragments utilized the terminal primers EcYRS NdeI-F and GsYRS-NcoI-R. This product was then digested with NdeI/NcoI and inserted into the pBK vector backbone, which was digested with the same restriction enzymes. The mutant chimeric tyrosyl-RS' were generated via site-directed mutagenesis with the following primers: GeobacYRS-D176G-R, GeobacYRS-GGFL-L180A-R, GeobacYRS-D176G-F, and GeobacYRS-GGFL-L180A-F. The H2 and H6 sequence were rationally designed based on previously reported *in vitro* work by Bedouelle et al.[20]

The pET22b-NtermHis-aaRS constructs were made by PCR amplifying the aaRS' from their respective pBK plasmids with the following primers: EcYRS-nterm10XHis-NdeI-F, EcYRS-HindIII-R, GsYRS-nterm-10XHis-NdeI-F, GsYRS-HindIII-R, MjYRS-nterm-10XHis-NdeI-F, and MjYRS-HindIII-R. This PCR product was digested with NdeI/HindIII and inserted into the pET22b vector backbone.

The pB1U-NtermHis-aaRS constructs were made by PCR amplifying the aaRS' from their respective pBK plasmids with the following primers: EcYRS-nterm10XHis-NdeI-F, EcYRS-XhoI-R, GsYRS-nterm-10XHis-NdeI-F, GsYRS-XhoI-R, MjYRS-nterm-10XHis-NdeI-F, and MjYRS-XhoI-R. This PCR product was digested with NdeI/XhoI and inserted into the pB1U vector backbone.

The pEvol T5 EcY-TAG sfGFP151-TAG plasmid was propagated in DH10B cells from a previous student's stocks.

Overlap extension was used to create the pET22-aaRS-GFP mut3 constructs. Two separate PCR products were amplified of the aaRS gene and GFPmut3 gene with Phusion HSII with the following primers: EcYRS-NdeI-F, EcYRS-linker-GFPmut3-iR, GsYRS-NdeI-F, GsYRS-GFPmut3-iR, Linker-GFPmut3-Nterm-iF, and GFPmut3-HindIII-R. The two products were joined together by primerless overlap extension followed by agarose gel purification (1% agarose gel, 150 V). The purified product was amplified with the terminal primers EcYRS-NdeI-F, GsYRS-NdeI-F and GFPmut3-HindIII-R and digested with NdeI/HindIII. This digested product was ligated by T4 DNA ligase into pET22b vector digested with the same restriction enzymes.

5.5.3 Unnatural amino acids

The o-methyl-l-tyrosine used in the present experiments was obtained from Fisher Scientific (catalog number AAH6309606).

para-benzoyl-l-phenylalanine was purchased from Chem-Impex International (catalog number 05110).

5.5.4 sfGFP-151-TAG fluorescence analysis and expression

The pBK aaRS and pEvol T5 EcY-TAG sfGFP151-TAG plasmids were co-transformed into ATMY6 cells for expression. An LB media culture (5 mL, 1x Spec/Kan/Chlor) was inoculated with a single colony and grown overnight. The next day, this overnight culture was used to inoculate another LB media culture (20 mL, 1X Spec/Kan/Chlor) and grown to an OD600 of 0.6. Once the cultures reached the optimal OD600, they were induced with IPTG (1 mM final concentration), UAA was added (1 mM final concentration), and they were incubated for 16 hours

at 30 ºC with shaking (250 rpm). Following induction, the cultures were spun down, the LB media was removed, and the cells were resuspended in 1X PBS. Fluorescence readings were collected in a 96-well plate using a SpectraMAX M5 (Molecular Devices) (ex = 488 nm, em = 534 nm). Mean of two independent experiments were reported, and error bars represent standard deviation.

5.5.5 EGFP-39-TAG fluorescence analysis, expression and purification

For EGFP-39-TAG fluorescence analysis, HEK293T cells were seeded at a density of 600,000 cells per 12-well dish 24 hours prior to transfection. A total amount of 1.5 µg DNA (0.75 µg of pAcBac1 EGFP-39-TAG and 0.75 µg of pB1U aaRS) + 3.5 µL PEI + 17.5 µL DMEM was used for transfection of each well once the cells reached ~70% confluence. Fluorescence images and readings were performed 48 hours post transfection. Fluorescence images were taken using a Zeiss Axio Observer fluorescence microscope. Fluorescence readings were collected in a 96-well plate using a SpectraMAX M5 (Molecular Devices) (ex = 488 nm, em = 510 nm). The mean of four independent experiments were reported, and error bars represent the standard deviation.

For EGFP-39-TAG expression and purification, HEK 293T cells were seeded at a density of 5 million cells per 100 mm dish 24 hours prior to transfection. A total amount of 12 µg DNA (6µg of pAcBac1 EGFP-39-TAG and 6 µg of pB1U aaRS) + 50 µL PEI MAX + 150 µL DMEM was used for transfection of each dish once the cells reached ~ 90% confluence. 48 hours post-transfection, fluorescence images were taken using a Zeiss Axio Observer fluorescence microscope.

5.5.6 CETSA assay in E. coli

For bacterial aaRS expression, the pET22b-Nterm10XHis-aaRS was transformed into *E. coli* cells. LB media cultures (5 mL, 2X Amp) were inoculated with a single colony and grown overnight at 37 ºC with shaking (250 rpm). The next day, 20 mL LB media cultures (2X Amp) were inoculated with 200 µL of the overnight culture, grown to an OD600 of 0.6 (37 ºC, 250 rpm), and then induced with IPTG (final concentration of 1 mM) for 15 minutes at 30 ºC with shaking. The cultures were then spun down, the LB media was removed, and the cell pellets were resuspended in 500 µL of sonication buffer (100 mM NaCl, 25 mM Tris HCl, pH 8.0). The cell pellets then underwent 3x freeze-thaw cycles followed by sonication (3x, 75% power, 20 pulses). The sonicated lysate was then spun down (4 ºC, 14,000 rpm, 10 min) and the supernatant was collected. Each supernatant was divided into 50 µL aliquots and heated at varying temperatures for 5 minutes on a Perkin Elmer Cetus DNA Thermal Cycler 480 and spun down at maximum speed for 10 minutes. Then, 3 µL of the supernatant was inoculated on a nitrocellulose membrane and treated to western blot analysis following previously described protocols.[33] Antibodies used for imaging include: mouse anti-Histidine 6X tag antibody (1:1000 dilution), chicken anti-mouse IgG secondary antibody-HRP conjugate (1:5000 dilution).

5.5.7 CETSA assay in HEK293T

One day before transfection, HEK293T cells were seeded at a density of 600,000 cells per well for a 6-well plate. A total amount of 1.6 µg of DNA + 8.3 µL MAX PEI + 30 µL DMEM was used for transfection of each well. The culture media was replaced 24 hours post-transfection. Lysate preparation, sample heat treatment, and western blot analysis were performed 48 hours following transfection using the same protocol described in the "CETSA assay in *E. coli*" procedure section.

5.5.8 Solubility western

E. coli DH10B cells were transformed with a pET22b-Nterm10XHis-aaRS plasmid. An overnight 5 mL LB media culture (2X Amp) was inoculated with one of the transformants. The subsequent day, the overnight culture was used to inoculate a 20 mL LB media culture (2X Amp) that was grown to an OD600 of 0.6 (37 ºC, 250 rpm). The 20 mL LB media cultures were then moved to 30 ºC and grown for an additional 4 hours with shaking (250 rpm). Cell lysate was prepped from these cultures following the protocol mentioned in the CETSA assay procedures. The clarified lysate was resolved on a 12% SDS-PAGE gel and worked up for a western following a previously described protocol.[33] The antibodies used for this protocol were the same as those used for the CETSA assay.

5.5.9 Mock FACS stability selection

The pET22b-*Ec*TyrRS wt-GFP mut3 and pET22b-*Ec*YRS-pBPA-Gen 1-GPF mut 3 were transformed into ATMY5 *E. coli* and grown at 37 ºC overnight. The next day, the plates were harvested and used to inoculate LB media cultures at an OD600 of 0.03 and were grown with shaking to an OD600 of 0.4 (37 ºC, 250 rpm). The cultures were induced with 1 mM IPTG at 30 ºC with shaking (250 rpm) for 1.5 hours. Following induction, the cultures were spun down (4,000 g, 10 min) and resuspended in 1X PBS. The cells were diluted to ~1.0 x 10^6 cells/mL and submitted for FACS analysis. 5.0 x 10^4 cells were recovered in a 2 mL 2X concentration LB media culture and grown at 37 ºC for 4 hours with shaking (250 rpm). The recovered cell culture was then plated on LB agar media supplemented with antibiotics (1X Spec/2X Amp) and grown overnight.

5.6 References

1. Edwards H., Schimmel P. A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase. *Mol. Cell. Biol.* 10, 1633 (1990).
2. Bedouelle, H. Tyrosyl-tRNA Synthetases. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience (2000-2013).
3. Yaremchuk A, Kriklivyi I, Tukalo M, Cusack S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. EMBO J. 21(14), 3829 (2002).
4. Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. Nat. Struct. Biol. 10(6), 425 (2003).
5. Chin, J. W. Expanding and reprogramming the genetic code. Nature 550, 53 (2017).
6. Dumas, A., Lercher, L., Spicer, C. D., Davis, B. G. Designing logical codon reassignment−Expanding the chemistry in biology. Chemical Science 6, 50 (2015).
7. Young, D. D., Schultz, P. G. Playing with the molecules of life. ACS Chem. Biol. 13, 854−870 (2018).
8. Chin, J.W., Cropp, T.A., Anderson, C., Mukherji, M., Schultz, P.G. An expanded eukaryotic genetic code. Science 301, 964 (2003).
9. Italia, J.S., Latour, C., Wrobel, C.J.J., Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. Cell Chemical Biology 25(10), 1304 (2018).
10. Luo, X., Fu, G., Wang, R. E., Zhu, X., Zambaldo, C., Liu, R., Liu, T., Lyu, X., Du, J., Xuan, W., Yao, A., Reed, S. A., Kang, M., Zhang, Y., Guo, H., Huang, C., Yang, P. Y., Wilson, I. A., Schultz, P. G., Wang, F. Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. Nat Chem Biol. 13(8), 845 (2017).
11. Wang, L., Brock, A., Herberich, B., Schultz, P.G. Expanding the Genetic Code of *Escherichia coli*. Science 292, 498 (2001).
12. Tokuriki, N., Tawfik, D. S. Stability effects of mutations and protein evolvability. Curr. Opin. Struct. Biol. 19, 596 (2009).
13. Tokuriki, N., Tawfik, D. S. Protein dynamism and evolvability. Science 324, 203 (2009).
14. Socha, R. D., Tokuriki, N. Modulating protein stability−directed evolution strategies for improved protein function. FEBS J. 280, 5582 (2013).
15. Trudeau, D. L., Tawfik, D. S. Protein engineers turned evolutionists: the quest for the optimal starting point. Curr. Opin. Biotechnol. 60, 46−52 (2019).
16. Martinez Molina, D., Nordlund, P. The cellular thermal shift assay: a novel biophysical assay for in situ drug target engagement and mechanistic biomarker studies. Annu. Rev. Pharmacol. Toxicol. 56, 141 (2016).
17. Martinez Molina, D., Jafari, R., Ignatushchenko, M., Seki, T., Larsson, E. A., Dan, C., Sreekumar, L., Cao, Y., Nordlund, P. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. Science 341, 84 (2013).

18. 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 Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 99, 11020 (2002).
19. Young, D. D., Young, T. S., Jahnz, M., Ahmad, I., Spraggon, G., Schultz, P. G. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry 50, 1894 (2011).
20. Guez-Ivanier, V., Hermann, M., Baldwin, D., Bedouelle, H. Mapping the stability determinants of bacterial tyrosyl transfer RNA synthetases by an experimental evolutionary approach. J. Mol. Biol. 234, 209 (1993).
21. Laowanapiban, P., Kapustina, M., Vonrhein, C., Delarue, M., Koehl, P., Carter, C. W., Jr. Independent saturation of three TrpRS subsites generates a partially assembled state similar to those observed in molecular simulations. Proc. Natl. Acad. Sci. U. S. A. 106, 1790 (2009).
22. Sakurama, H., Takita, T., Mikami, B., Itoh, T., Yasukawa, K., Inouye, K. (2009) Two crystal structures of lysyl-tRNA synthetase from Bacillus stearothermophilus in complex with lysyladenylate-like compounds: insights into the irreversible formation of the enzyme-bound adenylate of L-lysine hydroxamate. J. Biochem. 145, 555 (2009).
23. Brick, P., Bhat, T. N., Blow, D. M. Structure of tyrosyl-tRNA synthetase refined at 2.3 A resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate. J. Mol. Biol. 208, 83 (1989).
24. Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., Kern, D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat. Struct. Mol. Biol. 11, 945 (2004).
25. Li, Y., Drummond, D. A., Sawayama, A. M., Snow, C. D., Bloom, J. D., Arnold, F. H. A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. Nat. Biotechnol. 25, 1051 (2007).
26. Jermutus, L., Tessier, M., Pasamontes, L., van Loon, A. P., Lehmann, M. Structure-based chimeric enzymes as an alternative to directed enzyme evolution: phytase as a test case. J. Biotechnol. 85, 15 (2001).
27. Guez-Ivanier, V., Hermann, M., Baldwin, D., Bedouelle, H. Mapping the stability determinants of bacterial tyrosyl transfer RNA synthetases by an experimental evolutionary approach. J. Mol. Biol. 234, 209 (1993).
28. Stemmer, W.P. Rapid evolution of a protein in vitro by DNA shuffling. Nature 370, 389 (1994).
29. Meyer, A.J., Ellefson, J.W., Ellington, A.D. Library Generation by Gene Shuffling. Curr Protoc. Mol. Biol. 105, Unit-15.12 (2014).
30. Wrenbeck, E.E., Bedewitz, M.A., Klesmith, J.R., Noshin, S., Barry, C.S., Whitehead, T.A. An Automated Data-Driven Pipeline for Improving Heterologous Enzyme Expression. 8 (3), 474 (2019).
31. Waldo, G.S., Standish, B.M., Berendzen, J., Terwilliger, T.C. Rapid protein-folding assay using green fluorescent reporter protein. Nat. Biotechnol 17, 691 (1999).
32. Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A., Rohland, N., Schultz, P.G., Jacobsen, J.M., Rinehart, J., Church, G.M., Isaacs, F.J. Genomically Recoded Organisms Expand Biological Functions. Science 342 (6156), 357 (2013).

33. Zheng, Y., Lewis Jr., T.L., Igo, P., Polleux, F., Chatterjee, A. Virus-Enabled Optimization and Delivery of the Genetic Machinery for Efficient Unnatural Amino Acid Mutagenesis in Mammalian Cells and Tissues. ACS Synth Biol. 6 (1), 13 (2017).

Chapter 6

Various short stories

6.1 Construction of a T-box riboswitch controlled fluorescent reporter

6.1.1 Introduction

Directed evolution platforms for engineering bacterial aminoacyl-tRNA synthetase (aaRS)/tRNA pairs for eukaryotic genetic code expansion (GCE) predominantly rely on the expression of a catalytically active reporter protein.[1-4] A nonsense codon is introduced into the catalyst (e.g., chloramphenicol acetyltransferase) at a structurally permissible location resulting in cell survival in the presence of an active mutant aaRS. While this general selection methodology has been proven successful, it limits our ability to quantitatively observe differences in the suppression efficiencies of mutant aaRSs. This is because sufficient levels of catalyst translation for cell survival is not always analogous to how well the tRNA is charged with the UAA by the mutant aaRS. Identifying more quantitative platforms for bacterial aaRS engineering will enable us to more easily identify the dynamic ranges of active aaRSs for GCE applications.

The natural regulation of the expression of native aaRSs and their cognate tRNA are important for maintaining cell health. This is accomplished in gram-positive bacteria through an evolutionarily conserved mRNA sequence called a "T-box" that modulates aaRS production by measuring the amount of charged and un-charged tRNA present in the cell.[5-8] The T-box RNA regulates expression of the adjacent, downstream gene through two main different chemical inputs, the tRNA anticodon loop and aminoacylation of the 3'-adenosine.[5-8] All T-boxes are composed of two domains, a 5' (stem I) domain that recognizes the anticodon loop of the tRNA and a 3' domain that base-pairs with the uncharged 3'-adenosine end of the tRNA (Figure 6.1).[5-8] In the presence of uncharged tRNA, a Shine-Dalgarno antisequester (transcription antiterminator) is

formed enabling the translation of the downstream aaRS (Figure 6.1).[5-8] In the presence of charged tRNA, a Shine-Dalgarno sequester (transcription terminator) is formed and gene expression is stopped[5-7] (Figure 6.1).

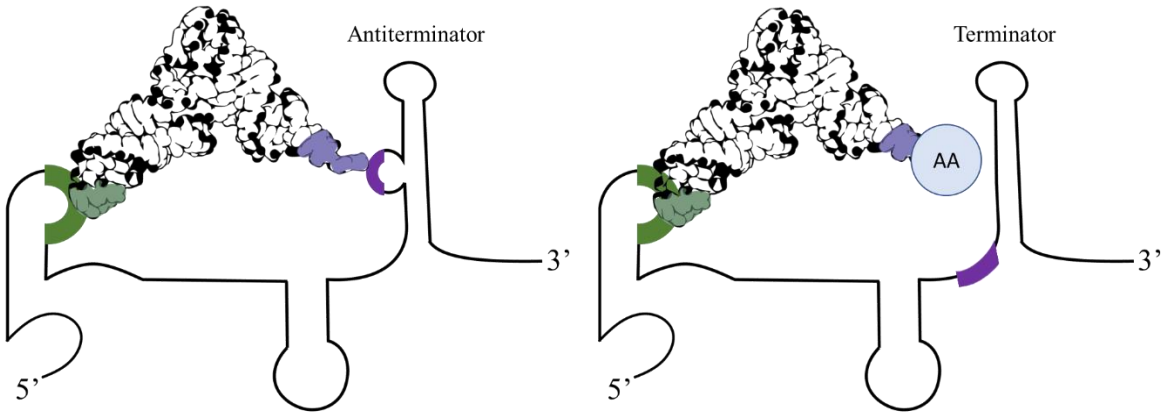


Figure 6.1. T-box structure and regulation mechanisms.[5-8] The tRNA structure is shown as a 3D surface outline. The general T-box RNA leader sequence stem-loop motifs are shown as a black outline from 5' to 3' end. On the left, uncharged tRNA experiences base pairing between the anticodon (green) and a specifier sequence on the RNA (green) in addition to the tRNA acceptor end (light purple) and a discriminatory sequence on the RNA (purple). The selective binding of these regions results in the formation of an antiterminator. On the right, an aminoacylated tRNA (light blue circle) cannot bind with the specifier sequence, resulting in the formation of a terminator in the RNA.

The innate ability of the T-box RNA sequence to recognize tRNA-acylation levels makes it an attractive candidate for developing a quantitative aaRS directed evolution platform. Ideally, a fluorescent reporter gene could be placed downstream of the T-box. Expression levels of the fluorescent reporter gene would then be directly tied to the amount of charged suppressor tRNA and provide access to quantitative suppression-efficiency measurements, as well as high-throughput selection methods (e.g., fluorescence activated cell-sorting). Efforts towards utilizing T-box RNA sequences for aaRS directed evolution platforms have been limited due to their natural lack of occurrence in organisms beyond gram-positive bacteria, and the challenges associated with engineering them to function with alternate aaRS/tRNA pairs. We believe that the ATM *E. coli*

strains we engineered could overcome this hurdle due to their inherent compatibility with gram-positive transcriptional and translational machinery components.[9]

6.1.2 Design and construction of a T-box riboswitch regulated reporter construct

The first T-box regulated selection construct that we sought to design would be used for the ATMY selection platform. Therefore, we would need a T-box that naturally regulates tyrosyl-tRNA synthetase (TyrRS) expression to ensure compatibility for engineering bacterial TyrRS/tRNA pairs. An attractive candidate for our selection construct is the *Geobacillus subtilis* tyrS T-box RNA regulation sequence. This T-box was first identified in 1992[5], where the T-box region was initially believed to be a 14-nucleotide sequence stretch upstream of the tyrS gene. However, by 1993 the full conserved primary sequence and structural elements of the tyrS T-box were identified (Figure 6.2).[10, 11] This extensive characterization made us confident in the identity of the necessary components for transferring the key T-box regulatory elements to our reporter construct.

To maintain the integrity of the T-box RNA sequence regulation mechanism, we aimed to conserve the naturally occurring T-box reporter sequence context when placing it in front of the fluorescent reporter. This was accomplished by ordering a custom designed gblock, supplied by IDT, that contained the *G. subtilis* tyrS sequence leader seqeunce[10, 11] (Figure 6.2). This sequence was PCR amplified with terminal primers and inserted into the pET22b sfGFPwt reporter plasmid through polymerase incomplete primer extension (PIPE) cloning. The sequence of the final reporter construct (Figure 6.3) was validated by Sanger sequencing.

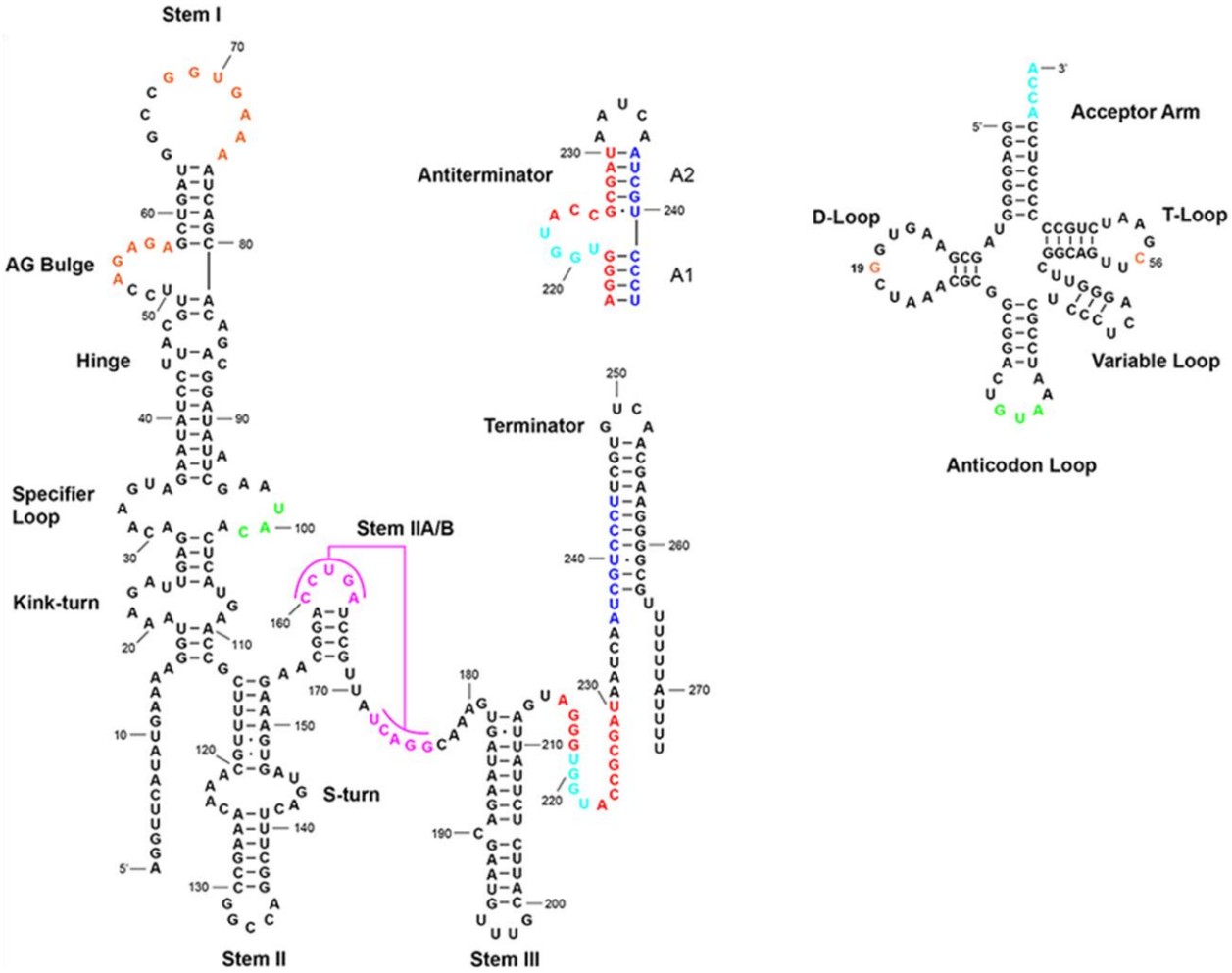


Figure 6.2. ***G. subtilis*** **tyrS RNA leader sequence.**[10, 11] On the left, is the tyrS leader sequence numbered from the transcriptional start site to end of the transcriptional terminator element. The terminator/terminator region are highlighted in red and blue. The discriminatory sequence is highlighted in teal and the anticodon recognition sequence is highlighted in green. Additional regions of interaction in the stem I loop and AG bulge of the RNA sequence that form binding interactions with the D-loop and T-loop of the *G. subtilis* tyrosyl-tRNA are highlighted in orange.

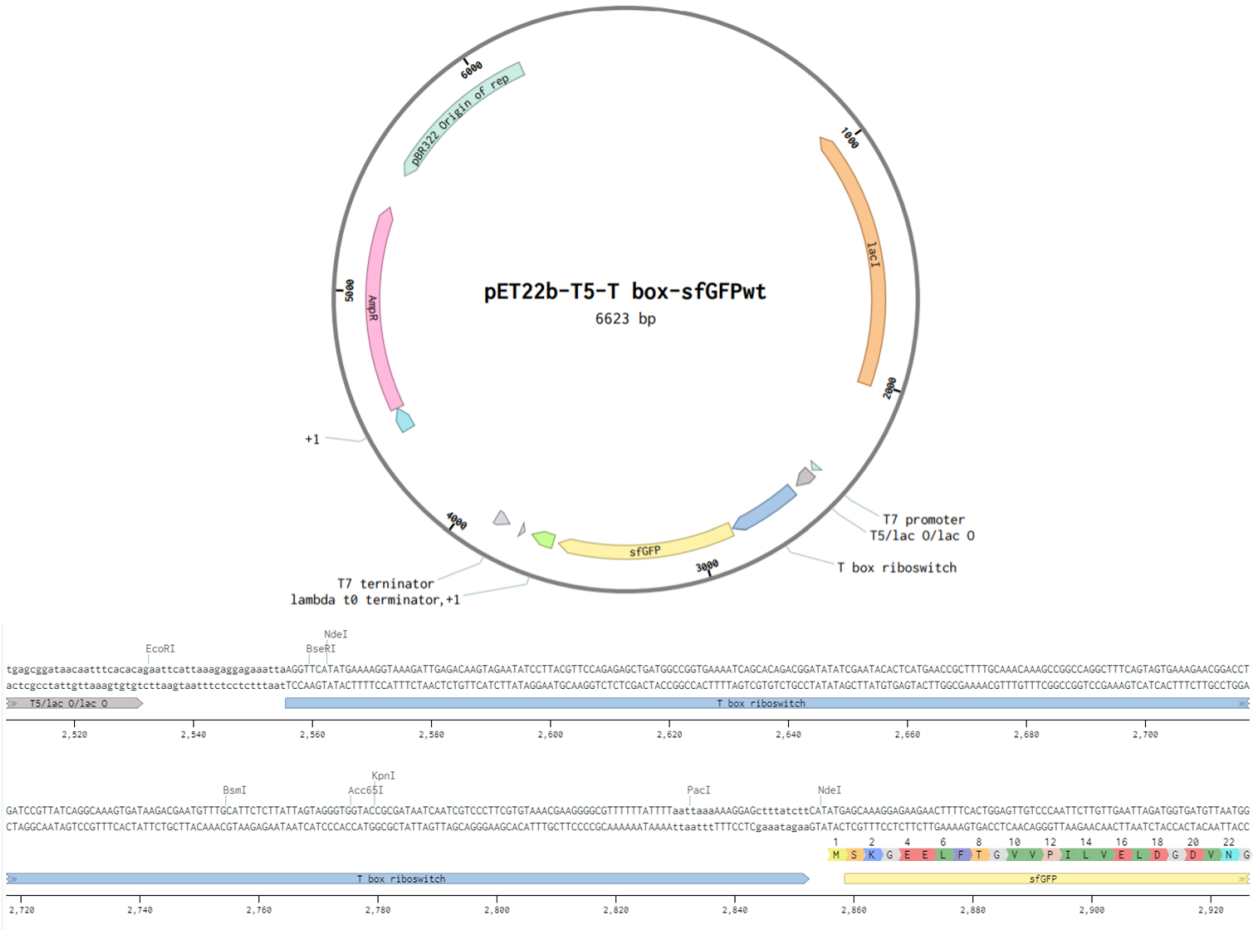


Figure 6.3. pET22b T-box sfGFPwt reporter construct. The T-box RNA leader sequence was cloned into the pET22b sfGFPwt construct following the T5/lac operon and immediately before the sfGFPwt gene sequence. The whole construct map is depicted (top) as well as the sequence of the T-box leader sequence before the sfGFPwt gene (bottom).

6.1.3 Characterization of the T-box riboswitch reporter construct

Following sequence validation of the reporter construct, we characterized the ability of the T-box RNA sequence to regulate expression of sfGFPwt. To carry this out, a wild-type *Escherichia coli* strain (EcNR1GT), a engineered *E. coli* strain that contains no tyrS (EcNR1GT pUltraBR MjY dtyrS), and a engineered *E. coli* strain that contains no tyrS or $tRNA^{Tyr}$ (pUltraBR MjY dtyrS tyrTV::inact tyrU::inact lambda::tolC) were transformed with pET22b T-box sfGFPwt.

After 20 hours of expression at 30 ºC with shaking (250 rpm) in the presence of IPTG (1 mM), sfGFPwt fluorescence was measured (Figure 6.4). As expected, no sfGFP fluorescence was observed in the *E. coli* wild-type strain, because the native *E. coli* TyrRS is suppressing the $tRNA^{EcTyr}$ at sufficient levels to impede antiterminator formation in the T-box leader sequence. Additionally, the engineered *E. coli* strain that has no tyrS or $tRNA^{Tyr}$ displayed ~10% of the fluorescence output of the pET22b sfGFPwt construct without a T-box. This was expected, since the direct interaction between the $tRNA^{Tyr}$ and T-box RNA leader sequence is required for stabilizing the terminator structure formation.[12, 13] The only strain that did not behave as anticipated was the *E. coli* tyrS deletion strain. In the absence of the tyrS all of the $tRNA^{Tyr}$ should be uncharged causing formation of the antiterminator loop upon binding with the T-box leader sequence, resulting in high levels of sfGFPwt fluorescence. Instead, we observed no fluorescence signal when this construct was expressed in this intermediate strain. We believe that this could be due to the loss of genomic integrity of this strain over time, since it was made by a former student more than 5 years ago.

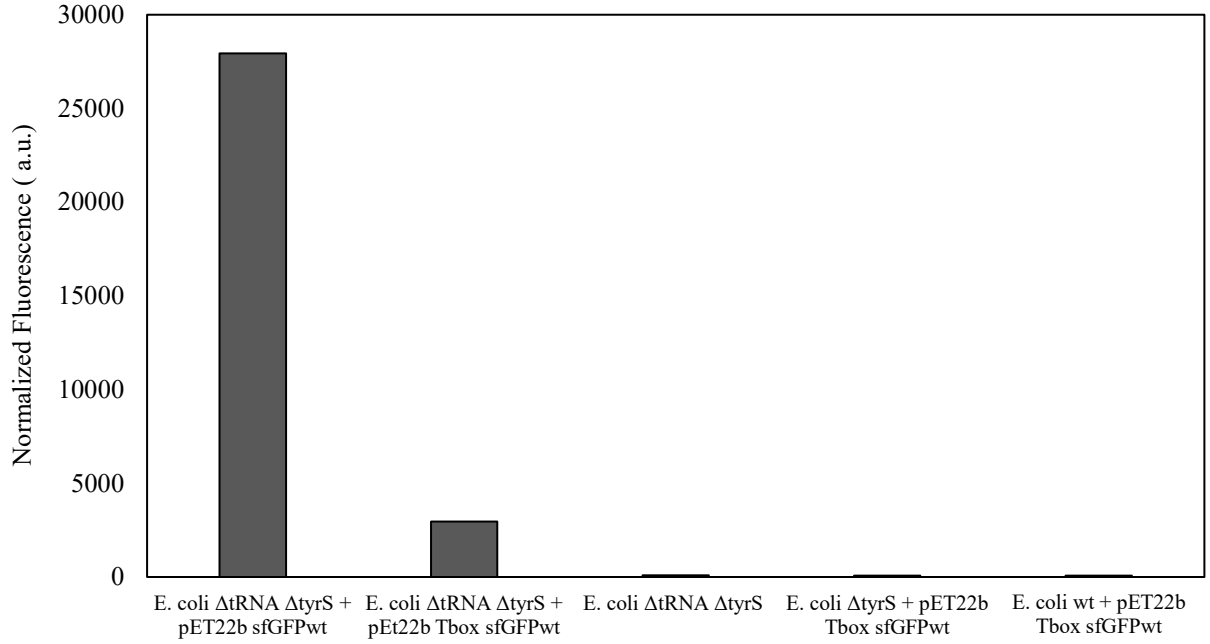


Figure 6.4. Fluorescence analysis of the pET22b T-box sfGFPwt reporter construct. Measured sfGFP fluorescence in resuspended *E. coli* cell cultures was used to determine if the T-box leader sequence was regulated by tRNA$^{Ec\text{Tyr}}$.

6.1.4 Conclusions

In summary, we have designed and characterized a T-box regulated reporter construct that can be used in ATMY *E. coli*. We demonstrated that expression levels of the fluorescent reporter can be tuned by the presence and absence of tRNA$^{Ec\text{Tyr}}$. Further optimization is required for this to become a viable selection construct.

Ongoing and future directions

Moving forward, more experiments characterizing the pET22b T-box sfGFP wt construct in the presence and absence of tRNA$^{Ec\text{Tyr}}$ will be carried out. To circumvent the issue of working with previously engineered ATMY *E. coli* intermediate strains, new plasmids containing the tRNA$^{Ec\text{Tyr}}$ and *E. coli* TyrRS will be constructed and co-transformed into ATMY5 to collect these

data points. Additionally, experiments will be carried out in more recently engineered C321-ATMY strains. Future work towards developing this construct into a bacterial TyrRS selection platform will be carried out by one of my current lab members, Elise Ficaretta.

6.2 Synthesis and proteomic characterization of photocaged aryl diazonium probes

6.2.1 Activity-based protein profiling

The extensive genomic characterization of organisms has pressed researchers towards developing new technologies to understand the cryptic field of proteomics.[14] Presently, over 2,000 genomic projects are underway, however, assigning the functional relationships of the proteome encoded by these genomes remains a daunting task.[14] Relative to the genome, the proteome experiences greater temporal variability leading to an exponential increase in complexity, veiling the purposes of these protein-interaction networks (Figure 6.5). Fortunately, some genomic technologies such as transcriptional profiling and RNA-interference-based gene silencing have provided insight into the effects of protein expression on physiological and pathological processes.[14-18] However, a large gap still remains in our ability to link transcriptional products to their protein function.

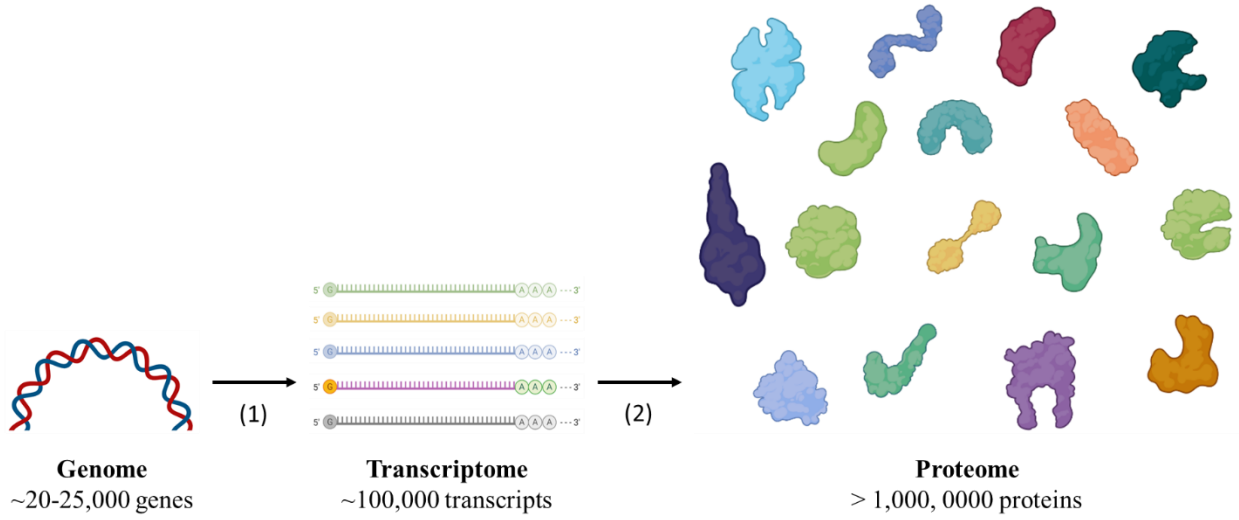


Figure 6.5. Increasing complexity of the proteome relative to the genome.[19] In the human genome there are roughly 20-25,000 genes whose transcriptional products are altered, step (1) depicted above, through the use of alternative promoters, alternative splicing, and mRNA editing. The protein products of these transcripts are further modified following translation through the addition of post-translational modifications (PTMs), step (2) above. This leads to a proteome of exponentially greater complexity than the genome that encodes it.

Activity-based protein profiling (ABPP) provides an attractive solution to bridge this gap. ABPP relies on the use of a probe that targets a subset of functionally active proteins in the proteome that usually have shared catalytic features. Since the probe exclusively targets catalytically active proteins, it can illustrate how the state of an enzyme is regulated by various post-translational modifications (PTMs).[20, 21] The probe has three main components: (i) a reactive/selective warhead, (ii) a tag for visualization and characterization of the proteome, and (iii) a linker that provides sufficient space between the warhead and the tag (Figure 6.6).[14] In conjunction with isotopic labeling and tandem mass spectrometry, this affinity and activity-based

probe can be used to distinguish dynamic changes in the proteome under varying conditions (Figure 6.7).[14]

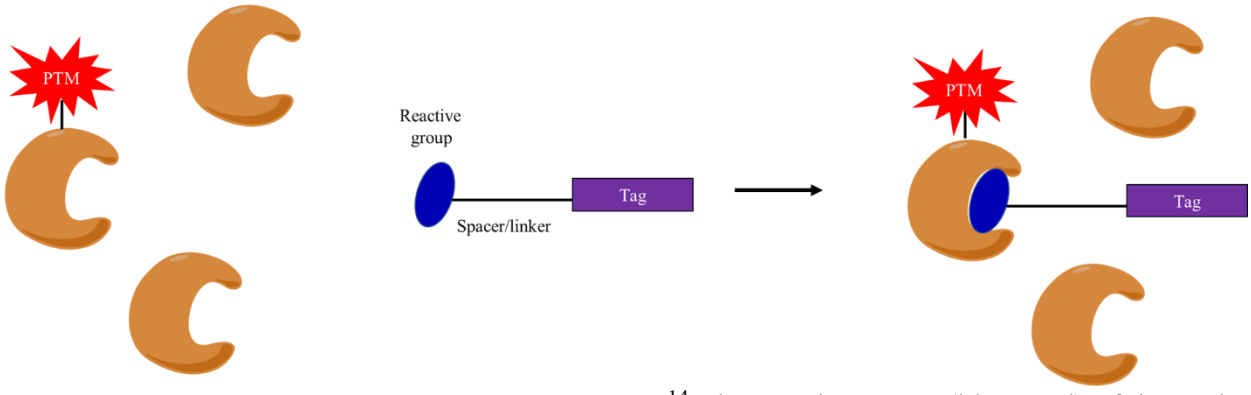


Figure 6.6. General ABPP probe characteristics.[14] The reactive group (blue oval) of the probe will selectively target active forms of a protein, that may be modulated through the addition of PTMs. The linker (black line) connects the reactive group of the probe to a tag (purple rectangle) that can be subsequently used for pull-down methods to isolate the labeled protein in the proteome.

ABPP has established itself as a robust technology for the discovery of enzymatic activities associated with a range of diseases (e.g., cancer, malaria).[22, 23] Its ability to interrogate the proteome based on functional space, rather than abundance, enables our ability to probe previously inaccessible areas of biomolecular space.[14] While this technology has provided numerous breakthroughs in our understanding of proteome dynamics, it is still greatly limited by the chemical warheads currently designed and used for ABPP studies. So far, these warheads are generally electrophilic handles that selectively couple with catalytic active sites.[14, 24] Among these handles, there is a disparity between the types of catalytic residues that are easily targeted, limiting our access to the proteome. We aim to provide additional probes to the repertoire of ABPP to expand its application to new enzyme classes.

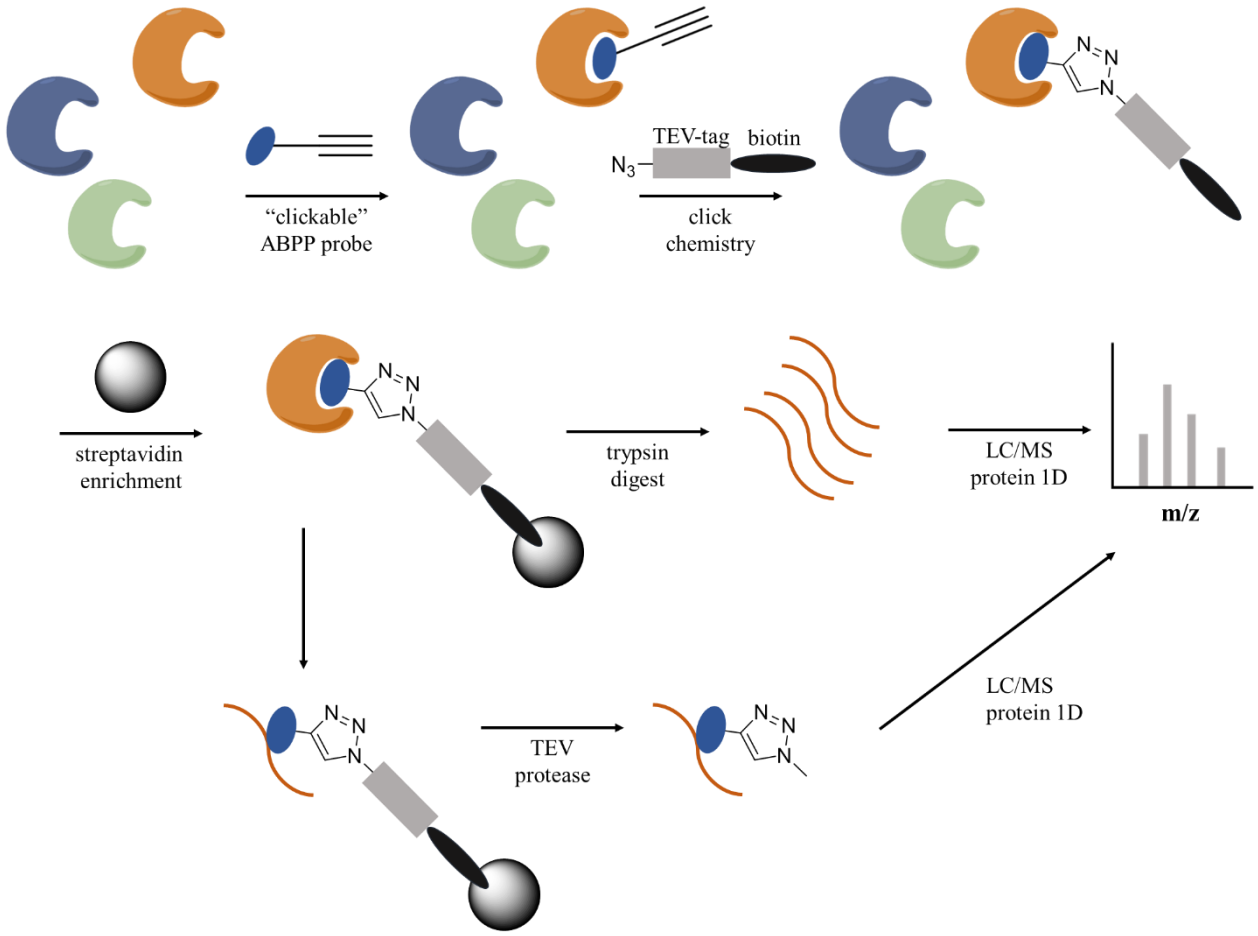


Figure 6.7. TOP-ABPP pull-down approach.[14] In tandem-orthogonal proteolysis (TOP)-ABPP the simultaneous characterization of the protein targets and the sites of protein labeling is achieved. Introduction of the alkyne handle on the ABPP probe enables access to click chemistry tools, which can be utilized to easily append a biotin handle following proteome labeling. The addition of a *Tobacco etch virus* (TEV) protease cleavage site between the azide reactive handle and the biotin tag enables downstream cleavage of the labeled active site peptide following tryptic digest.

6.2.2 Aryl-diazonium ions for selective labeling

Azo-coupling between aryl diazonium ions and activated aromatic residues is a well-characterized, selective reaction.[25, 26] The Francis[27, 28] and Jewett[29] labs, amongst others[30, 31], have demonstrated the power of this reaction to label tyrosine residues in a protein at basic pH (>9) with highly-activated electrophilic analogs of the aryl-diazonium (e.g., 4-nitrobenzenediazonium)

(Figure 6.8). This reaction is especially advantageous for pull-down studies due to the selective cleavage of the azobenzene product through reducing conditions.[27, 28, 32] Prior work in our lab has demonstrated the utility of tuning the electrophilicity of the aryl diazonium to gain different reactivity profiles, ultimately resulting in the design of a new chemoselective probe (Figure 6.9).[33] During these studies, it was observed that the selective coupling of 4-carboxybenzenediazonium (4CDz) with 5-hydroxytryptophan (5-HTP) diminishes as the pH becomes more basic (>9), because it begins to label tyrosine. Based on these observations, we believe that 4CDz would be able to selectively target activated tyrosine residues in the proteome, and hopefully a new enzymatic class. We sought to further tune the electrophilicity of the aryl diazonium probe by modifying the 4-position. We believe, that it can be altered to exclusively label activated aromatic residues providing access to a new range of catalytic sites and enzyme classes for ABPP.

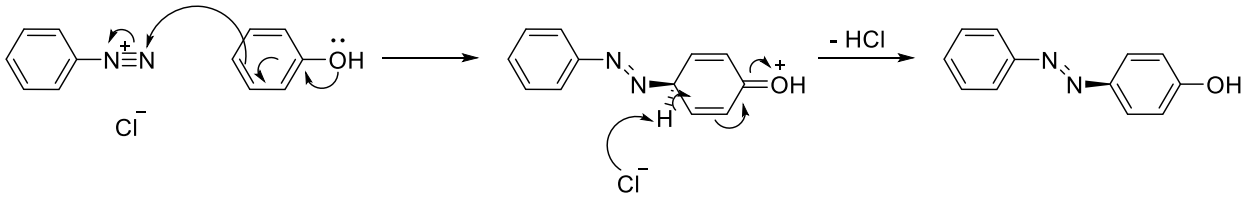


Figure 6.8. Azo-coupling mechanism between aryl diazonium ions and aromatic nucleophiles.

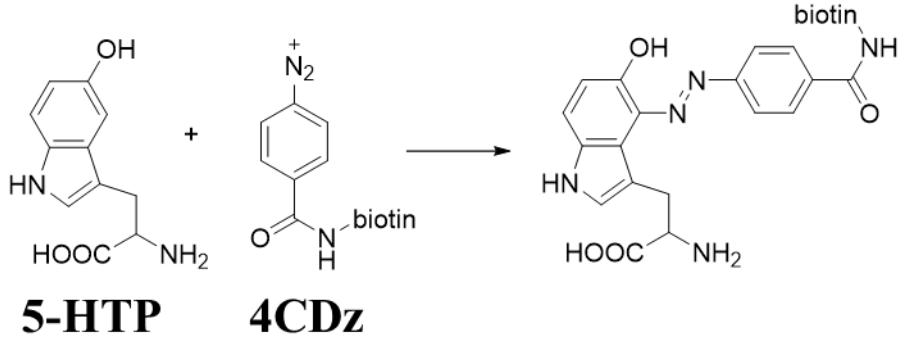


Figure 6.9. The chemoselective rapid azo-coupling reaction (CRACR) developed in the Chatterjee lab. At physiological pH, 5-HTP is the only nucleophilic aromatic amino acid capable of reacting with the 4CDz probe.

6.2.3 Synthesis of photocaged aryl diazonium probes

Aryl-diazonium ion integrity is dominated by the thermodynamic and entropically favorable release of dinitrogen gas (Figure 6.10). To mitigate the spontaneous release of dinitrogen, and loss of electrophilic handle, we generated aryl-triazabutadiene groups to protect the diazonium. Addition of the triazabutadiene group enables the selective release of the diazonium ion under light or low pH (Figure 6.11) therefore, giving us stringent control over the release of the diazonium ion.[34, 35] Protocols for the synthesis of the aryl diazonium ions characterized in this chapter are outlined below. All ^{1}H NMR and ^{13}C NMR can be found in appendix III and appendix IV of this dissertation.

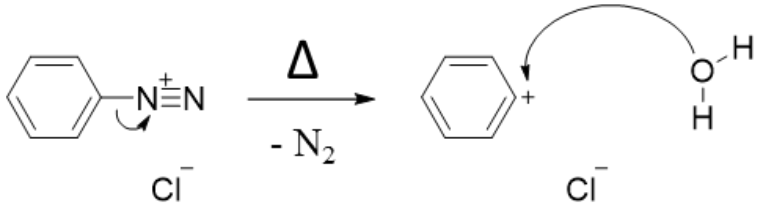


Figure 6.10. Dediazotization of aryl diazonium ions.

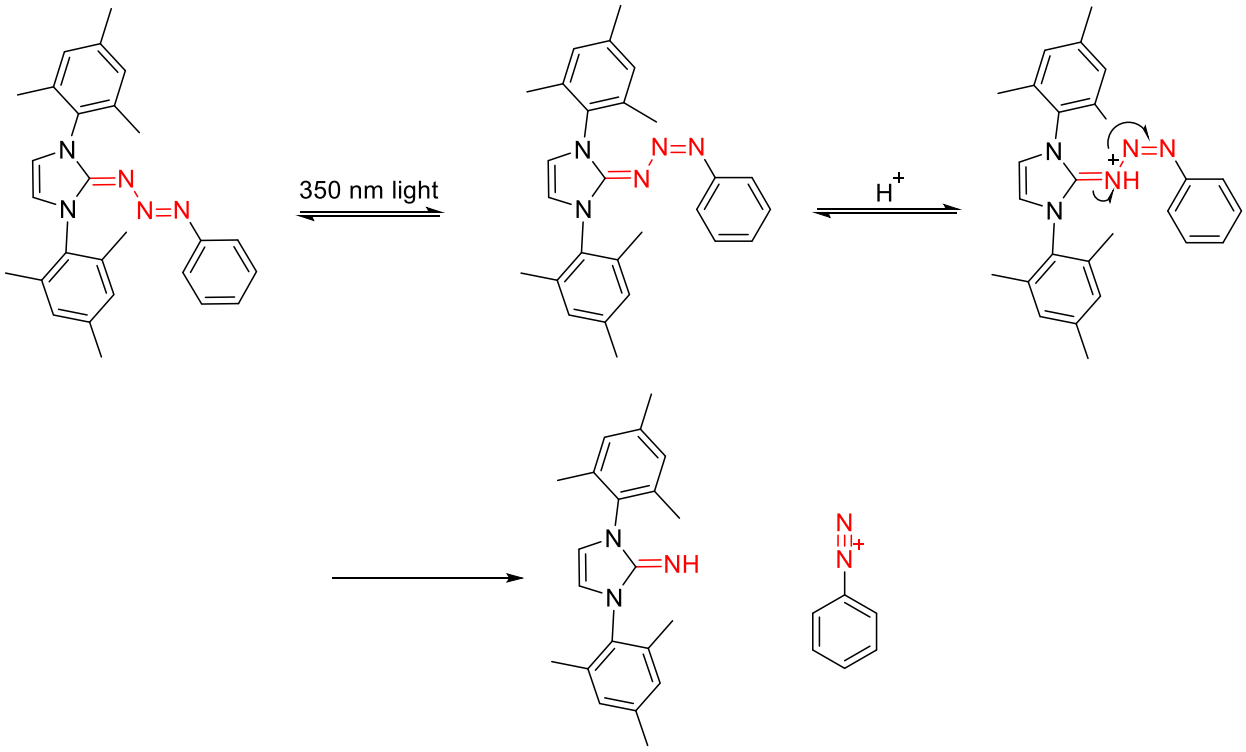
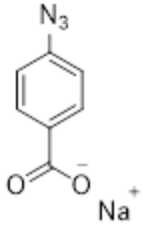
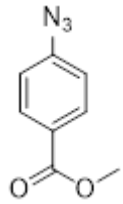


Figure 6.11. Mechanism for the photorelease of the aryl diazonium ion.[34, 35] Photoinduced E/Z isomerization increases the kinetics for the protonation of the triazabutadiene backbone (highlighted red) leading the release of the aryl diazonium ion.



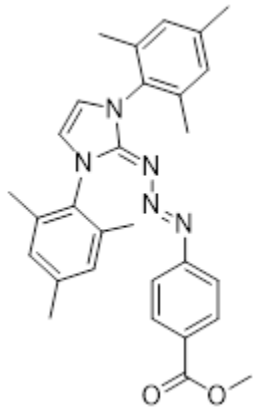
Synthesis of 4-azidobenzoic acid. **(1)**

4-aminobenzoic acid (5.016 g, 36.58 mmol, 1.000 eq) was dissolved in 25 mL of H_2O in a 500 mL round bottom flask. 5 mL of sulfuric acid was added to the solution dropwise. Then, it was left to stir until it reached 0 ºC (30 min). In a separate 125 mL Erlenmeyer flask $NaNO_2$ (2.600 g, 37.68 mmol, 1.030 eq) was dissolved in 6 mL H_2O. In another separate 125 mL Erlenmeyer flask, NaN_3 (2.900 g, 44.61 mmol, 1.224 eq) was dissolved in 4 mL of H_2O. The $NaNO_2$ aqueous solution was then added dropwise to the 4-aminobenzoic acid aqueous solution (stirring, 0 ºC). Next, the NaN_3 aqueous solution was added dropwise to the 4-aminobenzoic acid solution (stirring, 0 ºC). The white precipitate was vacuum filtered and desiccated, yielding 4-azidobenzoic acid (5.407 g, 33.14 mmol, 90.60%).



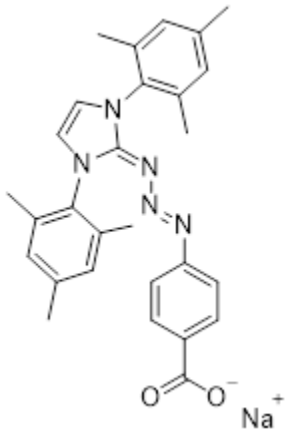
Synthesis of methyl-4-azidobenzoate. (**2**)

4-azidobenzoic acid (1.020 g, 6.129 mmol, 1.000 eq) was dissolved in 15 mL of dry DMF. Cs_2CO_3 (4.093 g, 12.26 mmol, 2.000 eq) was added to the stirring solution (0 ^oC). Iodomethane (0.8 mL, 12.26 mmol, 2.000 eq) was then added dropwise to the stirring solution (0 ^oC). The reaction was monitored by TLC until it reached completion. Then it was washed with H_2O three times followed by EtOAc. The organic layer was then dried over $MgSO_4$ and concentrated in vacuo yielding methyl 4-azidobenzoate (0.7732 g, 3.731 mmol, 60.88%).



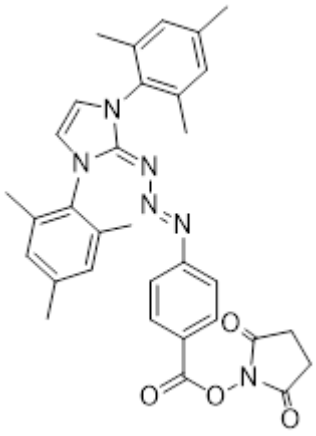
Synthesis of methyl (E)-4-((((1,3-dimesityl- 1,3-dihydro-2H-imidazol-2-ylidene)amino)methylene)amino) benzoate. (**3**)

Methyl 4-azidobenzoate (0.599 g, 3.974 mmol, 1.200 eq) was dissolved in 6 mL of dry THF and left to stir on ice under nitrogen for 20 min. Then 1,3-Bis (2,4,6-trimethylphenyl) imidazolium chloride (1.132 g, 3.311 mmol, 1.000 eq) was added to the solution and left to stir for 5 min. NaH (0.1520 g, 3.300 mmol, 1.000 eq) was then added. The reaction was left stirring overnight at 0 ^oC under nitrogen. The progress of the reaction was monitored by TLC. Once it had reached completion the solution was diluted with EtOAc, washed with H_2O and concentrated in vacuo. This yielded methyl (*E*)-4-((((1,3-dimesityl- 1,3-dihydro-2*H*-imidazol-2-ylidene)amino)methylene)amino) benzoate (1.436 g, 2.981 mmol, 75.00%). ^1H NMR ($CDCl_3$, 400 MHz) δ 7.68 (d, 2H), δ 6.98 (s, 4H), δ 6.61 (s, 2H), δ 6.55 (d, 2H), δ 3.84 (s, 3H), δ 2.35 (s, 6H), δ 2.14 (s, 12H) ppm.



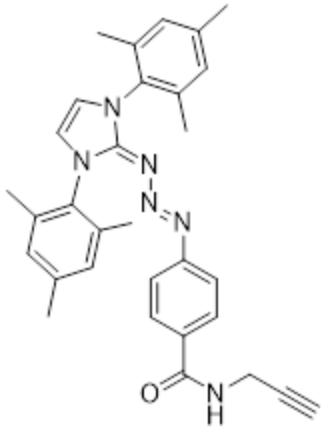
Synthesis of (E)-4-((1,3-dimesityl- 1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)benzoic acid. (**4**)

Methyl (*E*)-4-((((1,3-dimesityl- 1,3-dihydro-2*H*-imidazol-2-ylidene)amino)methylene)amino) benzoate (0.4720 g, 0.9880 mmol, 1.000 eq) was dissolved in a solution of Methanol: DCM (8 mL: 1 mL). Sodium hydroxide was added to the stirring solution (0.1610 g, 1.9601 mmol, 2.000 eq). The reaction was then heated to reflux. Once the reaction reached completion based on TLC, it was cooled to 25 ^{o}C and concentrated en vacuo to yield (E)-4-((1,3-dimesityl-1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)benzoic acid (0.3822 g, 0.8176 mmol, 82.75 %).



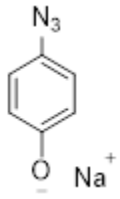
Synthesis of 2,5-dioxopyrrolidin-1-yl (E)-4(E)-4-((1,3-dimesityl- 1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)benzoate. (**5**)

(E)-4-((1,3-dimesityl- 1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)benzoic acid (0.2040 g, 0.4281 mmol, 1.000 eq) was dissolved in DCM under inert conditions at 0 ^{o}C. 1-Ethyl-3-(30dimethylaminopropyl)carbodiimde (0.1400 g, 0.6421 mmol, 1.500 eq) was added dropwise followed by N-hydroxysuccinimide (0.0840 g, 0.6421 mmol, 1.500 eq). The progress of the reaction was monitored by TLC. Upon completion the solution was diluted with EtOAc, washed with water then saturated sodium bicarbonate. It was then dried and concentrated in vacuo to yield 2,5-dioxopyrrolidin-1-yl (E)-4(E)-4-((1,3-dimesityl- 1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)benzoate (0.1567 g, 0.2775 mmol, 64.82 %). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.76 (d, 2H), δ 6.99 (s, 4H), δ 6.64 (s, 2H), δ 6.59 (d, 2H), δ 2.85 (s, 4H), δ 2.35 (s, 6H), δ 2.13 (s, 12H) ppm.



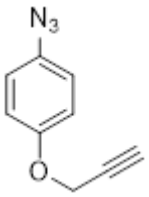
Synthesis of (E)-4-((1,3-dimesityl-1,3-dihydro- 2H -imidazol - 2-ylidene)triaz-1-en-1-yl)-N-(prop-2-yn-1-yl)benzamide. (**6**)

2,5-dioxopyrrolidin-1- yl (E)-4(E)-4-((1,3-dimesityl- 1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)benzoate (0.09050 g, 0.1603 mmol, 1 eq) was dissolved in DMF (3.000 mL, anhydrous) and stirred at 0 ºC. Triethylamine (0.2305 g, 2.2778 mmol, 14.00 eq) was added dropwise to the solution followed by prop-2-yne-1-amine (0.01400 g, 0.01230 mmol, 1.200 eq). Progress of the reaction was monitored by TLC. Once the reaction reached completion, the solution was washed with EtOAc, water, then saturated sodium bicarbonate followed by concentration in vacuo yielding (E)-4-((1,3-dimesityl-1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)-N-(prop-2-yn-1-yl)benzamide (0.06782 g, 0.1344 mmol, 83.86 %). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.43 (d, 2H), δ 6.94 (s, 4H), δ 6.58 (s, 2H), δ 6.51 (d, 2H), δ 2.26 (s, 6H), δ 2.20 (s, 1H), δ 2.11 (s, 12H) ppm.



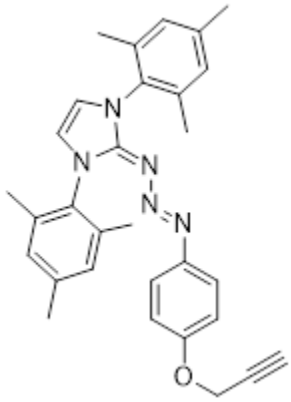
Synthesis of 4-azidophenol. (**7**)

4-aminophenol (2.045 g, 18.33 mmol, 1 eq) was dissolved in water (50 mL) at 4 ºC. Hydrochloric acid (15 mL) was added dropwise to the aqueous solution followed by $NaNO_2$ (2.532 g, 36.66 mmol, 2.000 eq). NaN_3 (1.471 g, 21.99 mmol, 1.200 eq) was added to the solution and the progress of the reaction was monitored by TLC. Following completion, the solution was washed with EtOAc and dried over Na_2SO_4 yielding 4-azidophenol (1.610 g, 11.91 mmol, 65.00%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 9.92 (s, 1H), δ 7.26 (s, 2H), δ 6.67 (s, 2H) ppm.



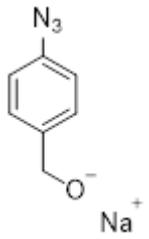
Synthesis of 1-azido-4-(prop-2-yn-1-yloxy)benzene. (**8**)

4-azidophenol (1.621 g, 11.91 mmol, 1.000 eq) was dissolved in anhydrous DMF at 4 ºC. Cs_2CO_3 (11.41 g, 35.742 mmol, 3.000 eq) was added to the solution followed by 3-bromopropyne (1.686 g, 14.29 mmol, 1.2 eq). The reaction was monitored by TLC until completion. It was subsequently quenched with H_2O, diluted with EtOAc, and washed with saturated sodium chloride. The crude product was purified by flash chromatography (4:1, Hexanes: EtOAc) and concentrated in vacuo yielding 1-azido-4-(prop-2-yn-1-yloxy)benzene (1.304 g, 7.535 mmol, 63.27%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 6.93 (s, 4H), δ 4.62 (s, 2H), δ 2.56 (s, 1H) ppm.



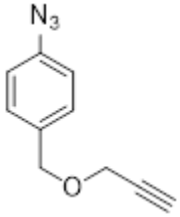
Synthesis of (E)-1,3-dimesityl-2-((4-prop-2-yn-1-yloxy)phenyl)triaz-2-en- 1-ylidene)-2,3-dihydro-1 H imidazole. (**9**)

1-azido-4-(prop-2-yn-1-yloxy)benzene (0.2000 g, 1.156 mmol, 1.200 eq) was dissolved in THF at 0 ºC under inert conditions. 1,3-bis(2,4,6-trimethylphenyl) imidazolium chloride (0.3220 g, 0.9630 mmol, 1.000 eq) was added to the solution followed by NaH (0.03850 g, 0.9630 mmol, 1.000 eq) and monitored by TLC. Purified through flash chromatography with an alumina column (DCM: MeOH, 10: 1) and concentrated in vacuo to yield (E)-1,3-dimesityl-2-((4-prop-2-yn-1-yloxy)phenyl)triaz-2-en-1-ylidene)-2,3-dihydro-1 H imidazole (0.2270 g, 0.4756 mmol, 41.14%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 6.98 (s, 4H), δ 6.63 (d, 2H), δ 6.54 (s, 2H), δ 6.51 (s, 2H), δ 4.61 (s, 2H), δ 2.49 (s, 1H), δ 2.35 (s, 6H), δ 2.16 (s, 12H) ppm.



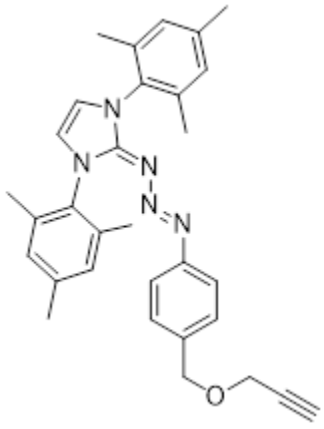
Synthesis of (4-azidophenyl)methanolate. (**10**)

4-aminobenzyl alcohol (0.5140 g, 4.060 mmol, 1.000 eq) was dissolved in H_2O: HCl (10: 1, v/v). $NaNO_2$ (0.4350 g, 6.090 mmol, 1.500 eq) was dissolved in H_2O and added dropwise to the solution. Then, NaN_3 (1.435 g, 16.24 mmol, 4.000 eq) was added quantitatively. The reaction was quenched with $NaHCO_3$, extracted with EtOAc, washed with saturated NaCl, dried over Na_2SO_4 and concentrated in vacuo to yield (4-azidophenyl)methanolate (0.4430 g, 3.006 mmol, 74.04%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.24 (d, 2H), δ 6.99 (d, 2H), δ 4.61 (s, 2H), δ 2.14 (s, 1H) ppm.



Synthesis of 1-azido-4-((prop-2-yn-1-yloxy)methyl)benzene. (**11**)

(4-azidophenyl)methanolate (0.05000 g, 0.3350 mmol, 1.000 eq) was dissolved in anhydrous DMF. 3-bromopropyne (0.04700 g, 0.4020 mmol, 1.200 eq) was added to the solution dropwise followed by NaH (0.0120 g, 0.5025 mmol, 1.500 eq). Progress of the reaction was monitored by TLC. Upon completion, the solution was washed with H_2O, dried over Na_2SO_4, and concentrated in vacuo to yield 1-azido-4-((prop-2-yn-1-yloxy)methyl) benzene (0.039 g, 0.2250 mmol, 67.16%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.34 (d, 2H), δ 7.00 (d, 2H), δ 4.56 (s, 2H), δ 4.15 (s, 2H), δ 2.46 (s, 1H) ppm.



Synthesis of (E)-1,3-dimesityl-2((4-((prop-2 yn-1yloxy) methyl)phenyl)triaz-2-en-1-ylidene)-2,3-dihydro-1 H-imidazole. (**12**)

1-azido-4-((prop-2-yn-1-yloxy)methyl) benzene (0,01000 g, 0.05781 mmol, 1.000 eq) was dissolved in anhydrous THF and stirred at 0 ºC. 1,3-bis(2,4,6-trimethylphenyl) imidazolium chloride (0.01900 g,

0.05781 mmol, 1.000 eq) was added to the solution followed by NaH (0.002000 g, 0.05781 mmol, 1.000 eq). Progress of the reaction was monitored by TLC. Upon completion, the solution was washed with H_2O, saturated NaCl, dried over Na_2SO_4 and concentrated in vacuo to yield (E)-1,3-dimesityl-2((4-((prop-2 yn-1yloxy) methyl)phenyl)triaz-2-en-1-ylidene)-2,3-dihydro-1 H-imidazole (0.0073 g, 0.01487 mmol, 25.73%). ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.24 (d, 2H), δ 7.01 (s, 4H), δ 6.58 (s, 2H), δ 6.55 (d, 2H), δ 4.48 (s, 2H), δ 4.09 (s, 2H), δ 2.42 (s, 1H), δ 2.34 (s, 6H), δ 2.14 (s, 12H) ppm.

6.2.4 HEK proteome labeling profiles of the aryl diazonium triazabutadiene probes

To better understand which aryl diazonium triazabutadiene probe would be the best candidate for ABPP, their labeling profiles of the HEK293T cell proteome were characterized relative to an established cysteine-selective ABPP probe, iodoacetamide alkyne.[36] We wanted to identify the probe that would demonstrate selective labeling of a different part of the proteome than iodoacetamide alkyne at a similar intensity, which could be roughly visualized through the banding patterns present on an SDS-PAGE gel.

To generate HEK293T cell lysate, the cells were first scraped from their cell culture dishes, spun down (4,500 rpm, 7 min, 4 ºC) and re-suspended in a CelLytic M solution (10 mL CelLytic M: 100 µL protease inhibitor: 1 µL universal nuclease, v/v). The aryl diazonium triazabutadiene probes (dissolved in MeOH) were subsequently added to the HEK293T cell lysate (2 mg/mL) to final concentrations of 20 µM and 50 µM, followed by UV light (365 nm, 10 sec) irradiation. The iodoacetamide alkyne was dissolved in DMSO and added to the HEK293T cell lysate (final concentration 10 µM). The samples were left to incubate for one hour at 25 ºC. Following labeling of the lysate with each probe, a rhodium-azide fluorescent marker was clicked to the alkyne handle

of each probe with a previously optimized copper-click protocol.[37] The labeling patterns of each probe were observed through SDS-PAGE fluorescent gel imaging (Figure 6.13). Notably, the 4CDz-alkyne at 50 μM demonstrated similar labeling intensity of the proteome relative to the iodoacetamide alkyne at 10 μM. Based on this observation, we believe the 4CDz-alkyne was the best option for proteomic studies moving forward.

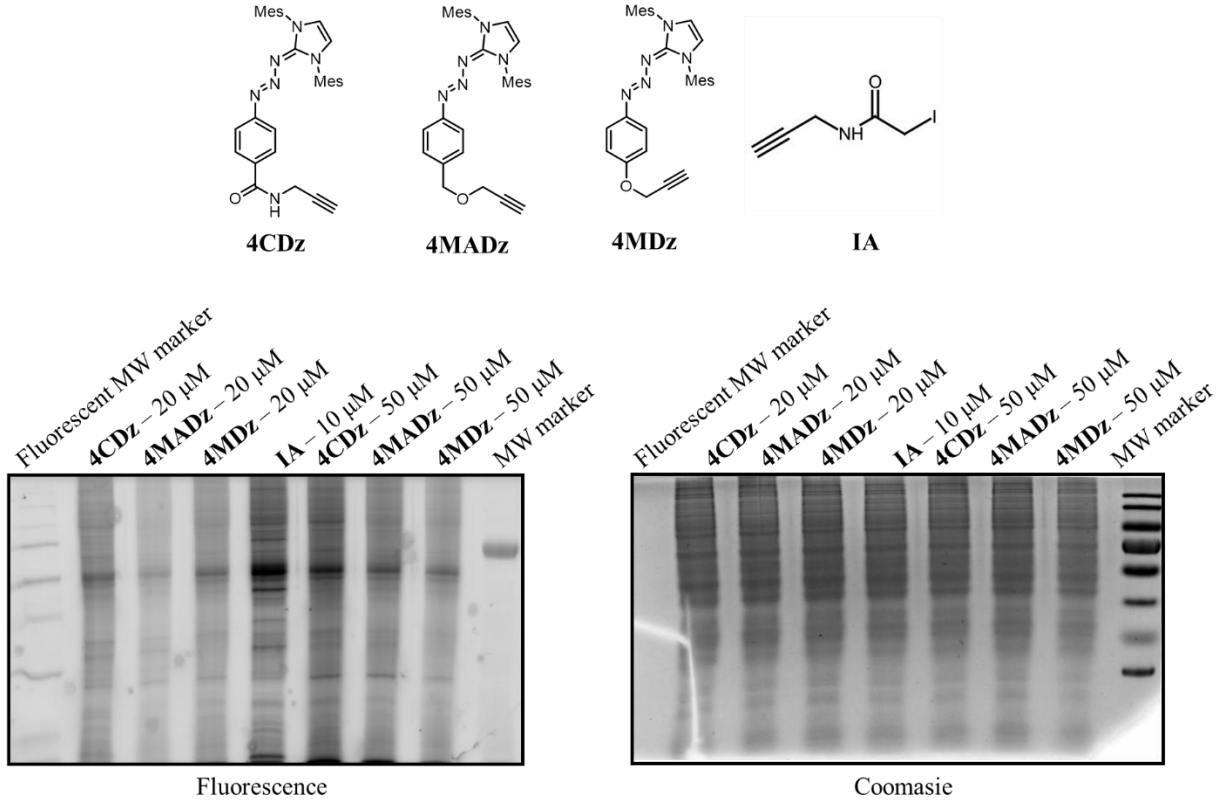


Figure 6.12. SDS-PAGE analysis of the HEK293T proteome. Reactivity profiles of the 4-carboxydiazonium (4CDz), 4-methylalcoholdiazonium (4MADz), and 4-methoxydiazonium (4MDz) were compared relative to the established iodoacetamide alkyne (IA). Based on fluorescence imaging of the SDS-PAGE gel (left) the 4CDz at 50 μM showed similar labeling intensity to the standard control of IA at 10 μM.

6.2.5 Labeling patterns of 4CDz-alkyne relative to 4CDz-biotin

Prior to carrying out the proteomic studies, we sought to further validate that the addition of the alkyne handle to 4CDz did not alter the reactivity profile of this probe relative to its previously reported biotinylated analog.[33] This was accomplished by attempting to selectively label sfGFP-151-5HTP in HEK lysate with both the 4CDz-alkyne and 4CDz-biotin. Purified sfGFP-151-HTP (6 μM final concentration) was spiked into HEK293T cell lysate (2mg/mL) and subsequently labeled with both 4CDz probes (dissolved in MeOH, 100 μM final concentration) for one hour at 4 °C. The reaction was quenched with excess 5-HTP followed by copper click of biotin-azide to the 4CDz-alkyne samples.[37] The samples were then analyzed based on SDS-PAGE gel western blot analysis (Figure 6.13). Upon examination of the gel, it was obvious that the 4CDz-alkyne was not behaving similarly to the 4CDz-biotin. With the 4CDz-alkyne we were unable to observe the sfGFP-151-5HTP band in the lysate + sfGFP reaction.

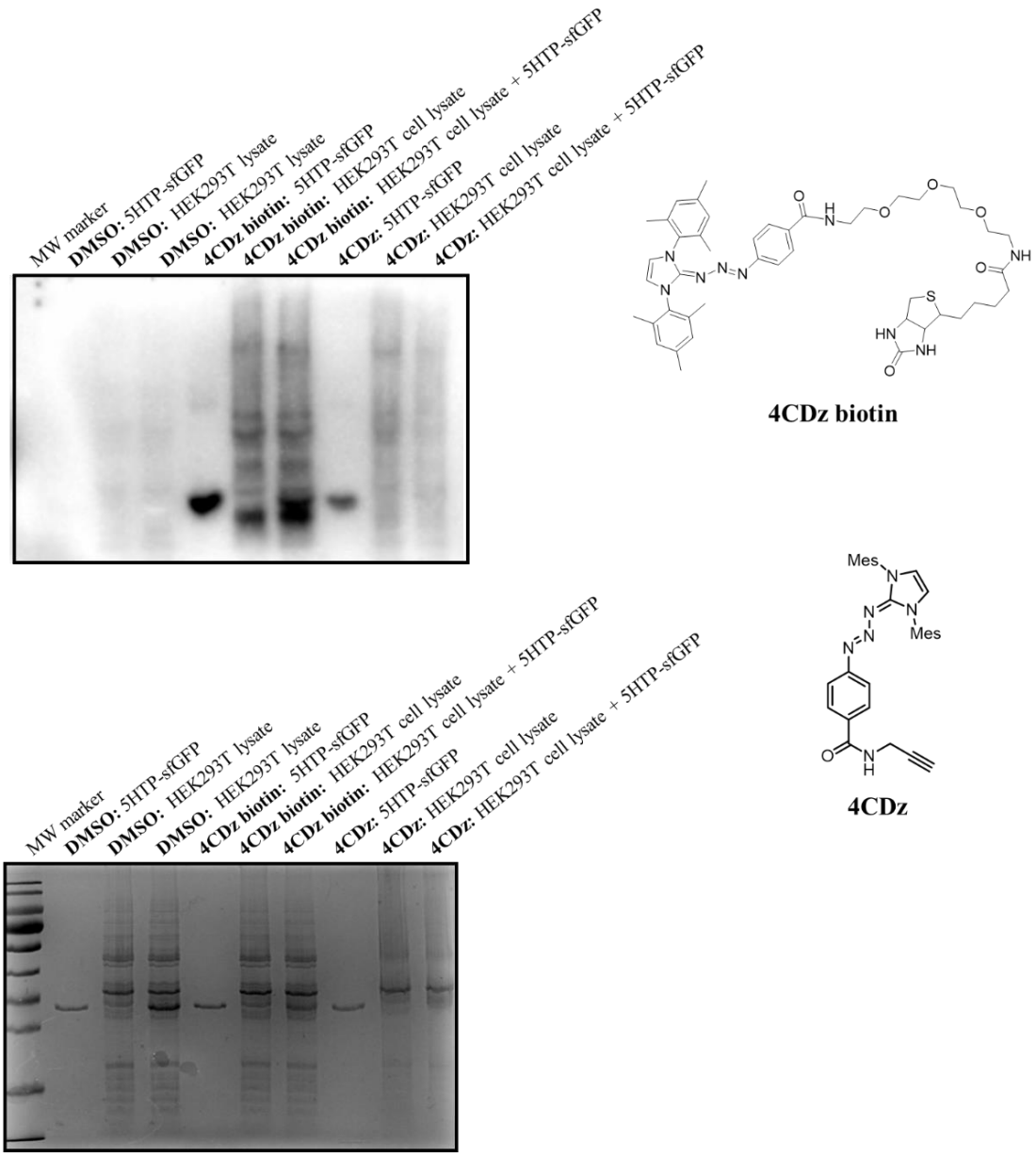
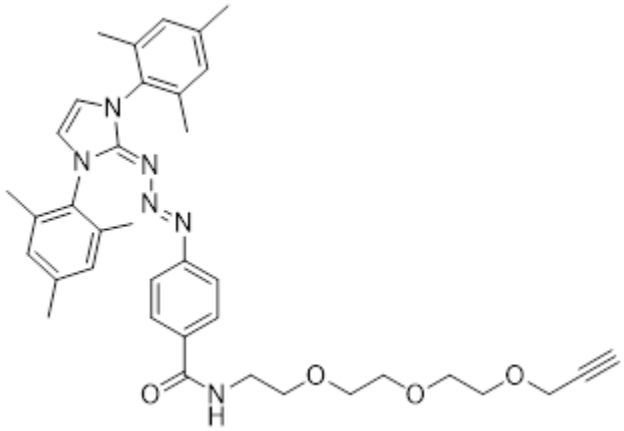


Figure 6.13. SDS-PAGE analysis of 4CDz-alkyne relative to 4CDz-biotin. Based on the western blot analysis of the SDS-PAGE gel (top) the 4CDz probe was unable to selectively label 5HTP-sfGFP in the lysate mixture. This contrasted with the previously characterized reactivity profile of the 4CDz biotin probe (top).

The electrophilicity of the two diazonium probes should be comparable due to their fundamental structural similarities. The only observable structural difference between the two diazonium probes is that the 4CDz-alkyne is more hydrophobic relative to 4CDz biotin, which has a PEG3-linker to help solubilize it in aqueous solutions. We believed that the addition of a PEG3

linker between the 4-carboxy group and alkyne handle would remedy the differences in reactivity. Synthesis of the new probe is described in the following paragraph.



Synthesis of (E)-4-((1, 3-dimesityl- 1,3-dihydro- 2H-imidazol- 2- ylidene)triaz- 1-en- 1-yl)- N-(2-(2-(2-(prop-2-yn- 1-yloxy)ethoxy)ethoxy)ethyl)benzamide. (**13**)

2,5-dioxopyrrolidin-1-yl (E)-4(E)-4-((1,3-dimesityl- 1,3-dihydro-2H-imidazol-2-ylidene)triaz-1-en-1-yl)benzoate (0.05000 g, 0.08850 mmol, 1.000 eq) was dissolved in anhydrous DMF at 4 ºC. Triethylamine (0.1250 g, 1.239 mmol, 14.00 eq) was added dropwise to the solution followed by propargyl-PEG3-amine (0.02000 g, 0.1063 mmol, 1.200 eq). Progress of the reaction was monitored by TLC. Once the reaction reached completion, the solution was washed with EtOAc, water, then saturated sodium bicarbonate followed by concentration in vacuo yielding ((E)-4-((1, 3-dimesityl- 1,3-dihydro- 2H-imidazol- 2- ylidene)triaz- 1-en- 1-yl)- N-(2-(2-(2-(prop-2-yn- 1-yloxy)ethoxy)ethoxy)ethyl)benzamide (0.4610 g, 0.07262 mmol, 82.06 %).^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.44 (d, 2H), δ 6.97 (s, 4H), δ 6.59 (s, 2H), δ 6.54 (d, 2H), δ 4.11 (d, 2H), δ 3.64 (dd, 12H), δ 2.37 (s, 6H), δ 2.13 (s, 12H) ppm.

With the new 4CDz-PEG3-alkyne probe in hand, we repeated the HEK293T/sfGFP-151-5HTP labeling experiment under the previously used conditions. SDS-PAGE and western blot analysis showed that the addition of the PEG3 linker restored the expected reactivity profile of our

4CDz-alkyne (Figure 6.14). This restored our confidence in the ability to use the 4CDz-PEG3-alkyne for proteomic experiments.

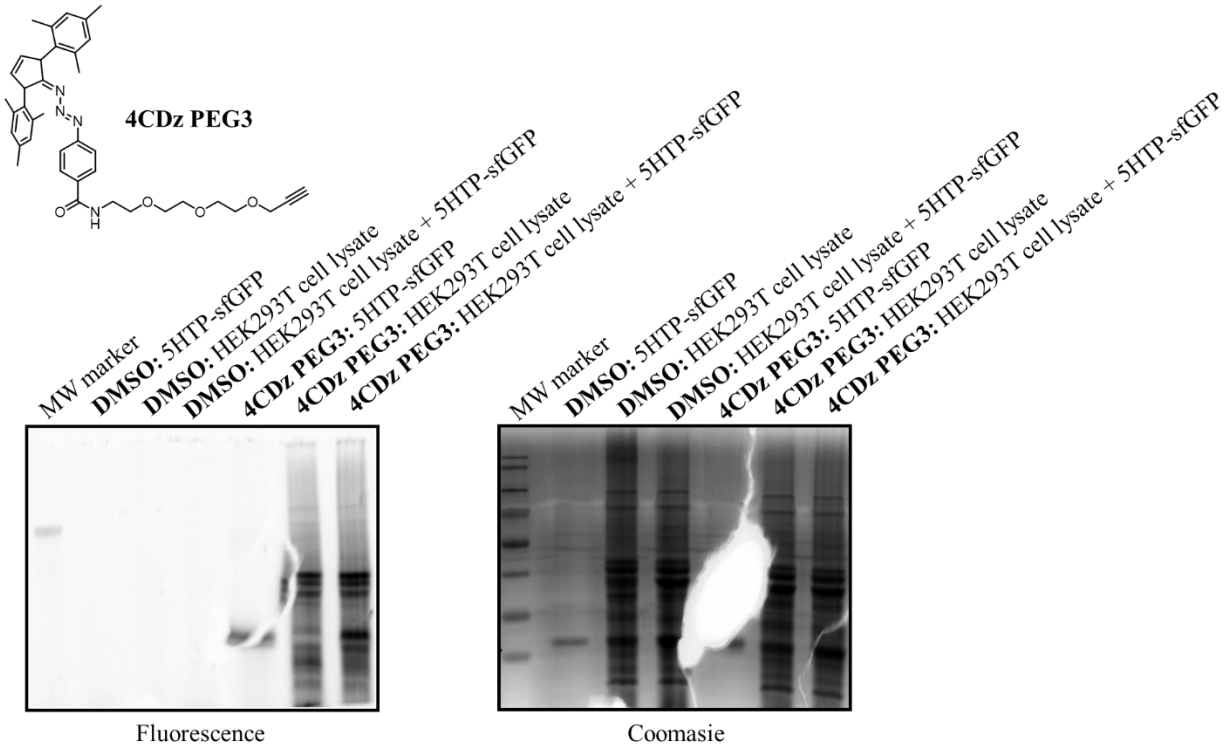


Figure 6.14. SDS-PAGE gel analysis of the 4CDz-PEG3-alkyne probe. The new 4CDz-PEG3-alkyne probe (100 μM) was incubated with 5HTP-sfGFP, HEK293T cell lysate, and HEK293T cell lysate + 5HTP-sfGFP. Addition of the PEG3 linker restored selective labeling of 5HTP-sfGFP with this aryl diazonium probe (far right lane in fluorescence image).

6.2.6 Selective labeling of tyrosine residues in a synthetic peptide with 4CDz-PEG3 alkyne

Prior to HEK293T cell proteome analysis, we sought to identify the distinctive MS1 and MS2 labeling mass patterns of our 4CDz-PEG3-alkyne conjugated to tyrosine residues. This would make downstream proteomic analysis easier since we would have a specific numerical value to search for. To set this standard, we obtained a synthetic peptide (75 μM) (Figure 6.15) from the Gao lab and labeled it with 4CDz-alkyne (150 μM) for 4 hours at 25 °C in 1X PBS (pH 9). The labeled peptide was characterized by MS analysis (Figure 6.16), where we observed that the

reaction had not gone to completion. This labeled sample was further characterized by tandem mass spectrometry (Figure 6.17) where we observed that both of the tyrosine resides in the peptide were being labeled. Moving forward, to better understand the MS2 labeling patterns the labeled peptide needs to be purified by HPLC, or the reaction conditions need to be optimized to observe 100% labeling efficiency.

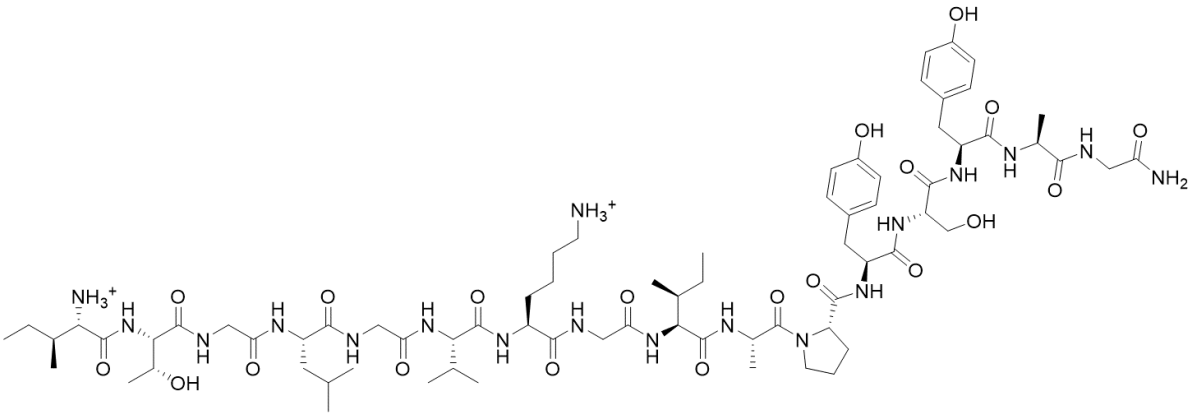


Figure 6.15. Synthetic peptide structure.

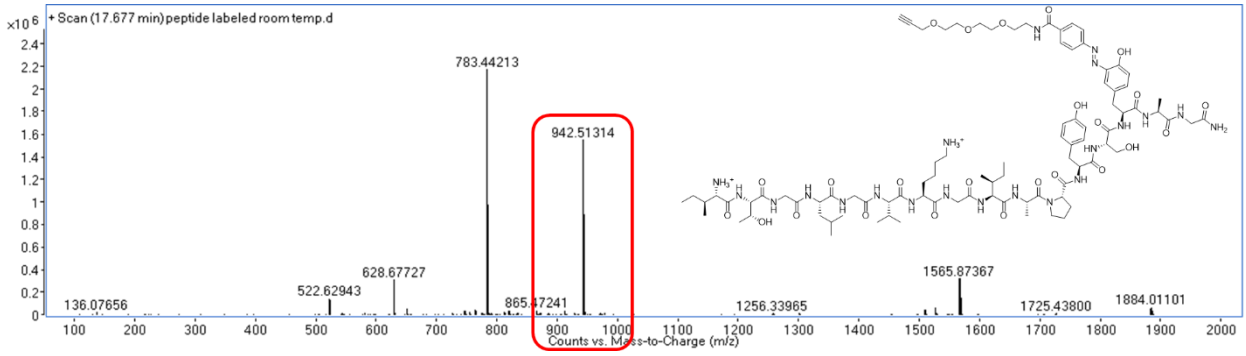


Figure 6.16. MS analysis of 4CDz-PEG3-alkyne labeled peptide.

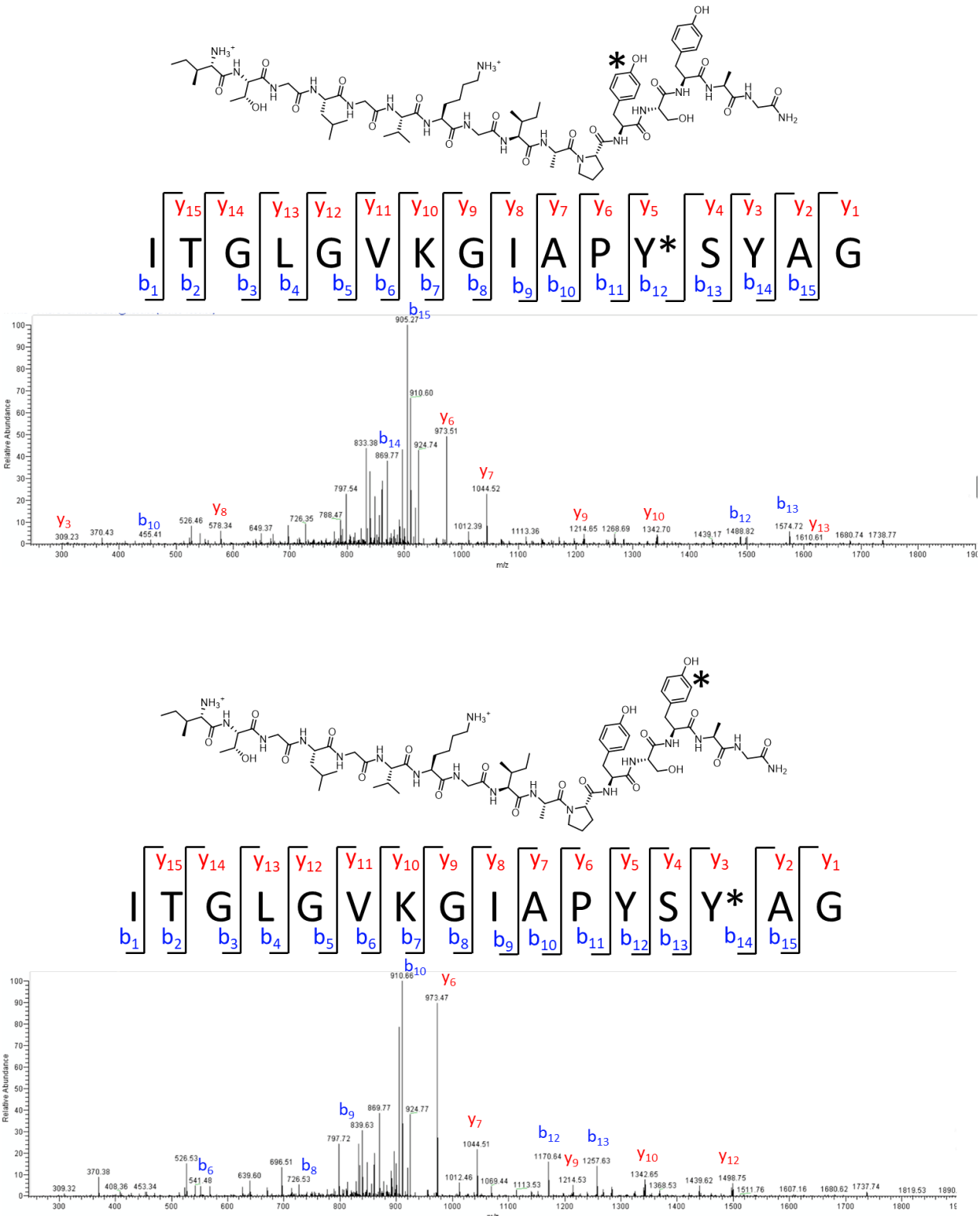


Figure 6.17. MS2 analysis of the 4CDz-PEG3-alkyne labeled peptide.

6.2.7 Conclusions

In summary, we created an aryl diazonium probe that can be used for ABPP proteomic analysis. We demonstrated that the electrophilicity of the probe can be easily tuned through modification of the 4- position further moderating the biomolecular space accessed by the probe. The current limitation to this technology is identifying the MS2 peaks. This issue can be rectified by further optimization of the MS2 ionization energies.

Ongoing and future directions

Further functionalization of the 4CDz structure provides promising avenues to alter its selectivity for analyzing different biochemical groups in the proteome. Two directions that have been discussed for future work would be to (i) introduce steric bulk at the ortho position of the diazonium and (ii) electronically tuning the reactivity of the probe by decorating either the protecting group mesitylene rings, or the core benzyl ring, with fluorine.

6.2.8 Acknowledgements

Dr. Partha Sarathi Addy was mentor and guided me through the synthesis of all of these probes. Dr. Kyle Cole and Dr. Jennifer Peeler were my collaborators in the Weerapana lab. They helped carry out the tandem mass spectrometry sample prep and analysis. Dr. Kaicheng Li supplied me with the synthetic peptide used in these experiments.

6.3 References

1. Italia, J.S., Addy, P.S., Wrobel, C.J.J., Crawford, L.A., Lajoie, M.J., Zheng, Y., Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. Nat Chem Biol. 13, 446 (2017).
2. Italia, J.S., Latour, C., Wrobel, C.J.J., Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. Cell Chemical Biology 25(10), 1304 (2018).
3. Chin, J.W., Cropp, T.A., Anderson, C., Mukherji, M., Schultz, P.G. An expanded eukaryotic genetic code. Science 301, 964 (2003).
4. Dumas, A., Lercher, L., Spicer, C.D., Davis, B.G. Designing logical codon reassignment - Expanding the chemistry in biology. Chem Sci. 6(1), 50 (2015).
5. Henkin, T.M., Glass, B.L., Grundy, F.J. Analysis of the *Bacillus subtilis tyrS* Gene: conservation of a Regulatory Seqeunce in Multiple tRNA Synthetase Genes. J. Bacteriol. 174 (4), 1299 (1992).
6. Li, S., Su, Z., Lehmann, J., Stamatopoulou, V., Giarimoglou, N., Henderson, F.E., Fan, L., Pintilie, G.D., Zhang, K., Chen, M., Ludtke, S.J., Wang, Y., Stathopoulos, C., Chiu, W., Zhang, J. Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. Nat Struct Mol Biol 26, 1094 (2019).
7. Zhang, J., Suddala, K.C. High-affinity recognition of specific tRNAs by an mRNA anticodon-binding groove. Nat Struct Mol Biol 26, 1114 (2019).
8. Weaver, J.W., Serganov, A. T-box RNA gets boxed. Nat Struct Mol Biol 26, 1081 (2019).
9. Grasso, K.T., Yeo, M.J.R., Hillenbrand, C.M., Ficaretta, E.D., Italia, J.S., Huang, R.L., Chatterjee, A. Structural robustness affects the engineerability of aminoacyl-tRNA synthetases for genetic code expansion. Biochemistry 60 (7), 489 (2021).
10. Kreuzer, K.D., Henkin, T.M. The T box riboswitch: tRNA as an effector to modulate gene regulaton. Microbiol Spectr. 6 (4), (2018).
11. Grundy, F.J., Henkin, T.M. tRNA as a positive regulator of transcription antitermination in B. subtilis. Cell 74 (3), 475 (1993).
12. Henkin, T.M. The T box riboswitch: A novel regulatory RNA that utilizes tRNA as its ligand. Biochimica et Biophysica Acta 1839 (10), 959 (2014).
13. Green, N. J., Grundy, F. J., Henkin, T. M. The T box mechanism: tRNA as a regulatory molecule. FEBS letters, 584(2), 318 (2010).
14. Cravatt, B.F., Wright, A.T., Kozarich, J.W. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. Annu. Rev. Biochem. 77, 383 (2008).
15. Hanahan, D., Weinberg, R.A. The Hallmarks of Cancer. Cell 100, 57 (2000).
16. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 10 (8), 789 (2004).
17. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, Cordón-Cardo C, Guise TA, Massagué J. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell. 3(6), 537 (2003).
18. Mootha, V.K., Lindgren, C.M., Eriksson, K., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., Houstis, N., Daly, M.J., Patterson, N., Mesirov, J.P., Golub, T.R., Tamayo, P., Spiegelman, B., Lander, E.S., Hirschhorn, J.N., Altshuler, D., Groop, L.C. PGC-1α-responsive genes involved in

oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 34, 267 (2003).
19. Proteomics background: Proteomics Center: ERASMUS MC. (2019, December 04). Retrieved March 08, 2021, from https://www.proteomicscenter.nl/background-information/.
20. Kobe, B., Kemp, B. Active site-directed protein regulation. Nature 402, 373 (1999).
21. Khan, A.R., James, M.N. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Protein Sci. 7 (4), 815 (1998).
22. Jessani, N., Humphrey, M., McDonald, W.H., Niessen, S., Masuda, K., Gangadharan, B., Yates, J.R. 3rd, Mueller, B.M., Cravatt, B.F. Carcinoma and stromal enzyme activity profiles associated with breast tumor growth in vivo. Proc Natl Acad Sci U S A. 101 (38), 13756 (2004).
23. Greenbaum, D.C., Baruch, A., Grainger, M., Bozdech, Z., Medzihradszky, K.F., Engel, J., DeRisi, J., Holder, A.A., Bogyo, M. A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. Science 298 (5600), 2002 (2002).
24. Shannon, D.A., Weerapana, E. Covalent protein modification: the current landscape of residue-specific electrophiles. Curr Opin Chem Biol. 24, 18 (2015).
25. Pauly, H. Hoppe-Seyler's Z. physiol. Chem. 94, 284 (1915).
26. Tracey, B.M., Shuker, D.E.G. Characterization of Azo Coupling Adducts of Benzenediazonium Ions with Aromatic Amino Acids in Peptides and Proteins. Chem. Res. Toxicol. 10 (12), 1378 (1997).
27. Hooker, J.M., Kovacs, E.W., Francis, M.B. Interior surface modification of bacteriophage MS2. J Am Chem Soc. 126, 3718 (2004).
28. Schlick, T.L., Ding, Z., Kovas, E.W., Francis, M.B. Dual-surface modification of the tobacco mosaic virus. J Am Chem Soc. 127, 3718 (2005).
29. Jensen, S.M., Kimani, F.W., Jewett, J.C. Light-Activated Triazabutadienes for the Modification of a Viral Suface. ChemBioChem 17, 2216 (2016).
30. Gavrilyuk, J., Ban, H., Nagano, M., Hakamata, W, Barbas III, C.F. Formylbenzene Diazonium Hexafluorophosphate Reagent for Tyrosine0Selective Modification of Proteins and the Introduction of a Bioorthognal Aldehyde. Bioconj Chem 23, 2321 (2012).
31. Jones, M.W., Mantovani, G., Blindauer, C.A., Ryan, S.M., Wang, X., Brayden, D.J., Haddleton, D.M. Direct peptide bioconjugation/PEGylation at tyrosine with linear and branched polymeric diazonium salts. J Am Chem Soc 134 (17), 7406 (2012).
32. Verhelst, S.H., Fonović, M., Bogyo, M. A mild chemically cleavable linker system for functional proteomic applications. Angew Chem Int Ed Engl 46 (8), 1284 (2007).
33. Addy, P.S., Erickson, S.B., Italia, J.S., Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. J Am Chem Soc 139, 11670 (2017).
34. Kimani, F. W., Jewett, J. C. Water-soluble triazabutadienes that release diazonium species upon protonation under physiologically relevant conditions. Angew. Chem. Int. Ed. 54, 4051 (2015).
35. He, J., Kimani, F. W., Jewett, J. C. A Photobasic Functional Group. J Am Chem Soc 137, 9764 (2015).
36. Abo, M., Li, C., Weerapana, E. Isotopically-Labeled Iodoacetamide-Alkyne Probes for Quantitative Cysteine-Reactivity Profiling. Mol. Pharmaceutics 15 (3), 743 (2018).

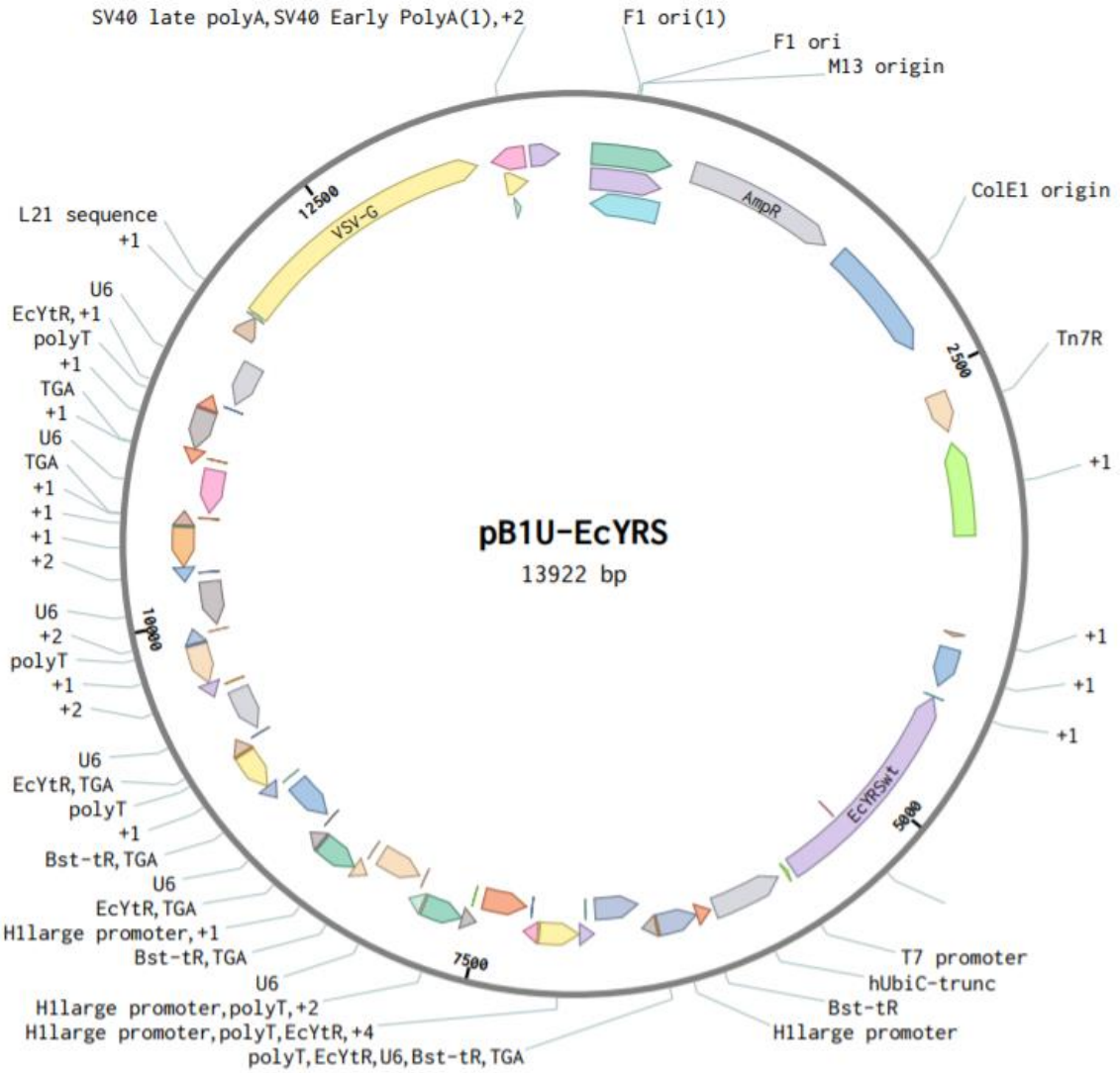
37. Weerapana, E., Speers, A.E., Cravatt, B.F. Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)—a general method for mapping sites of probe modification in proteomes. Nat. Protoc 2 (6), 1414 (2007).

Appendix I.
Plasmid maps and sequences

List of plasmid sequences

Chapter 2 Plasmids

pB1U-EcYRS variant



Sequence color-coding key

feature	Color
VSV-G	text
M13 ori	text
AmpR	text
ColE1 ori	text
Tn7R	text
EcYRS wt	text
EcYtRNA	text
Bst-tRNA	text
H1large promoter	text
U6 promoter	text
T7 promoter	text
hUbiC-trunc	text
Gentamicin resistance	text
BgH PolYA	text

Ec-pBPARS Active site mutations

residue	Mutation
Y37	G
D182	G
L186	A

Sequence

tctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaa
tatcaacgcttatttgcagcctgaatggcgaatgggacgcgccctgtagcggcgcattaagcgc
ggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcct
ttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggg
ggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattaggg

tgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtcc
acgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctatt
cttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaaca
aaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcgggga
aatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatga
gacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacattt
ccgtgtcgcccttattccctttttgcggcattttgccttcctgttttgctcacccagaaacg
ctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatc
tcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttt
taaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgc
cgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacgg
atggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaa
cttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggat
catgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtg
acaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttac
tctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctg
cgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctc
gcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgac
ggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgatt
aagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatt
tttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacg
tgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcct
ttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtt
tgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatacc
aaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcct
acatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtctta
ccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttc
gtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcat
tgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcg
gaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgg
gtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatgg
aaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt
tctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgatac
cgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctg
atgcggtattttctccttacgcatctgtgcggtatttcacaccgcagaccagccgcgtaacctg
gcaaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgtgggcggacaata
aagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactagacagaatag
ttgtaaactgaaatcagtccagttatgctgtgaaaaagcatactggacttttgttatggctaaa
gcaaactcttcattttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaagggc
atggtaaagactatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccg
atctcggcttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttctt
cccgtatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcacgta
gatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtggcaatg
ccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcggctgct

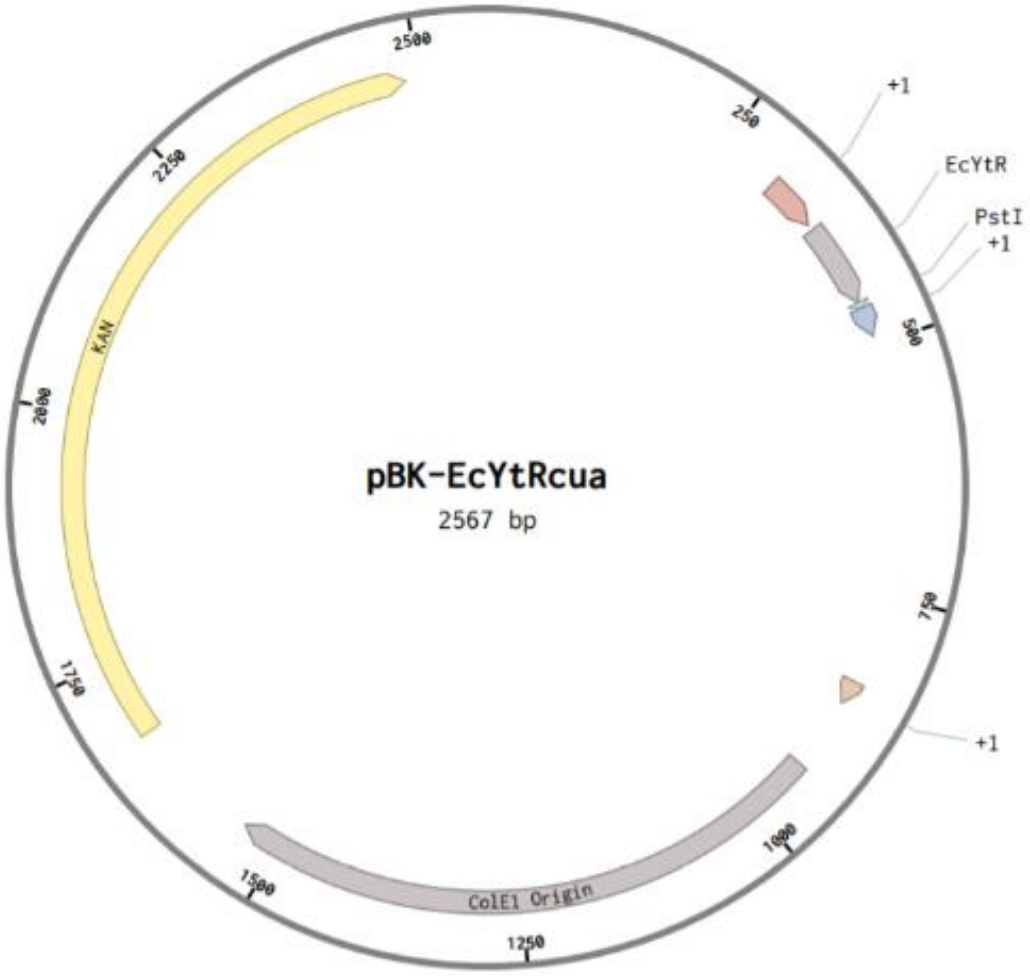
caaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttggtcgaaggcagcaag
cgcgatgaatgtcttactacggagcaagttcccgaggtaatcggagtccggctgatgttgggag
taggtggctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcatggatttgactt
ggtcagggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagg
gcgactgccctgctgcgtaacatcgttgctgctgcgtaacatcgttgctgctccataacatcaa
acatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatagactgtacaaaaaaa
cagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttc
tggaccagttgcgtgagcgcatacgctacttgcattacagtttacgaaccgaacaggcttatgt
caactgggttcgtgccttcatccgtttccacggtgtgcgtcacccggcaaccttgggcagcagc
gaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtctccacgcatcgtc
aggcattggcggccttgctgttcttctacggcaaggtgctgtgcacggatctgccctggcttca
ggagatcggtagacctcggccgtcgcggcgcttgccggtggtgctgaccccggatgaagtggtt
cgcatcctcggttttctggaaggcgagcatcgtttgttcgcccaggactctagctatagttcta
gtggttggctacgtacccgtagtggctatggcagggcttgcgcttaatgcgccgctacagggcg
cgtggggataccccctagagccccagctggttctttccgcctcagaagccatagagcccaccgc
atccccagcatgcctgctattgtcttcccaatcctccccttgctgtcctgccccaccccaccc
cccagaatagaatgacacctactcagacaatgcgatgcaatttcctcattttattaggaaagga
cagtgggagtggcaccttccagggtcaaggaaggcacgggggaggggcaaacaacagatggctg
gcaactagaaggcacagtcgaggctgatcagcgggtttaaacgggccctctagactcgagttaa
agtcgacttaacgcgttgaattcttaTTTCCAGCAAATCAGACAGTAATTCTTTTTACCGCGAC
GCAGTAAGGTAAAACGACCAAACAGACGATCTTCTTCTTTAAAGAAGTATTCAGGATCGGACTG
TTTTTCACCGTTAATGGTGATGGCATTGGAGGCGATAGTTTTACGTGCCTGACCACGGGAAGGT
TGCAGTTCAGAATCGACCAGTGCCTGCATCAGGTCTGCGCCCTTTTCCATCTCAACCATCGGTA
CGCCGTCCTGCGCCAGCTGTTCGAAGTCCGCTTCACTCAGCGCACTCAAAGAACCGCTGAACAG
GCATTCGGTAATACGTTTTGCCGCCTGTAAACCTTCTTCACCGTGAACCAGACGAGTCACCTGC
TCCGCCAGTACATACTGGGCGCGCGGTGCTTTACCGCTGTTTTTATCTTCTTCTTCCAGGGCGT
TGATCTCTTCAATGCTCATAAAGGTGAAGAACTTCAGGAAGCGGTAAACGTCGGCATCCGCAGT
GTTGATCCAGAACTGGTAGAATTTGTACGGGCTGGTTTTCTTCGGATCtAACCAGACTGCGCCG
CCTTCAGTTTTACCAAATTTGGTGCCATCTGCTTTAGTGATCAGCGGAACGGTCAGGCCAAACA
CCTGATTCTGATGCAGACGACGGGTCAGGTCGATACCAGAAGTGATGTTACCCCACTGGTCAGA
ACCACCAATTTGCAGCACCACACCGTACTGTTTGTTCAGACAGGCGAAGTCATAACCCTGCAAC
AGGTTGTAGGAAAACTCAGTGAACGAAATCCCCTGATCTTCACGGTTGAGACGCTGCTTAACCG
CTTCTTTGTTGATCATCTGGTTAACGGAGAAGTGTTTGCCAATATCGCGCAGGAAGGTCAGCAC
ATTCATATTGCCGAACCAGTCATAGTTGTTCGCCGCGATAGCAGAGTTTTCTCCACAGTCGAAA
TCGAGGAACGGGGCAACCTGCTTACGGATTTTGTCCACCCACTCCTGAACAGTTTCTTCGGTGT
TCAGCTTACGCTCGGCAGCTTTGAAGCTCGGGTCGCCAATCAGACCCGTCGCGCCGCCTACCAG
CGCAACCGGCTTGTGGCCCGCCTGCTGGAAGCGTTTCAGGCATAACAATGGAACAAGATGCCCC
AAATGCAAGCTGTCAGCGGTAGGATCGAAGCCGCAATAGAGCGCGATCGGGCCTTGCGCCAGTC
GCTCTGCTAACGCTTCCTCGTCCGTCACCTGGGCTACCAGCCCCCGCTCTTGCAATTGTTTAAT
CAAGTTACTGCTTGCcatGGTGGCgctagccagcttgggtctccctatagtgagtcgtattaat
ttcgataagccagtaagcagtgggttctctagttagccagagagctCtagaccaagtgacgatc
acagcgatccacaaacaagaaccgcgacccaaatcccggctgcgacggaactagctgtgccaca
cccggcgcgtccttatataatcatcggcgttcaccgccccacggagatccctccgcagaatcgc
cgagaagggactacttttcctcgcctgttccgctctctggaaagaaaaccagtgccctagagtc

acccaagtcccgtcctaaaatgtccttctgctgatactggggttctaaggccgagtcttatgag
cagcgggccgctgtcctgagcgtccgggcggaaggatcaggacgctcgctgcgcccttcgtctg
acgtggcagcgctcgccgtgaggaggggggcgcccgcgggaggcgccaaaacccggcgcggagg
ccttcgaacggCCActagcaaaaaaTGGAgGGGGACGGATTCGAACCGCCGAACCCAAAGGGAG
CGGATTtgaAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCgggGATCTGTGGTCTCATACA
GAACTTATAAGATTCCCAAATCCAAAGACATTTCACGTTTATGGTGATTTCCCAGAACACATAG
CGACATGCAAATATTGCAGGGCGCCACTCCCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGC
ACGCGCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGAC
GTCAGCGTTCGAATTCCTAGCAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGAC
GGCAGATTTgaAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCAC
AAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTT
TAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTG
TACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGT
ATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGcaaaaaaTGG
AgGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTtgaAGTCCGCCGCGTTTAGCCA
CTTCGCTACCCCTCCgggGATCTGTGGTCTCATACAGAACTTATAAGATTCCCAAATCCAAAGA
CATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTGCAGGGCGCCACT
CCCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGA
CACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCCTAGCAAAAAAT
GGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTgaAGTCTGCTCCCTTTGGC
CGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAAT
ACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTA
CCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAA
TTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAA
ATAGGCCCTCTTCCTGCCCGACctagcaaaaaaTGGAgGGGGACGGATTCGAACCGCCGAACCC
AAAGGGAGCGGATTtgaAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCgggGATCTGTGGT
CTCATACAGAACTTATAAGATTCCCAAATCCAAAGACATTTCACGTTTATGGTGATTTCCCAGA
ACACATAGCGACATGCAAATATTGCAGGGCGCCACTCCCCTGTCCCTCACAGCCATCTTCCTGC
CAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGG
TTGATGACGTCAGCGTTCGAATTCCTAGCAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAG
TCTGTGACGGCAGATTTgaAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTC
CTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATA
GTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCAC
GTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCT
TGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGca
aaaaaTGGAgGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTtgaAGTCCGCCGCG
TTTAGCCACTTCGCTACCCCTCCgggGATCTGTGGTCTCATACAGAACTTATAAGATTCCCAAA
TCCAAAGACATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTGCAGG
GCGCCACTCCCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCG
CCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCCTAG
CAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTgaAGTCTGCTC
CCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAA
ATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACT
GCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTA

CAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAAT
CATGGGAAATAGGCCCTCTTCCTGCCCGACctagcaaaaaaTGGAgGGGGACGGATTCGAACCG
CCGAACCCAAAGGGAGCGGATTtgaAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCgggGA
TCTGTGGTCTCATACAGAACTTATAAGATTCCCAAATCCAAAGACATTTCACGTTTATGGTGAT
TTCCCAGAACACATAGCGACATGCAAATATTGCAGGGCGCCACTCCCCTGTCCCTCACAGCCAT
CTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTT
GGAGCGGGTTGATGACGTCAGCGTTCGAATTCCTAGCAAAAAATGGTGGGGGAAGGATTCGAAC
CTTCGAAGTCTGTGACGGCAGATTTgaAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGT
GTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGC
ATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTC
TACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCT
AACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCG
ACCTAGcaaaaaaTGGAgGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTtgaAGT
CCGCCGCGTTTAGCCACTTCGCTACCCCTCCgggGATCTGTGGTCTCATACAGAACTTATAAGA
TTCCCAAATCCAAAGACATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAAT
ATTGCAGGGCGCCACTCCCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGG
TGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGA
ATTCCTAGCAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTgaA
GTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAA
GCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATT
TTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTT
TGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATAT
GAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACctagcaaaaaaTGGAgGGGGACGGAT
TCGAACCGCCGAACCCAAAGGGAGCGGATTtgaAGTCCGCCGCGTTTAGCCACTTCGCTACCCC
TCCgggGATCTGTGGTCTCATACAGAACTTATAAGATTCCCAAATCCAAAGACATTTCACGTTT
ATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTGCAGGGCGCCACTCCCCTGTCCCTC
ACAGCCATCTTCCTGCCAGGGCGCACGCGCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCCGC
GATTCCTTGGAGCGGGTTGATGACGTCAGCGTTCGAATTCCTAGCAAAAAATGGTGGGGGAAGG
ATTCGAACCTTCGAAGTCTGTGACGGCAGATTTgaAGTCTGCTCCCTTTGGCCGCTCGGGAACC
CCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTA
CGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTA
TTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATT
ATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTT
CCTGCCCGACCTAGcaaaaaaTGGAgGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGA
TTtgaAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCgggGATCTGTGGTCTCATACAGAAC
TTATAAGATTCCCAAATCCAAAGACATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGAC
ATGCAAATATTGCAGGGCGCCACTCCCCTGTCCCTCACAGCCATCTTCCTGCCAGGGCGCACGC
GCGCTGGGTGTTCCCGCCTAGTGACACTGGGCCCGCGATTCCTTGGAGCGGGTTGATGACGTCA
GCGTTCGAATTCCTAGCAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCA
GATTTgaAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGA
TATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAA
ACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACT
AATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATAT
GCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACctagtcaataatcaatgt

caacgcgtatatctggcccgtacatcgcgaagcagcgcaaaacGGATCCtgcaggtatttGCGG
CCGCggtccgtatactccggaatattaatagatcatggagataattaaaatgataaccatctcg
caaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattcc
ggattattcataccgtcccaccatcgggcgcgAACTCCTAAAAAACCGCCACCatgaagtgcct
tttgtacttagccttttttattcattggggtgaattgcaagttcaccatagtttttccacacaac
caaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaa
attggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctat
tcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgtgatttccgctggtat
ggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgcaaggaaa
gcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggata
tgcaactgtgacggatgccgaagcagtgattgtccaggtgactcctcaccatgtgctggttgat
gaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgcc
ccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaa
cctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggag
ggcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaat
actgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatct
ctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacc
tcagtggatgtaagtctaattcaggacgttgagaggatcttggattattccctctgccaagaaa
cctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcctaa
aaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccaga
tacatcagagtcgatattgctgctccaatcctctcaagaatggtcggaatgatcagtggaacta
ccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaattggacccaatgg
agttctgaggaccagttcaggatataagtttcctttatacatgattggacatggtatgttggac
tccgatcttcatcttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgctt
cgcaacttcctgatgatgagagtttatttttggtgatactgggctatccaaaaatccaatcga
gcttgtagaaggttggttcagtagttggaaaagctctattgcctctttttctttatcataggg
ttaatcattggactattcttggttctccgagttggtatccatctttgcattaaattaaagcaca
ccaagaaaagacagatttatacagacatagagatgaaccgacttggaaagtgataaggccaggc
cggccaagcttgtcgagaagtactagaggatcataatcagccataccacatttgtagaggtttt
acttgctttaaaaaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaattgtt
gttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttca
caaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatctta
tcatgtctggatctgatcactgcttgagcctaggagatccgaaccagataagtgaaatctagtt
ccaaactattttgtcatttttaattttcgtattagcttacgacgctacacccagttcccatcta
ttttgtcactcttcccctaaataatccttaaaaactccatttccacccctcccagttcccaacta
ttttgtccgcccacagcggggcatttttcttcctgttatgttttttaatcaaacatcctgccaac
tccatgtgacaaaccgtcatcttcggctacttt

pBK $EcYtR_{CUA}$



Sequence color-coding key

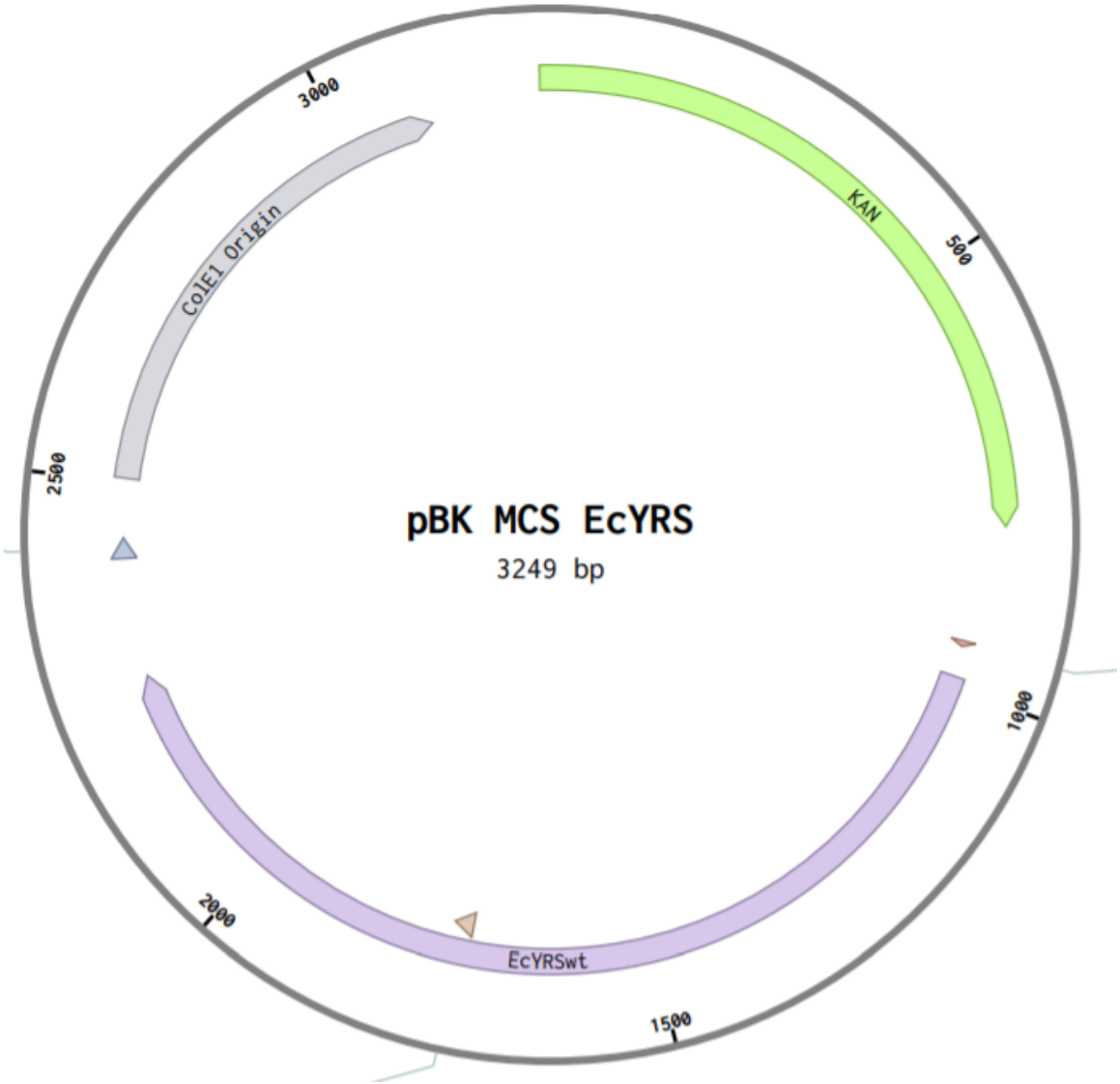
feature	Color
ColE1 ori	text
EcYtRNA$_{CUA}$	text
KAN resistance	text
lpp promoter	text
rrnC terminator	text

Sequence

ctttgctgagttgaaggatccgcgcttgtttcggcgtgggtatggtggcaggccccgtggccg
gggactgttgggcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcataag
tgcggcgacgatagtcatgccccgcgcccaccggaaggagctaccggcagcggtgcggactgtt
gtaactcagaataagaaatgaggccgctcatggcgttctgttgcccgtctcactggtgaaaaga
aaaacaaccctggcgccgcttctttgagcgaacgatcaaaaataagtggcgcccatcaaaaaa
atattctcaacataaaaaactttgtgtaatacttgtaacgctgaattcGGTGGGGTTCCCGAGC
GGCCAAAGGGAGCAGACTCTAAATCTGCCGTCATCGACTTCGAAGGTTCGAATCCTTCCCCCAC
CACCActgcagatccttagcgaaagctaaggatttttttaagcttggcactggccgtcgtttt
acaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacGAATTCat
ccccctttcgccagacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggt
tgaggccgttgagcaccgccgccgcaaggaatggtCCTAGGaAGATCTaaggagcccatggctg
cagtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatctctgcaatat
attgagtttgcgtgcttttgtaggccggataaggcgttcacgccgcatccggcaagaaacagca
aacaatccaaaacgccgcgttcagcggcgtttttctgcttttcttcgcgaattaattccgctt
cgcacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtt
tttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga
aacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctg
ttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttc
tcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtg
cacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacc
cggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggta
tgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagta
tttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccg
gcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaa
aaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaac
tcacgttaagggattttggtcatgaacaataaaactgtctgcttacataaacagtaatacaagg
ggtgttatgagccatattcaacgggaaacgtcttgctcgaggccgcgattaaattccaacatgg

atgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatcta
tcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgcc
aatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgacca
tcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccgggaaaac
agcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtg
ttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttc
gtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacga
gcgtaatggctggcctgttgaacaagtctggaaagaaatgcataagcttttgccattctcaccg
gattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaa
taggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatg
gaactgcctcggtgagttttctccttcattacagaaacggcttttcaaaaatatggtattgat
aatcctgatatgaataaattgcagtttcatttgatgctcgatgagttttctaatcagaattgg
ttaattggttgtaacactggcagagcattacgctgacttgacgggacggcggctttgttgaata
aatcgaa

pBK-MCS EcYRS D265R library



Sequence color-coding key

feature	Color
ColE1 ori	text
EcYRS wt	text
KAN resistance	text

EcYRS D265R library mutations

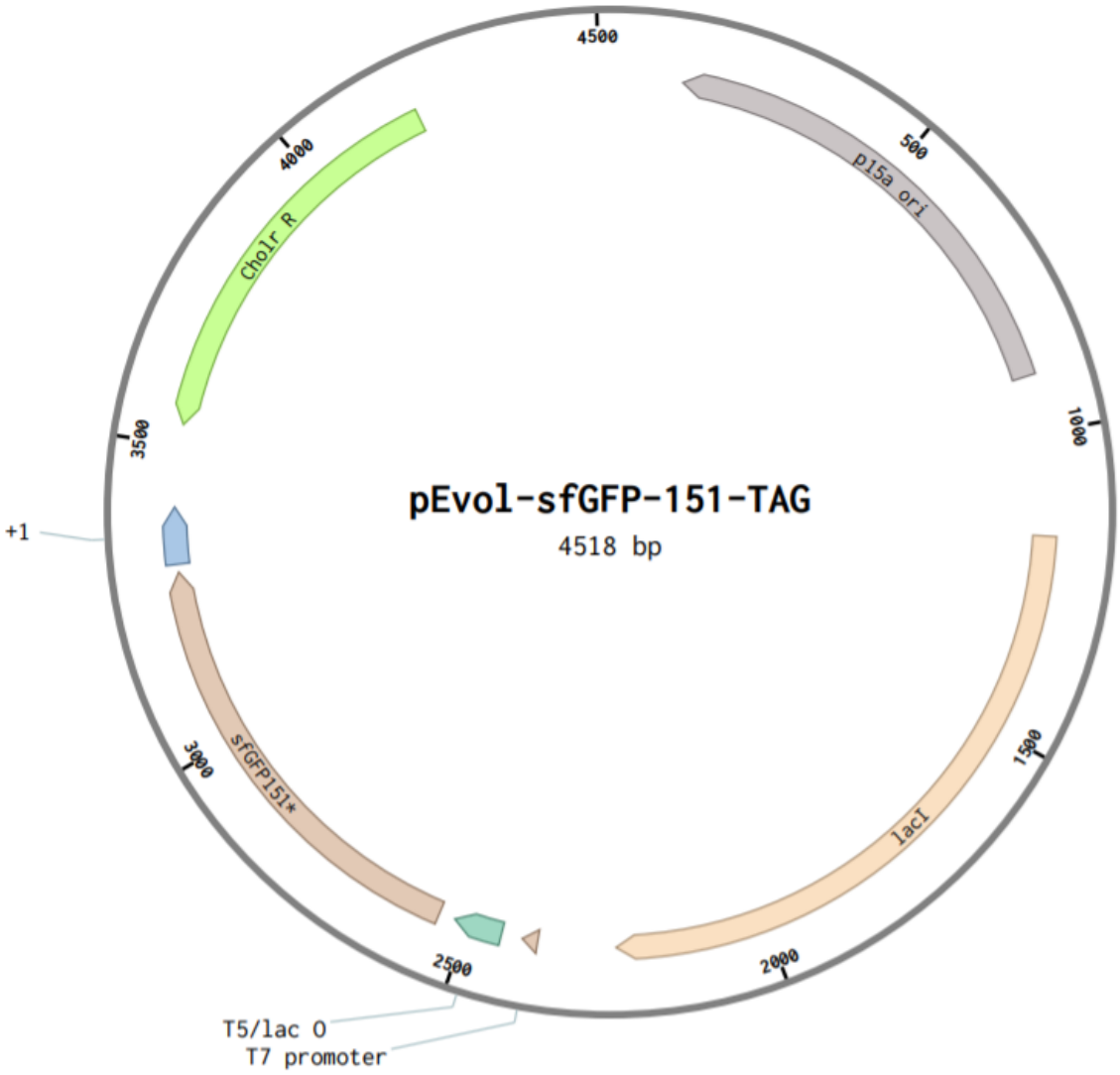
residue	Mutation
Y37	FLIMVSTAYHCG
L71	NBT
N126	NSPTACGDH
D182	NST
F183	NNK
L186	NNK
D265	R

Sequence

caacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgatttatatgggt
ataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagcc
cgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgag
atggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgta
ctcctgatgatgcatggttactcaccactgcgatccccgggaaaacagcattccaggtattaga
agaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcat
tcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaat
cacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgt
tgaacaagtctggaaagaaatgcataagcttttgccattctcaccggattcagtcgtcactcat
ggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttg
gacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagtt
ttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaa
ttgcagtttcatttgatgctcgatgagttttttctaatcagaattggttaattggttgtaacact
ggcagagcattacgctgacttgacgggacggcggctttgttgaataaatcgaacttttgctgag
ttgaaggatccGCGGCCGCtcgggttgtcagcctgtcccgcttataagatcatacgccgttata

cgttgtttacgctttgaggaatcccatatgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGC
GGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGAT
CGCGCTCTATTGCGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTG
TTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGG
GTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCA
GGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAAC
TCTGCTATCGCGGCGAACAACTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCG
ATATTGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAA
CCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTGTTGCAGGGTTATGACTTC
GCCTGTCTGAACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACA
TCACTTCTGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCC
GCTGATCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCAGTCTGGTTaGAT
CCGAAGAAAACCAGCCCGTACAAATTCTACCAGTTCTGGATCAACACTGCGGATGCCGACGTTT
ACCGCTTCCTGAAGTTCTTCACCTTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGA
TAAAAACAGCGGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGACTCGTCTGGTT
CACGGTGAAGAAGGTTTACAGGCGGCAAAACGTATTACCGAATGCCTGTTCAGCGGTTCTTTGA
GTGCGCTGAGTGAAGCGGACTTCGAACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGA
AAAGGGCGCAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCCGTGGTCAGGCA
CGTAAAACTATCGCCTCCAATGCCATCACCATTAACGGTGAAAAACAGTCCGATCCTGAATACT
TCTTTAAAGAAGAAGATCGTCTGTTTGGTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTA
CTGTCTGATTTGCTGGAAAtaaccatggctgcagtttcaaacgctaaattgcctgatgcgctac
gcttatcaggcctacatgatctctgcaatatattgagtttgcgtgcttttgtaggccggataag
gcgttcacgccgcatccggcaagaaacagcaaacaatccaaaacgccgcgttcagcggcgtttt
ttctgcttttcttcgcgaattaattccgcttcgcacatgtgagcaaaaggccagcaaaaggcca
ggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccctgacgagcatca
caaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgttt
ccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccg
cctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggt
gtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgcc
ttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcag
ccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtg
gcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacc
ttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttt
ttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttc
tacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgaacaataaa
actgtctgcttacataaacagtaatacaaggggtgttatgagccatatt

pEvol-sfGFP-151-TAG



Sequence color-coding key

feature	Color
p15a ori	text
lacI	text
Chlor resistance	text
sfGFP151-TAG	text
lac O	text
lambda t0 terminator	text

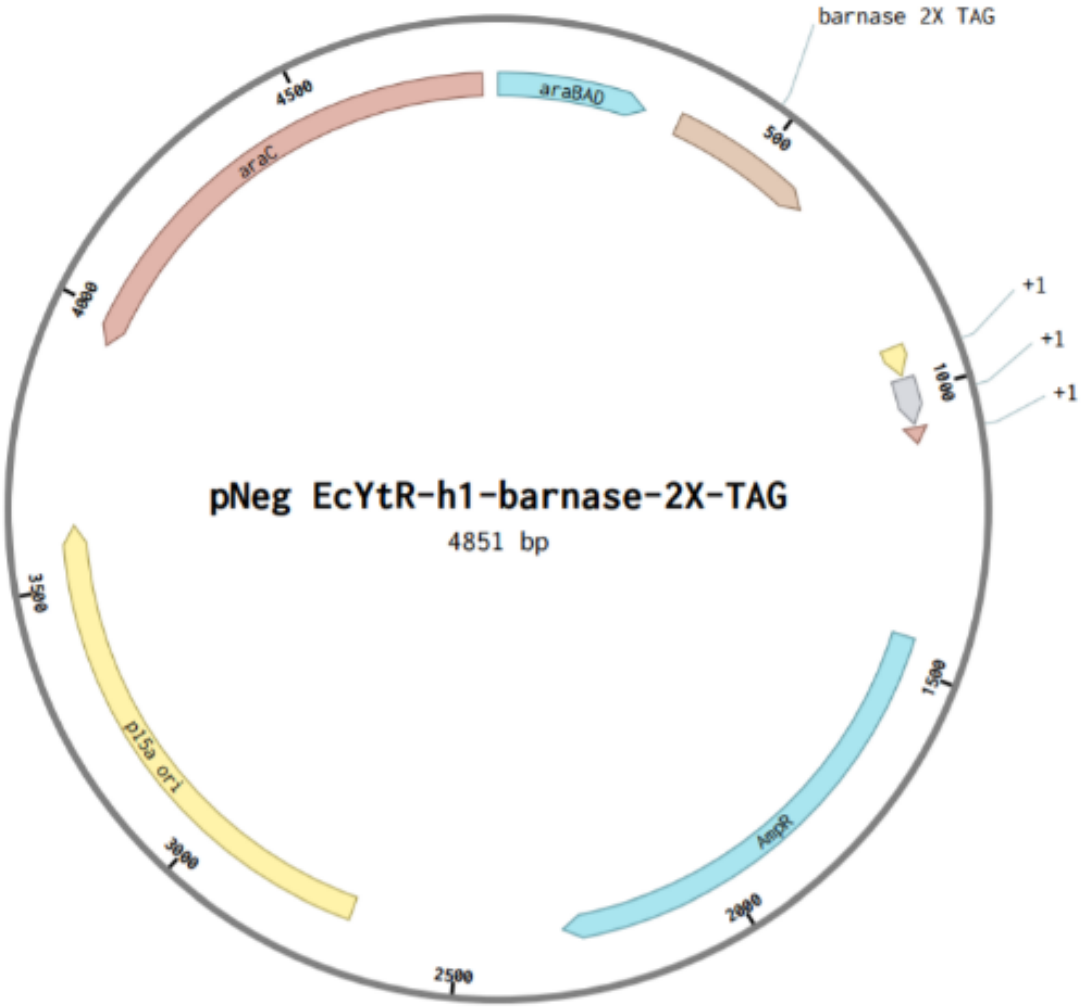
Sequence

gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttac
tatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcacc
ggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgc
tcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatg
ccaggaagatacttaacagggaagtgagagggccgcggcaaagccgtttttccataggctccgc
ccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactat
aaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtt
taccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgg
gtaggcagttcgctccaagctggactgtatgcacgaacccccgttcagtccgaccgctgcgcc
ttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcag
ccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaa
ggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctca
gagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttttcagagcaagagattacgcgc
agaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgc
aatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatg
tttgacagcttatcatcgataagcttgcaatttatctcttcaaatgtagcacctgaagtcagcc
ccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgact
gggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacat
taattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatg
aatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcgg
tccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataac
atgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccgga
ctcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtggga
acgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgcctt
cccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcag

acgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgacc
agatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtct
ggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatc
ctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcacc
gccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagtt
gatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggt
ggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaa
ttcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggt
tcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttac
tggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaag
gttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagg
ctcactatagggaattgtgagcggataacaattcccctctagagtttgacagcattatcatcg
atctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataatagattcaat
tgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatATGAGCAAAGGA
GAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACA
AATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTAT
TTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTT
CAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCG
AAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGA
AGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAA
GATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATAGATCACGG
CAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTC
CGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCA
GACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACA
TGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGG
ATCCCACCACCACCACCACCACTAAggcGCTGAGCtTGgactcctgttgatagatccagtaatg
acctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtg
agaatCCAAGCTAGCTTGGCGCTGCACTCGAGCAGCTCAGGGTCGAATTTGCTTTCGAatttct
gccattcatccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataa
ctgccttaaaaaaattacgccccgccctgccactcatcgcagtactgttgtaattcattaagca
ttctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcac
cttgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattg
gccacgtttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaacatattct
caataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatat
gtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgc
tcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcattg
ccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaa
cttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggtta
taggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgccattgggatatat
caacggtggtatatccagtgatttttttctccatttagcttccttagctcctgaaaatctcga
taactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacg
tgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacac
caggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtc

gggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttc
tgtttctatcagctgtccctcctgttcagctactgacg

pNegEcYtR-h1-barnase-2X-TAG



Sequence color-coding key

feature	Color
p15a ori	text
araC	text
Amp resistance	text
araBAD	text
EcYtR-h1	text
rrnC terminator	text
lpp promoter	text
barnase-2X-TAG	text

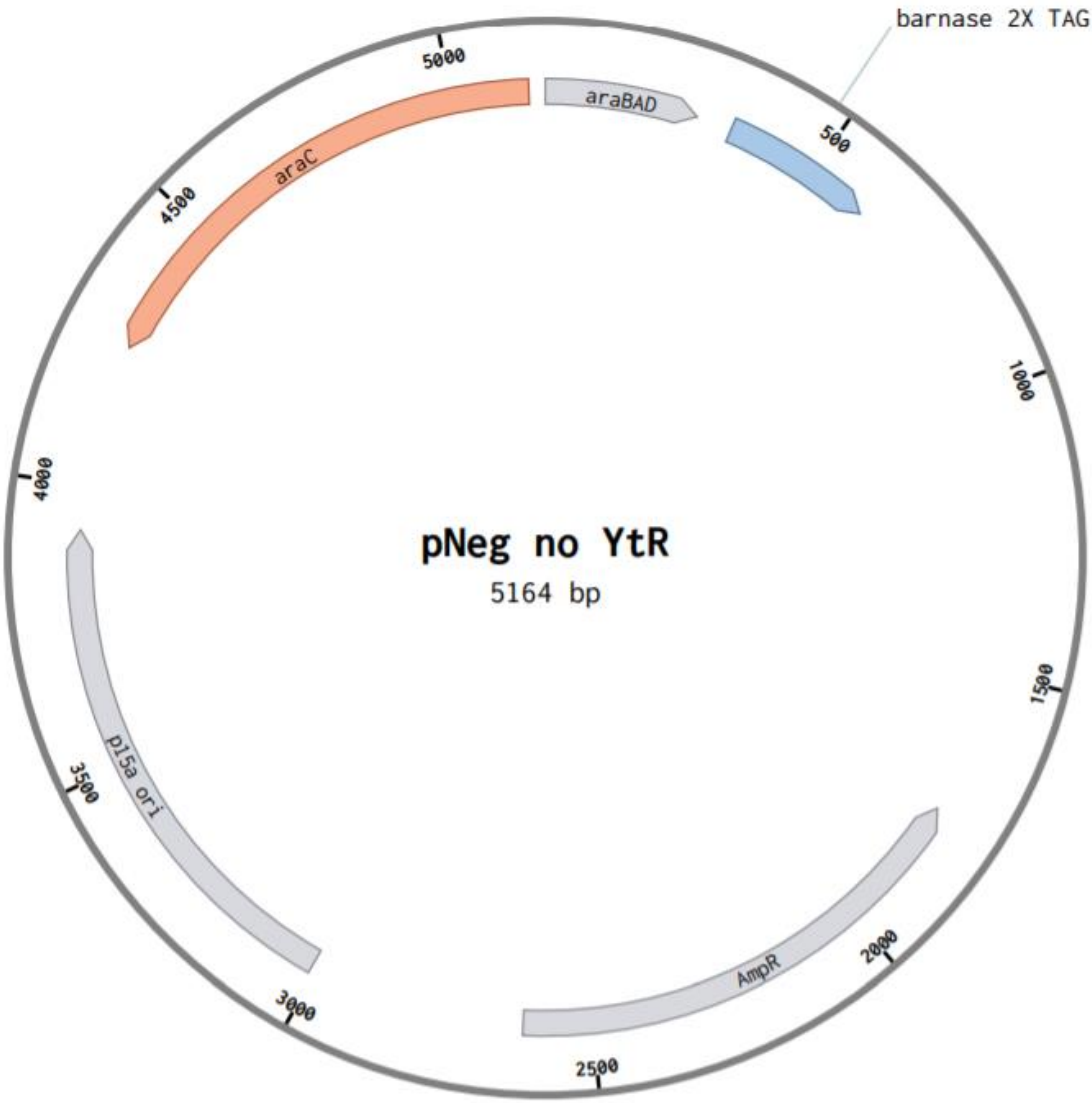
Sequence

aaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactggctcttctcgct
aaccaaaccggtaaccccgcttattaaaagcattctgtaacaaagcgggaccaaagccatgaca
aaaacgcgtaacaaaagtgtctataatcacggcagaaaagtccacattgattatttgcacggcg
tcacactttgctatgccatagcatttttatccataagattagcggatcctacctgacgctttt
atcgcaactctctactgtttctccatacccgtttttttgggctagaaataattttgtttaactt
taagaaggagatatacatagttgacggggttgcggattatcttcagacatatcataagctacct
gataattacattacaaaatcagaagcacaagccctcggctgggtggcatcaaaagggaaccttg
cataggtcgctccggggaaaagcatcggcggagacatcttctcaaacagggaaggcaaactccc
gggcaaaagcggacgaacatggcgtgaagcggatattaactatacatcaggcttcagaaattca
gaccggattctttactcaagcgactggctgatttgcatggcatgcgGATCCgcgcttgtttcgg
cgtgggtatggtggcaggccccgtggccgggggactgttgggcgctgccggcacctgtcctacg
agttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccgga
aggagctaccggcagcggtgcggactgttgtaactcagaataagaaatgaggccgctcatggcg
ttctgttgcccgtctcactggtgaaaagaaaaacaaccctggcgccgcttctttgagcgaacga
tcaaaaataagtggcgccccatcaaaaaaatattctcaacataaaaaactttgtgtaatacttg
taacgctgaattcGAAGTGGTTCCCGAGCGGCCAAAAGGAGCAGACTCTAAATCTGCCGTCATC
GACTTCGAAGGTTCGAATCCTTCCCACTTCACCActgcagatccttagcgaaagctaaggattt
tttttaagcttggcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttaccc
aacttaatcgccttgcagcacGAATTCatccccctttcgccagacgctctcccttatgcgactc
ctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggt
CCTAGGaAGATCTaaggagcccatggccgagcatggggaacgaaaactcacgttaagggatttt
ggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaa
tcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcac
ctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataac

tacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctca
ccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctg
caactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgcc
agttaatagtttgcgcaacgttgttgccattgctgcaggcatcgtggtgtcacgctcgtcgttt
ggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgt
gcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgtt
atcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttt
tctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgct
cttgcccggcgtcaacacgggataataccgcgccacatagcagaactttaaaagtgctcatcat
tggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatg
taacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgag
caaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatact
catactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatac
atatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgtacgtgcc
gatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccagg
atttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggt
gatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtt
tctatcagctgtccctcctgttcagctactgacggggtggtgcgtaacggcaaaagcaccgccg
gacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaa
gtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggat
atattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatg
gcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagag
ggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgc
tcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggct
ccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcg
tttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtat
gcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaac
ccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagttagtc
ttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaa
gccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggc
ggtttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatctt
attaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaag
tcagccccatacgatataagttgtaattctcatgtttgacagcattatcatcgataagctttaa
tgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgtgtatgaaatctaacaatg
cgctcatcgtcatcctcggcaccgtcaccctggatgctgtaggcataggcttggttatgccggt
actgccgggcctcttgcgggatggccacgatgcgtccggcgtagaggatctgctcatgtttgac
agcttatcatcgatgcataatgtgcctgtcaaatggacgaagcagggattctgcaaaccctatg
ctactccgtcaagccgtcaattgtctgattcgttaccaattatgacaacttgacggctacatca
ttcacttttttcttcacaaccggcacggaactcgctcgggctggccccggtgcatttttttaaata
cccgcgagaaatagagttgatcgtcaaaaccaacattgcgaccgacggtggcgataggcatccg
ggtggtgctcaaaagcagcttcgcctggctgatacgttggtcctcgcgccagcttaagacgcta
atccctaactgctggcggaaaagatgtgacagacgcgacggcgacaagcaaacatgctgtgcga
cgctggcgatatcaaaattgctgtctgccaggtgatcgctgatgtactgacaagcctcgcgtac
ccgattatccatcggtggatggagcgactcgttaatcgcttccatgcgccgcagtaacaattgc

tcaagcagatttatcgccagcagctccgaatagcgcccttcccccttgcccggcgttaatgattt
gcccaaacaggtcgctgaaatgcggctggtgcgcttcatccgggcgaaagaaccccgtattggc
aaatattgacggccagttaagccattcatgccagtaggcgcgcggacgaaagtaaacccactgg
tgataccattcgcgagcctccggatgacgaccgtagtgatgaatctctcctggcgggaacagca
aaatatcacccggtcggcaaacaaattctcgtccctgatttttcaccaccccctgaccgcgaat
ggtgagattgagaatataacctttcattcccagcggtcggtcgataaaaaaatcgagataaccg
ttggcctcaatcggcgttaaacccgccaccagatgggcattaaacgagtatcccggcagcaggg
gatcattttgcgcttcagccatactttcatactcccgccattcagagaag

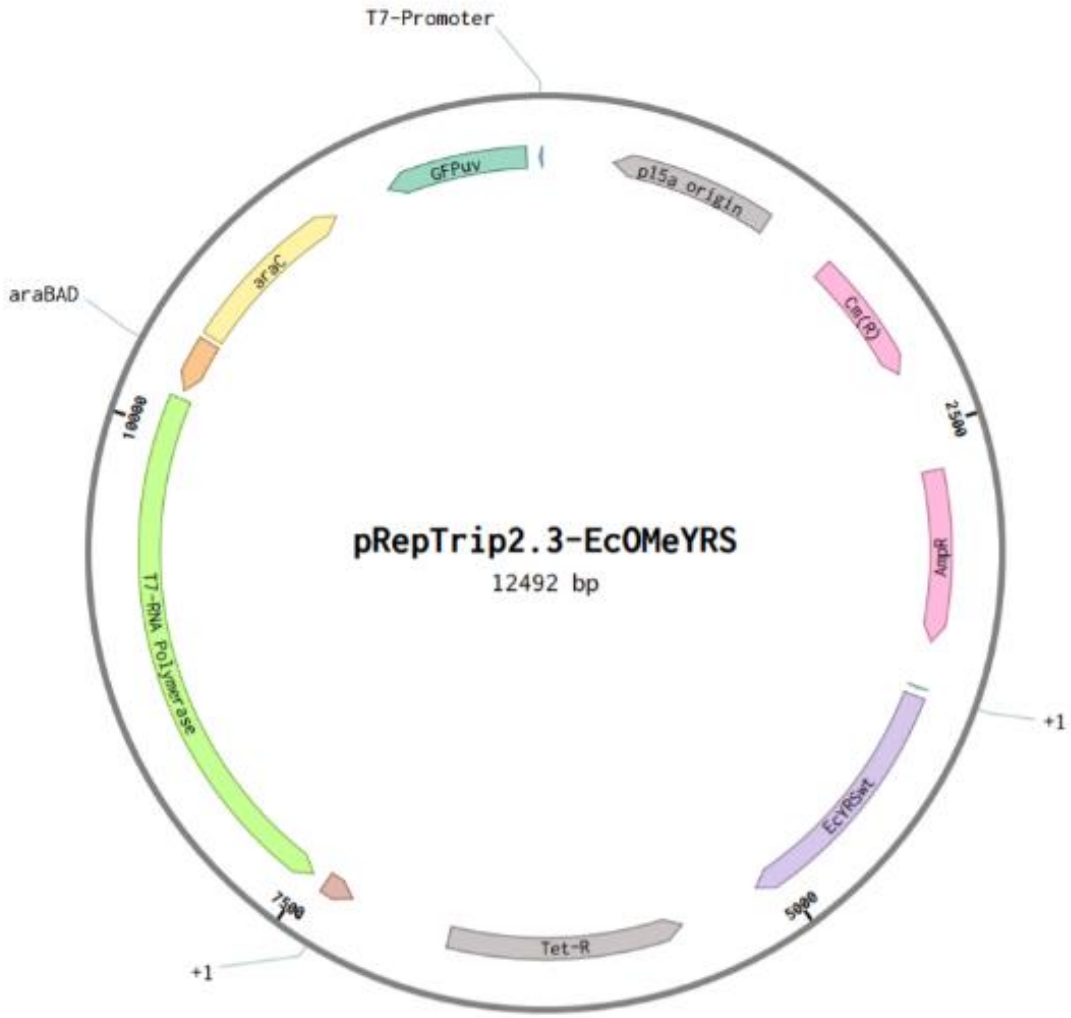
pNeg-No YtR



Sequence

This construct has the same sequence as pNeg EcYtR-h1-barnase-2X-TAG, but is missing the lpp-EcYtR-h1 rrnC terminator portion.

pRepTrip2.3-EcOMeYRS



Sequence color-coding key

feature	Color
p15a ori	text
Chlor resistance	text
Amp resistance	text
GFPuv	text
EcYRSwt	text
glnS promoter	text
Tet resistance	text
T7 RNA polymerase	text
araBAD	text
araC	text

Ec-OMeYRS active site mutations

residue	Mutation
Y37	V
D182	S
F183	M
L186	A

Sequence

gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggta
tcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggag
tcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcat
tggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaat
ttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagtt
ttcgttccactgagcgtcagaccccttaataagatgatcttcttgagatcgttttggtctgcgc
gtaatctcttgctctgaaaacgaaaaaaccgccttgcagggcggtttttcgaaggttctctgag
ctaccaactctttgaaccgaggtaactggcttggaggagcgcagtcaccaaaacttgtcctttc

agtttagccttaaccggcgcatgacttcaagactaactcctctaaatcaattaccagtggctgc
tgccagtggtgcttttgcatgtctttccgggttggactcaagacgatagttaccggataaggcg
cagcggtcggactgaacgggggttcgtgcatacagtccagcttggagcgaactgcctacccgg
aactgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaacc
gaaaggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctggtatctttatag
tcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcgg
agcctatggaaaaacggctttgccgcggccctctcacttccctgttaagtatcttcctggcatc
ttccaggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgt
agcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtg
cagccttttttctcctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaa
gccagtatacactccgctagcgctgatgtccggcggtgcttttgccgttacgcaccaccccgtc
agtagctgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacaccat
catacactaaatcagtaagttggcagcatcacccgacgcactttgcgccgaataaatacctgtg
acggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccgggaagccctggg
ccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatga
aataagatcactaccgggcgtatttttgagttatcgagattttcaggagctaaggaagctaaa
atggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacatt
ttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggc
ctttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcc
cgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagctggtgatatggg
atagtgttcacccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggag
tgaataccactaggatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggt
gaaaacctggcctatttccctaaagggtttattgagaatatgttttcgtctcagccaatccct
gggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgtttt
caccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcat
catgccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatg
agtggcagggcggggcgtaattttttttaaggcagttattggtgcccttaaacgcctggttgcta
cgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacccggtcg
tcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttattgac
taccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggccagtttgctcaggctctcccc
gtggaggtaataattgacgatatgatcatttattctgcctcccagagcatgataaaaacggtta
gcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatc
cggaacataatggActagcggtggcacttttcggggaaatgtgcgcggaacccctatttgttta
tttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaat
aatattgaaaaaggaagagtatggcgtagagtattcaacatttccgtgtcgcccttattccctt
ttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgct
gaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttg
agagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgc
ggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaat
gacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaat
tatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcgg
aggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgt
tgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaa
tggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaatt

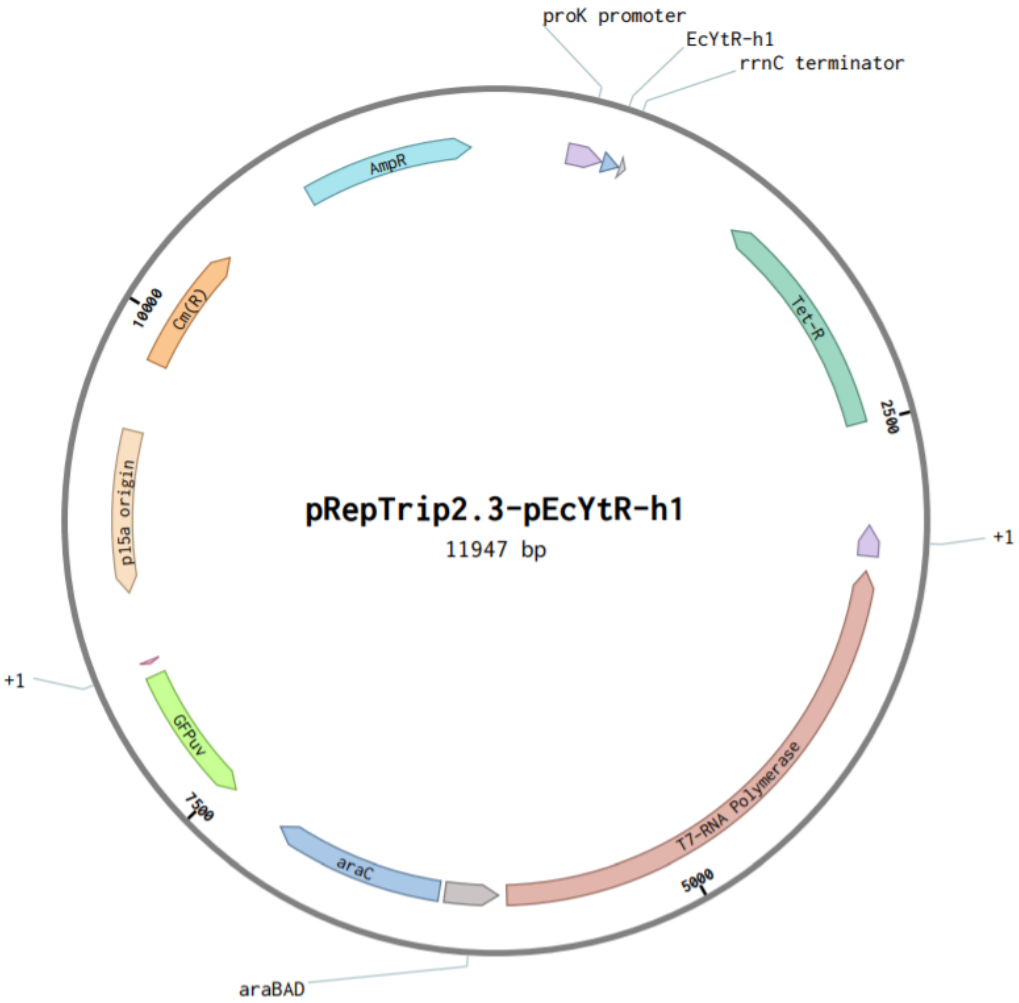
aatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggc
tggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactgg
ggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatgga
tgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagac
caagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctagg
tgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcaCTAGTaattac
gctgacttgacgggacggcggctttgttgaataaatcgaacttttgctgagttgaaggatccGC
GGCCGCtcgggttgtcagcctgtcccgcttataagatcatacgccgttatacgttgtttacgct
ttgaggaatcccaTATGatgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGT
AGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTAT
TGCGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGA
AACGCTTCCAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGG
CGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTG
GACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCG
CGGCGAACAACTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAA
ACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGAT
CAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTGTTGCAGGGTTATGACTTCGCCTGTCTGA
ACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGG
TATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACT
AAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCAGTCTGGTTaGATCCGAAGAAAA
CCAGCCCGTACAAATTCTACCAGTTCTGGATCAACACTGCGGATGCCGACGTTTACCGCTTCCT
GAAGTTCTTCACCTTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAACAGC
GGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGACTCGTCTGGTTCACGGTGAAG
AAGGTTTACAGGCGGCAAAACGTATTACCGAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAG
TGAAGCGGACTTCGAACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCGCA
GACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCCGTGGTCAGGCACGTAAAACTA
TCGCCTCCAATGCCATCACCATTAACGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGA
AGAAGATCGTCTGTTTGGTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATT
TGCTGGAAAtaaggtaccatggctgcagtttcaaacgctaaattgcctgatgcgctacgcttat
caggcctacatgatctctgcaatatattgagtttgcgtgcttttgtaggccggataaggcgttc
acgccgcatccggcaagaaacagcaaacaatccaaaacgccgcgttcagcggcgtttttttctgc
ttttcttcgcgaattaattccgcttcgcacatgtgagcaaaaggccagcaaaaggccaggaacc
gtAGATCTaaggagcccgagatgcgccgcgtgcggctgctggagatggcggacgcgatggatat
gttctgccaagggttggtttgcgcattcacagttctccgcaagaattgattggctccaattctt
ggagtggtgaatccgttagcgaggtgccgccggcttccattcaggtcgaggtggcccggctcca
tgcaccgcgacgcaacgcggggaggcagacaaggtatagggcggcgcctacaatccatgccaac
ccgttccatgtgctcgccgaggcggcataaatcgccgtgacgatcagcggtccaatgatcgaag
ttaggctggtaagagccgcgagcgatccttgaagctgtccctgatggtcgtcatctacctgcct
ggacagcatggcctgcaacgcgggcatcccgatgccgccggaagcgagaagaatcataatgggg
aaggccatccagcctcgcgtcgcgaacgccagcaagacgtagcccagcgcgtcggccgccatgc
cggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagc
gagggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaag
cggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaaga
agacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggtt

gaaggctctcaagggcatcggtcgacgctctcccttatgcgactcctgcattaggaagcagccc
agtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgc
ccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagccc
gaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacct
gtggcgccggtgatgccggccacgatgcgtccggcgtagaggatccacaggacgggtgtggtcg
ccatgatcgcgtagtcgatagtggctccaagtagcgaagcgagcaggactgggcggcggccaaa
gcggtcggacagtgctccgagaacgggtgcgcatagaaattgcatcaacgcatatagcgctagc
agcacgccatagtgactggcgatgctgtcggaatggacgatatcccgcaagaggcccggcagta
ccggcataaccaagcctatgcctacagcatccagggtgacggtgccgaggatgacgatgagcgc
attgttagatttcatacacggtgcctgactgcgttagcaatttaactgtgataaactaccgcat
taaagcttatcgatgataagctgtcaaacatgagaattacaacttatatcgtatggggctgact
tcaggtgctacattgctcaaagatgcaggggtaaaagctaaccgcatctttaccgacaaggcat
ccggcagttcaacagatcgggaagggctggatttgctgaggatgaaggtggaggaaggtgatgt
cattctggtgaagaagctcgaccgtcttggccgcgacaccgccgacatgatccaactgataaaa
gagtttgatgctcagggtgtagcggttcggtttattgacgacgggatcagtaccgacggtgata
tggggcaaatggtggtcaccatcctgtcggctgtggcacaggctgaacgccggaggatcctccg
ggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttca
ccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatg
cctggcagttccctactctcgcatggggagaccccacactaccatcggcgctacggcgtttcac
ttctgagttcggcatggggtcaggtgggaccaccgcgctactgccgccaggcaaattctgtttt
atcagaccgcttctgcgttctgatttaatctgtatcaggctgaaaatcttctctcatccgacgt
cttaggcgaaggcgaagtccgactctaagatgtcacggaggttcaagttacctttagccggaag
tgctggcattttgtccaattgagactcgtgcaactggtcagcgaactggtcgtagaaatcagcc
agtacatcacaagactcatatgtgtcaaccatagtttcgcgcactgctttgaacaggttcgcag
cgtcagccggaatggtaccgaaggagtcgtgaatcagtgcaaaagattcgattccgtacttctc
gtgtgcccacactacagtcttacgaaggtggctaccgtcttggctgtgtacaaagttaggagcg
ataccagactcctgtttgtgtgcatcaatctcgctatctttgttggtgttaatggtaggctgta
agcggaactgaccgaggaacatcaggttcaagcgcgtctgaataggcttcttgtattcctgcca
cacagggaaaccatcaggagttacccaatgcacagcgcaacgcttgcgaagaatctctccagtc
ttcttatctttgacctcagcagccagcagcttagcagcagacttaagccagttcattgcttcaa
ccgcagctaccaccgtcacgctcacagattcccaaatcagcttagccatgtatccagcagcctg
attcggctgagtgaacatcagacccttgccggaatcaatagctggctgaatggtatcttccagc
acttgttgacggaagccgaactctttggacccgtaagccagcgtcatgactgaacgcttagtca
cactgcgagtaacaccgtaagccagccattgaccagccagtgccttagtgcccagcttgacttt
ctcagagatttcaccagtgttctcatcggtcacggtaactacttcgttatcggtcccattgatt
gcgtctgcttgtagaatctcgttgactttcttagcaacaatcccgtagatgtcctgaacggttt
cactaggaagcaagttaaccgcgcgaccacctacctcatctcggagcatcgcggagaagtgctg
gatgccagagcaagacccgtcaaacgccagcggaagggagcagttatagctcaggccgtggtgc
tgtaccccagcgtactcaaagcagaacgcaaggaagcagaacggagaatcttgctcagcccacc
aagtgttctccagtggagacttagcgcaagccatgatgttctcgtggttttcctcaatgaactt
gatgcgctcagggaacggaaccttatcgacacccgcacagtttgcaccgtggattttcagccag
tagtaaccttccttaccgattggtttacctttcgccagcgtaagcagtcctttggtcatatcgt
taccttgcgggttgaacattgacacagcgtaaacacgaccgcgccagtccatgttgtaagggaa
ccagatggccttatggttagcaaacttattggcttgctcaagcatgaactcaaggctgatacgg

cgagacttgcgagccttgtccttgcggtacacagcagcggcagcacgtttccacgcggtgagag
cctcaggattcatgtcgatgtcttccggtttcatcgggagttcttcacgctcaatcgcagggat
gtcctcgaccggacaatgcttccacttggtgattacgttggcgaccgcAaggactttcttgttg
attttccatgcggtgttttgcgcaatgttaatcgctttgtacacctcaggcatgtaaacgtctt
cgtagcgcatcagtgctttcttactgtgagtacgcaccagcgccagaggacgacgaccgttagc
ccaatagccaccaccagtaatgccagtccacggcttaggaggaactacgcaaggttggaacatc
ggagagatgccagccagcgcacctgcacgggttgcgatagcctcagcgtattcaggtgcgagtt
cgatagtctcagagtcttgacctactacgccagcattttggcggtgtaagctaaccattccggt
tgactcaatgagcatctcgatgcagcgtactcctacatgaatagagtcttccttatgccacgaa
gaccacgcctcgccaccgagtagacccttagagagcatgtcagcctcgacaacttgcataaatg
ctttcttgtagacgtgccctacgcgcttgttgagttgttcctcaacgttttttcttgaagtgctt
agcttcaaggtcacggatacgaccgaagcgagcctcgtcctcaatggcccgaccgattgcgctt
gctacagcctgaacggttgtattgtcagcactggttaggcaagccagagtggtcttaatggtga
tgtacgctacggcttccggcttgatttcnnnncaggaactggaaggctgtcgggcgcttgccgc
gcttagctttcacttcctcaaaccagtcgttgatgcgtgcaatcatcttagggagtagggtagt
gatgagaggcttggcggcagcgttatccgcaacctcaccagctttaagttgacgctcaaacatc
ttgcggaagcgtgcttcacccatctcgtaagactcatgctcaagggccaactgttcgcgagcta
aacgctcaccgtaatggtcagccagagtgttgaacgggatagcagccagttcgatgtcagagaa
gtcgttcttagcaatgttaatggtattnnnngtgcacggtaatcatggtcatggttaattcctc
ctgttagcccaaaaaacgggtatggagaaacagtagagagttgcgataaaaagcgtcaggtagg
atccgctaatcttatggataaaaatgctatggcatagcaaagtgtgacgccgtgcaaataatca
atgtggacttttctgccgtgattatagacacttttgttacgcgtttttgtcatggctttggtcc
cgctttgttacagaatgcttttaataagcggggttaccggtttggttagcgagaagagccagta
aaagacgcagtgacggcaatgtctgatgcaatatggacaattggtttcttctctgaatggcggg
agtatgaaaagtatggctgaagcgcaaaatgatcccctgctgccgggatactcgtttaatgccc
atctggtggcgggtttaacgccgattgaggccaacggttatctcgatttttttatcgaccgacc
gctgggaatgaaaggttatattctcaatctcaccattcgcggtcaggggtggtgaaaaatcag
ggacgagaatttgtttgccgaccgggtgatattttgctgttcccgccaggagagattcatcact
acggtcgtcatccggaggctcgcgaatggtatcaccagtgggtttactttcgtccgcgcgccta
ctggcatgaatggcttaactggccgtcaatatttgccaatacggggttctttcgcccggatgaa
gcgcaccagccgcatttcagcgacctgtttgggcaaatcattaacgccgggcaaggggaagggc
gctattcggagctgctggcgataaatctgcttgagcaattgttactgcggcgcatggaagcgat
taacgagtcgctccatccaccgatggataatcgggtacgcgaggcttgtcagtacatcagcgat
cacctggcagacagcaattttgatatcgccagcgtcgcacagcatgtttgcttgtcgccgtcgc
gtctgtcacatcttttccgccagcagttagggattagcgtcttaagctggcgcgaggaccaacg
tatcagccaggcgaagctgcttttgagcaccacccggatgcctatcgccaccgtcggtcgcaat
gttggttttgacgatcaactctatttctcgcgggtatttaaaaaatgcaccggggccagcccga
gcgagttccgtgccggttgtgaagaaaaagtgaatgatgtagccgtcaagttgtcataattggt
aacgaatcagacaattgacggcttgacggagtagcatagggtttgcagaatccctgcttcgtcc
atttgacaggcacattatgcatgccgcttcgccttcgcgcgcgaattgatctgctgcctcgcgc
gtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcctgcag
caaaaaaccccctcaagacccgtttagaggccccaaggggttatgctagttattgctcagcggtg
gcagcagccaactcagcttcctttcgggctttgttatttgtagagctcatccatgccatgtgta
atcccagcagcagttacaaactcaagaaggaccatgtggtcacgcttttcgttgggatctttcg

aaagggcagattgtgtcgacaggtaatggttgtctggtaaaaggacagggccatcgccaattgg
agtattttgttgataatggtctgctagttgaacggatccatcttcaatgttgtggcgaattttg
aagttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagttatagt
tgtactcgagtttgtgtccgagaatgtttccatcttctttaaaatcaatacctttttaactcgat
acgattaacaagggtatcaccttcaaacttgacttcagcacgcgtcttgtagttcccgtcatct
ttgaaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaagtcatgcc
gtttcatatgatccggataacgggaaaagcattgaacaccataagagaaagtagtgacaagtgt
tggccatggaacaggtagttttccagtagtgcaaataaatttaagggtaagttttccgtatgtt
gcatcaccttcaccctctccactgacagaaaatttgtgcccattaacatcaccatctaattcaa
caagaattgggacaactccagtgaaaagttcttctcctttactcatatgtatatctccttcttta
aagttaaacaaaattatttctagagggaaaccgttgtggtctccctatagtgagtcgtattaat
ttcgcgggatcg

pRepTrip2.3-pEcYtR-h1



Sequence

This construct has the same sequence as pRepTrip2.3-EcOMeYRS, except the EcYRS/glns portion of the sequence is replaced with the following sequence:

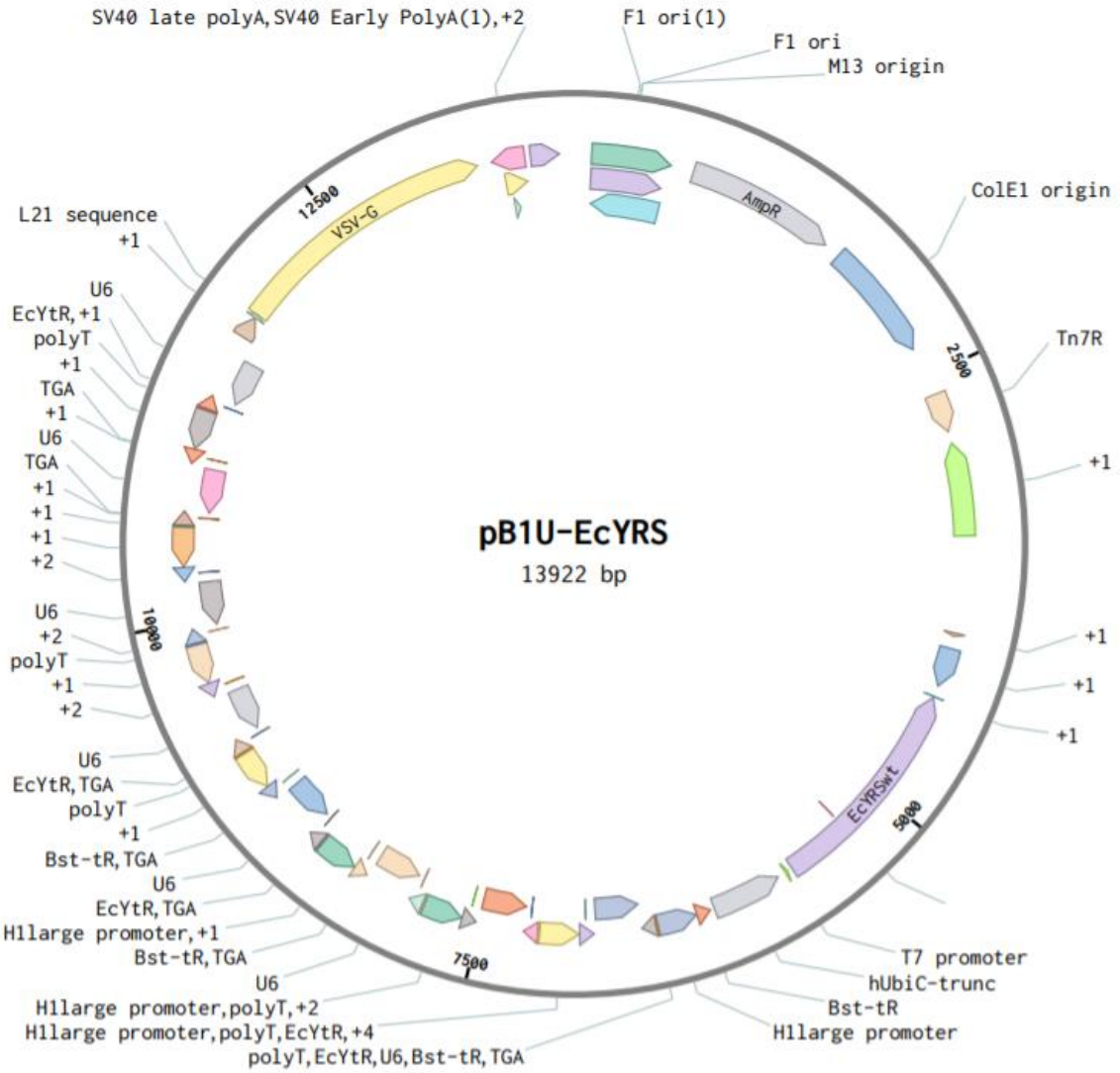
Sequence color-coding key

feature	Color
proK promoter	text
EcYtR-h1	text
rrnC terminator	text

cactggtgaaaagaaaaacaaccctggcgccgcttctttgagcgaacgatcaaaaataagtggc
gcgtgtgcttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattga
cgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattt
tgctattaagggattgacgagggcgTATCTgcgcagtaagatgcgccccgcattgaattcGAAG
TGGTTCCCGAGCGGCCAAAGGGAGCAGACTCTAAATCTGCCGTCATCGACTTCGAAGGTTCGAA
TCCTTCCCACTTCCCActgcagatccttagcgaaagctaaggatttttttaagcttggcactg
gccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcag
cacGAATTCatccccc

Chapter 3 Plasmids

pAcBac1 EGFP39-TAG



Sequence color-coding key

feature	Color
VSV-G	text
M13 ori	text
AmpR	text
ColE1 ori	text
Tn7R	text
Gentamicin resistance	text
WPRE	text
CMV	text
EGFP39-TAG	text

Sequence

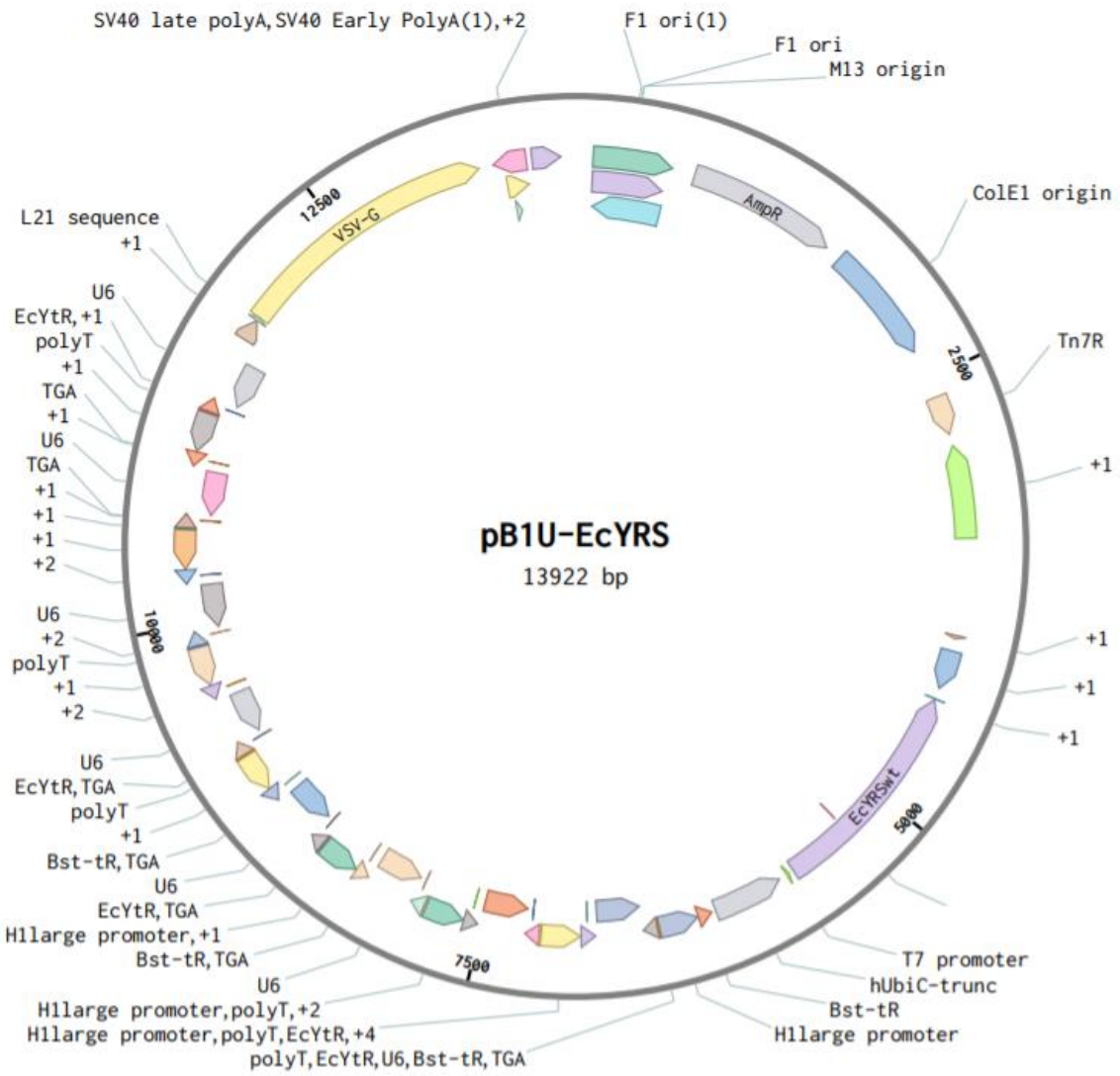
ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactga
atatcaacgcttatttgcagcctgaatggcgaatgggacgcgccctgtagcggcgcattaagcg
cggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcc
tttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgg
gggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg
gtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtc
cacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctat
tcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaac
aaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcgggg
aaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatg
agacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatt
tccgtgtcgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaac
gctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggat
ctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcactt
ttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcg
ccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacg
gatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggcca
acttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggga
tcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgt
gacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactactta
ctctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttct
gcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtct
cgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacga

cggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgat
taagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcat
ttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaac
gtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcc
ttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgt
ttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac
caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcc
tacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtctt
accgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggtt
cgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagca
ttgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtc
ggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcg
ggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatg
gaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatg
ttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgata
ccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcct
gatgcggtattttctccttacgcatctgtgcggtatttcacaccgcagaccagccgcgtaacct
ggcaaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgtgggcggacaat
aaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactagacagaata
gttgtaaactgaaatcagtccagttatgctgtgaaaaagcatactggacttttgttatggctaa
agcaaactcttcattttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaaggg
catggtaaagactatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagcc
gatctcggcttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttct
tcccgtatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcacgt
agatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtggcaat
gccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcggctgc
tcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttggtcgaaggcagcaa
gcgcgatgaatgtcttactacggagcaagttcccgaggtaatcggagtccggctgatgttggga
gtaggtggctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcatggatttgact
tggtcagggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttag
ggcgactgccctgctgcgtaacatcgttgctgctgcgtaacatcgttgctgctccataacatca
aacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatagactgtacaaaaaa
acagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggtt
ctggaccagttgcgtgagcgcatacgctacttgcattacagtttacgaaccgaacaggcttatg
tcaactgggttcgtgccttcatccgtttccacggtgtgcgtcacccggcaaccttgggcagcag
cgaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtctccacgcatcgt
caggcattggcggccttgctgttcttctacggcaaggtgctgtgcacggatctgccctggcttc
aggagatcggtagacctcggccgtcgcggcgcttgccggtggtgctgaccccggatgaagtggt
tcgcatcctcggttttctggaaggcgagcatcgtttgttcgcccaggactctagctatagttct
agtggttggcctacgtacccgtagtggctatggcagggcttgcgcttaatgcgccgctacaggg
cgcgtggggataccccctagagccccagctggttctttccgcctcagaagccatagagcccacc
gcatccccagcatgcctgctattgtcttcccaatcctccccttgctgtcctgccccaccccac
cccccagaatagaatgacacctactcagacaatgcgatgcaatttcctcattttattaggaaag
gacagtgggagtggcaccttccagggtcaaggaaggcacgggggagggcaaacaacagatggc

tggcaactagaaggcacagtcgaggctgatcagcgggtttaaacgggccctctagactcgagtt
aaagtcgacgcggggaggcggcccaaagggagatccgactcgtctgagggcgaaggcgaagacg
cggaagaggccgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaa
gggacgtagcagaaggacgtcccgcgcagaatccaggtggcaacacaggcgagcagccaaggaa
aggacgatgatttccccgacaacaccacggaattgtcagtgcccaacagccgagcccctgtcca
gcagcgggcaaggcaggcggcgatgagttccgccgtggcaatagggaggggaaagcgaaagtc
ccggaaaggagctgacaggtggtggcaatgccccaaccagtgggggttgcgtcagcaaacacag
tgcacaccacgccacgttgcctgacaacgggccacaactcctcataaagagacagcaaccagga
tttatacaaggaggagaaaatgaaagccatacgggaagcaatagcatgatacaaaggcattaaa
gcagcgtatccacatagcgtaaaaggagcaacatagttaagaataccagtcaatctttcacaaa
ttttgtaatccagaggttgattgtcgacttaacgcgttgaattcTCAATGGTGATGGTGATGAT
GACCGGTATGCATATTCAGATCCTCTTCTGAGATGAGTTTTTGTTCGAAGGGCCCCTTGTACAG
CTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGC
TTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCAGCA
CGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTC
GATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATATAG
ACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGT
CGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGT
CTTGTAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCG
GACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGG
TCAGGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAG
GGTCAGCTTGCCGTAGGCTAGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGT
TTACGTCGCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCT
CACCATGgTGGCGGCgctagccagcttgggtctccctatagtgagtcgtattaatttcgataag
ccagtaagcagtgggttctctagttagccagagagctctgcttatatagacctcccaccgtaca
cgcctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccgttgat
tttggtgccaaaacaaactcccattgacgtcaatggggtggagacttggaaatccccgtgagtc
aaaccgctatccacgcccattgatgtactgccaaaaccgcatcaccatggtaatagcgatgact
aatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgccagg
cgggccatttaccgtcattgacgtcaataggggggcgtacttggcatatgatacacttgatgtac
tgccaagtgggcagtttaccgtaaatagtccacccattgacgtcaatggaaagtccctattggc
gttactatgggaacatacgtcattattgacgtcaatgggcgggggtcgttgggcggtcagccag
gcgggccatttaccgtaagttatgtaacgcggaactccatatatgggctatgaactaatgaccc
cgtaattgattactattaataactagtcaataatcaatgtcaacgcgtatatctggcccgtaca
tcgcgaagcagcgcaaaacGGATCCtgcaggtatttGCGGCCGCggtccgtatactccggaata
ttaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtattttactgtt
ttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccaccat
cgggcgcgAACTCCTAAAAAACCGCCACCatgaagtgcctttgtacttagccttttttattcat
tggggtgaattgcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgtt
ccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgacttaataggca
cagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtca
tgcttccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattcc
atccgatccttcactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaacaaggaa
cttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccgaagc

agtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgat
tcacagttcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaacct
ggcattctgactataaggtcaaagggctatgtgattctaacctcatttccatggacatcacctt
cttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactac
tttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagac
tcccatcaggtgtctggttcgagatggctgataaggatctctttgctgcagccagattccctga
atgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcag
gacgttgagaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtc
ttccaatctctccagtggatctcagctatcttgctcctaaaaacccaggaaccggtcctgcttt
caccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgct
ccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatg
actgggcaccatatgaagacgtggaaattggacccaatggagttctgaggaccagttcaggata
taagtttccttatacatgattggacatggtatgttggactccgatcttcatcttagctcaaag
gctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtt
tatttttggtgatactgggctatccaaaaatccaatcgagcttgtagaaggttggttcagtag
ttggaaaagctctattgcctctttttttctttatcatagggttaatcattggactattcttggtt
ctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacag
acatagagatgaaccgacttggaaagtgataaggccaggccggccaagcttgtcgagaagtact
agaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccaca
cctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagct
tataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgc
attctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgct
tgagcctaggagatccgaaccagataagtgaaatctagttccaaactattttgtcatttttaat
tttcgtattagcttacgacgctacacccagttcccatctattttgtcactcttccctaaataat
ccttaaaaactccatttccacccctcccagttcccaactattttgtccgcccacagcggggcat
ttttcttcctgttatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttc
ggctacttt

pB1U-EcYRS variants



Ec-pBPARS- Gen 1 mutations

residue	Mutation
Y37	G
D182	G
L186	A
D265	R

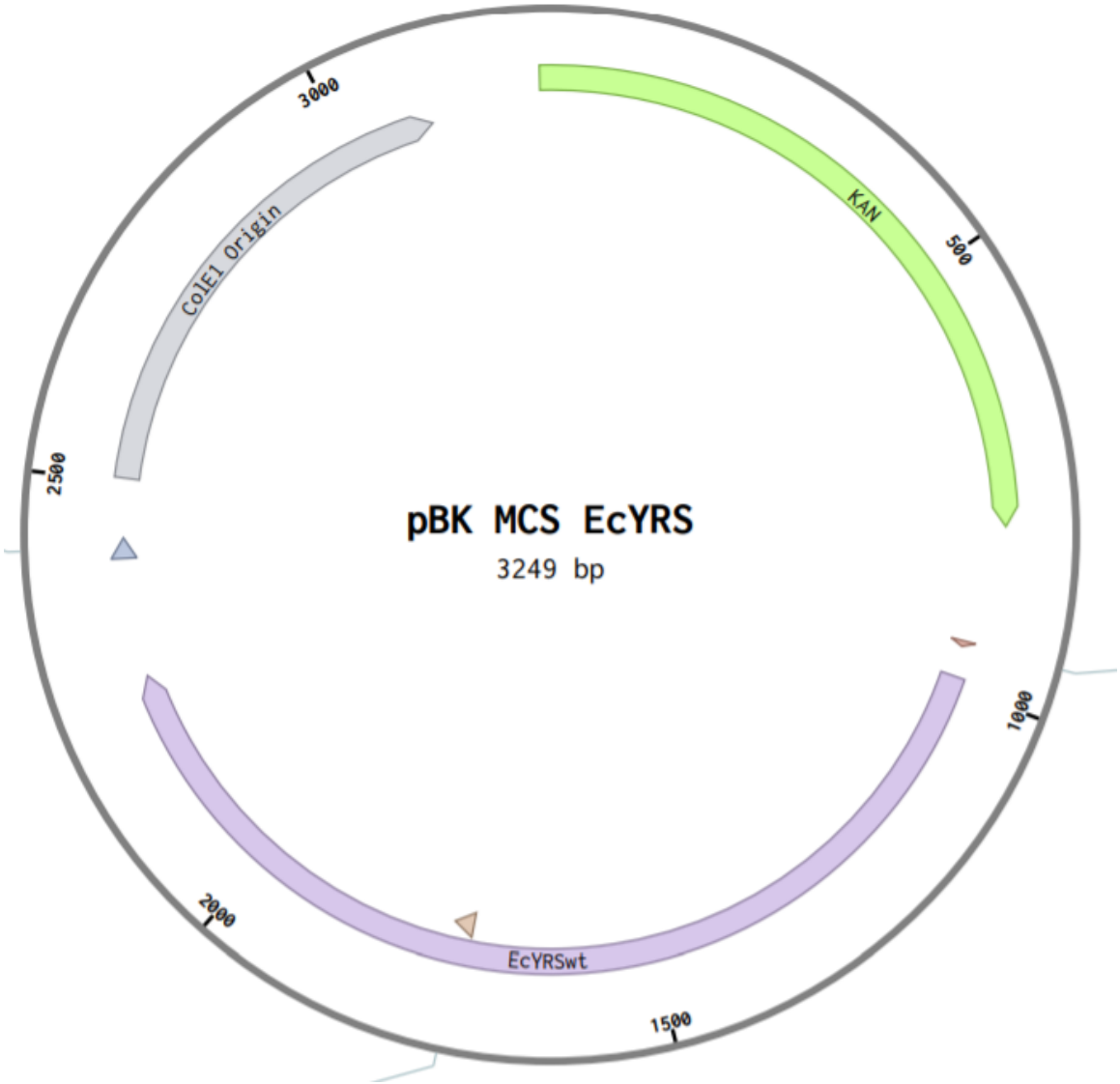
Ec-pBPARS- Gen 3.1 mutations

residue	Mutation
I7	F
Y37	G
G180	S
D182	G
L186	A

Sequence

The sequence for all of these constructs are the same as the pB1U-EcYRS variant shown on page X of this appendix.

pBK EcYRS variants



Ec-pBPARS- Gen 1 mutations

residue	Mutation
Y37	G
D182	G
L186	A
D265	R

Ec-pBPARS- Gen 2 mutations

residue	Mutation
Y37	G
D182	G
L186	A

Ec-pBPARS- Gen 3.1 mutations

residue	Mutation
I7	F
Y37	G
G180	S
D182	G
L186	A

Ec-pBPARS- Gen 3.2 mutations

residue	Mutation
Y37	G
D182	G
L186	A
S366	F

Ec-pBPARS- Gen 3.1-F7I mutations

residue	Mutation
Y37	G
G180	S
D182	G
L186	A

Ec-pBPARS- Gen 3.1-S180G mutations

residue	Mutation
I7	F
Y37	G
D182	G
L186	A

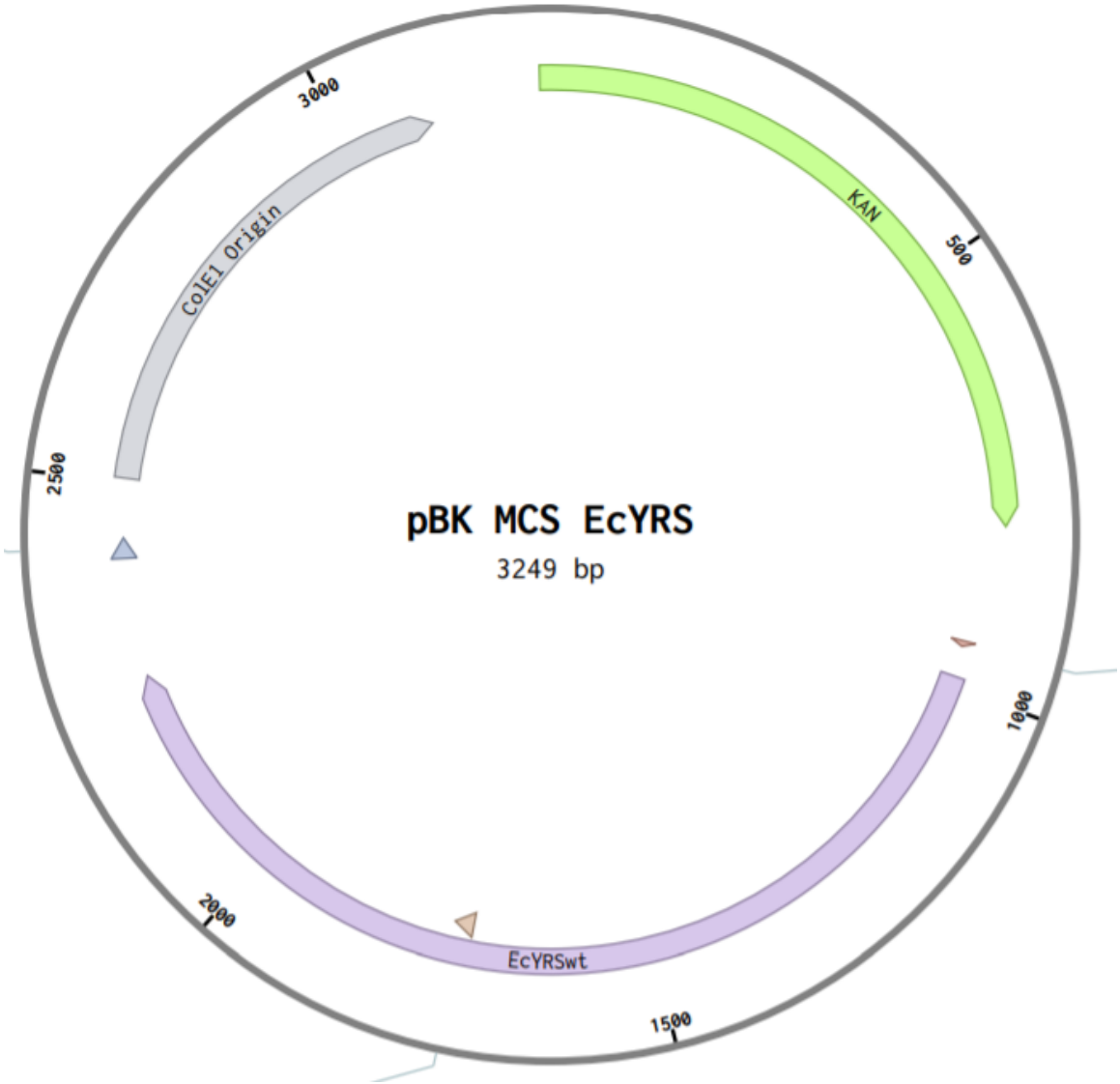
Ec-pBPARS- Gen 3.1-S366F mutations

residue	Mutation
I7	F
Y37	G
G180	S
D182	G
L186	A
S366	F

Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix. The amino acid mutations for the different EcYRS variants are outlined in tables above.

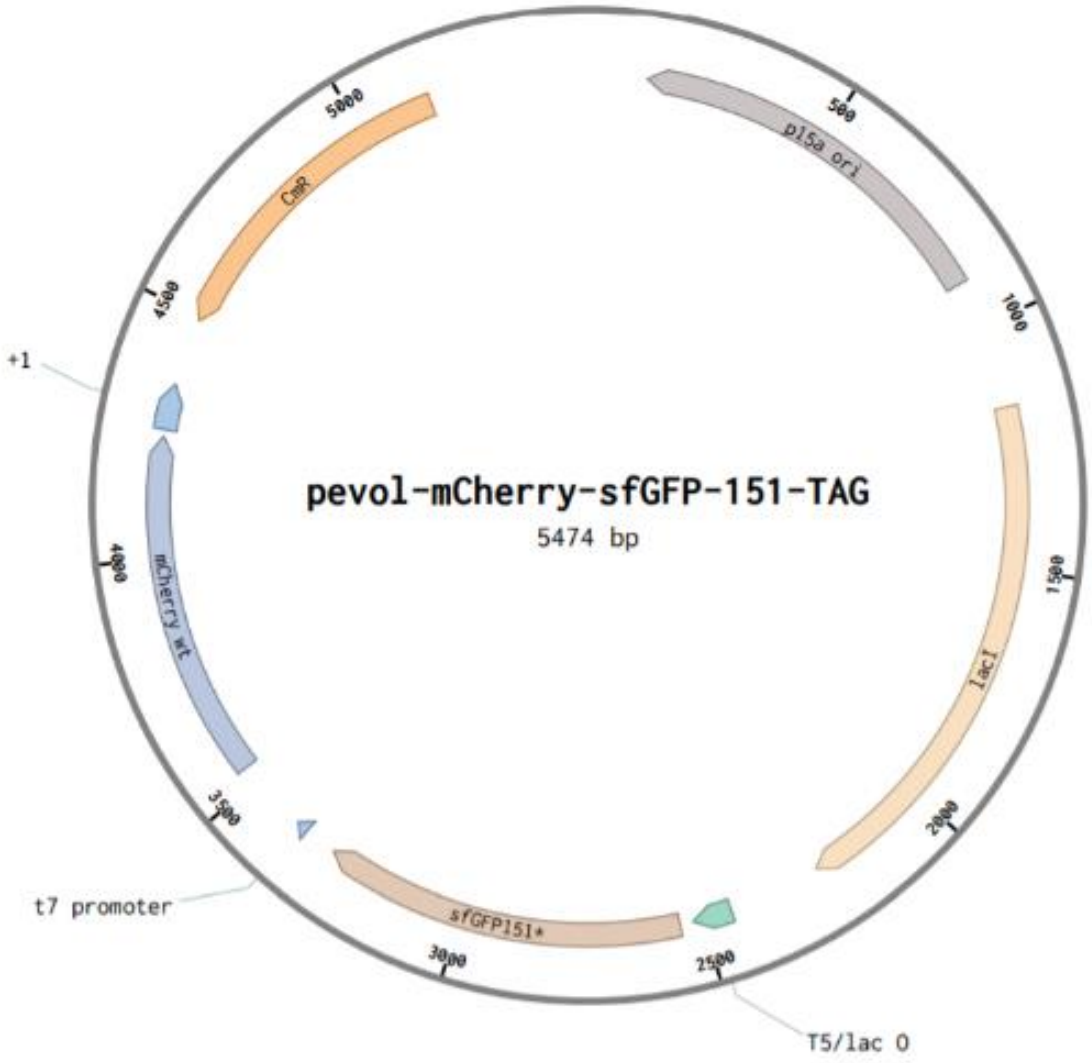
pBK-MCS EcYRS D265R library



Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix.

pEvol-mCherry-sfGFP-151-TAG



Sequence color-coding key

feature	Color
sfGFP151-TAG	text
mCherry WT	text
lac O	text
lambda t0 terminator	text

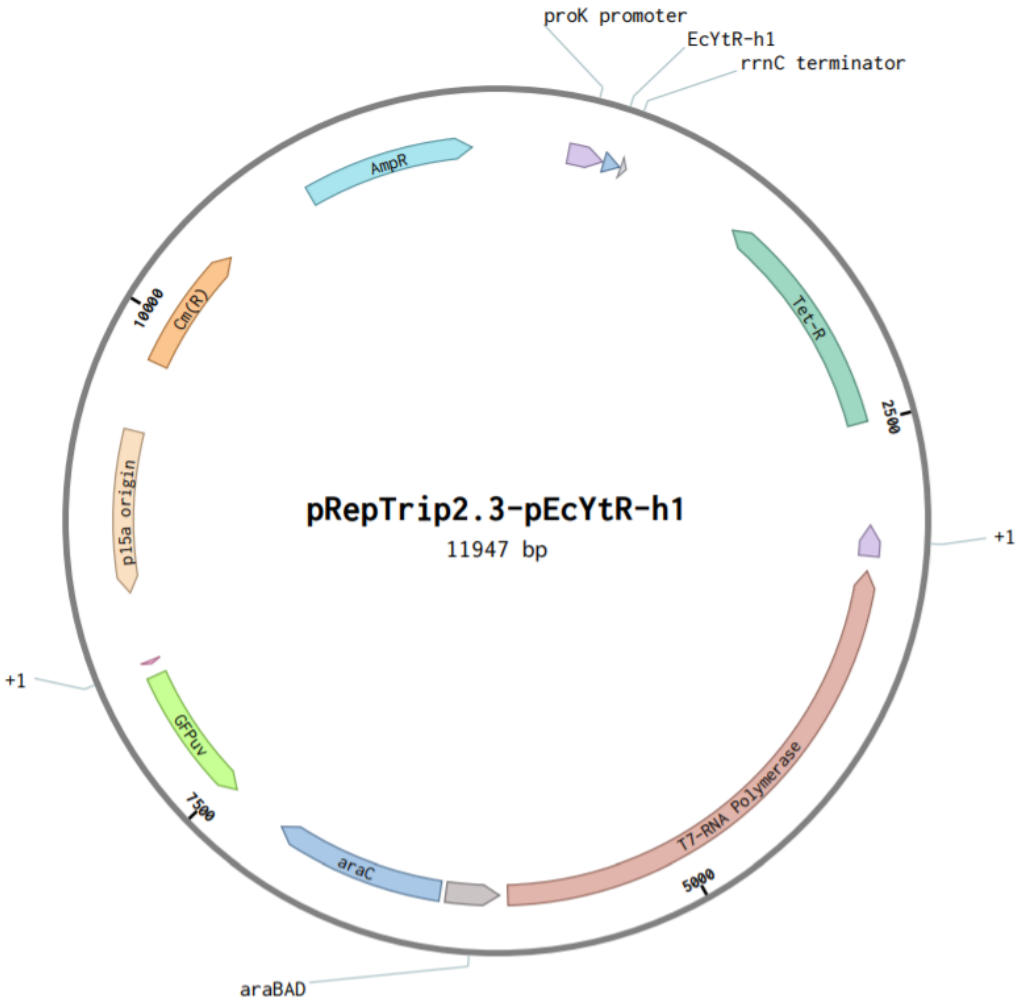
Sequence

This reporter has the same sequence as the previously detailed construct on page X of this appendix. The sfGFP-151-TAG has been swapped out with an sfGFP-151-TAG and mCherry wt reporter sequence shown below.

aaatcataaaaaatttatttgctttgtgagcggataacaattataatagattcaattgtgagcg
gataacaatttcacacagaattcattaaagaggagaaattacatATGAGCAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCT
GTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTA
CTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTT
TTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTAT
GTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGT
TTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAA
CATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATAGATCACGGCAGACAAA
CAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAAC
TAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCA
TTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTT
CTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCCACC
ACCACCACCACCACTAAaagcttaattagctgagctttgttcagaacgctcggttgccgccggg
cgtttttttattggtgagaatggaattgtgagcggataacaattcccctctagagtttgacagca
ttGtcatcgatctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataat
agattcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatATG
GCCgcCaccatggtgagcaagggcgaggaggataacatggccatcatcaaggagttcatgcgct
tcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgagggcgaggg
ccgcccctacgagggcacccagaccgccaagctgaaggtgaccaagggtggcccctgcccttc
gcctgggacatcctgtcccctcagttcatgtacggctccaaggcctacgtgaagcaccccgccg
acatccccgactacttgaagctgtccttccccgagggcttcaagtgggagcgcgtgatgaactt
cgaggacggcggcgtggtgaccgtgacccaggactcctccctgcaggacggcgagttcatctac
aaggtgaagctgcgcggcaccaacttcccctccgacggccccgtaatgcagaagaagaccatgg

gctgggaggcctcctccgagcggatgtaccccgaggacggcgccctgaagggcgagatcaagca
gaggctgaagctgaaggacggcggccactacgacgctgaggtcaagaccacctacaaggccaag
aagcccgtgcagctgcccggcgcctacaacgtcaacatcaagttggacatcacctcccacaacg
aggactacaccatcgtggaacagtacgaacgcgccgagggccgccactccaccggcggcatgga
cgagctgtacaagggcGCTGAGCtTGgactcctgttgatagatccagtaatgacctcagaactc
catctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaat

pRepTrip2.3-pEcYtR-h1

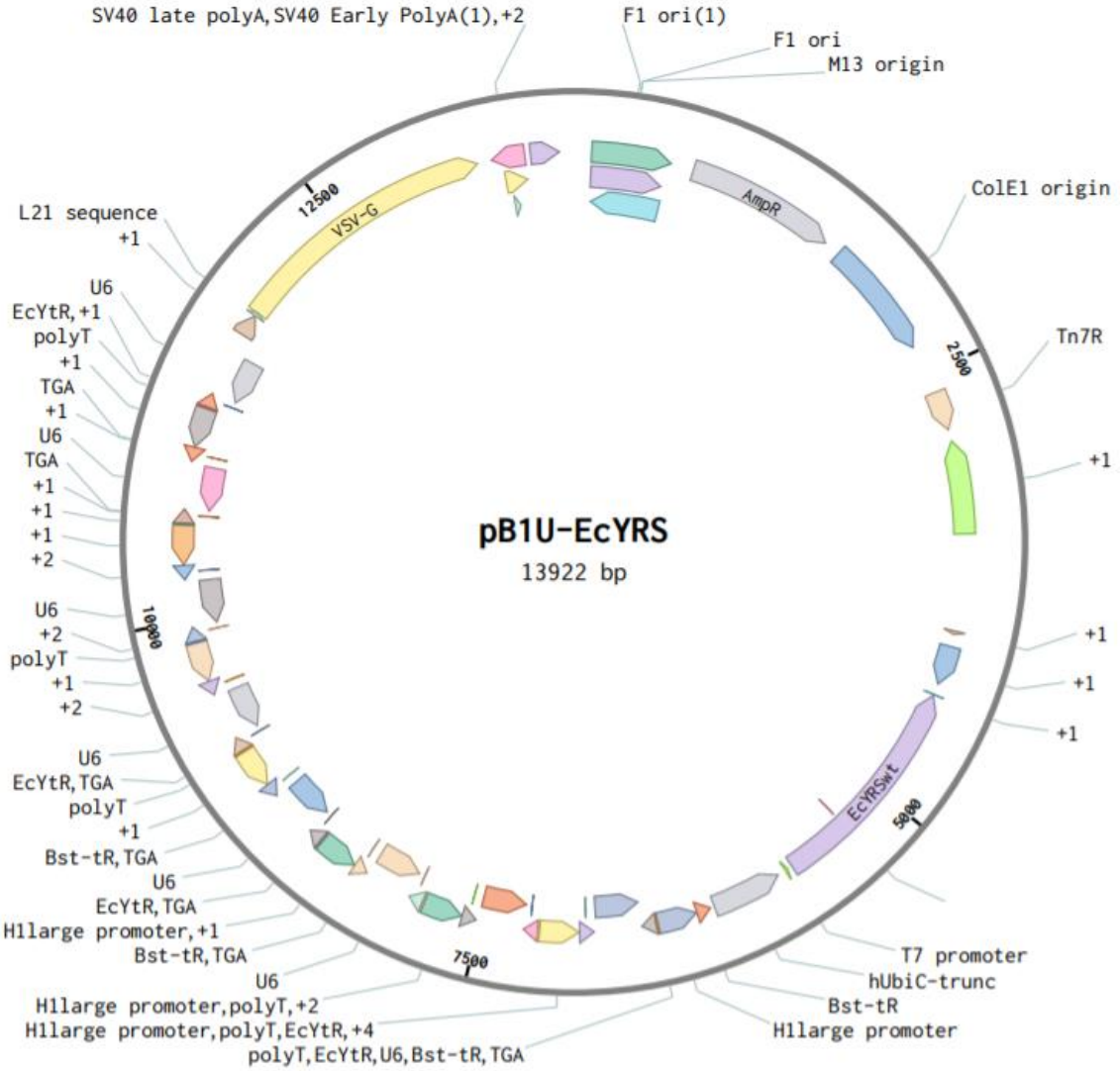


Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix.

Chapter 4 Plasmids

pB1U-EcYRS variants



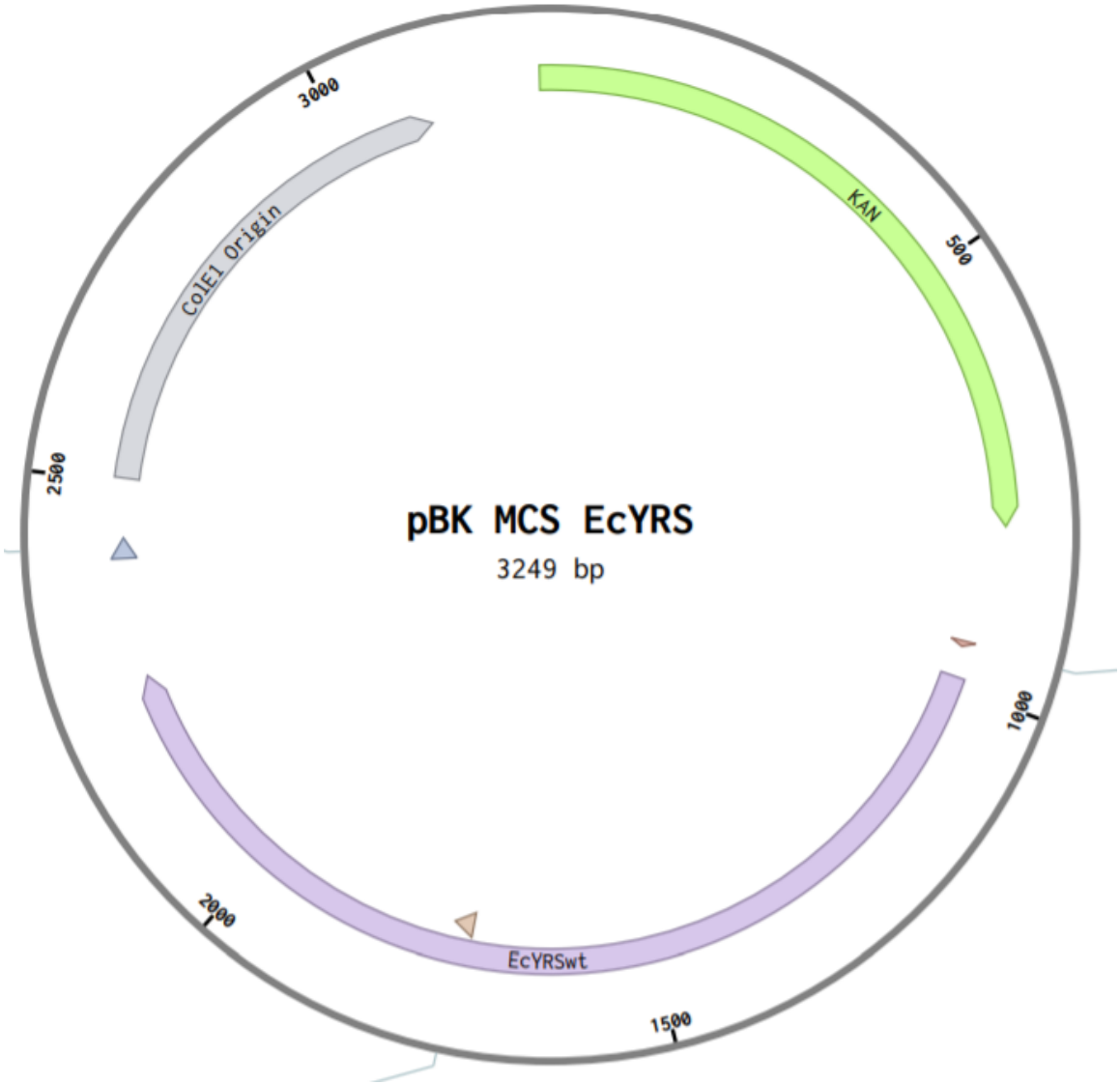
EcYRS mutations

variant	Y37	L71	D182	F183	L186	D265
pAEYRS-11	G	V	S	M	C	R
pAcrFRS-9	G	V	C	Y	C	R
sTyr-RS-1	Y	V	G	Y	I	D
sTyr-RS-2	Y	V	G	Y	L	D
sTyr-RS-3	Y	V	G	Y	L	R
sTyr-RS-4	Y	V	G	F	V	R

Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix. The amino acid mutations for the different EcYRS variants are outlined in tables above.

pBK-MCS EcYRS variants



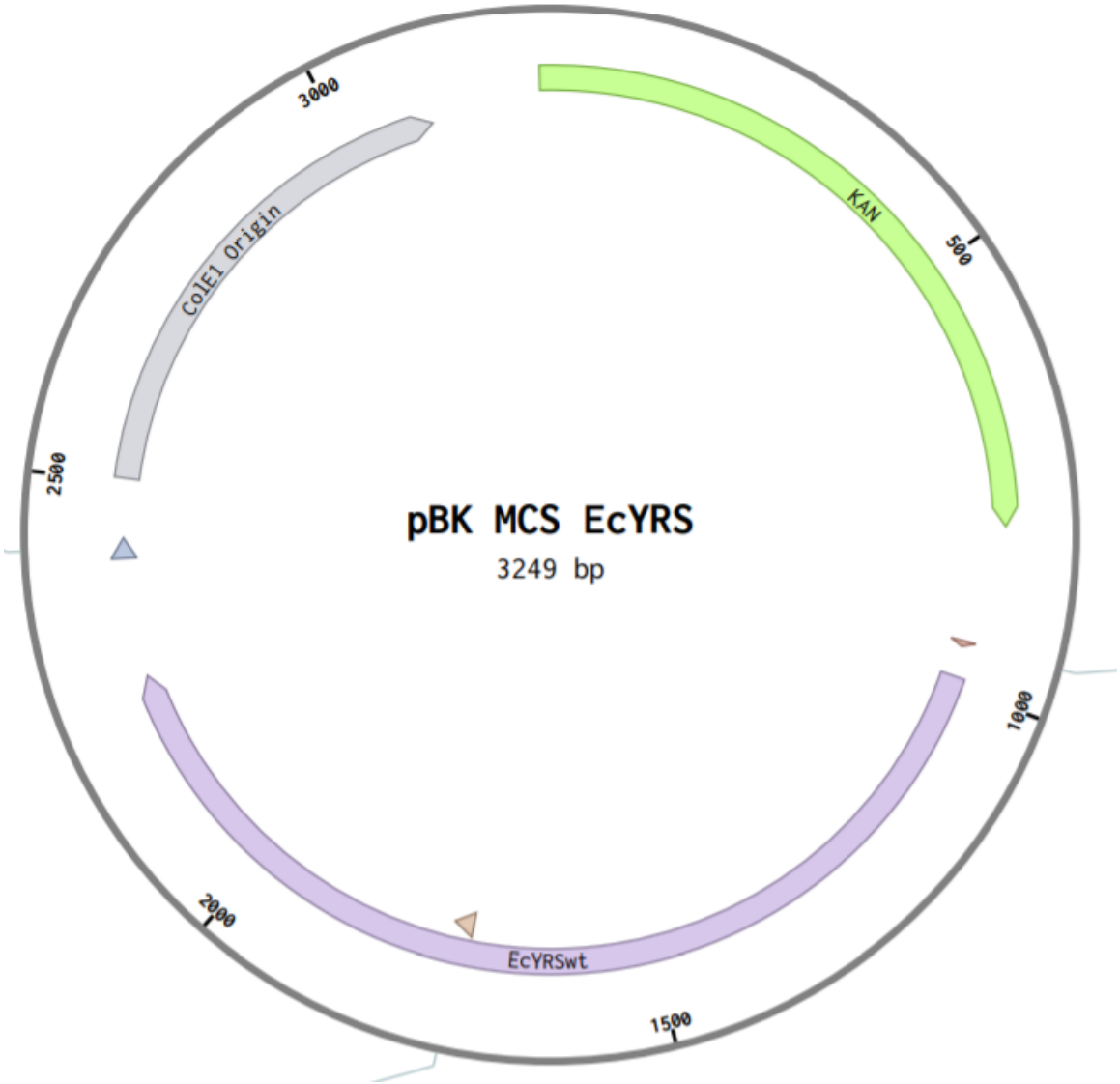
EcYRS mutations

variant	Y37	L71	D182	F183	L186	D265
pAEYRS-11	G	V	S	M	C	R
pAcrFRS-9	G	V	C	Y	C	R
sTyr-RS-1	Y	V	G	Y	I	D
sTyr-RS-2	Y	V	G	Y	L	D
sTyr-RS-3	Y	V	G	Y	L	R
sTyr-RS-4	Y	V	G	F	V	R

Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix. The amino acid mutations for the different EcYRS variants are outlined in tables above.

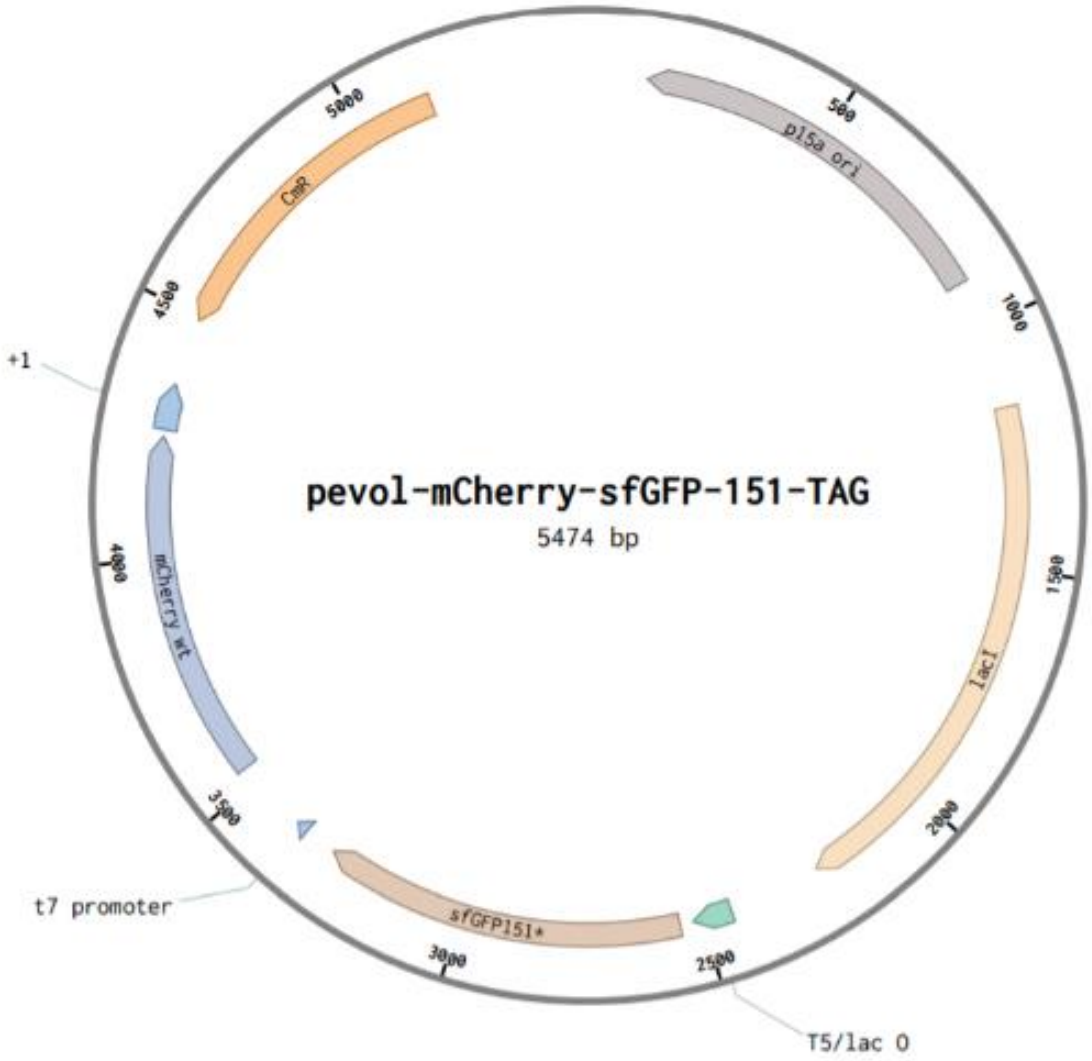
pBK-MCS EcYRS-D265R library



Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix

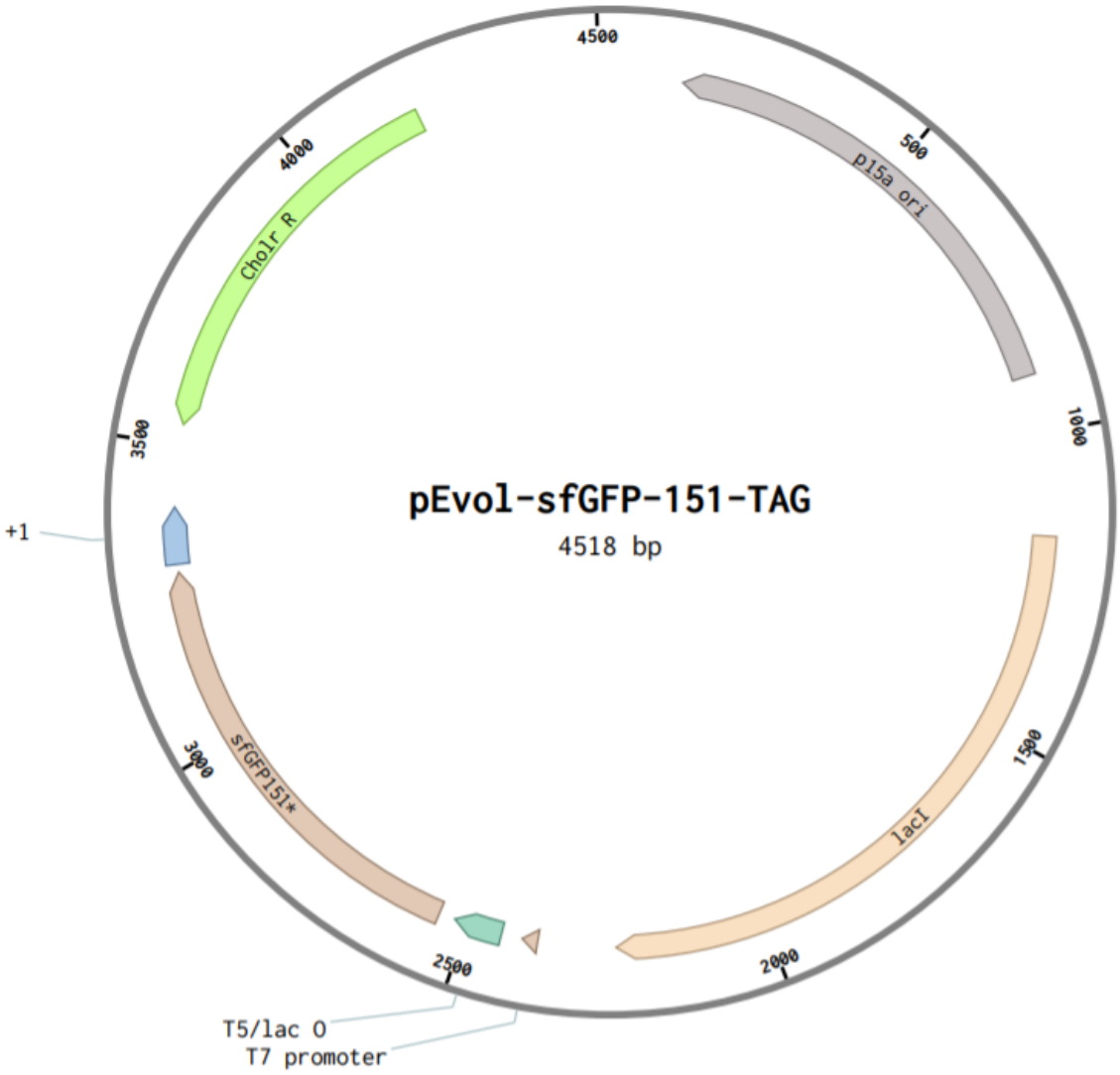
pEvol-mCherry-sfGFP-151-TAG



Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix

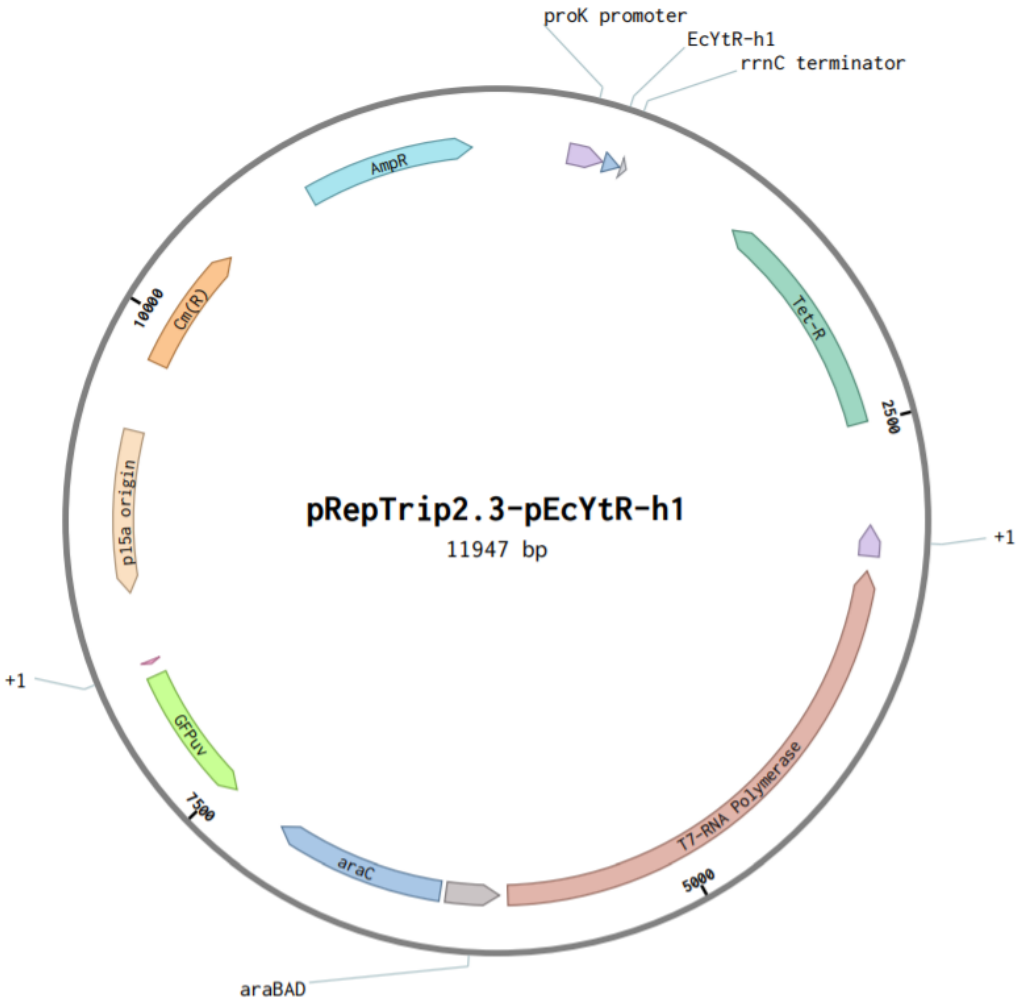
pEvol-sfGFP-151-TAG



Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix

pRepTrip2.3-pEcYtR-h1

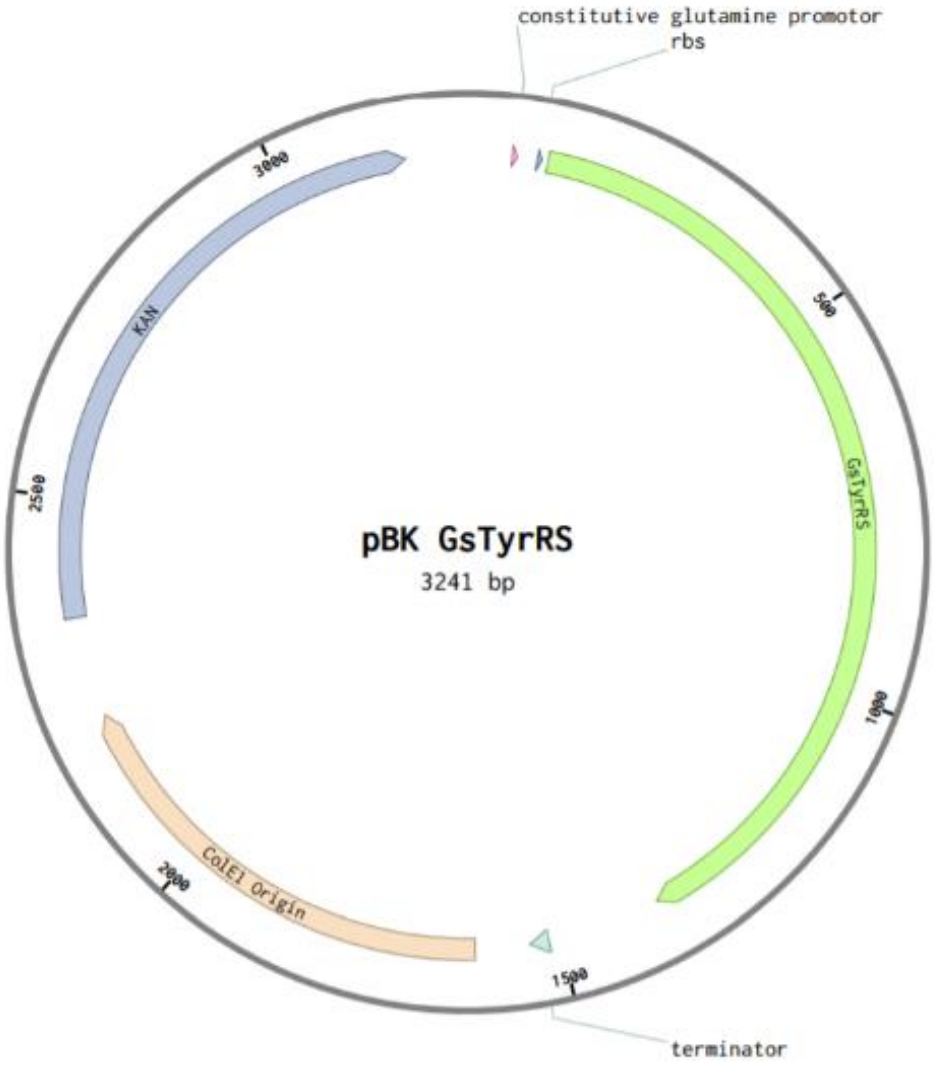


Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix

Chapter 5 Plasmids

pBK GsTyrRS



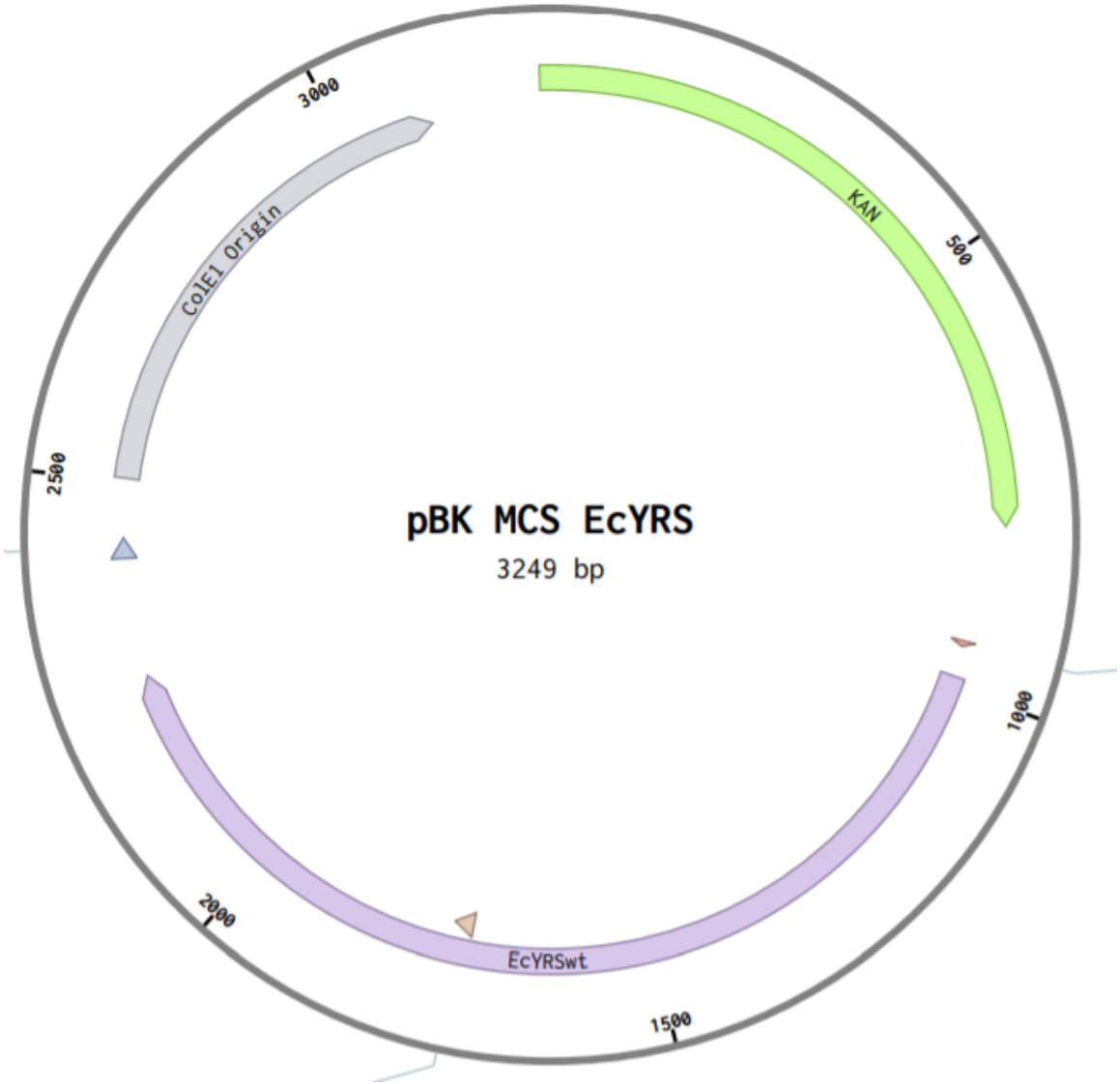
Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix. The only modification is the replacement of the EcTyrRS by the GsTyrRS (sequence below).

ATGGACCTGCTGGCGGAACTGCAATGGCGTGGCCTGGTTAATCAGACCACCGACGAAGATGGCC
TGCGTAAACTGCTGAACGAGGAACGTGTGACCCTGTATTGCGGTTTCGACCCGACCGCGGATAG
CCTGCACATCGGCAACCTGGCGGCGATTCTGACCCTGCGTCGTTTTCAGCAAGCGGGTCACCGT

CCGATCGCGCTGGTTGGTGGTGCGACCGGTCTGATTGGCGACCCGAGCGGCAAGAAAAGCGAGC
GTACCCTGAACGCGAAGGAAACCGTTGAAGCGTGGAGCGCGCGTATCAAAGAACAGCTGGGTCG
TTTCCTGGACTTTGAGGCGGATGGCAACCCGGCGAAGATTAAAAACAACTATGACTGGATCGGT
CCGCTGGATGTGATTACCTTCCTGCGTGATGTGGGCAAGCACTTTAGCGTTAACTACATGATGG
CGAAAGAGAGCGTTCAGAGCCGTATCGAAACCGGTATTAGCTTCACCGAGTTTAGCTACATGAT
GCTGCAAGCGTATGACTTCCTGCGTCTGTACGAAACCGAAGGCTGCCGTCTGCAGATCGGTGGC
AGCGATCAATGGGGTAACATCACCGCGGGCCTGGAACTGATTCGTAAGACCAAAGGTGAAGCGC
GTGCGTTTGGCCTGACCATCCCGCTGGTGACCAAGGCGGACGGTACCAAGTTTGGCAAAACCGA
AAGCGGTACCATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTCTATCAGTTTTGGATC
AACACCGACGATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTCTGAGCAAAGAGGAAA
TCGAAGCGCTGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGCGCAAAAAGCGCTGGC
GGAGGAAGTGACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCGATCCGTATTAGCGAA
GCGCTGTTTAGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAACAAGGCTTCAAGGACG
TGCCGAGCTTTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCTGGTTAGCGCGGGTAT
CAGCCCGAGCAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATTTACGTGAACGGCGAG
CGTCTGCAAGATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAGGTCGTTTTACCGTTA
TCCGTCGTGGCAAGAAGAAATACTATCTGATTCGTTATGCGTAA

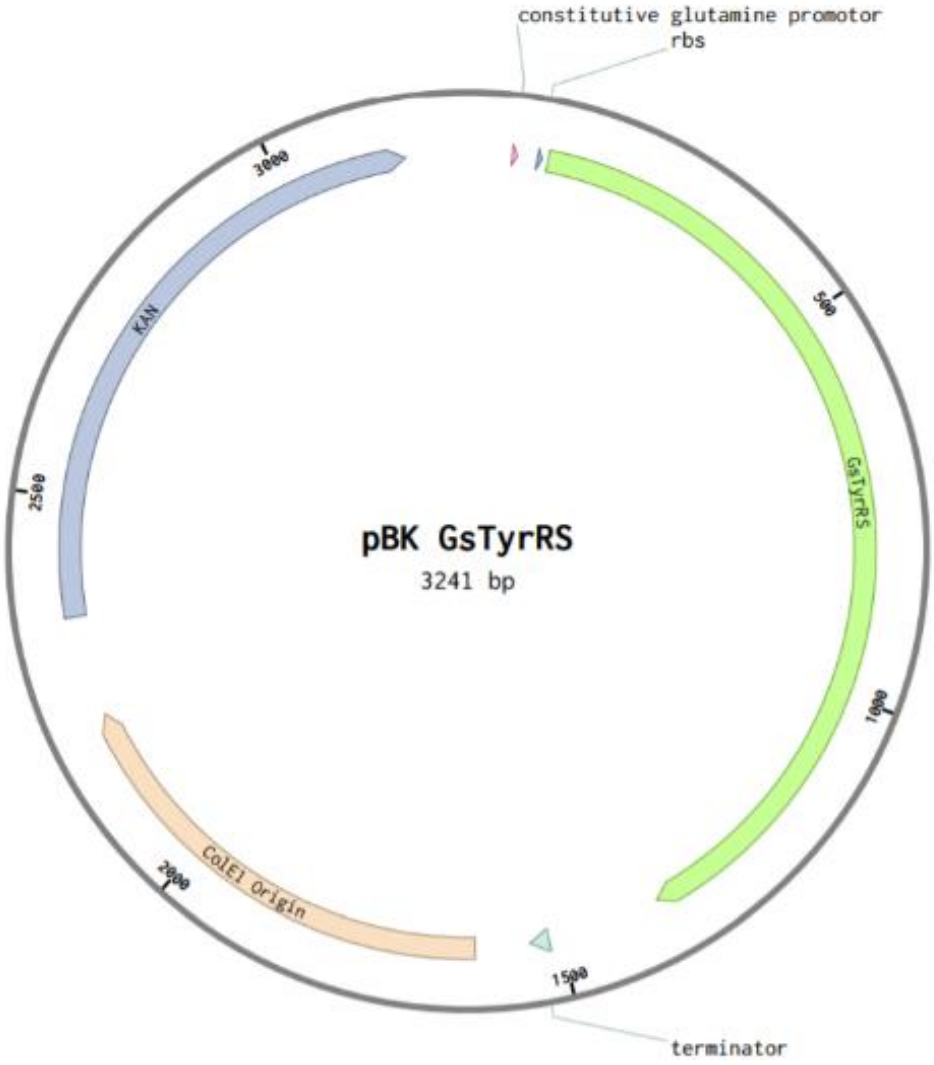
pBK MCS EcYRS



Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix

pBK H2TyrRS



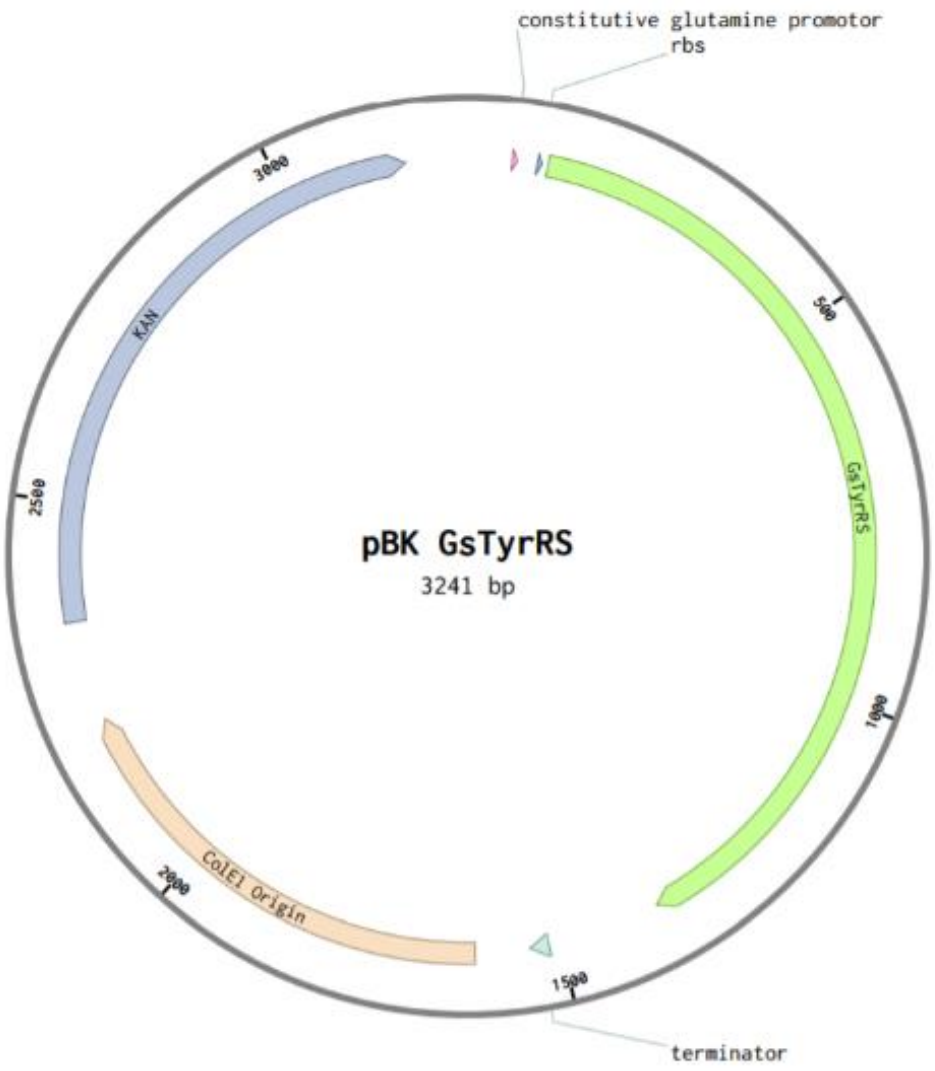
Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix. The only modification is the replacement of the EcTyrRS by the H2TyrRS (sequence below).

atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGTCACCGTCCGATCGCGCTGGTTGGTGGTGCGACCGGTCTGATTGGCGACCCGAGCGGCAAGA
AAAGCGAGCGTACCCTGAACGCGAAGGAAACCGTTGAAGCGTGGAGCGCGCGTATCAAAGAACA
GCTGGGTCGTTTCCTGGACTTTGAGGCGGATGGCAACCCGGCGAAGATTAAAAACAACTATGAC

TGGATCGGTCCGCTGGATGTGATTACCTTCCTGCGTGATGTGGGCAAGCACTTTAGCGTTAACT
ACATGATGGCGAAAGAGAGCGTTCAGAGCCGTATCGAAACCGGTATTAGCTTCACCGAGTTTAG
CTACATGATGCTGCAAGCGTATGATTTCCTGCGTCTGTACGAAACCGAAGGCTGCCGTCTGCAG
ATCGGTGGCAGCGATCAATGGGGTAACATCACCGCGGGCCTGGAACTGATTCGTAAGACCAAAG
GTGAAGCGCGTGCGTTTGGCCTGACCATCCCGCTGGTGACCAAGGCGGACGGTACCAAGTTTGG
CAAAACCGAAAGCGGTACCATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTCTATCAG
TTTTGGATCAACACCGACGATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTCTGAGCA
AAGAGGAAATCGAAGCGCTGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGCGCAAAA
AGCGCTGGCGGAGGAAGTGACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCGATCCGT
ATTAGCGAAGCGCTGTTTAGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAACAAGGCT
TCAAGGACGTGCCGAGCTTTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCTGGTTAG
CGCGGGTATCAGCCCGAGCAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATTTACGTG
AACGGCGAGCGTCTGCAAGATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAGGTCGTT
TTACCGTTATCCGTCGTGGCAAGAAGAAATACTATCTGATTCGTTATGCGTAA

pBK H6TyrRS



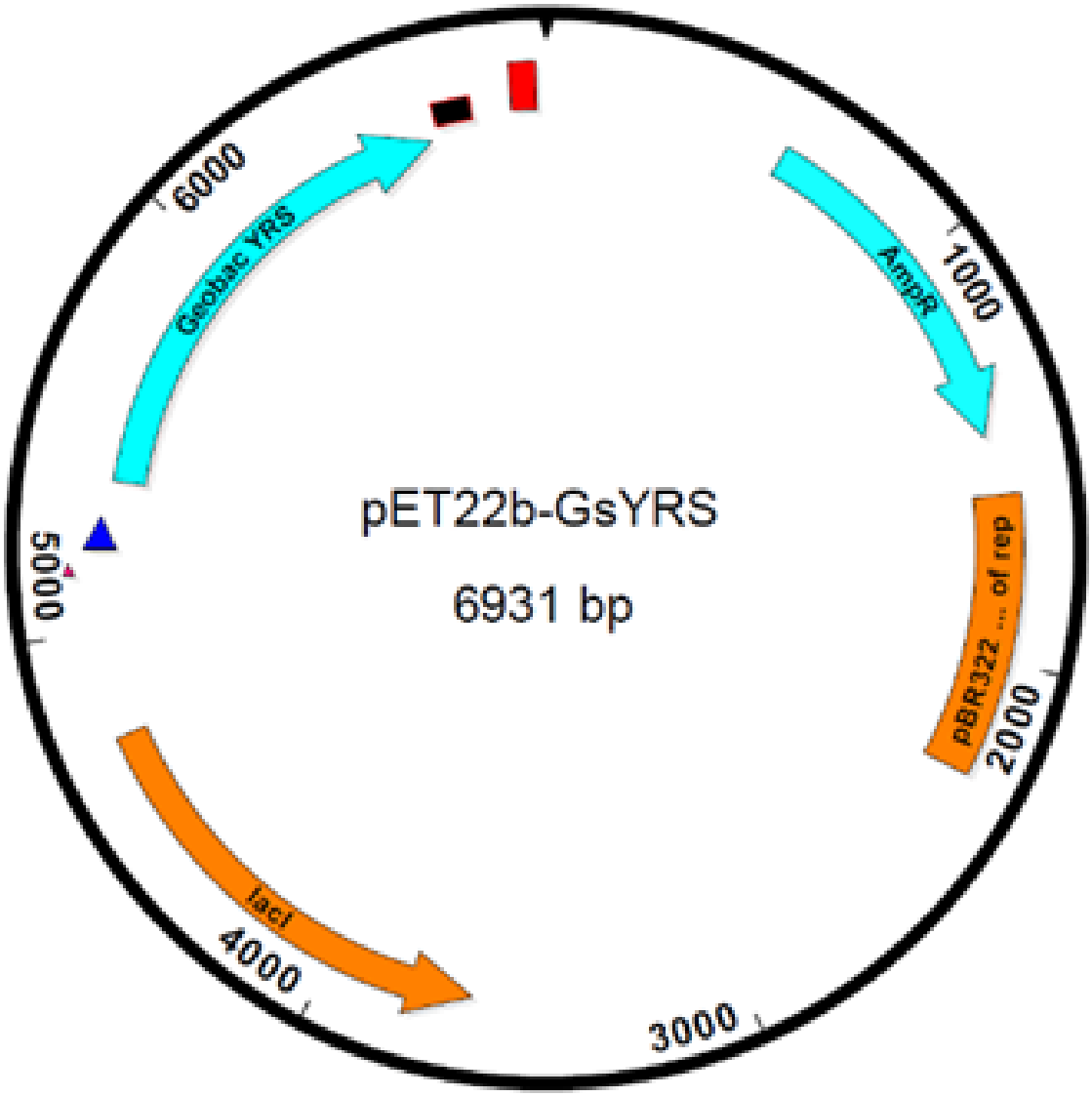
Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix. The only modification is the replacement of the EcTyrRS by the H6TyrRS (sequence below).

atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGTTTCAAAG

CTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCA
GGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAACAACTATGAC
TGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACC
AGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCAC
TGAGTTTTCCTACAGCCTGTTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTG
GTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTC
GTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACTAAAGCAGATGGCACCAA
ATTTGGCAAAACCGAAAGCGGTACCATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTC
TATCAGTTTTGGATCAACACCGACGATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTC
TGAGCAAAGAGGAAATCGAAGCGCTGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGC
GCAAAAAGCGCTGGCGGAGGAAGTGACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCG
ATCCGTATTAGCGAAGCGCTGTTTAGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAAC
AAGGCTTCAAGGACGTGCCGAGCTTTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCT
GGTTAGCGCGGGTATCAGCCCGAGCAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATT
TACGTGAACGGCGAGCGTCTGCAAGATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAG
GTCGTTTTACCGTTATCCGTCGTGGCAAGAAGAAATACTATCTGATTCGTTATGCGTAA

pET22bNtermHis-aaRS



Sequence color-coding key

feature	Color
Amp resistance	text
pBR322 ori	text
lacI	text
GsTyrRS	text

Sequence

tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagc
gtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcg
ccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttag
tgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcg
ccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgt
tccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgcc
gatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaa
atattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtt
tatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttca
ataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttt
gcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaag
atcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagag
ttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggta
ttatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgact
tggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatg
cagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagga
ccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttggg
aaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggc
aacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaata
gactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggt
ttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggcc
agatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaa
cgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaag
tttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaa
gatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtca
gaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgct
tgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactct
ttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccg

tagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgt
taccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagtt
accggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga
acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaag
ggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagct
tccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgt
cgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttt
tacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattc
tgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgag
cgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatc
tgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatag
ttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgcc
aacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtg
accgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagc
tgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtc
cagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagg
gcggtttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaa
tgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggtt
actggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcact
cagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatc
ctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacga
aacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcg
cttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagccta
gccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataat
ggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgc
aagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgc
cgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcat
aagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctc
aagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgcg
tggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagac
cgtttcccgcgtggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcg
gcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgt
tgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgcaaattgtcgcggcgat
taaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtc
gaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcattaact
atccgctggatgaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatt
tcttgatgtctctgaccagacacccatcaacagtattattttctcccatgaagacggtacgcga
ctgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggcccattaa
gttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaatcaaattca
gccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatg
ctgaatgagggcatcgttcccactgcgatgctggttgccaacgatcagatggcgctgggcgcaa
tgcgcgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacga
taccgaagacagctcatgttatatcccgccgttaaccaccatcaaacaggattttcgcctgctg
gggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatcagc

tgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaatacgcaaaccgcctctcc
ccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcag
tgacctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagt
aggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaaca
gtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtg
gcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcg
ccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaat
acgactcactatagggaattgtgagcggataacaattcccctctagagtttgacagcattatc
atcgatctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataatagatt
caattgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatatgcatca
tcaccatcaccatcatcatcatcacgcaagcagtaacttgattaaacaattgcaagagGACCTG
CTGGCGGAACTGCAATGGCGTGGCCTGGTTAATCAGACCACCGACGAAGATGGCCTGCGTAAAC
TGCTGAACGAGGAACGTGTGACCCTGTATTGCGGTTTCGACCCGACCGCGGATAGCCTGCACAT
CGGCAACCTGGCGGCGATTCTGACCCTGCGTCGTTTTCAGCAAGCGGGTCACCGTCCGATCGCG
CTGGTTGGTGGTGCGACCGGTCTGATTGGCGACCCGAGCGGCAAGAAAAGCGAGCGTACCCTGA
ACGCGAAGGAAACCGTTGAAGCGTGGAGCGCGCGTATCAAAGAACAGCTGGGTCGTTTCCTGGA
CTTTGAGGCGGATGGCAACCCGGCGAAGATTAAAAACAACTATGACTGGATCGGTCCGCTGGAT
GTGATTACCTTCCTGCGTGATGTGGGCAAGCACTTTAGCGTTAACTACATGATGGCGAAAGAGA
GCGTTCAGAGCCGTATCGAAACCGGTATTAGCTTCACCGAGTTTAGCTACATGATGCTGCAAGC
GTATGACTTCCTGCGTCTGTACGAAACCGAAGGCTGCCGTCTGCAGATCGGTGGCAGCGATCAA
TGGGGTAACATCACCGCGGGCCTGGAACTGATTCGTAAGACCAAAGGTGAAGCGCGTGCGTTTG
GCCTGACCATCCCGCTGGTGACCAAGGCGGACGGTACCAAGTTTGGCAAAACCGAAAGCGGTAC
CATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTCTATCAGTTTTGGATCAACACCGAC
GATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTCTGAGCAAAGAGGAAATCGAAGCGC
TGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGCGCAAAAAGCGCTGGCGGAGGAAGT
GACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCGATCCGTATTAGCGAAGCGCTGTTT
AGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAACAAGGCTTCAAGGACGTGCCGAGCT
TTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCTGGTTAGCGCGGGTATCAGCCCGAG
CAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATTTACGTGAACGGCGAGCGTCTGCAA
GATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAGGTCGTTTTACCGTTATCCGTCGTG
GCAAGAAGAAATACTATCTGATTCGTTATGCGTAAaagcttaattagctgagcttggactcctg
ttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgcc
gggcgtttttattggtgagaatccaagctagcttggcggcggccgcactcgagcaccaccacc
accaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgc
tgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttgctgaaa
ggaggaactatatccggat

In addition to the GsTyrRS, the following aaRS sequences, and their mutant pBPA variants, were also put into this construct:

EcTyrRS:

atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAAG
CTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCA
GGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAACAACTATGAC
TGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACC
AGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCAC
TGAGTTTTCCTACAACCTGTTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTG
GTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTC
GTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACTAAAGCAGATGGCACCAA
ATTTGGTAAAACTGAAGGCGGCGCAGTCTGGTTaGATCCGAAGAAAACCAGCCCGTACAAATTC
TACCAGTTCTGGATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCACCTTTA
TGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAACAGCGGTAAAGCACCGCGCGC
CCAGTATGTACTGGCGGAGCAGGTGACTCGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCA
AAACGTATTACCGAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCGAAC
AGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCGCAGACCTGATGCAGGCACT
GGTCGATTCTGAACTGCAACCTTCCCGTGGTCAGGCACGTAAAACTATCGCCTCCAATGCCATC
ACCATTAACGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTG
GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGCTGGAAAtaa

H2TyrRS:

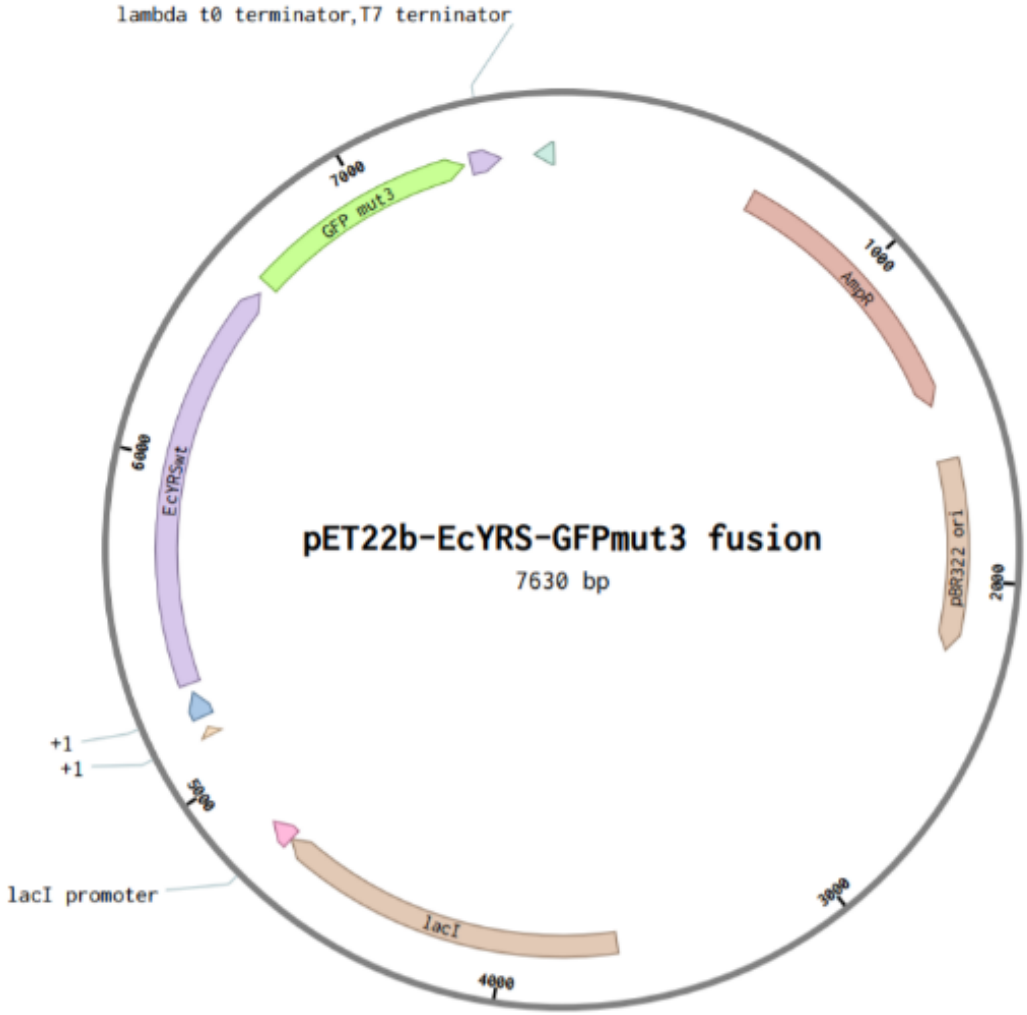
atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGTCACCGTCCGATCGCGCTGGTTGGTGGTGCGACCGGTCTGATTGGCGACCCGAGCGGCAAGA
AAAGCGAGCGTACCCTGAACGCGAAGGAAACCGTTGAAGCGTGGAGCGCGCGTATCAAAGAACA
GCTGGGTCGTTTCCTGGACTTTGAGGCGGATGGCAACCCGGCGAAGATTAAAAACAACTATGAC
TGGATCGGTCCGCTGGATGTGATTACCTTCCTGCGTGATGTGGGCAAGCACTTTAGCGTTAACT
ACATGATGGCGAAAGAGAGCGTTCAGAGCCGTATCGAAACCGGTATTAGCTTCACCGAGTTTAG
CTACATGATGCTGCAAGCGTATGATTTCCTGCGTCTGTACGAAACCGAAGGCTGCCGTCTGCAG
ATCGGTGGCAGCGATCAATGGGGTAACATCACCGCGGGCCTGGAACTGATTCGTAAGACCAAAG
GTGAAGCGCGTGCGTTTGGCCTGACCATCCCGCTGGTGACCAAGGCGGACGGTACCAAGTTTGG
CAAAACCGAAAGCGGTACCATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTCTATCAG
TTTTGGATCAACACCGACGATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTCTGAGCA
AAGAGGAAATCGAAGCGCTGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGCGCAAAA
AGCGCTGGCGGAGGAAGTGACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCGATCCGT
ATTAGCGAAGCGCTGTTTAGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAACAAGGCT
TCAAGGACGTGCCGAGCTTTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCTGGTTAG
CGCGGGTATCAGCCCGAGCAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATTTACGTG

AACGGCGAGCGTCTGCAAGATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAGGTCGTT
TTACCGTTATCCGTCGTGGCAAGAAGAAATACTATCTGATTCGTTATGCGTAA

H6TyrRS:

atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGTTTCAAAG
CTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCA
GGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAACAACTATGAC
TGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACC
AGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCAC
TGAGTTTTCCTACAGCCTGTTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTG
GTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTC
GTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACTAAAGCAGATGGCACCAA
ATTTGGCAAAACCGAAAGCGGTACCATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTC
TATCAGTTTTGGATCAACACCGACGATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTC
TGAGCAAAGAGGAAATCGAAGCGCTGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGC
GCAAAAAGCGCTGGCGGAGGAAGTGACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCG
ATCCGTATTAGCGAAGCGCTGTTTAGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAAC
AAGGCTTCAAGGACGTGCCGAGCTTTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCT
GGTTAGCGCGGGTATCAGCCCGAGCAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATT
TACGTGAACGGCGAGCGTCTGCAAGATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAG
GTCGTTTTACCGTTATCCGTCGTGGCAAGAAGAAATACTATCTGATTCGTTATGCGTAA

pET22b-aaRS-GFP mut3



Sequence color-coding key

feature	Color
Amp resistance	text
pBR322 ori	text
lacI	text
EcTyrRS	text
GFPmut3	text
linker sequence	text

Sequence

tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagc
gtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcg
ccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttag
tgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcg
ccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgt
tccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgcc
gatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaa
atattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtt
tatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttca
ataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttt
gcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaag
atcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagag
ttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggta
ttatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgact
tggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatg
cagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagga
ccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttggg
aaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggc
aacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaata
gactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggt
ttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggcc
agatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaa
cgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaag
tttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaa
gatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtca
gaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgct
tgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactct
ttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccg

tagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgt
taccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagtt
accggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcga
acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaag
ggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagct
tccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgt
cgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttt
tacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattc
tgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgag
cgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatc
tgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatag
ttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgcc
aacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtg
accgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagc
tgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtc
cagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagg
gcggtttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaa
tgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggtt
actggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcact
cagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatc
ctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacga
aacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcg
cttcacgttcgctcgcgtatcggtgattcattctgctaaccagtaaggcaaccccgccagccta
gccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgccggcgataat
ggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgc
aagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgc
cgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaagacagtcat
aagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctc
aagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgct
cactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgc
ggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaa
cagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgc
cccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcgg
tatcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcg
cattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattc
agcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcg
gctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacaga
acttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgccc
agtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaa
gaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcgg
atagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggct
tcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatt
taatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcag
caacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatc

gccgcttccacttttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaa
cggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcac
caccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagt
aggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaaca
gtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtg
gcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcg
ccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaat
acgactcactatagggaattgtgagcggataacaattcccctctagagtttgacagcattatc
atcgatctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataatagatt
caattgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatatgGCAAG
CAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCG
TTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTACCGCTGACA
GCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAA
GCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGAG
CGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCC
CGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAACAACTATGACTGGTTCGG
CAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAGATGATC
AACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTT
CCTACAACCTGTTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTGGTGCTGCA
AATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTGCAT
CAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACTAAAGCAGATGGCACCAAATTTGGTA
AAACTGAAGGCGGCGCAGTCTGGTTaGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAGTT
CTGGATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCACCTTTATGAGCATT
GAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAACAGCGGTAAAGCACCGCGCGCCCAGTATG
TACTGGCGGAGCAGGTGACTCGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGTAT
TACCGAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCGAACAGCTGGCG
CAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCGCAGACCTGATGCAGGCACTGGTCGATT
CTGAACTGCAACCTTCCCGTGGTCAGGCACGTAAAACTATCGCCTCCAATGCCATCACCATTAA
CGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTGGTCGTTTT
ACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGCTGGAAAGGATCCGCTGGCTCCG
CTGCTGGTTCTGGCGAATTCCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGT
TGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCA
ACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAA
CACTTGTCACTACTTTCGGTTATGGTGTTCAATGCTTTGCGAGATACCCAGATCACATGAAACA
GCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAA
GATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAA
TCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAATTGGAATACAA
CTATAACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTC
AAAATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTC
CAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTAAACTTTC
GAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATT
ACACATGGCATGGATGAACTATACAAATAAaagcttaattagctgagcttggactcctgttgat
agatccagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcg

```
ttttttattggtgagaatccaagctagcttggcggcggccgcactcgagcaccaccaccaccac
cactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagc
aataactagcataaccccttggggcctctaaacgggtcttgaggggtttttgctgaaaggagg
aactatatccggat
```

In addition to the EcTyrRS, the following variants were cloned into this construct.

Ec-pBPARS- Gen 1 mutations

residue	Mutation
Y37	G
D182	G
L186	A
D265	R

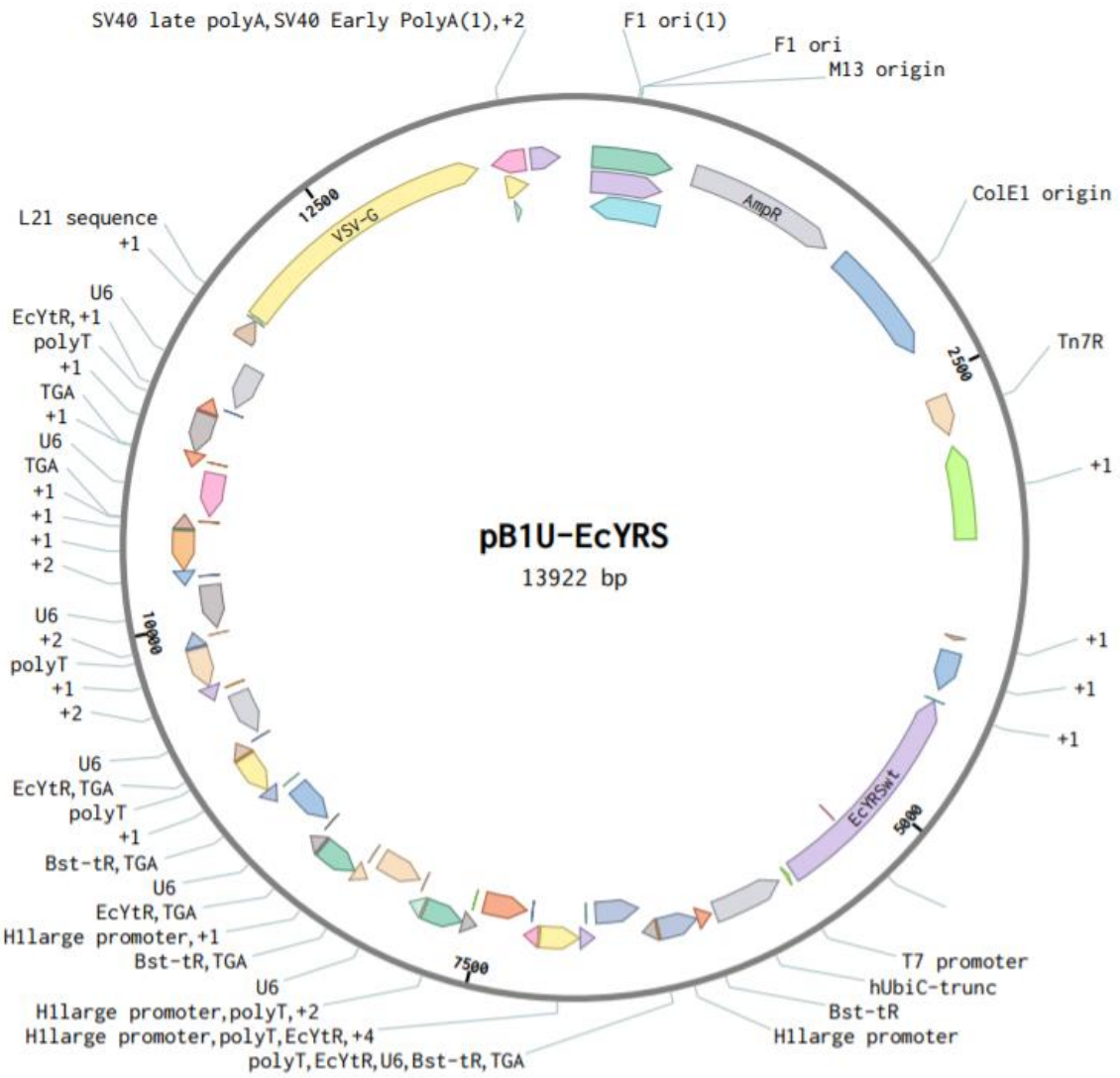
Ec-pBPARS- Gen 3.1 mutations

residue	Mutation
I7	F
Y37	G
G180	S
D182	G
L186	A

Ec-pBPARS- Polyspecific mutations

residue	Mutation
Y37	V
D182	S
F183	M
L186	A

pB1U-NtermHis-aaRSvariant



Sequence

The sequence for this construct is the exact same as the one illustrated on page X of this appendix. The only modification is the addition of an N-term Histidine tag:

`atgCATCATCACCATCACCATCATCATCATCAC`

The following aaRS sequences, and their mutant pBPA variants, were also put into this construct:

EcTyrRS:

atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAAG
CTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCA
GGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAACAACTATGAC
TGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACC
AGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCAC
TGAGTTTTCCTACAACCTGTTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTG
GTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTC
GTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACTAAAGCAGATGGCACCAA
ATTTGGTAAAACTGAAGGCGGCGCAGTCTGGTTaGATCCGAAGAAAACCAGCCCGTACAAATTC
TACCAGTTCTGGATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCACCTTTA
TGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAACAGCGGTAAAGCACCGCGCGC
CCAGTATGTACTGGCGGAGCAGGTGACTCGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCA
AAACGTATTACCGAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCGAAC
AGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCGCAGACCTGATGCAGGCACT
GGTCGATTCTGAACTGCAACCTTCCCGTGGTCAGGCACGTAAAACTATCGCCTCCAATGCCATC
ACCATTAACGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTG
GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGCTGGAAAtaa

GsTyrRS:

atgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgttt
ttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggg
ttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttt
ccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggc
aagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcac
agaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagt
gataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttt
tgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccat
accaaaacgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactatta
actggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaag
ttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagc
cggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatc
gtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgaga
taggtgcctcactgattaagcattggtaa

H2TyrRS:

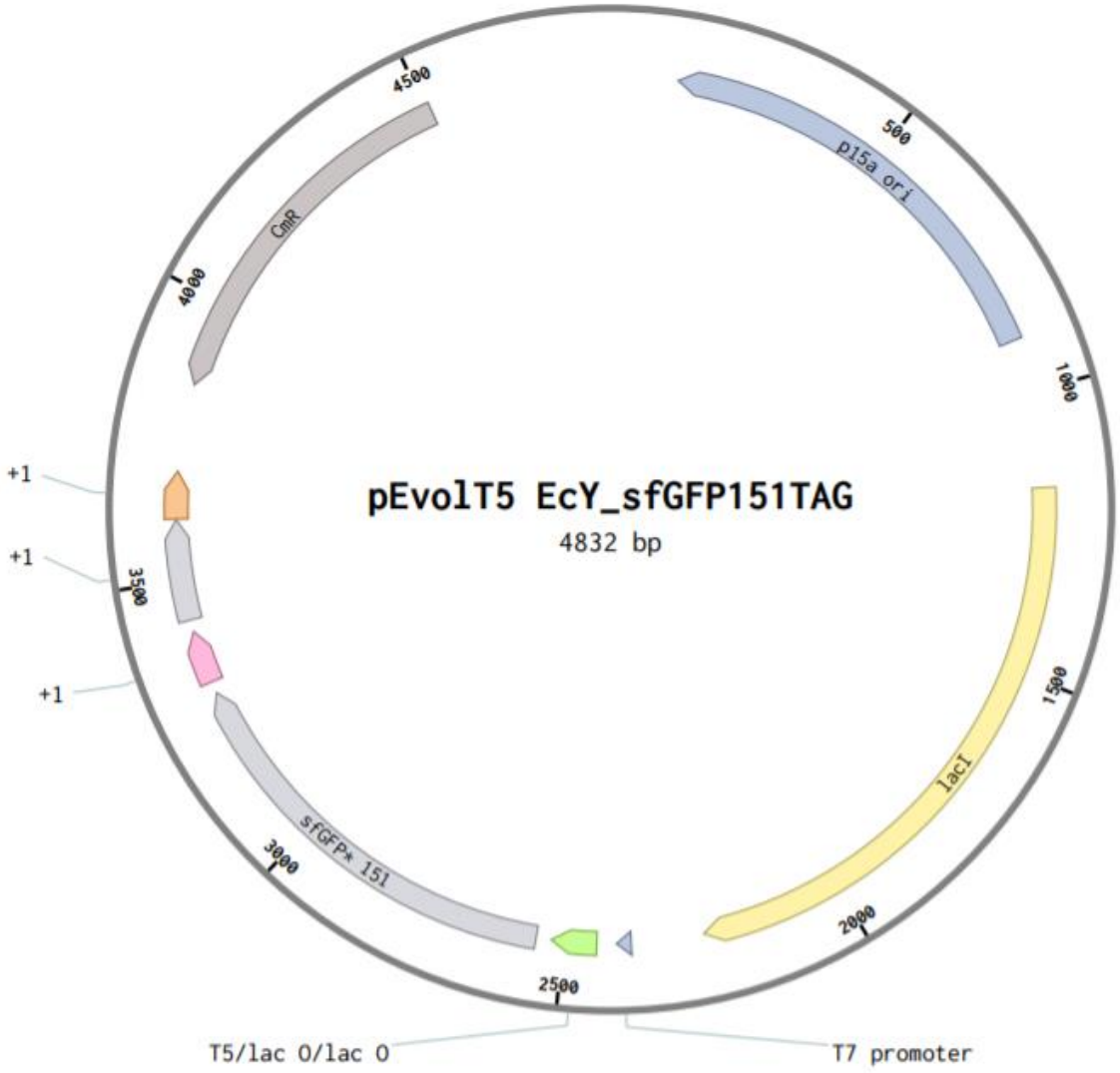
atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGTCACCGTCCGATCGCGCTGGTTGGTGGTGCGACCGGTCTGATTGGCGACCCGAGCGGCAAGA
AAAGCGAGCGTACCCTGAACGCGAAGGAAACCGTTGAAGCGTGGAGCGCGCGTATCAAAGAACA
GCTGGGTCGTTTCCTGGACTTTGAGGCGGATGGCAACCCGGCGAAGATTAAAAACAACTATGAC
TGGATCGGTCCGCTGGATGTGATTACCTTCCTGCGTGATGTGGGCAAGCACTTTAGCGTTAACT
ACATGATGGCGAAAGAGAGCGTTCAGAGCCGTATCGAAACCGGTATTAGCTTCACCGAGTTTAG
CTACATGATGCTGCAAGCGTATGATTTCCTGCGTCTGTACGAAACCGAAGGCTGCCGTCTGCAG
ATCGGTGGCAGCGATCAATGGGGTAACATCACCGCGGGCCTGGAACTGATTCGTAAGACCAAAG
GTGAAGCGCGTGCGTTTGGCCTGACCATCCCGCTGGTGACCAAGGCGGACGGTACCAAGTTTGG
CAAAACCGAAAGCGGTACCATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTCTATCAG
TTTTGGATCAACACCGACGATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTCTGAGCA
AAGAGGAAATCGAAGCGCTGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGCGCAAAA
AGCGCTGGCGGAGGAAGTGACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCGATCCGT
ATTAGCGAAGCGCTGTTTAGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAACAAGGCT
TCAAGGACGTGCCGAGCTTTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCTGGTTAG
CGCGGGTATCAGCCCGAGCAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATTTACGTG
AACGGCGAGCGTCTGCAAGATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAGGTCGTT
TTACCGTTATCCGTCGTGGCAAGAAGAAATACTATCTGATTCGTTATGCGTAA

H6TyrRS:

atgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGCCCAGGTGACGGACG
AGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCGCTCTATTGCGGCTTCGATCCTAC
CGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCG
GGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGTTTCAAAG
CTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCA
GGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAACAACTATGAC
TGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACC
AGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCAC
TGAGTTTTCCTACAGCCTGTTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTG
GTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTC
GTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACTAAAGCAGATGGCACCAA
ATTTGGCAAAACCGAAAGCGGTACCATTTGGCTGGATAAGGAGAAAACCAGCCCGTACGAATTC
TATCAGTTTTGGATCAACACCGACGATCGTGACGTTATTCGTTACCTGAAGTATTTCACCTTTC
TGAGCAAAGAGGAAATCGAAGCGCTGGAGCAGGAACTGCGTGAGGCGCCGGAAAAGCGTGCGGC
GCAAAAAGCGCTGGCGGAGGAAGTGACCAAACTGGTTCACGGTGAGGAAGCGCTGCGTCAGGCG
ATCCGTATTAGCGAAGCGCTGTTTAGCGGTGATATCGCGAACCTGACCGCGGCGGAGATTGAAC
AAGGCTTCAAGGACGTGCCGAGCTTTGTTCACGAAGGTGGCGATGTGCCGCTGGTTGAGCTGCT
GGTTAGCGCGGGTATCAGCCCGAGCAAACGTCAGGCGCGTGAAGACATCCAAAACGGTGCGATT

TACGTGAACGGCGAGCGTCTGCAAGATGTTGGCGCGATTCTGACCGCGGAACACCGTCTGGAAG
GTCGTTTTACCGTTATCCGTCGTGGCAAGAAGAAATACTATCTGATTCGTTATGCGTAA

pEvolT5-EcY-TAG-sfGFP-151-TAG



Sequence color-coding key

feature	Color
sfGFP151-TAG	text
chlor resistance	text
lac O	text
lambda t0 terminator	text
lacI	text
EcYtR	text
p15a ori	text
proK promoter	text

Sequence

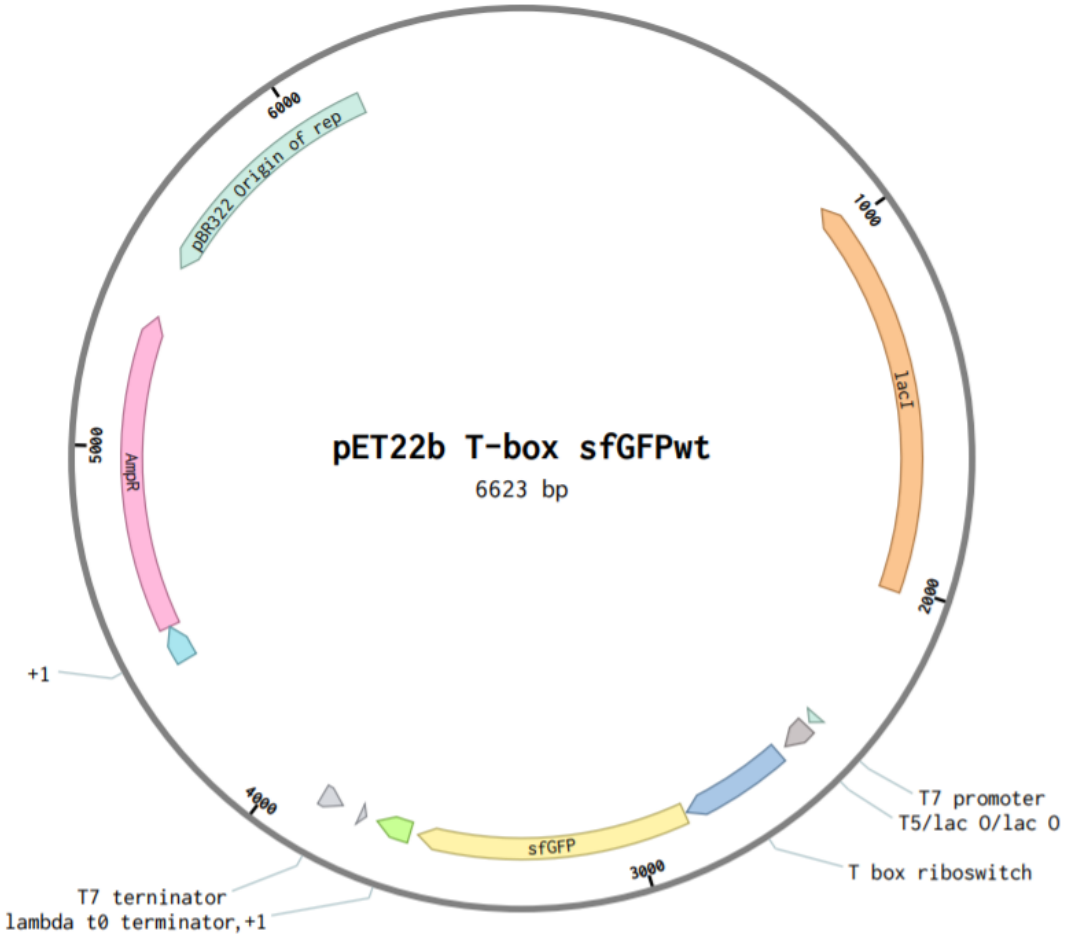
gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttac
tatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcacc
ggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgc
tcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatg
ccaggaagatacttaacagggaagtgagagggccgcggcaaagccgttttccataggctccgc
ccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactat
aaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtt
taccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgg
gtaggcagttcgctccaagctggactgtatgcacgaacccccgttcagtccgaccgctgcgcc
ttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcag
ccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaa
ggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctca
gagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagagcaagagattacgcgc
agaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgc
aatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatg
tttgacagcttatcatcgataagcttgcaatttatctcttcaaatgtagcacctgaagtcagcc
ccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgact
gggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacat
taattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatg
aatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcgg

tccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataac
atgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccgga
ctcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtggga
acgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgcctt
cccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcag
acgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgacc
agatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtct
ggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatc
ctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcacc
gccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagtt
gatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggt
ggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaa
ttcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggt
tcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttac
tggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaag
gttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagg
ctcactatagggaattgtgagcggataacaattcccctctagagtttgacagcattatcatcg
atctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataatagattcaat
tgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatATGAGCAAAGGA
GAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACA
AATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTAT
TTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTT
CAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCG
AAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGA
AGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAA
GATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATAGATCACGG
CAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTC
CGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCA
GACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACA
TGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGG
ATCCCACCACCACCACCACCACTAAaagcttaattagctgagcttggactcctgttgatagatc
cagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgttttt
tattggtgagaatccaagctagcttggcgCTGCAgtgtgcttctcaaatgcctgaggccagttt
gctcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggc
ctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagggcgTATCTgc
gcagtaagatgcgccccgcatttGGTGGGGTTCCCGAGCGGCCAAAGGGAGCAGACTCTAAATC
TGCCGTCATCGACTTCGAAGGTTCGAATCCTTCCCCCACCACCAAATTCGAAAAGCCTGCTCAA
CGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCTTTCGAatttctgccatt
catccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgcct
taaaaaaattacgccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgc
cgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtc
gccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacg
tttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaacatattctcaataa
acccttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtag

aaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatgg
aaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcattgccatac
ggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtg
cttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggta
cattgagcaactgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacgg
tggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactc
aaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccga
tcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggat
ttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtga
tgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttc
tatcagctgtccctcctgttcagctactgacg

Chapter 6 Plasmids

pET22b T-box sfGFPwt



Sequence color-coding key

feature	Color
sfGFPwt	text
amp resistance	text
lac O	text
lambda t0 terminator	text
lacI	text
T-box	text
pBR322 ori	text

Sequence

tcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtct
ggcttctgataaagcgggccatgttaagggcggtttttcctgtttggtcactgatgcctccgt
gtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgata
cgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtat
ggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaatacagatgt
aggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggc
gctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctc
aggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctg
ctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgc
acccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggac
cagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcat
cgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcct
acgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccacc
ggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagt
gagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgc
cagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtg
gtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagaga
gttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaa
cggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcacca
acgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaacca
gcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggc
actccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccag
ccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggt
gacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatact
gttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttcc

acagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcga
gaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccac
gctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagg
gccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgc
ggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaac
gtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgaca
tcgtataacgttactggtttcacattcaccacctgaattgactctcttccgggcgctatcatg
ccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcg
actcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaa
tggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacg
ccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcga
tataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagag
gatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataaca
attcccctctagatcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataa
tagattcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattaAGGTT
CATATGAAAAGGTAAAGATTGAGACAAGTAGAATATCCTTACGTTCCAGAGAGCTGATGGCCGG
TGAAAATCAGCACAGACGGATATATCGAATACACTCATGAACCGCTTTTGCAAACAAAGCCGGC
CAGGCTTTCAGTAGTGAAAGAACGGACCTGATCCGTTATCAGGCAAAGTGATAAGACGAATGTT
TGCATTCTCTTATTAGTAGGGTGGTACCGCGATAATCAATCGTCCCTTCGTGTAAACGAAGGGG
CGTTTTTTATTTTaattaaaAAAGGAGctttatcttCATATGAGCAAAGGAGAAGAACTTTTCA
CTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCG
TGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGA
AAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCtatGGTGTTCAATGCTTTTCCC
GTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACA
GGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAA
GGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTC
TTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAA
GAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCA
GACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACC
TGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGA
GTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCCACCACCAC
CACCACCACTAAaagcttaattagctgagcttggactcctgttgatagatccagtaatgacctc
agaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttattggtgagaat
ccaagctagcttggcggcggccgcactcgagcaccaccaccaccaccactgagatccggctgct
aacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccc
ttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggc
gaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtga
ccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccac
gttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgct
ttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccct
gatagacggttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttcca
aactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatt
tcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatat
taacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatt

tttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataa
tattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcgg
cattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatca
gttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagtttt
cgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattat
cccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggt
tgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagt
gctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccga
aggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaacc
ggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggcaaca
acgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagact
ggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttat
tgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagat
ggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaa
atagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagttta
ctcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatc
ctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagacc
ccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgca
aacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttt
ccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagt
taggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttacc
agtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccg
gataaggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacga
cctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggag
aaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttcca
gggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgat
ttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacg
gttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtg
gataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgca
gcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtg
cggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaa
gccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaaca
cccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccg
tctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcg
gtaaagctcatcagcgtggtcgtgaagcgat

Appendix II.
Oligonucleotide primers

Chapter 2 Oligos

Primer name	Oligonucleotide sequence (5'→3')
BKrep-BglII-R	AATAATAagatctGCTCCTTAGATCTTCCT AGGACCATTCC
BKrep-NcoI-R	AATAATAccatggGCTCCTTAGATCTTCCT AGGACCATTCC
BKrep-SpeI-F	AATAATAactagtATTACGCTGACTTGACG GGACGG
BKrep-BamHI-F	AATAATAggatccGCGCTTGTTTCGGCGTG GGTATG
EcYRS-D265R-F	GATCAACACTGCGCGTGCCGACGTTTACCG CTTCCTGAAGTTCTTCAC
EcYRS-D265R-R	GTGAAGAACTTCAGGAAGCGGTAAACGTCG GCACGCGCAGTGTTGATC
EcYRS-NheI-F	TTTGAGGAATCCGCTAGCGCAAGCAGTAAC TTGATTAAACAATTGCAAGAG
EcYRS-XhoI-R	AATTCTCGAGTTATTTCCAGCAAATCAGAC ACTAATTC
JI MCS sqR	GAGATCATGTAGGCCTGATAAGCGTAGC
mutiF	cAAAGGGAGCAGACTCTAAATCTGCCGTCA TCGACTTCGAAGGTTCGAATCCTTCC**NNNN N**CACCACTGCAGATCCTTAGCGAAAGCTAA G
mutiR	cTTCGAAGTCGATGACGGCAGATTTAGA GTCTGCTCCCTTTGGCCGCTCGGGAAC**N NNNN**CGAATTCAGCGTTACAAGTATTAC ACAAAGTTTTTTATG
NEGrep-SphI-F	AATAATAgcatg**c**TCGAACTTTTGCTGAGT TGAAGG
pBKseqtF	AATAATA**gcatgc**TCGAACTTTTGCTGAGT TGAAGG

Chapter 3 Oligos

Primer name	Oligonucleotide sequence (5'→3')
pBPARS-F7I-F	ATGGCAAGCAGTAACTTGATTAAACAATTG CAAGAGCGGG
pBPARS-S180G-F	GTTTTCCTACAACCTGTTGCAGGGTTATGG GTTC
pBPARS-S366F-F	GATGCAGGCACTGGTCGATTTGAACTGCAA CCTTCCCGTG
EcoNI-t7 promoter-F	ATATTATATcctgcattaggTCGATCCCGC GAAATTAATACGACTCACTATAG
EcoNI-lambda t0 term-R	ATATTATATcctgcattaggGCCAAGCTAG CTTGGATTCTCACCAATAAAAAACGC

Chapter 4 Oligos

Same as chapter 2.

Chapter 5 Oligos

Primer	Oligonucleotide sequence (5'→3')
EcYRS-HindII-R	AATTAAGCTTTTATTTCCAGCAAATCAGAC ACTAATTC
EcYRS-NdeI-F	TTTGAGGAATCCCATATGGCAAGCAGTAAC TTGATTAAACAATTGCAAGAG
EcYRS-nterm10XHis-NdeI-F	GAAATTACATATGCATCATCACCATCACCA TCATCATCATCACGCAAGCAGTAACTTGAT TAAACAATTGCAAGAG
EcYRS-XhoI-R	AATTCTCGAGTTATTTCCAGCAAATCAGAC ACTAATTC
GeobacYRS-D176G-F	GTTTAGCTACATGATGCTGCAAGCGTATGG TTTCCTGCGTCTGTACGAAACCGAAGGCTG
GeobacYRS-D176G-R	CAGCCTTCGGTTTCGTACAGACGCAGGAAA CCATACGCTTGCAGCATCATGTAGCTAAAC
GeobacYRS-GGFL-L180A-F	CAAGCGTATGGTTTCCTGCGTGCCTACGAA ACCGAAGGCTGCCGTCTG
GeobacYRS-GGFL-L180A-R	CAGACGGCAGCCTTCGGTTTCGTAGGCACG CAGGAAACCATACGCTTG
GeobacYRS-Y34G-R	GCTATCCGCGGTCGGGTCGAAACCGCAACC CAGGGTCACACGTTCCTCGTTCAGC
GeobacYS-D176G-F	CAAGCGTATGGTTTCCTGCGTGCCTACGAA ACCGAAGGCTGCCGTCTG
GsYRS NcoI-R	GCAGCCATGGTACCTTACGCATAACGAATC AGATAGTATTTC
GsYRS NdeI-F	CAAGCGTATGGTTTCCTGCGTGCCTACGAA ACCGAAGGCTGCCGTCTG
GsYRS-HindIII-R	AATTAAGCTTTTACGCATAACGAATCAGAT AGTATTTC
GsYRSnterm-10XHis-NdeI-F	GAAATTACATATGCATCATCACCATCACCA TCATCATCATCACGACCTGCTGGCGGAACT GCAATGG
GsYRS-XhoI-R	AATTCTCGAGTTACGCATAACGAATCAGAT AGTATTTC
H2 EcYRS-iR	GCGATCGGACGGTGACCCGCCTGCTGGAAG CGTTTCAGGCATAACAATG
H2 GsYRS-iF	CCTGAAACGCTTCCAGCAGGCGGGTCACCG TCCGATCGCGCTGGTTG
H6 EcYRS-iR	CCGCTTTCGGTTTTGCCAAATTTGGTGCCA TCTGCTTTAGTGATCAGCGGAACG
H6 GsYRS-iF	CACTAAAGCAGATGGCACCAAATTTGGCAA AACCGAAAGCGGTACCATTTG

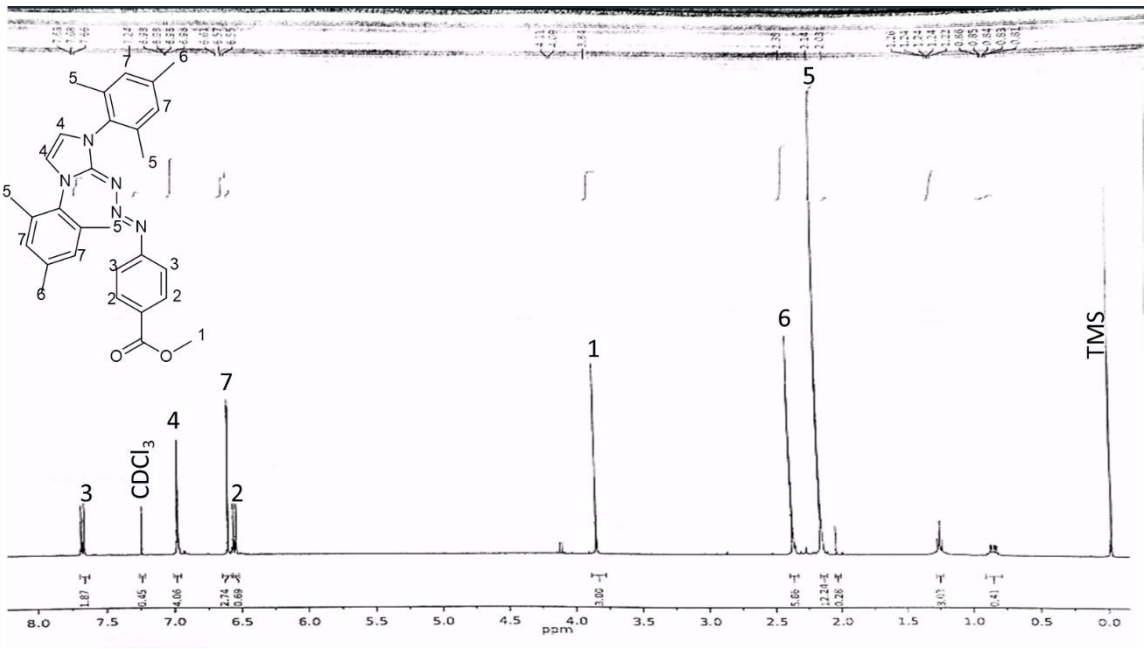
MjYRS-HindIII-R	AATTAAGCTTTTATAATCTCTTTCTAATTG GCTCTAAAATC
MjYRS-nterm-10XHis-NdeI-F	GAAATTACATATGCATCATCACCATCACCA TCATCATCATCACgacgaatttgaaatgat aaagagaaacacatctg
MjYRS-XhoI-R	ATTAACTCGAGttataatctctttctaatt ggctctaaaatctttataagttcttc
EcYRS-linker-GFPmut3-iR	CAGCAGCGGAGCCAGCGGATCCTTTCCAGC AAATCAGACAGTAATTC
GsYRS-GFPmut3-iR	GCAGCGGAGCCAGCGGATCCCGCATAACGA ATCAGATAGTATTTCTTC
Linker-GFPmut3-Nterm-iF	GATCCGCTGGCTCCGCTGCTGGTTCTGGCG AATTCCGTAAAGGAGAAGAACTTTTCACTG
GFPmut3-HindIII-R	AATAATAAAGCTTTTATTTGTATAGTTCAT CCATGCCATGTGTAATC

Chapter 6 Oligos

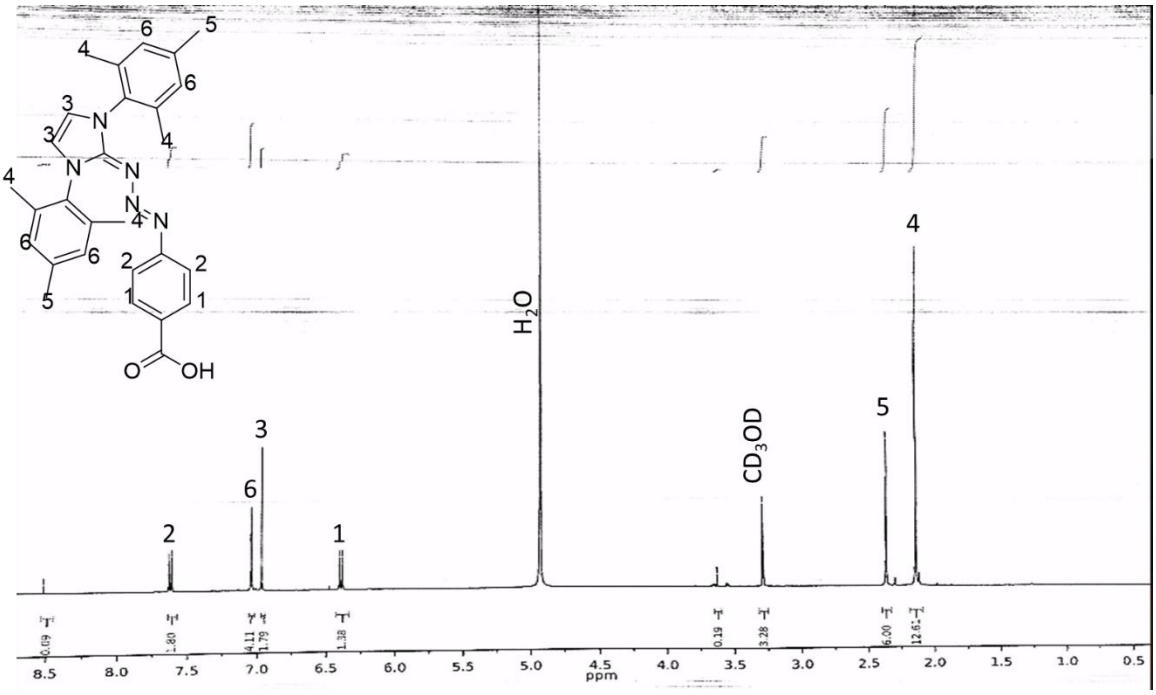
Primer	Oligonucleotide sequence (5'→3')
pET22b-vector-F	GTTTTTTATTTTaattaaaAAAGGAGctttatcttcatATGAGCAAAGGAGAAGAACTTT
pET22b-vector-R	CTCAATCTTTACCTTTTCATATGAACCTtaatttctcctctttaatgaattctgtgtg
Tbox-F	cacacagaattcattaaagaggagaaattaAGGTTCATATGAAAAGGTAAAGATTGAG
Tbox-R	AAGTTCTTCTCCTTTGCTCATatgaagataaagCTCCTTTtttaattAAAATAAAAAAC

Appendix III. ^{1}H NMR spectra

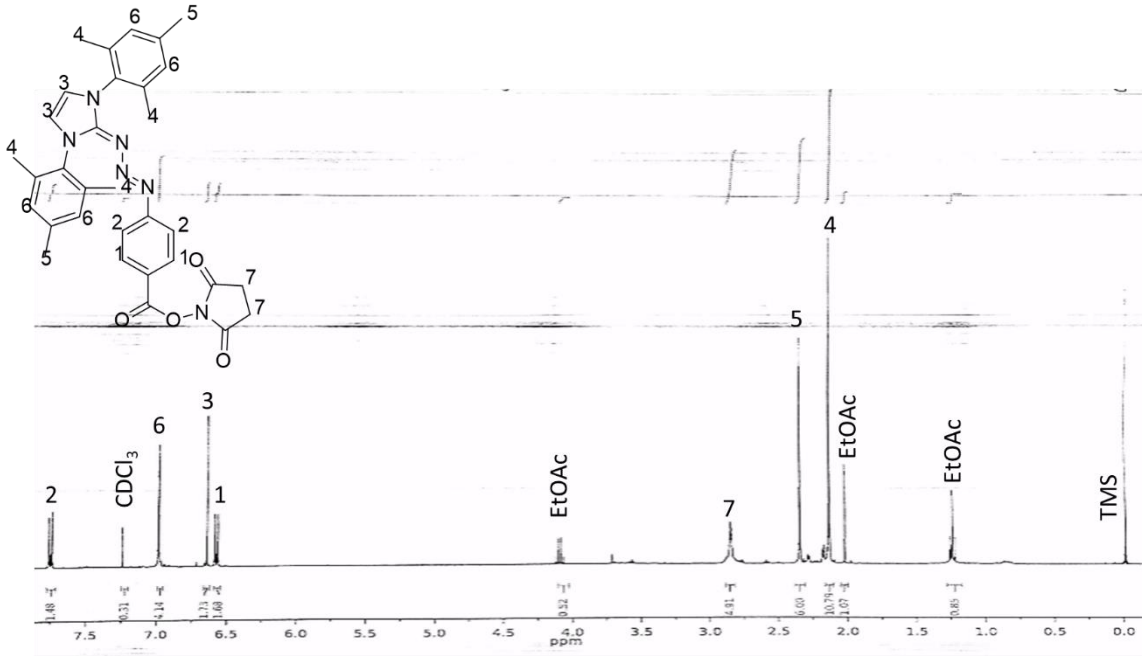
Chapter 6, **3**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



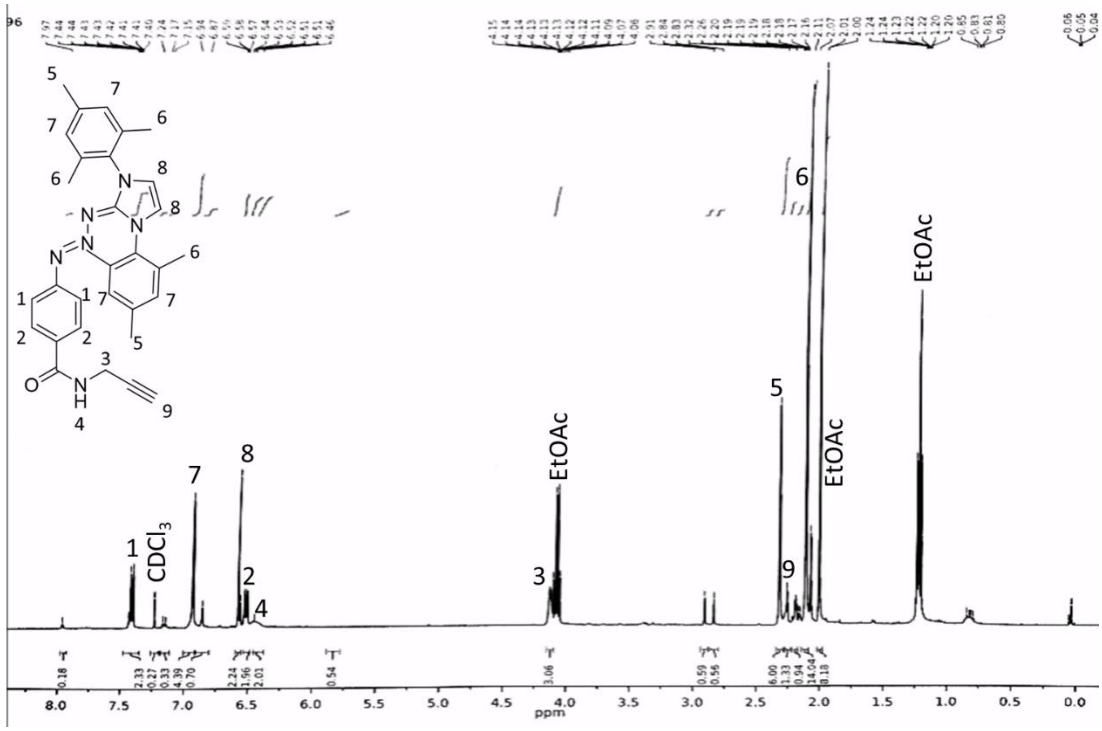
Chapter 6, **4**
^{1}H NMR spectrum
400 MHz, CD_3OD



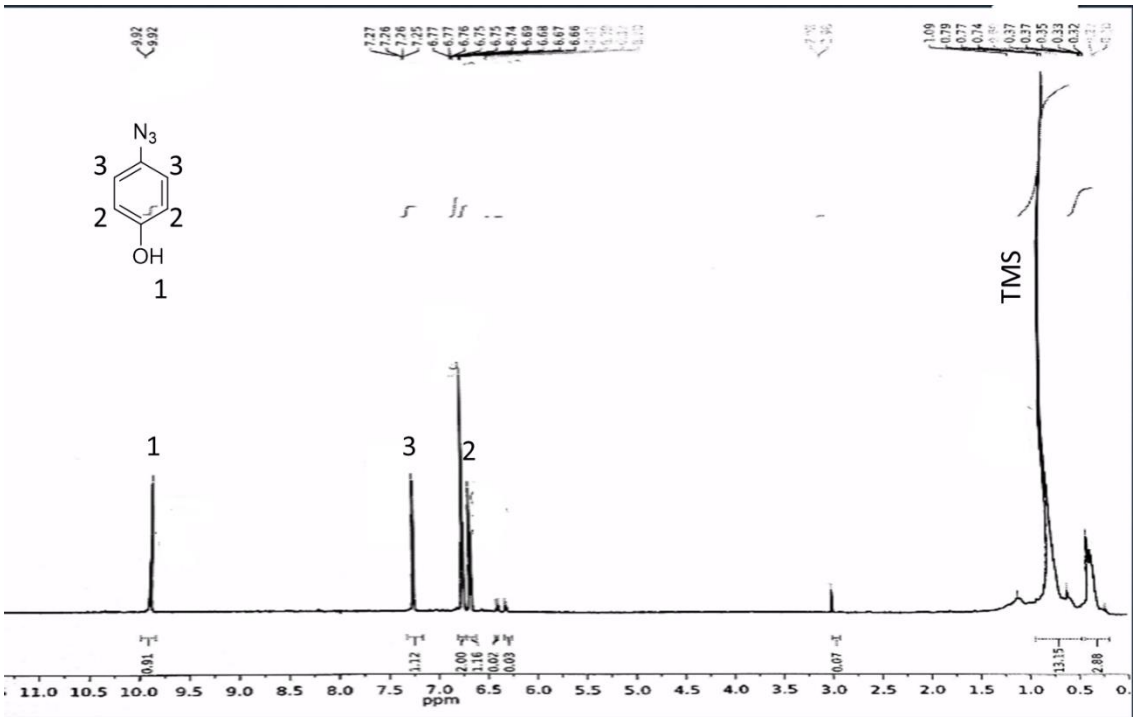
Chapter 6, **5**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



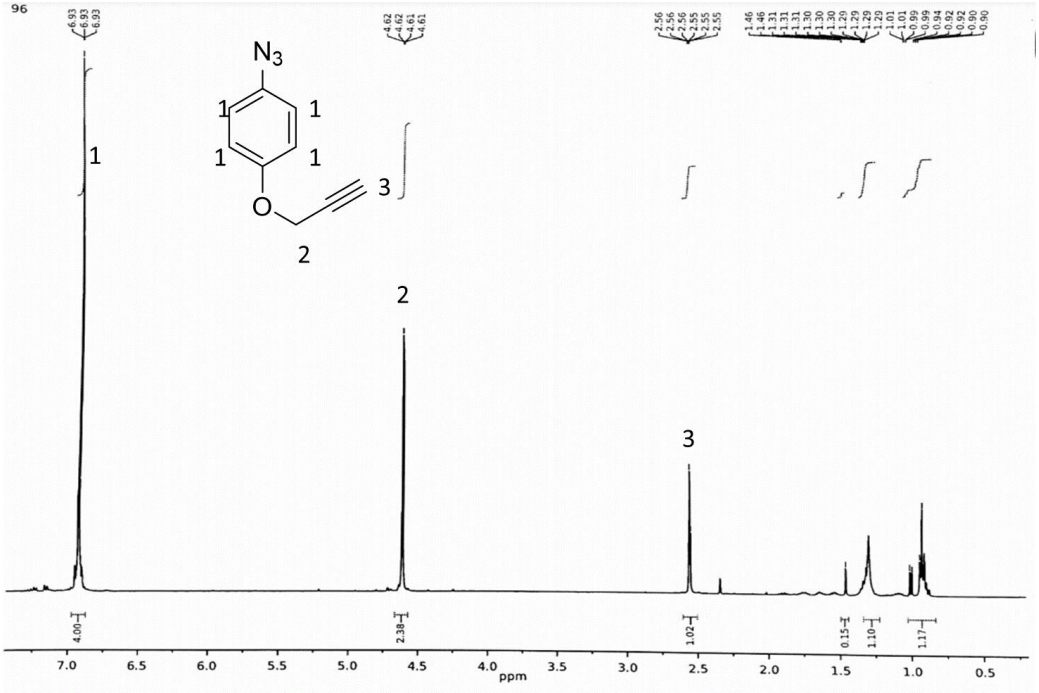
Chapter 6, **6**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



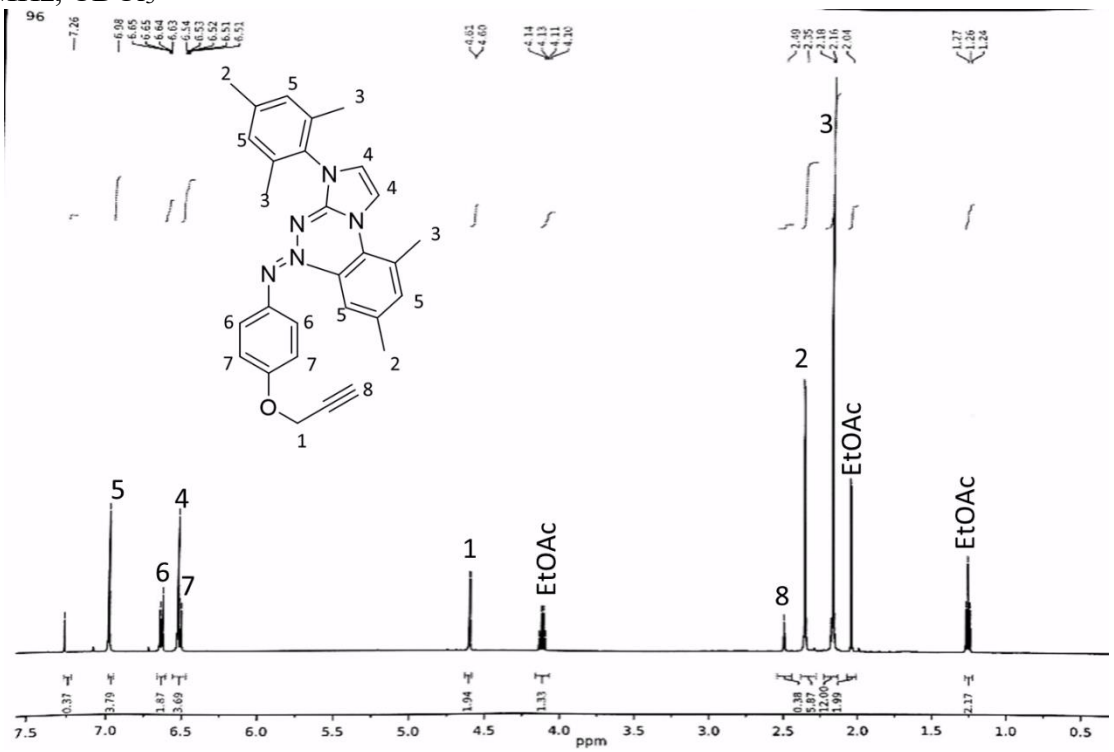
Chapter 6, **7**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



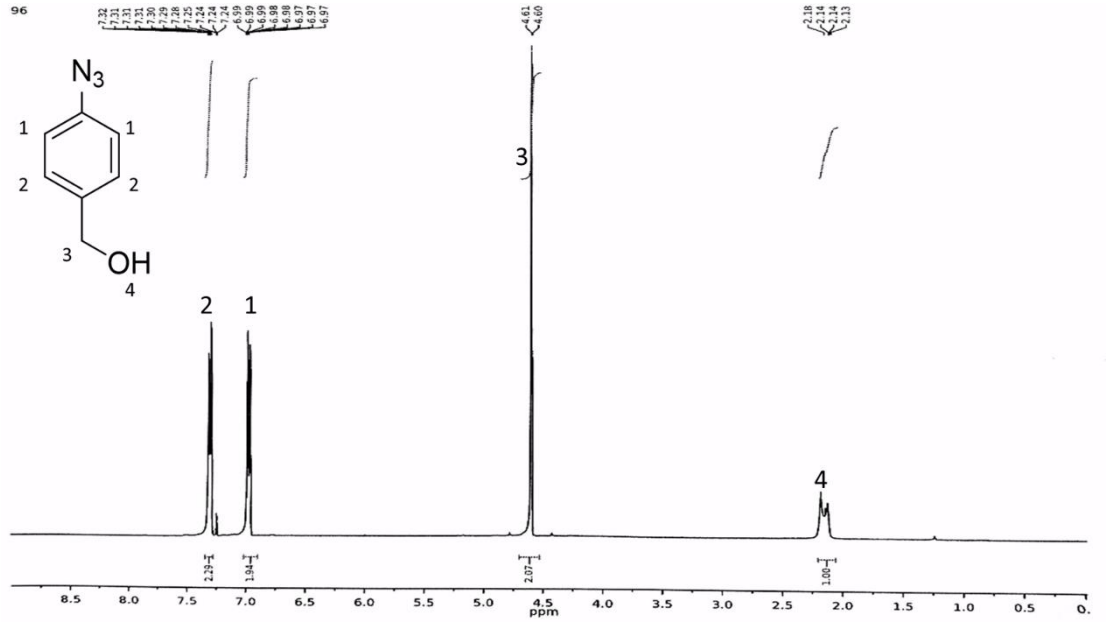
Chapter 6, **8**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



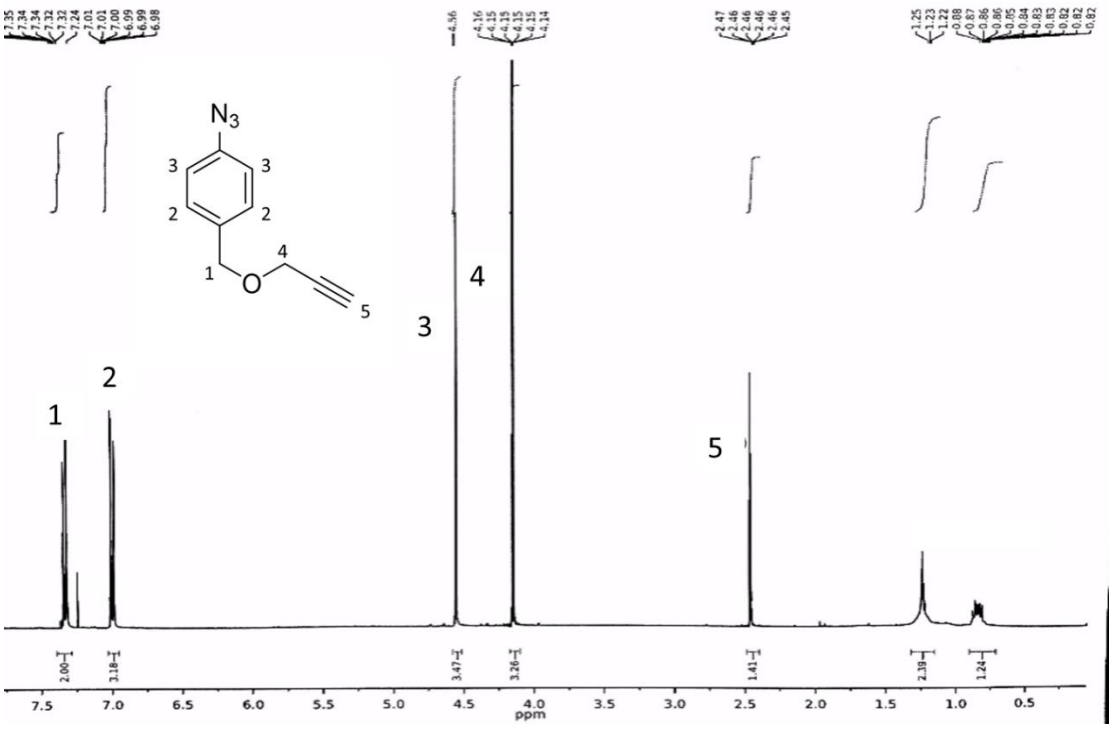
Chapter 6, **9**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



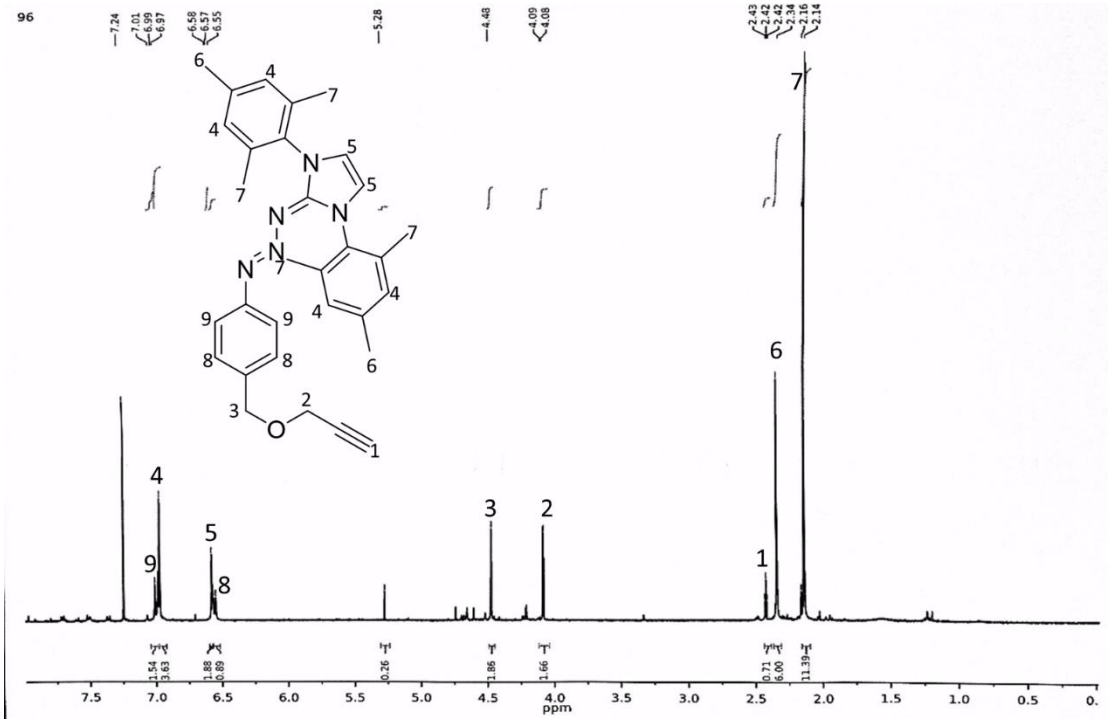
Chapter 6, **10**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



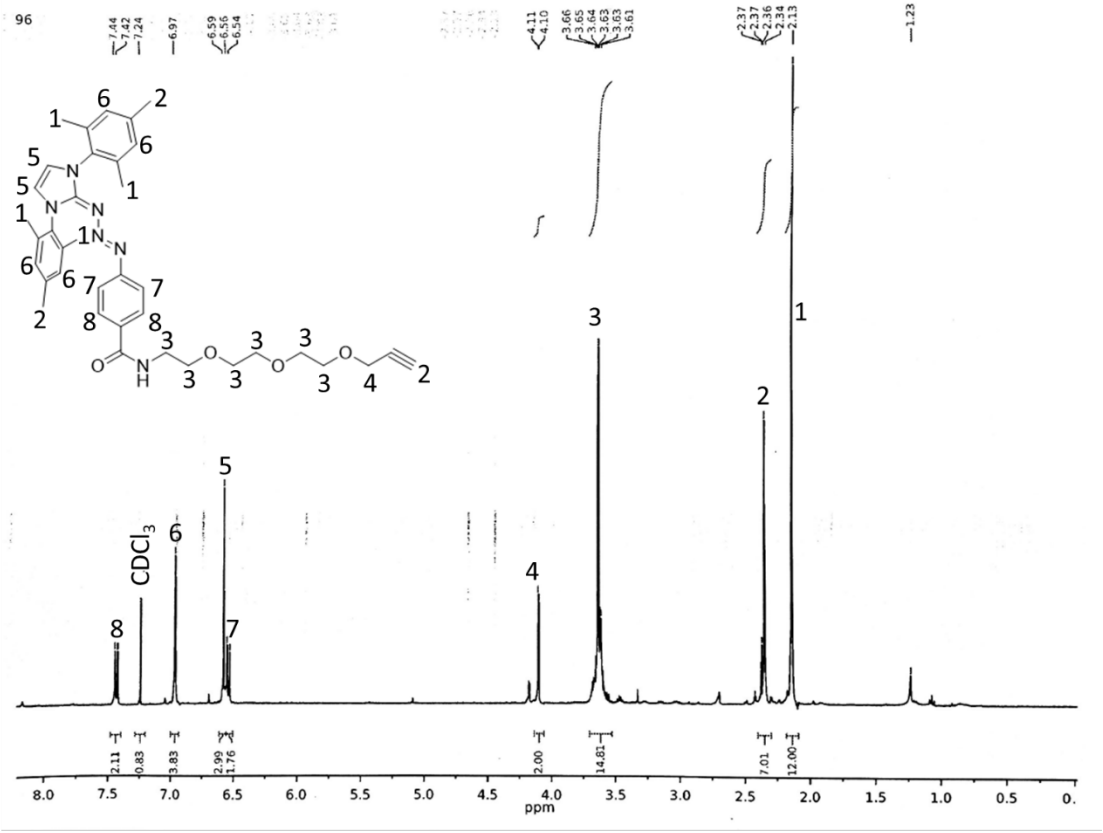
Chapter 6, **11**
^{1}H NMR spectrum
400 MHz, $CDCl_3$



Chapter 6, **12**
^{1}H NMR spectrum
400 MHz, $CDCl_3$

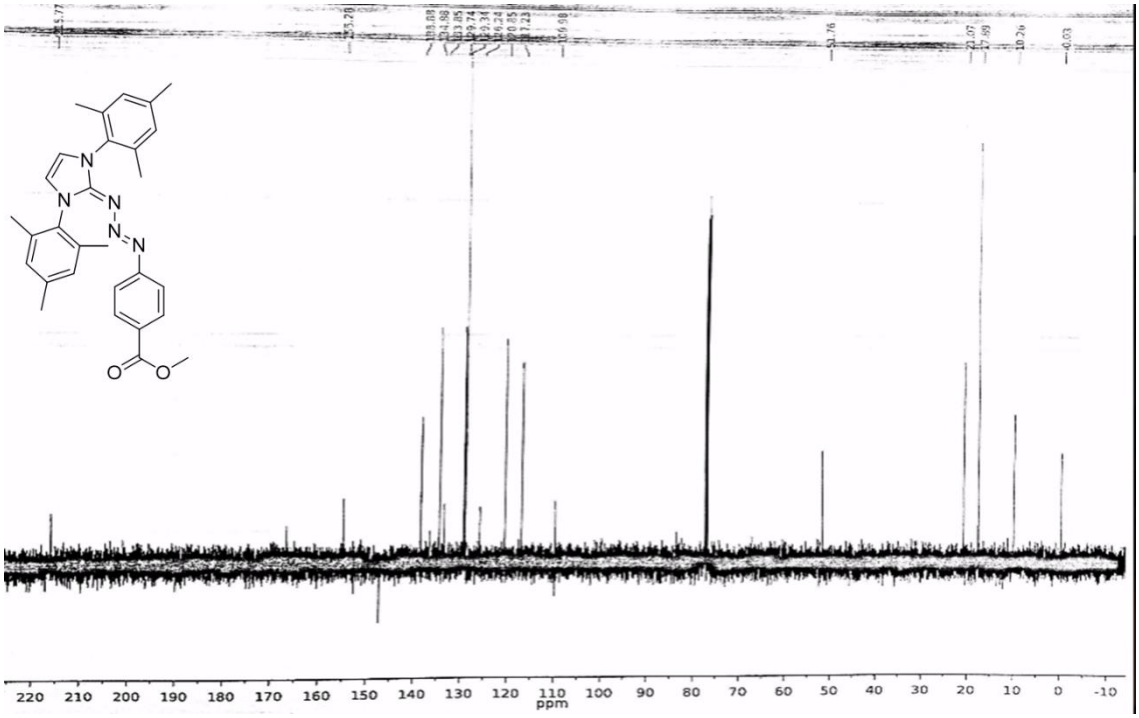


Chapter 6, **13**
^{1}H NMR spectrum
400 MHz, $CDCl_3$

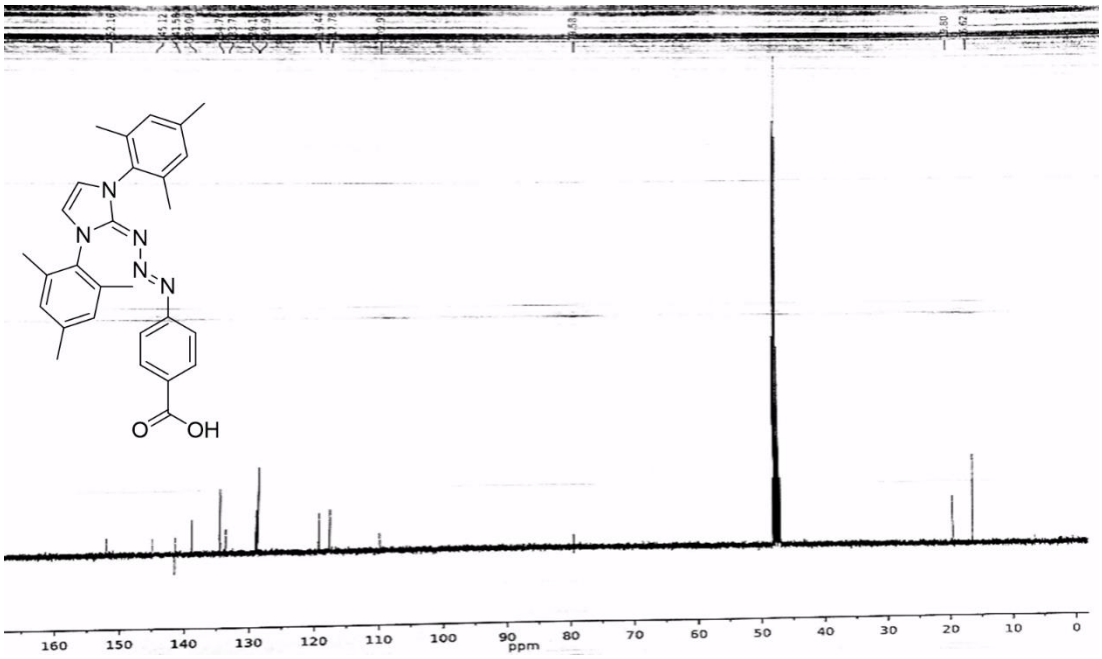


Appendix IV. ^{13}C NMR spectra

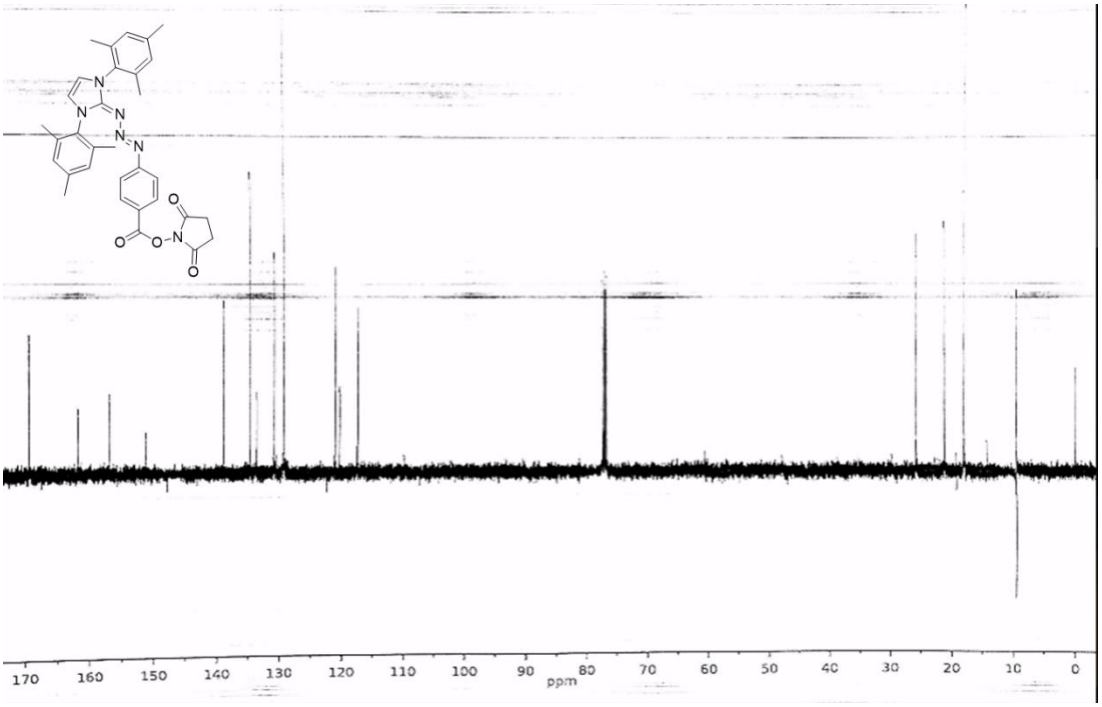
Chapter 6, **3**
^{13}C NMR spectrum
400 MHz, $CDCl_3$



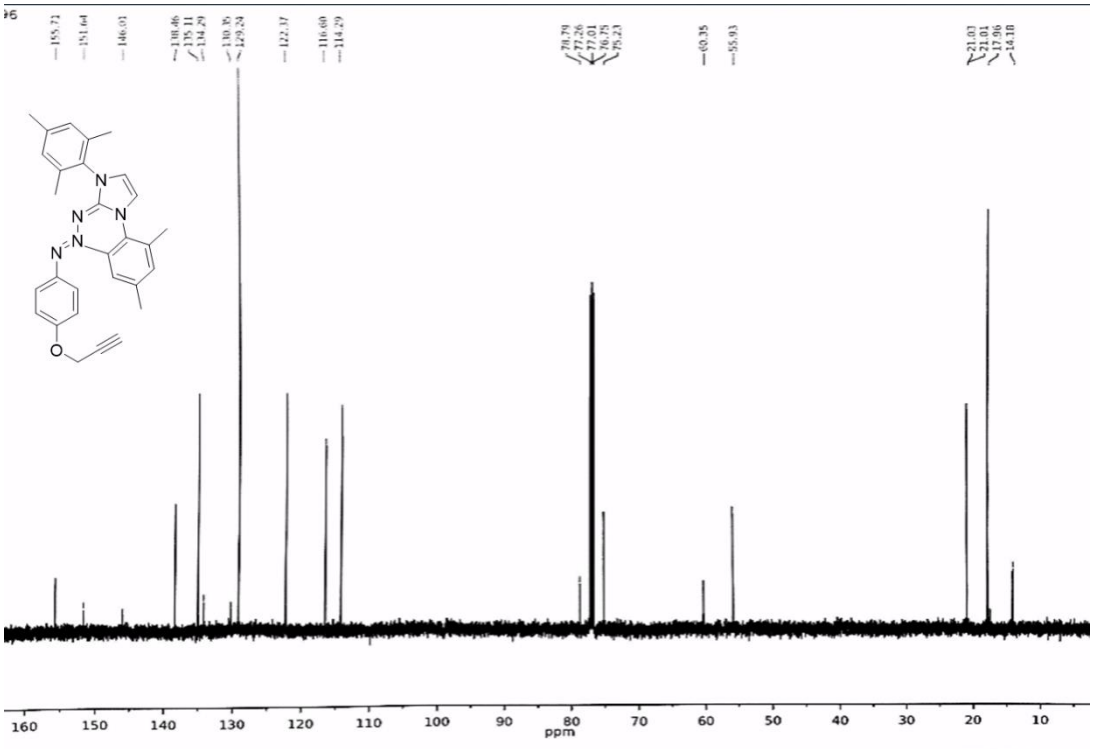
Chapter 6, **4**
^{13}C NMR spectrum
400 MHz, $CDCl_3$



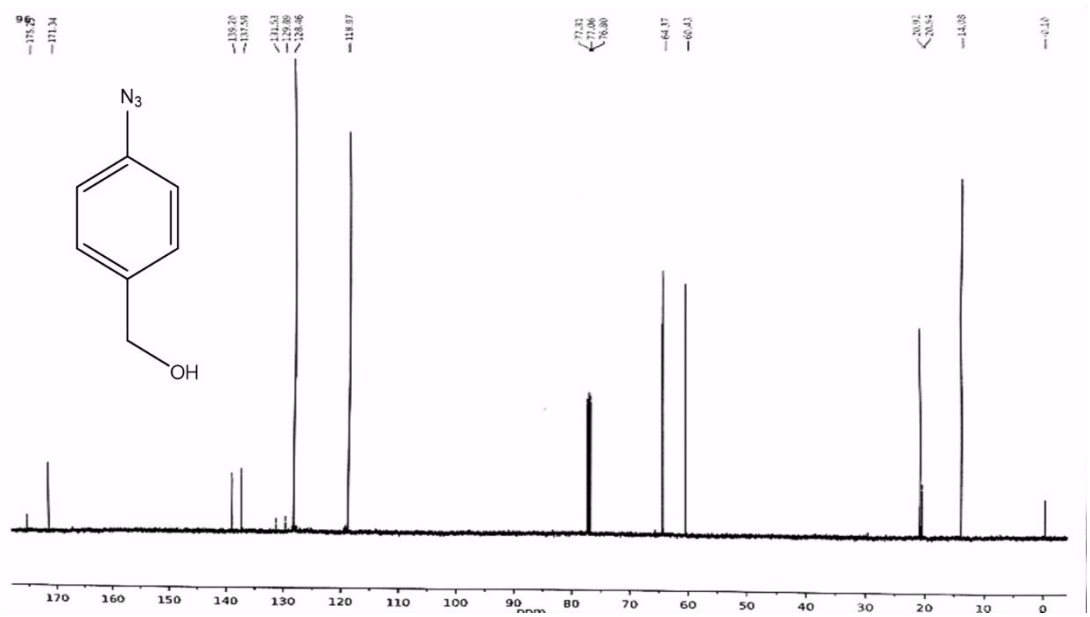
Chapter 6, **5**
^{13}C NMR spectrum
400 MHz, $CDCl_3$



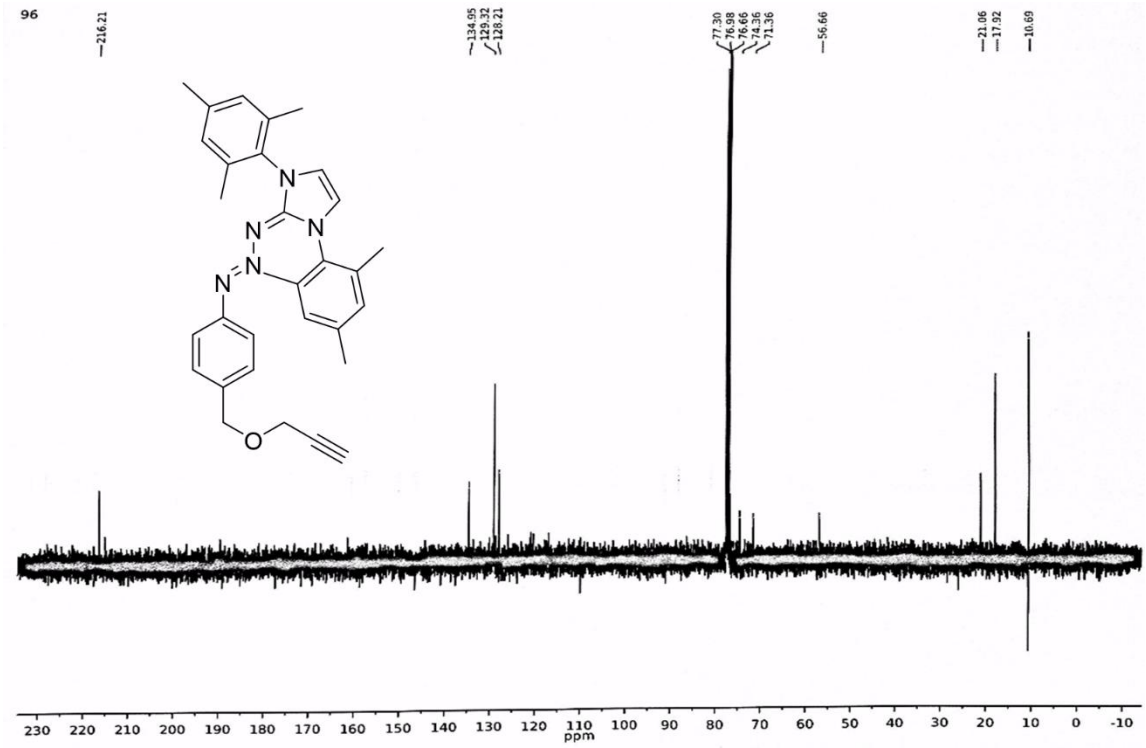
Chapter 6, **9**
^{13}C NMR spectrum
400 MHz, $CDCl_3$



Chapter 6, **10**
^{13}C NMR spectrum
400 MHz, $CDCl_3$



Chapter 6, **12**
^{13}C NMR spectrum
400 MHz, $CDCl_3$



Chapter 6, **13**
^{13}C NMR spectrum
400 MHz, $CDCl_3$