DEVELOPMENT AND APPLICATIONS OF GENETIC CODE EXPANSION PLATFORMS FOR EUKARYOTES

Dissertation

by

Shu Wang

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Boston College
The Graduate School of Arts and Sciences
Department of Chemistry

Aug 2022

DEVELOPMENT AND APPLICATIONS OF GENETIC CODE EXPANSION PLATFORMS FOR EUKARYOTES

by

Shu Wang

Thesis Advisor: Abhishek Chatterjee

Abstract

The genetic codon expansion (GCE) is a technique that uses an orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pair to incorporate noncanonical amino acids (ncAA) into proteins, to enable more protein-based chemistry. In the past two decades, more than 200 ncAAs have been site-specifically introduced into proteins in *E. coli,* and facilitated studies of protein structures, functions and interaction with other molecules. Although a large variety of ncAAs are available for incorporation in the bacterial systems, significantly fewer ncAAs are accessible for incorporation in eukaryotic cells. An expanded GCE toolbox will be beneficial for numerous applications in eukaryotic systems. Currently, introducing ncAAs in eukaryotes predominantly relies on the archaeal pyrrolysyl tRNA/aaRS pair. Such a strong dependence on a single platform has precluded genetic encoding of many desirable ncAAs, including structural mimics of many important post-translational modifications. The work presented in this thesis first developed an engineered *E. coli* leucyl

tRNA/aaRS pair to enable site-specific incorporation of citrulline, an important PTM, into proteins expressed in mammalian cells. This technology was used to reveal the role of citrullination on site R372 and R374 of PAD4. Additionally, aiming at genetically encoding more diverse ncAAs, all 20 *E. coli* derived tRNA/aaRS pairs were screened for their ability to suppress TAG and TGA in mammalian cells. This study revealed several tRNA/aaRS pairs that are suitable for ncAA incorporation in mammalian cells, including those selective for phenylalanine, lysine, arginine, serine and glutamine. Efforts are currently under way to engineer these pairs to genetically encode new structural classes of ncAAs.

Dedication

I dedicate this thesis to my parents,

Zhixin Wang and Qian Feng,

as well as my fiancée Nathchar Naowarojna.

Thank you for teaching me to face our challenges with positivity, strength, and determination

Acknowledgment

Like the classic Chinese novel, *Journey to the west*, wrote, the path to seek knowledge and truth is long, tough, and unpredictable. My path to pursue science is no difference. Monk Tang Sanzang could not obtain Buddhist sacred texts without the support from the emperor, guidance of Buddha, and company of his apprentice; I cannot finish this journey without support from family, guidance of mentors and help from friends and colleagues. I truly grateful to the helps I received in along the way of my Ph.D. and would like to thank them with my very limited English skills.

I'd like to start by thanking my mentor and PhD advisor, Abhishek Chatterjee. You are a true scientist with passion and dedication to contribute to humanity. I am continually impressed by your apparent ease of coming up with new ideas, your work ethic, your availability to your students, and your quality of characters. I appreciate the opportunity and guidance provided to me. And thank you for patience, understanding and support you have in these years. I have learned valuable lifelong lessons from you other than science. They will become a part of my life. Even though the world is big, I hope that our paths will cross again in the future.

Regarding the lab, I am so thankful for the highly collaborative and friendly environment we have developed. I am always confident that I could find someone to ask for help if I needed assistance with a new protocol, equipment, just needed to chat through an idea, or vent about frustrations. While everyone deserves acknowledgement,

I’d like to thank Dr. Yunan Zheng, Dr. Sarah Erickson and Dr. James Italia for teaching me some techniques. I’d like to thank Delilah Jewel for the mental support, healthy peer pressure, discussion on aspects of society that I have never experienced. Additionally, I’d like to thank Dr. Soumya Jyoti Singha Roy for his help with synthesis project and organic chemistry discussions. I also thank my undergraduate, Yichen Weng, for all her help and give me opportunity to learn to be a mentor. I could not have been nearly as productive without her and expect great things in years to come from her. Lastly, thanks Elise, Rachel, Conor for being great friends, and everyone in the lab. And good luck on your future endeavors!

Thank you to Eranthie Weerapana and Jianmin Gao for their guidance in graduate school. I’d like to thank all members of the Weerapana and Gao labs for all their help over the years. The accessibility and collegial atmosphere in the department is inspiring. I would also thank my collaborator from UMass medical school, Dr. Paul Thompson and Dr. Santanu Mondal, for their insightful knowledge, and the opportunity to get exposure to different field.

Thank you to Dale Mahoney, Steve Quinn, Ian Parr, and the rest of the support staff at Boston College. A lot of our work is simplified by their efforts.

I’d like to thank my family and friends who have supported me throughout these years.

Lastly, I’d like to thank the people whom this thesis is dedicated. Thank my parents for mental and finical supports. My father gives me full freedom to choose my life path and career. I am thankful for my mom, a wise woman who teaches and shows me how to endure challenges and hardships in life. Finally, I’d like to thank my fiancée, Nathchar, for sticking with me through the ups and downs of our Ph.D., and for sharing our passion in good foods.

Table of Contents

List of Figures

List of Abbreviations

Standard 3- letter and 1-letter codes are used for the 20 natural amino acids.

5HTP	5-hydroxytryptophan
aaRS	aminoacyl-tRNA synthetase (aa can be replaced with 3- letter code or UAA shorthand)
Amp	ampicillin
arab	arabinose
ATM	Altered Translational Machinery
ATMW	Altered Translational Machinery Tryptophanyl Strain
ATMF	Altered Translational Machinery Phenylalanine Strain
ATMK	Altered Translational Machinery Lysine Strain
ATP	adenosine tri-phosphate
BocK	Boc-lysine
Chlor	chloramphenicol
Cit	L-citrulline
CAP	2-aminocaplyric acid
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethylsulfoxide
EGFP	enhanced green fluorescent protein
ESI-MS	electrospray ionization mass spectrometry

FBS	Fetal bovine serum
FRET	Förster resonance electron transfer
Fxa	Factor X protease
GCE	genetic code expansion
Gent	gentamycin
$Gent^R$	gentamycin resistance
HEK293T	Human embryonic kidney cell line
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	luria broth
Mj	*M. jannaschii*
MS	mass spectrometry
ncAA	non-canonical amino acid
nbcit	o-nitrobenzoyl citrulline
pAzF	4-azido-L-phenylalanine
PEI	Polyethylamine
Pen	Penicillin
PTM	post-translational modification
Pyl	pyrrolysyl

Ph	*P. horikoshii*
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sGFP	Splitting GFP
sfGFP	super-folder green fluorescent protein
Spec	spectinomycin
Strep	Streptinomycin
tRNA	transfer RNA
lysS	lysine-tRNA ligase S
pheST	phenylalanine-tRNA ligase αβ
lysU	lysine-tRNA ligase U
UAA	unnatural amino acid
Zeo	zeocin
Zeo^R	zeocin resistance

Chapter 1. Introduction

A portion of this chapter has been discussed in:

Wang S, Osgood AO, Chatterjee A. Uncovering post-translational modification-associated protein-protein interactions. *Curr Opin Struct Biol.* 2022 Jun;74:102352.

1.1 Genetic code expansion in mammalian system --- current scope and limitations:

Proteins play critical roles in all aspects of our biology from building structural components and performing catalytic reactions, to enabling transport of key molecules in and out of cells and signal transduction. To achieve these varied and diverse functions, protein not only rely on 20 canonical amino acids they are primarily composed of, but also numerous post-translational modifications (PTMs) to further expand its chemical space.[1] It has also been possible to introduce non-natural chemistries into proteins using the genetic code expansion (GCE) technology. This technology relies on an engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pair to incorporate a noncanonical amino acid (ncAA) in response to a nonsense or frameshift codon.[2-6] This technology has enabled numerous powerful applications to either probe or engineer protein function. For example, by using GCE, a noncanonical amino acid (ncAA) mimicking the PTM can be incorporated into the desired site to compare with non-modified protein.[7, 8] A photocaged amino acid can be introduced as a conditional switch to evaluate its essentiality.[9, 10] Key residues within a catalytic site can be replaced by analogs with related but varied properties to probe reaction mechanism.[9, 11] Proteins containing a photo-crosslinker can be used to capture transient protein-protein interaction.[12, 13] Bioconjugation handle-containing proteins can be site-specifically labeled to probe and engineer their function.[14, 15] (Figure 1-1)

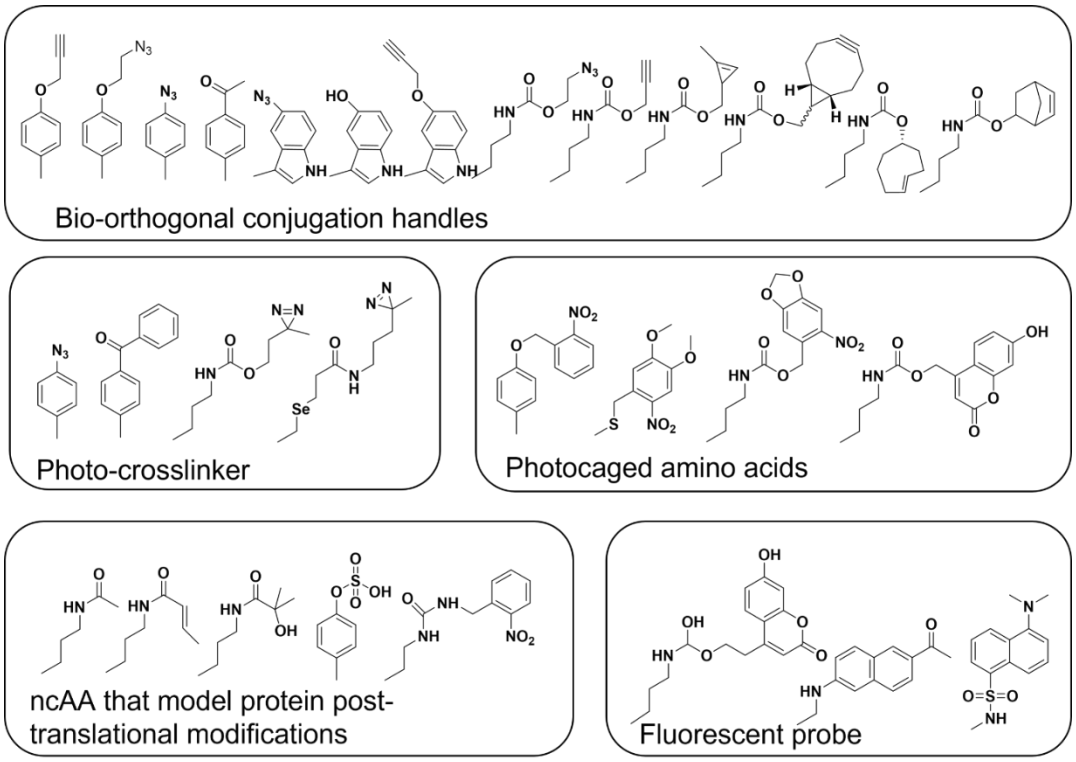


Figure 1-1. Selected ncAAs with different function group that has been incorporated into eukaryotic system.

Despite its great potential, GCE technology in higher eukaryotes is currently limited by the limited structural diversity of ncAAs that are currently available. This is largely due to the fact that the GCE in higher eukaryotes has been exclusively dependent on the archaea-derived pyrrolysyl pair.[8, 16, 17] This pair is functional in both eukaryotic and bacterial cells, because of its unique structural features, enabling its facile engineering in *E. coli* followed by application in eukaryotes. Although additional pairs from *E. coli*, including Leu, Tyr, Trp have been developed for GCE in eukaryotes, many fewer ncAAs have been genetically encoded using these platforms.[18-21] (Figure 1-2) Additional platforms are needed to enable facile incorporation of a more diverse toolbox of ncAAs into proteins in higher eukaryotes. Additionally, access to different

pairs are needed to incorporate multiple different ncAAs into the same protein. Such ability to simultaneously incorporate multiple distinct ncAAs into one protein has numerous applications, from being able to install different biophysical probes (e.g., a FRET pair) to study protein function, and to construct next-generation biotherapeutics.[22-24] The Chatterjee group has also explored this strategy to investigate crosstalk between two PTMs on the same protein, and to trap a PTM-dependent protein-protein interaction.[23, 25]

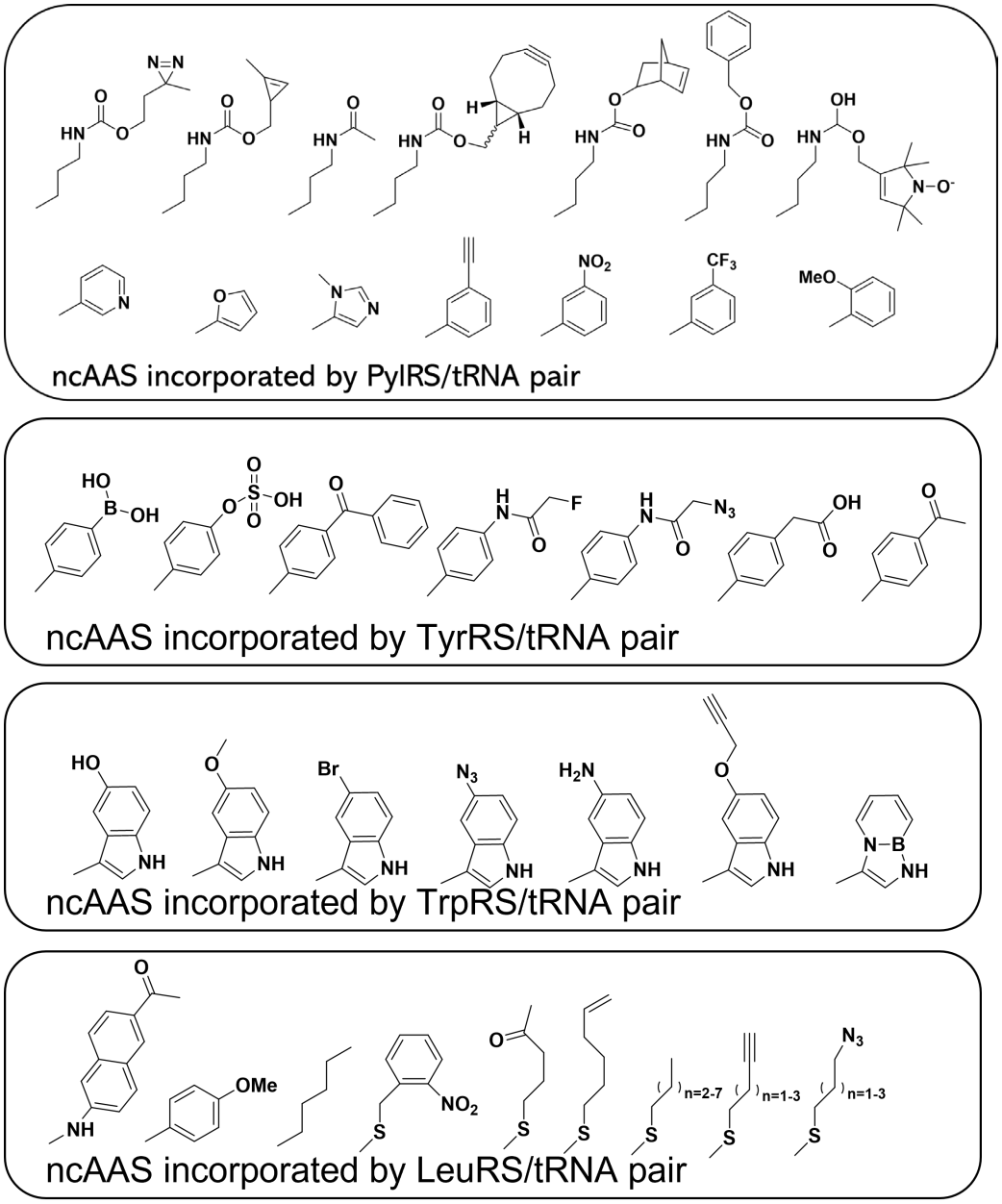


Figure 1-2. Examples of ncAAs that incorporated by different pairs in eukaryotic system.

Access to additional tRNA/aaRS pairs, which can be readily engineered to genetically encode diverse structural classes of ncAAs, is critical to overcome these limitations. In addition to the bacterial pairs mentioned above, Lin's group developed

chimeric tRNA/aaRS pairs, which are obtained by fusing the pyrrolysyl pair with various bacterial pairs, which retain the substrate specificity of the bacterial pair but are universally orthogonal like the pyrrolysyl pair.[26, 27] One of such chimeric pairs was further evolved to genetically encode new ncAAs, including tryptophan analogs. However, the efficiency and robustness of these systems remain a concern. These pairs are also not compatible with popular pyrrolysyl system for facilitating simultaneous incorporation of multiple ncAAs. Furthermore, Chin's group recently discovered and evolved novel PylRS form other species (e.g., *Methanomethylophilus alvus*) that are orthogonal to the original $tRNA^{Pyl}$/PylRS pairs (from *M. barkeri* and *M. mazei*). [28] These two pyrrolysyl pairs can be used together to site-specifically incorporate two distinct ncAAs into the same protein.[29] However, the active sites of the two different PylRS enzymes are essentially identical, and both enable genetically encoding similar types of ncAAs.

As previously noted, the GCE technology relies on orthogonal (does not cross-react with host counterparts) tRNA/aaRS pairs. Traditionally, bacterial pairs are orthogonal in eukaryotes, and eukaryotic/archaeal pairs are orthogonal in bacterial cells.[16] Of course, the pyrrolysyl pair is an exceptional case, as it is orthogonal in all domains of life due to its unique structure.[17, 30, 31] Consequently, to expand the repertoire of tRNA/aaRS in higher eukaryotes, the most logical place to start from are other bacteria-derived pairs. Indeed, three *E. coli* derived pairs have already been demonstrated to be suitable for ncAA incorporation in eukaryotes: Tyr, Leu, and Trp. Exploring the

potential of the other bacterial pairs for ncAA incorporation in eukaryotes may significantly expand the scope of this technology.

1.2 Altered Translational Machinery strategy for expansion of genetic toolkit in bacteria and mammalian cells

To engineer the aaRS substrate specificity, and charge a ncAA, a directed evolution approach is typically required. Orthogonality of the target protein, size of the library, availability of high throughput selection scheme and difficulty of active member recovery are the main aspects to consider while choosing or designing a directed evolution platform. Among these aspects, the orthogonality is the foundation, and it is naturally formed by distance on the evolutional tree. The prokaryotic and eukaryotic/archaeal tRNA/aaRS systems in general are orthogonal to each other; therefore, their directed evolution of eukaryotic/archaeal pairs are typically performed in prokaryotic cells, whereas the evolution of prokaryotic pairs requires a eukaryotic host. As the model organism of bacteria, *E. coli* is a preferred platform due to its fast doubling time, high transformation efficiency, easy plasmid amplification and extraction, and optimized high throughput selection scheme such as antibiotic based selection, and phage-assisted continuous evolution (PACE).[32] However, *E. coli* derived tRNA/aaRS pairs such as Leu, Trp and Tyr cannot be evolved using a bacteria-based selection system. The alternative is to evolve these bacteria-derived pairs in

eukaryotic cells. However, such systems are significantly less efficient and has experienced limited success.

The Chatterjee lab envisioned a universal system, that allowed us to evolve bacteria-derived aaRSs in *E. coli*. The challenge, of course, was that a bacteria-derived pair will cross-talk to the corresponding endogenous wild-type pair in *E. coli*. Such a cross-talk would result in false positives during the selection, as well as proteome-wide charging of a ncAA that would be lethal. Our lab has overcome this challenge by replacing the endogenous tRNA/aaRS pair for a particular amino acid with an imported orthogonal pair from an evolutionarily distant host. While the imported orthogonal pair substitute the wild-type function, the original pair can be reintroduced into the resulting strain (called altered-translational machinery or ATM strains) as an orthogonal nonsense suppressor and engineered. The tRNA/aaRS pairs evolved in this manner are not only functional for ncAA incorporation in mammalian cells, but also in the ATM *E. coli* strain. With this method, our lab has created two *E. coli* strains ATMW and ATMY, to evolve *E. coli* tryptophan and tyrosine pairs, respectively (Figure 1-3).

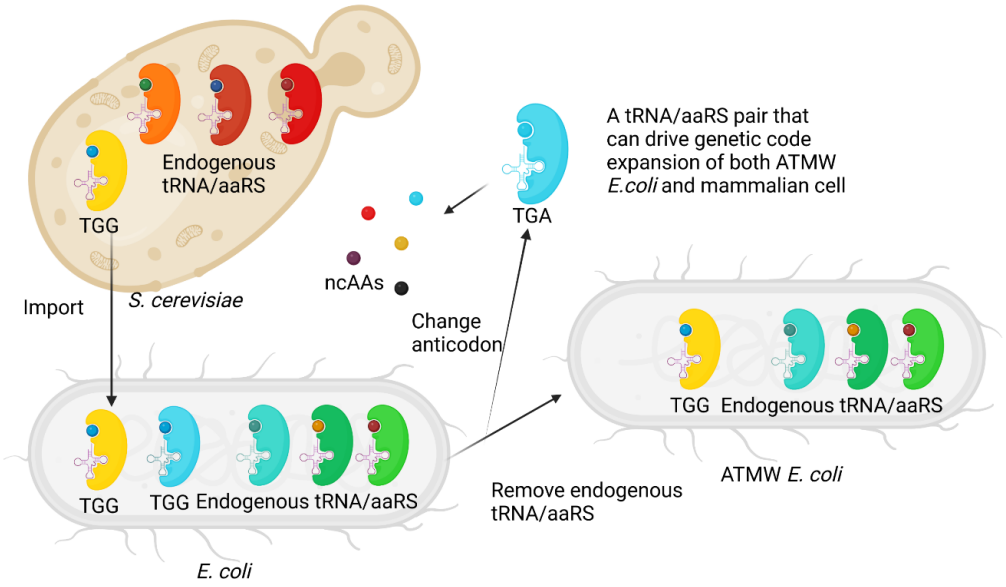


Figure 1-3. Illustration of function and construction of the ATMW strain.

Despite the success of the above described strategy using ATM system for evolving the bacterial tyrosine and tryptophan pairs, this methodology can be challenging to extend to additional amino acids. The core design of the ATM strategy is that the desired endogenous tRNA/aaRS pair is liberated and evolved to a nonsense suppressor. To achieve this, an orthogonal alien pair is essential that can effectively substitute the function of the endogenous tRNA/aaRS pair being liberated. For ATMY and ATMW, the Tyr pair from *M. jannaschii* and Trp pair from yeast were imported, respectively. However, finding such an orthogonal alien pair for other systems can be challenging. Moreover, the difficulty of engineering an ATM strain for each amino acid is vastly different, which is partly related to the number of different codons and tRNAs

associated with the particular amino acid in the host. For instance, development of ATMY and ATMW required the deletion of only three and one tRNA genes from the genome, respectively. To create a similar strain for *E. coli* leucyl pair would be significantly more challenging. In *E. coli*, there are eight different leucyl tRNA genes to decode the six leucyl codons,[33] and knocking out all eight copies is already technically demanding. Furthermore, not all leucine codons are decoded with the same efficiency; balancing and optimizing such nuanced function with a foreign tRNA can be particularly challenging. Chapters 4 and 5 in this thesis describe the efforts to build ATM strains for phenylalanine and lysine which ran into significant challenges. The key challenge in both cases were identifying an orthogonal tRNA/aaRS pair that could functionally replace the endogenous ones. In conclusion, ATM system is powerful and beneficial for developing tRNA/aaRS pairs that can be used in both eukaryotic and E. coli expression systems, but it can be challenging to develop.

1.3 Directed evolution platform for evolving tRNA/aaRS in the eukaryotic system.

The ATM system is designed to "create" orthogonality in *E. coli* for its endogenous pairs. Evolving an orthogonal bacterial pair in a eukaryotic host is another strategy. The *E. coli* $tRNA^{Leu}$/LeuRS pair has contributed many distinct ncAAs into the mammalian protein, such as α-aminocaprylic acid, Anap, O-nitrobenzyl cysteine, *O*-methyl tyrosine and more. The pair was first evolved in the yeast and then imported into mammalian

system.[8, 9, 21, 34-36] *S. cerevisiae* is a unicellular eukaryote and uses tRNA/aaRS pair homologous to those in higher eukaryotes. However, it also represents an established platform for in-cell directed evolution. These characters make *S. cerevisiae* a good candidate for the direct evolution of the bacterial tRNA/aaRS pairs. Indeed, the Schultz group developed such a directed evolution platform in yeast, taking advantage of the conditional expression of the GAL4 transcription factor.[21] Two stop codons were introduced into the GAL4 sequence at permissive sites to conditionally control its expression by the activity of the nonsense suppressing tRNA/aaRS pair. However, we found that bacterial pairs that are orthogonal in the mammalian system are not always orthogonal in *S. cerevisiae*. In chapter 6 of this thesis, my work demonstrated that the bacterial arginyl-tRNA is relatively orthogonal in mammalian cells, but it is not adequately orthogonal in *S. cerevisiae*. These challenges are described in further detail in the Chapter 6, and our group is working on further optimize it.

Directed evolution of aaRSs directly in mammalian cells is not ideal, because mammalian cells are unable to maintain plasmid, has low cell counts in unit volume, has slow growth rate, and does not have practical protocol for library recovery. These factors have made it difficult to perform directed evolution in mammalian cells. However, a limited number of engineered aaRS mutants can be individually screened in mammalian cells for novel activities. For instance, Dr. Zheng in our lab rational designed and screened mutants of the *E. coli* leucyl-tRNA synthetase (EcLeuRS), to genetically encode novel ncAA substrates. In chapter 2, I discuss that using the same

strategy, it was possible to find a mutant EcLeuRS that can incorporate a photocaged citrulline for the first time in mammalian cells.[7, 37] Additionally, in chapter 6, I proposed and tested several mutations of ArgRS for citrulline incorporation. This method is restricted to only a small number of candidates and is not always successful.

The mammalian cell itself is not a promising directed evolution platform, but with the help of AAV2, our lab developed a virus-assisted directed evolution of tRNAs (VADER) system in mammalian cells, which employs a double-sieve selection scheme to facilitate single-step enrichment of active-yet-orthogonal tRNA mutants from naïve libraries.[38] Although this system is currently designed to evolve better tRNAs, with proper adjustment of the enrichment step, it may enable the directed evolution of aaRSs in mammalian cells.

1.4 Genetic code expansion facilitates the study of biological functions of PTMs.

Advancing of GCE revolutionized the study of PTMs, by enabling site-specific incorporation of ncAAs that structurally mimic various PTMs. Prior to the development of the GCE technology, PTM studies were traditionally performed by three different methods, each with critical drawbacks:

The first and easiest way is mutating the site of the PTM with a different canonical amino acid, which either mimic the unmodified amino acid but cannot be actually modified (loss of the function), or the modified amino acid (gain of the function). To

mimic the loss of the function, the target amino acid is replaced by an amino acid that is structurally or chemically similar, but one that cannot be subjected to the same PTM. Examples of this strategy include replacing a tyrosine with phenylalanine to resemble the structure, or replacing an arginine with lysine to retain the charge on the side chain. For the gain of the function, the target amino acid is replaced by an amino acid that share similar properties with the PTM. For instance, negatively charged aspartic acid and glutamic acid residues are often used to mimic the effect of phosphorylation, charge-neutral glutamine has been used to model lysine acetylation, and positively charged arginine is used to mimic an unacylated lysine residue.[39-42] This method is operationally simple, can take advantage of the chemical diversity of the twenty canonical amino acids, and has been extensively used in the research community. On the other hand, the canonical amino acids are not perfect structural mimics for the PTMs, which raises concerns about whether the observed functional effects are truly meaningful. Furthermore, there are many PTMs which cannot adequately be mimicked by canonical amino acids.

Another approach to model PTMs involves taking advantage of the natural mechanisms responsible for installing them. Incubating the target protein with the "writer" protein that introduces the PTM have been frequently used. The advantage of this method is that it recreates the exact modification and it is also operationally straightforward. However, this method has several limitations. For example, biochemical origin of many PTMs have not yet been identified. Additionally, many

writer proteins are conditionally activated *in vivo*, and their activities are hard to reconstitute *in vitro*. Writers often modify multiple residues, making it hard to assess effects at specific sites.

Solid-phase synthesis of proteins and peptides is another effective method for incorporating ncAAs into proteins with precise site control. A major advantage of this method is that there is no limitation on ncAA selection, and one or more ncAAs can be introduced into chosen sites. The weaknesses for this method include the inability to synthesize larger proteins, need to refold the protein that can be difficult, and inability to test *in vivo* effects. Native chemical ligation has been used to attach synthetic peptides to full-length proteins, but such attachments are restricted to protein termini, and the folding issues and the inability to do *in vivo* activity still apply.

The GCE technology allowed ncAA to be incorporated into protein site-specifically. Moreover, the ncAA containing protein is made within living cells, which facilitates its correct folding/localization, and enables *in vivo* studies.[2] However, only a small subset of known PTMs can be currently incorporated using this technology in eukaryotes (Figure 1-4). PTMs with large and complex structures remain particularly challenging to encode, as it has been difficult to engineer existing aaRSs to recognize such ncAAs. The unpredictable ncAA incorporation efficiency is another challenge commonly encountered with the GCE approach. Therefore, evolving new pairs to expand the PTMs scope and enhance the efficiency for existing pairs are the current focus for the field and our lab.

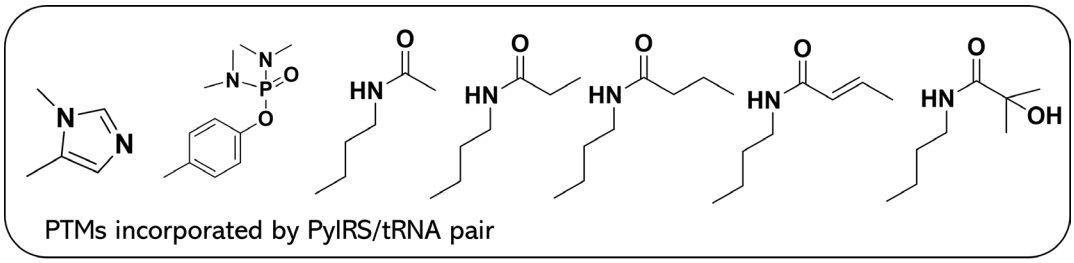


Figure 1-4. PTMs mimicking ncAAs that can be incorporated by PylRS/tRNA Pairs.

1.5 Conclusion

Overall, this thesis covers works from developing new GCE platforms to enable incorporation of new classes of ncAAs into proteins in eukaryotes, as well as examples of how such tools could be employed for useful applications. Chapter 2 discusses site-specific incorporation of citrulline into proteins expressed in mammalian cells and studies on how citrullination of specific residues affect the function of the PAD4 protein. Auto-citrullination of PAD4 was previously studied by other methods, and our work demonstrated the advantages of the GCE technology. Our technology facilitated site-specific incorporation of citrulline, revealing the effects of this PTM at key residues, and helped settle a long-standing argument on how auto-citrullination affects the activity of PAD4. With the same method, we also attempted to decipher the impact of citrullination on GSK3 beta, which is discussed in Chapter 7.

With the success of genetically encoding citrulline in eukaryotes using a pair other than pyrrolysyl, we are motivated to develop additional pairs, which may facilitate the incorporation of other PTMs that remain currently inaccessible to GCE. In Chapter 3,

we tested all *E. coli* tRNA/aaRS pairs to evaluate their orthogonality and activity as TAG and TGA suppressors in mammalian cells. Through this work, we identified His, Lys, Arg, and Phe pairs as promising candidates for further development for genetic code expansion of mammalian cells. To this end, we attempted the development of *E. coli* Lys, Phe, and Arg pairs using different strategies, which are discussed in Chapters 4, 5, and 6, respectively.

1.6 Reference

1. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, Jr., *Protein posttranslational modifications: the chemistry of proteome diversifications.* Angew Chem Int Ed Engl, 2005. **44**(45): p. 7342-72.
2. Chin, J.W., *Expanding and reprogramming the genetic code.* Nature, 2017. **550**(7674): p. 53-60.
3. Shandell, M.A., Z. Tan, and V.W. Cornish, *Genetic Code Expansion: A Brief History and Perspective.* Biochemistry, 2021. **60**(46): p. 3455-3469.
4. Ambrogelly, A., S. Palioura, and D. Söll, *Natural expansion of the genetic code.* Nature Chemical Biology, 2007. **3**(1): p. 29-35.
5. Noren, C.J., et al., *A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins.* Science, 1989. **244**(4901): p. 182-188.
6. Xiao, H. and P.G. Schultz, *At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code.* Cold Spring Harb Perspect Biol, 2016. **8**(9).
7. Mondal, S., et al., *Site-specific incorporation of citrulline into proteins in mammalian cells.* Nature Communications, 2021. **12**(1): p. 45.
8. Dumas, A., et al., *Designing logical codon reassignment - Expanding the chemistry in biology.* Chem Sci, 2015. **6**(1): p. 50-69.
9. Wu, N., et al., *A Genetically Encoded Photocaged Amino Acid.* Journal of the American Chemical Society, 2004. **126**(44): p. 14306-14307.
10. Deiters, A., et al., *A Genetically Encoded Photocaged Tyrosine.* Angewandte Chemie International Edition, 2006. **45**(17): p. 2728-2731.
11. Chen, L., et al., *Use of a Tyrosine Analogue To Modulate the Two Activities of a Nonheme Iron Enzyme OvoA in Ovothiol Biosynthesis, Cysteine Oxidation*

versus Oxidative C–S Bond Formation. Journal of the American Chemical Society, 2018. **140**(13): p. 4604-4612.
12. Grasso, K.T., et al., *A Facile Platform to Engineer Escherichia coli Tyrosyl-tRNA Synthetase Adds New Chemistries to the Eukaryotic Genetic Code, Including a Phosphotyrosine Mimic.* ACS Central Science, 2022. **8**(4): p. 483-492.
13. Chin, J.W., et al., *Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli.* Proceedings of the National Academy of Sciences, 2002. **99**(17): p. 11020-11024.
14. Sarathi Addy, P., J.S. Italia, and A. Chatterjee, *An Oxidative Bioconjugation Strategy Targeted to a Genetically Encoded 5-Hydroxytryptophan.* ChemBioChem, 2018. **19**(13): p. 1375-1378.
15. Lee, H.S., et al., *Genetic Incorporation of a Small, Environmentally Sensitive, Fluorescent Probe into Proteins in Saccharomyces cerevisiae.* Journal of the American Chemical Society, 2009. **131**(36): p. 12921-12923.
16. Italia, J.S., et al., *Expanding the genetic code of mammalian cells.* Biochem Soc Trans, 2017. **45**(2): p. 555-562.
17. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized tRNA.* Science, 2002. **296**(5572): p. 1459-1462.
18. Italia, J.S., et al., *Genetically encoded protein sulfation in mammalian cells.* Nat Chem Biol, 2020. **16**(4): p. 379-382.
19. Italia, J.S., et al., *Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes.* Cell Chem Biol, 2018. **25**(10): p. 1304-1312.e5.
20. Italia, J.S., et al., *An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes.* Nature Chemical Biology, 2017. **13**(4): p. 446-450.
21. Chin Jason, W., et al., *An Expanded Eukaryotic Genetic Code.* Science, 2003. **301**(5635): p. 964-967.
22. Mitchell, A.L., et al., *A Unique Genetically Encoded FRET Pair in Mammalian Cells.* Chembiochem, 2017. **18**(6): p. 511-514.
23. Zheng, Y., et al., *Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis.* ACS Chem Biol, 2018. **13**(5): p. 1137-1141.
24. Italia, J.S., et al., *Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites.* Journal of the American Chemical Society, 2019. **141**(15): p. 6204-6212.
25. Seet, B.T., et al., *Reading protein modifications with interaction domains.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 473-83.

26. Ding, W., et al., *Chimeric design of pyrrolysyl-tRNA synthetase/tRNA pairs and canonical synthetase/tRNA pairs for genetic code expansion.* Nat Commun, 2020. **11**(1): p. 3154.
27. Zhao, H., et al., *Directed-evolution of translation system for efficient unnatural amino acids incorporation and generalizable synthetic auxotroph construction.* Nature Communications, 2021. **12**(1): p. 7039.
28. Beránek, V., J.C.W. Willis, and J.W. Chin, *An Evolved Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells.* Biochemistry, 2019. **58**(5): p. 387-390.
29. Dunkelmann, D.L., et al., *Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids.* Nat Chem, 2020. **12**(6): p. 535-544.
30. Wang, L., J. Xie, and P.G. Schultz, *Expanding the genetic code.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 225-49.
31. Polycarpo, C., et al., *An aminoacyl-tRNA synthetase that specifically activates pyrrolysine.* Proceedings of the National Academy of Sciences, 2004. **101**(34): p. 12450-12454.
32. Esvelt, K.M., J.C. Carlson, and D.R. Liu, *A system for the continuous directed evolution of biomolecules.* Nature, 2011. **472**(7344): p. 499-503.
33. Chan, P.P. and T.M. Lowe, *GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes.* Nucleic acids research, 2016. **44**(D1): p. D184-D189.
34. Sakamoto, K., et al., *Genetic Encoding of 3-Iodo-l-Tyrosine in Escherichia coli for Single-Wavelength Anomalous Dispersion Phasing in Protein Crystallography.* Structure, 2009. **17**(3): p. 335-344.
35. Shen, B., et al., *Genetically encoding unnatural amino acids in neural stem cells and optically reporting voltage-sensitive domain changes in differentiated neurons.* Stem Cells, 2011. **29**(8): p. 1231-40.
36. Chatterjee, A., et al., *A Genetically Encoded Fluorescent Probe in Mammalian Cells.* Journal of the American Chemical Society, 2013. **135**(34): p. 12540-12543.
37. Zheng, Y., et al., *Expanding the Scope of Single- and Double-Noncanonical Amino Acid Mutagenesis in Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases.* Biochemistry, 2018. **57**(4): p. 441-445.
38. Kelemen, R.E., et al., *Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells.* bioRxiv, 2022: p. 2022.01.21.477302.
39. Wu, L., et al., *p50 mono-ubiquitination and interaction with BARD1 regulates cell cycle progression and maintains genome stability.* Nat Commun, 2020. **11**(1): p. 5007.

40. Cui, W., et al., *Phosphorylation Modulates the Coregulatory Protein Exchange of the Nuclear Receptor Pregnane X Receptor.* J Pharmacol Exp Ther, 2020. **373**(3): p. 370-380.
41. Zhang, W., et al., *SIRT1 modulates cell cycle progression by regulating CHK2 acetylation-phosphorylation.* Cell Death Differ, 2020. **27**(2): p. 482-496.
42. Wiese, M., et al., *Citrullination of HP1gamma chromodomain affects association with chromatin.* Epigenetics Chromatin, 2019. **12**(1): p. 21.

Chapter 2. Genetically encoding a photocaged citrulline in mammalian cells

A portion of this chapter has been discussed in:

Mondal S, Wang S, Zheng Y, Sen S, Chatterjee A, Thompson PR. Site-specific incorporation of citrulline into proteins in mammalian cells. Nat Commun. 2021 Jan 4;12(1):45

2.1 Introduction

Citrullination is a post-translational modification (PTM) that involves the hydrolysis of the positively-charged guanidium group on arginine to generate a neutral urea (Figure 2-1a).[1] Citrullination plays crucial roles in many physiological processes, including the epigenetic regulation of gene transcription, Neutrophil Extracellular Trap (NET)-formation or NETosis, and maintaining pluripotency.[1-7] Citrullination is catalyzed by the Protein Arginine Deiminases (PADs) (Figure 2-1a), a group of four catalytically active cysteine hydrolases (PAD1-4).[8] PADs are Ca^{2+}-dependent enzymes and the presence of calcium increases PAD activity by >10,000-fold. Calcium-binding leads to dramatic conformational rearrangements, particularly of the nucleophilic cysteine (C645 in PAD1, 4; C647 in PAD2; C646 in PAD3) to form a catalytically competent active site.[9]

Aberrant protein citrullination is a hallmark of multiple autoimmune disorders, including rheumatoid arthritis (RA), multiple sclerosis (MS), ulcerative colitis (UC) and lupus, as well as several neurodegenerative diseases and cancer. Of note, multiple pan- and isozyme-selective PAD inhibitors are known and these inhibitors show efficacy in animal models of RA, UC, MS, lupus and sepsis.[1, 8, 10-12] The contribution of protein hypercitrullination to the pathology of various diseases has been further established using the phenylglyoxal (PG)-based citrulline-specific probes, Rhodamine-PG (Rh-PG) and Biotin-PG.[1, 8] For example, Rh-PG enabled

visualization of extensive citrullination of serum proteins and a marked decrease upon treatment with pan-PAD inhibitor, Cl-amidine, in a mouse model of UC.[13] Using Biotin-PG and a chemoproteomic platform, we also identified various classes of novel citrullinated proteins, including serine protease inhibitors (SERPINs), serine proteases, transport proteins and complement system components along with known citrullinated proteins (e.g., vimentin, enolase, keratin and fibrin) in the serum, synovial fluid and synovial tissue of RA patients.[14] Although the list of citrullinated proteins is ever expanding, the effect of citrullination on the structure and activity of a given protein remains poorly understood.

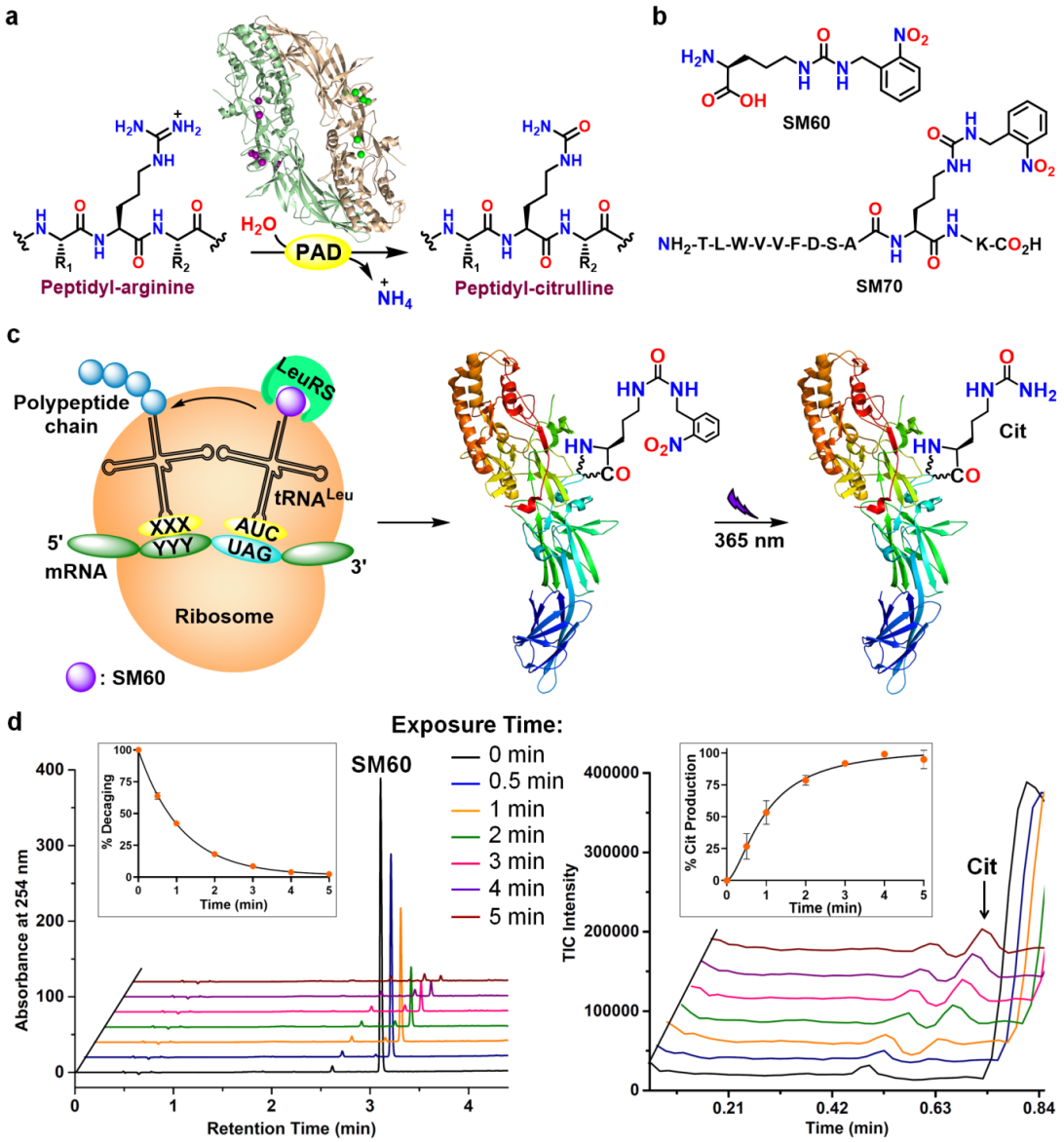


Figure 2-1. SM60/nb-Cit, a photocaged-citrulline and its conversion to citrulline with 365 nm UV light**.** a) Conversion of peptidyl-arginine to peptidyl-citrulline by the PADs. **b**) Chemical structures of **SM60** and **SM70**. **c)** Schematic representation of the incorporation of **SM60** into proteins by an engineered LeuRS-tRNA^{Leu} pair and subsequent conversion to citrulline. **d)** Decaging of **SM60** to citrulline with 365 nm UV irradiation. Left and right panels indicate the HPLC and ion chromatograms, showing the disappearance of **SM60** and the formation of citrulline (Cit), respectively,

with increasing UV exposure. Quantitative analyses of decaging and Cit formation are shown in the insets. Assay mixture: 1 mM **SM60**, 2 mM DTT, Phosphate-buffered saline pH 7.4.

Currently, the most commonly used strategy for generating a citrullinated protein involves its treatment with a PAD. However, this leads to citrullination at all sites that are available *in vitro*, which may not fully recapitulate the situation *in vivo*. Moreover, the degree of modification at each site is frequently partial, leading to a complex heterogeneous mixture.[14-16] Clearly, this strategy fails to provide information on the effect of individual citrullination events, underscoring the need for a method to site-specifically incorporate citrulline into proteins.

Although Gln mutations have been used as surrogates for citrulline (Cit),[17] Gln is smaller and does not accurately mimic the H-bonding patterns afforded by Cit. *In vitro* translation systems or post-translational mutagenesis approaches that have been used to incorporate Cit are limited by their cumbersome nature, the need for specialized equipment, and for the latter approach, the need to incorporate a dehydroalanine at the site of modification, which is itself challenging and generates a mixture of D- and L-stereoisomers.[18, 19] Additionally, these strategies preclude the expression of site-specifically citrullinated proteins in living cells, and therefore, are ineffective for interpreting the downstream implications of this PTM. By contrast genetic code expansion technologies enable the site-specific incorporation of unnatural amino acids (UAAs) into proteins using engineered aminoacyl-tRNA synthetase (aaRS)-tRNA

pairs.[20-24] This technology has been used to genetically encode many important PTMs, enabling the expression of homogeneously modified protein at desired sites in living cells.[25-28] However, genetically encoding Cit using this technology has remained elusive so far.

In this Chapter, we report a collaborated work with Dr. Mondal from the Thompson Lab. We report the facile site-specific incorporation of Cit into proteins in mammalian cells using an *E. coli*-derived engineered leucyl-tRNA synthetase (EcLeuRS)-EctRNA$^{Leu}_{CUA}$ pair. This pair, in response to a nonsense codon (UAG), charges a photocaged-citrulline, **nb-Cit** (Figure 2-1b, c), into proteins expressed in HEK293T or EXPI293F cells. Subsequently, the photocage is removed with 365 nm UV to generate Cit *in vitro* or in living cells (Figure 2-1c). As a proof-of-concept, we incorporated citrulline (Cit) at two well-known auto-citrullination sites, R372 and R374, in PAD4 and elucidated how these modifications impact enzyme activity.

2.2 Results

2.2.1 Development of nb-Cit: a photocaged-citrulline

Envisioning that it may be challenging to develop an engineered aaRS that would selectively charge Cit, while discriminating against a nearly isostructural arginine, we hypothesized that these challenges could be overcome through a caging strategy. As such, we designed a photocaged-citrulline (nb-Cit, comprising an *o*-nitrobenzyl

photocage on the Cit side chain), which is structurally distinct from the 20 canonical amino acids but can be efficiently converted to Cit post-translationally (Figures 2-1b, d). The nb-Cit used in this chapter is coming from two sources. 1) The published work was performed using material supplied by the Thompson group, namely SM60. [29] 2) Subsequently, I synthesized it in our lab under supervision of Dr. Roy with a different approach that was designed by Dr. Roy. (Figture 2-2) The nb-Cit is characterized by ^{1}HNMR spectroscopy and mass spectrometry to ensure the equivalency to SM60, supplied by Thompson group. Using LC-MS, Dr. Mondal found that **SM60** can be quantitatively converted to Cit in phosphate-buffered saline (PBS) supplemented with dithiothreitol (DTT) using 365 nm UV radiation for 5 min (Figures 2-1d). Quantitative conversion was further supported by ^{1}H NMR analysis, which shows the rapid disappearance of the benzylic protons of **SM60** at 4.5 ppm with increasing UV exposure and by the photo-decaging of **Fmoc-SM60** to **Fmoc-Cit**.

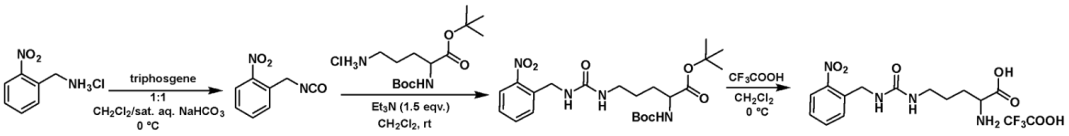


Figure 2-2. Synthetic scheme of nb-Cit.

2.2.2 Genetically encoding nb-Cit in eukaryotes

Four different aaRS/tRNA pairs have been successfully engineered for incorporating UAAs in eukaryotic cells: bacteria-derived tyrosyl, tryptophanyl, and leucyl pairs and the archaea-derived pyrrolysyl pair.[21-24, 30, 31] The first two pairs

are restricted to structural analogs of phenylalanine and tryptophan, respectively, precluding their use to genetically encode **nb-Cit**. However, both the archaeal pyrrolysyl (PylRS/tRNA^{Pyl}) and *E. coli* leucyl (EcLeuRS-EctRNA$^{Leu}_{CUA}$) pairs have been engineered to charge UAAs structurally similar to **nb-Cit**. Engineered aaRSs often exhibit substrate polyspecificity, i.e. the ability to use several structurally analogous UAAs, while discriminating against the canonical amino acids. This property has provided a facile route to rapidly expand the repertoire of genetically encoded UAAs without having to engineer new aaRS mutants for each distinct substrate. To explore if such a polyspecific aaRS can accept **nb-Cit** as a substrate, we screened several existing PylRS and EcLeuRS mutants using an EGFP-39-TAG expression assay in HEK293T cells in the presence of their cognate amber suppressor tRNA. This screen identified an EcLeuRS mutant (M40I, Y499I, Y527A and H529G in the active site, and T252A in the editing domain)[32] which enabled robust expression of the fluorescence reporter only when **nb-Cit** was supplemented in the medium (Figure 2-3a,b). Purification of the resulting full-length EGFP using a C-terminal polyhistidine tag, followed by mass spectrometry, showed a mass consistent with the successful incorporation of **nb-Cit** (Figures 2-3c,d). Furthermore, UV irradiation of cells expressing EGFP-39-**nb-Cit** before lysis followed by protein purification and MS analysis afforded a single protein mass consistent with the complete deprotection and incorporation of citrulline at position 39 of EGFP (Figure 2-3d). Notably, **nb-Cit** is nontoxic (EC_{50} > 10 mM) to HEK293T cells at the concentration used for nonsense suppression, i.e., 1 mM.

Furthermore, a combination of **nb-Cit** and 365 nm UV exhibits an EC_{50} of 4.5 ± 0.2 mM for the inhibition of cell proliferation, indicating that the products of the photodecaging reaction (nitrosobenzaldehyde and citrulline) have negligible cytotoxicity at the working concentration.

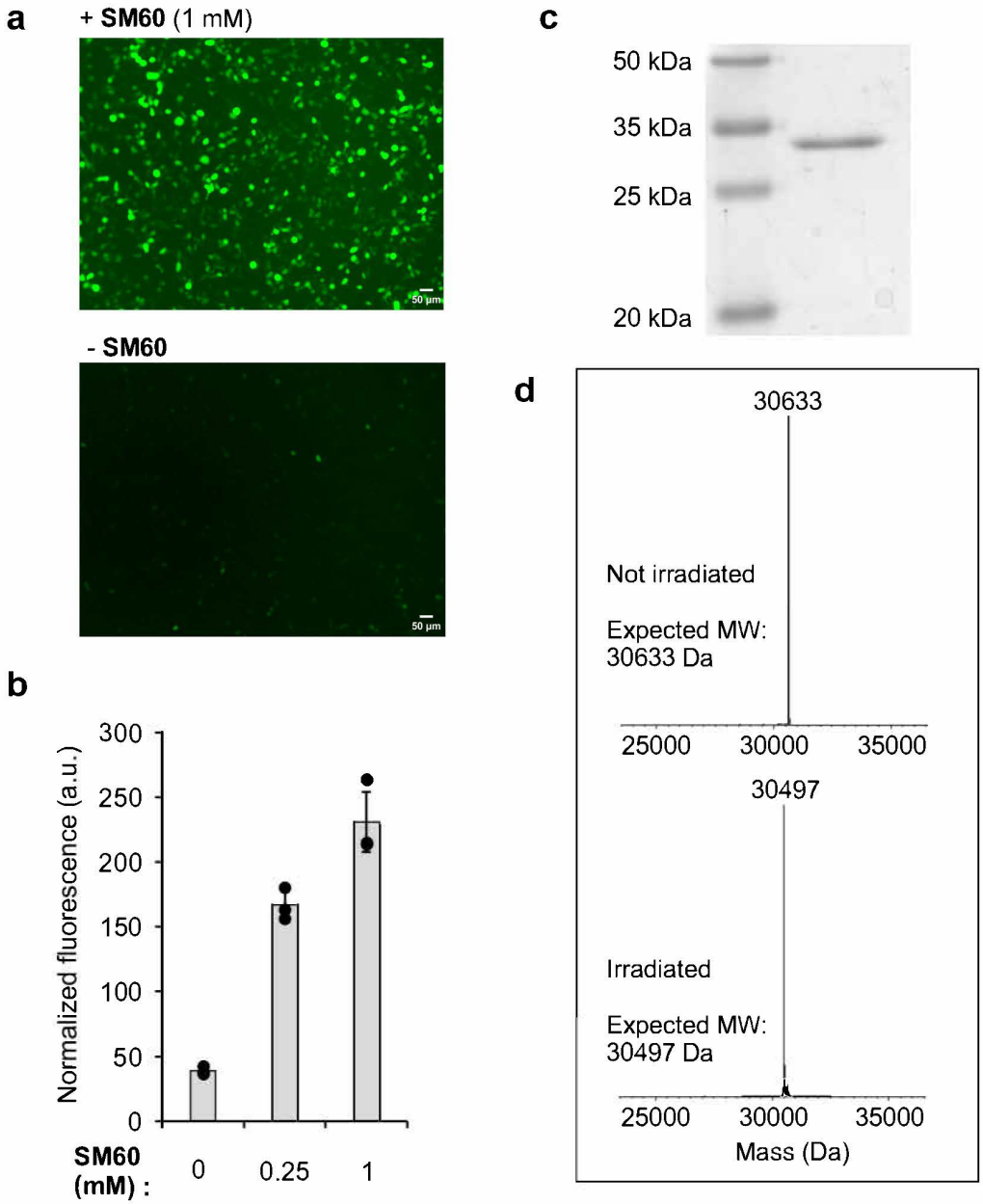


Figure 2-3. Site-specific incorporation of SM60 in EGFP and subsequent conversion to citrulline in HEK293T cells. a) EGFP-39-TAG reporter expression by EcLeuRS-

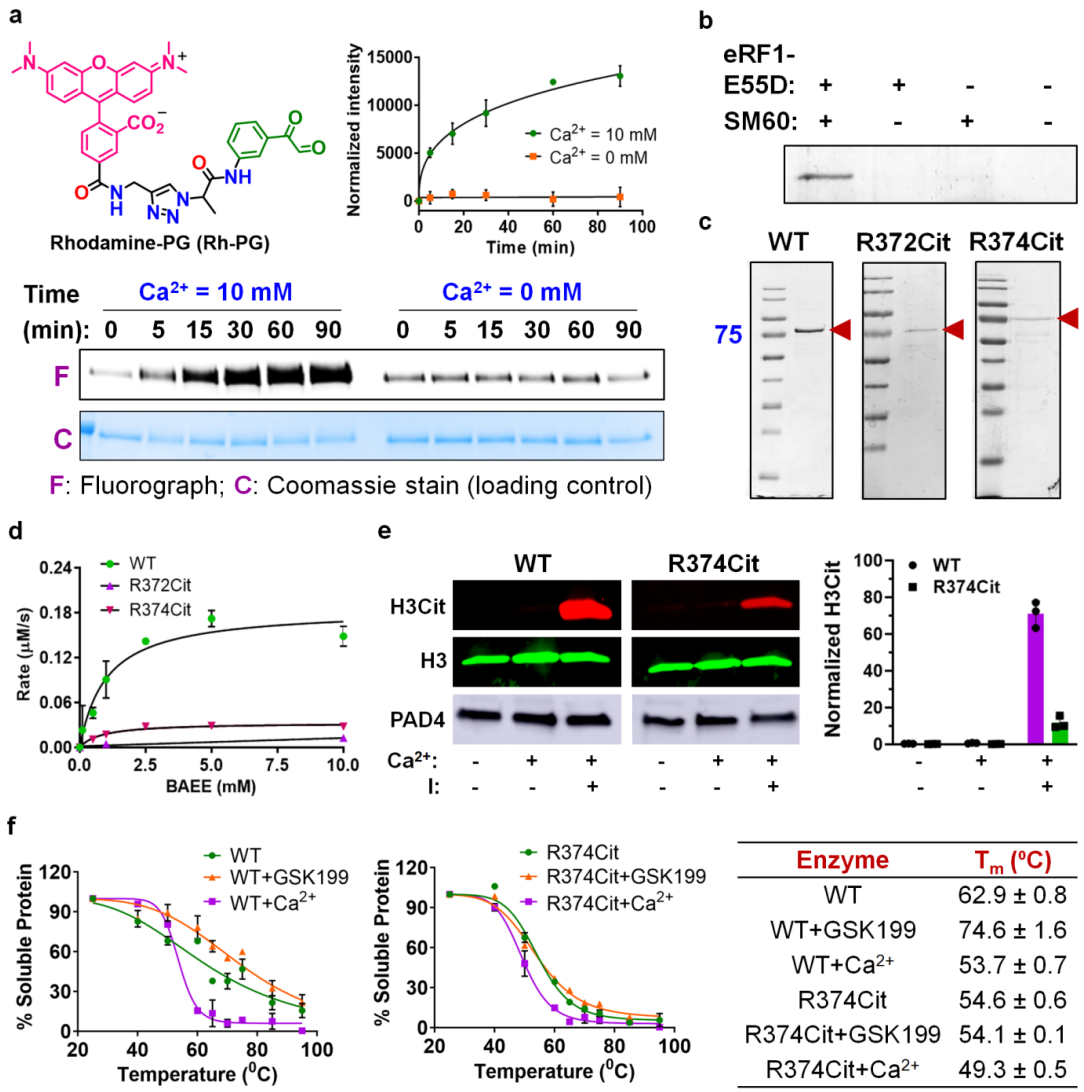
EctRNA$^{Leu}_{CUA}$ pair in HEK293T cells in the presence of **SM60** indicated by the fluorescence of EGFP. b) Quantification of EGFP-39-TAG reporter expression efficiency in the presence of increasing concentration of **SM60**. c) Coomassie stain of purified EGFP containing **SM60**. Full gel is given in Figure 2-6. d) Deconvoluted mass spectrum of EGFP before and after 365 nm UV irradiation (1 min), indicating the presence of **SM60** and citrulline, respectively, at 39 position. Non-deconvoluted spectra are given in Figures 2-7,8.

2.2.3 Site-specific incorporation of citrulline in PAD4

Having established our ability to site-specifically incorporate Cit into EGFP, we sought to exploit this technology to address the effect of autocitrullination on PAD4 activity. We focused on these studies because the effect of autocitrullination on PAD4 activity has been debated. While Andrade *et al.* reported that autocitrullination negatively impacts PAD4 activity, Thompson *et al*. showed that autocitrullination has little to no impact on PAD4 activity.[17, 33] Using a citrulline-specific fluorescent probe Rh-PG,[13] we confirmed that PAD4 autocitrullinates in the presence of Ca^{+2} in a time-dependent manner (Figure 2-3a). We and others have previously mapped several autocitrullination sites in PAD4.[17, 33] While most of these sites are on the surface, the frequently observed R372 and R374 sites are present in the active site. Notably, the guanidinium groups on these two residues are only 3.5 Å from each other and the expected electrostatic repulsions are delicately balanced by H-bonding and salt-bridge

interactions with D345. Moreover, R374 forms two H-bonds with the small molecule substrate, BAA.[34] Therefore, we expected that citrullination at these sites would significantly impact enzyme activity. To evaluate this hypothesis, we incorporated Cit at positions 372 and 374 in PAD4. Wild-type (WT) PAD4 and the 372 and 374 TAG mutants were separately cloned into a pAcBac3 plasmid,[32, 35] which also encodes the mutant EcLeuRS and 8 copies of the $EctRNA^{Leu}_{CUA}$. Subsequently, these plasmids were transfected separately into HEK293T cells, and the PAD4 protein or its mutants (after irradiation to remove the photocage), were purified using a C-terminal poly-histidine tag. While WT PAD4 expression was robust (10 μg/10^7 cells), yields for the mutants were very low, indicating poor suppression efficiency at these sites. The Chin group has recently reported a mutant eukaryotic release factor (eRF1 E55D) that can enhance TAG-suppression efficiency in mammalian cells upon overexpression.[36] To explore if this strategy can overcome the low suppression efficiency at the target sites in PAD4, we cloned eRF1-E55D mutant under a CMV promoter in a pIDTSMART vector. Indeed, co-transfection of this plasmid significantly improved the efficiency of nonsense suppression and enabled the purification of the desired mutants (2-4 μg/10^7 cells, Figure 2-4b, c). Since the +0.98 Da mass change upon citrullination is difficult to detect by intact MS analysis, we confirmed the incorporation of citrulline at the desired position by LC-MS/MS analysis of the peptides resulting from Lys-C/Glu-C digestion of PAD. Notably, we found that PAD4 expression does not cause any cytotoxicity in the HEK293T cells. Using a similar procedure, we also expressed and purified wild-

type PAD4 and the R374Cit mutant from EXPI293F cells that grow in suspension and are highly scalable expression system.



Enzyme	T_m (°C)
WT	62.9 ± 0.8
WT+GSK199	74.6 ± 1.6
WT+Ca^{2+}	53.7 ± 0.7
R374Cit	54.6 ± 0.6
R374Cit+GSK199	54.1 ± 0.1
R374Cit+Ca^{2+}	49.3 ± 0.5

Figure 2-4. Autocitrullination of PAD4, incorporation of citrulline in PAD4 and implications thereof. a) Chemical structure of Rh-PG and the fluorescence labeling of autocitrullinated PAD4. The bands at 0 min in the presence of calcium and at all the time-points in the absence of calcium correspond to the basal levels of autocitrullination during the expression and purification of PAD4 from *E. coli*. b) Expression of R372Cit PAD4 in the presence of engineered release factor, eRF1-E55D and **SM60** in HEK293T

cells, indicating the essential role of eRF1-E55D for efficient TAG-suppression. c) Coomassie stains for WT, R372Cit and R374Cit PAD4, indicating their purity. d) Michaelis-Menten kinetics for WT, R372Cit and R374Cit PAD4. e) Western blot analysis of histone H3 citrullination in live HEK293T cells by WT and R374Cit PAD4. The normalization procedure is given in the supporting information. f) Thermal shift profiles of WT and R374Cit PAD4. The table indicates the melting temperatures T_m. g) Effect of autocitrullination on the enzymatic activity of PAD4.

2.2.4 Activity of WT, R372Cit, R374Cit and autocitrullinated PAD4

Our collaborator, Dr. Mondal, first ensured that the biochemical activity and calcium dependence of WT PAD4 expressed from HEK293T cells ($PAD4_{Mam}$) and *E. coli* ($PAD4_{Bac}$) are similar. Next, he determined kinetic values for the R372Cit and R374Cit mutants. Notably, both WT and the R374Cit mutant exhibited a time-dependent increase in citrulline production. By contrast, the R372Cit mutant produced only a negligible amount of citrulline after 90 min. Furthermore, the rate of citrulline production, indicated by the slope of the straight line, is significantly lower for the R374Cit mutant than WT PAD4. In agreement, the k_{cat}/K_m values of the R374Cit and R372Cit mutants are 9- and 181-fold lower than that for WT PAD4 (Figure 2-4d). These *in vitro* results led us to investigate the activity of WT and R374Cit PAD4 in live cells. Specifically, I overexpressed the two enzymes in HEK293T cells and evaluated their ability to citrullinate histone H3. Treatment of PAD4-overexpressing HEK293T cells

with calcium and a calcium ionophore, i.e., ionomycin, followed by western blot analysis indicated that WT PAD4 is 6-times more active than the R374Cit mutant for the citrullination of histone H3, consistent with the *in vitro* results (Figure 2-4e).

Intrigued by these results, we sought to understand why the activity of these Cit-containing mutants is lower than WT PAD4. Kinetic studies indicated that R374Cit mutant possesses a similar K_m, but a 10-fold lower k_{cat} than WT PAD4, suggesting a slow conversion of substrate to product. Furthermore, RFA, a PAD-targeted activity-based probe that covalently modifies the active site cysteine, C645,[37] fluorescently labeled only WT PAD4 when tested with both purified enzymes and enzyme-containing cell lysates. Based on these observations, we hypothesized that citrullination may induce local conformational changes within the active site, leading to very slow or no reaction between C645 and the guanidium group of substrates, BAEE or the fluoroacetamidine warhead on RFA. To investigate this possibility, Dr. Mondal performed a thermal shift assay in the presence of a PAD4-selective ligand, GSK199, which binds to an allosteric pocket near the active site and H-bonds to both D473 and H471.[38] Using this assay, we found the melting temperatures (T_m) of WT and R374Cit PAD4 to be 62.9 and 54.6 ºC, respectively (Figures 2-4). Despite having a lower T_m, the R374Cit mutant is as stable as WT PAD4 at 37 ºC. As expected, GSK199 increased the T_m of WT PAD4, however, it did not increase the T_m of the R374Cit mutant (Figure 2-4). These results indicate that GSK199 binds poorly to the R374Cit mutant, likely due to local conformational changes around the active site. However, the

overall folding of the mutant is the same as WT since both the WT and R374Cit proteins exhibit a similar ΔT_m in the presence of calcium (Figure 2-4).

2.2.5 Quantitative proteomics of PAD4 autocitrullination

As discussed earlier, conflicting reports indicated that autocitrullination can either inactivate the enzyme or have no effect. Since our present results suggest that autocitrullination should decrease PAD4 activity, Dr. Mondal regenerated autocitrullinated PAD4 by incubating the enzyme in the presence of 10 mM $CaCl_2$. Consistent with their previous observations, autocitrullinated PAD4 exhibits similar activity to control PAD4 (incubated in the absence of $CaCl_2$). The activity loss for both autocitrullinated and control PAD4 in this experiment is likely due to the oxidation of C645 over time. Given that autocitrullination does not decrease enzyme activity, but citrullination of R372 and R374 does, we asked two questions. Are R372 and R374 the preferred sites of autocitrullination? Also, what fraction of PAD4 gets autocitrullinated? If only a small fraction of PAD4 is autocitrullinated, then this process should not impact the activity of uncitrullinated enzyme.

To answer these questions, we took a quantitative proteomics approach. PAD4 was autocitrullinated for various times, and digested with Glu-C and Lys-C to maximize peptide coverage. The resulting peptides were then labeled with tandem mass tags (TMT) and were subjected to tandem mass analysis. Enzyme incubated in the absence of calcium was used as the negative control. From this analysis, we identified 13 unique

citrullination sites on PAD4 (Figure 2-5a). Notably, these exclude the previously reported R156, R205, R383, R609 and R639 and include two new sites, R650 and R651. Although both R372 and R374 residues show a time-dependent increase in citrullination, citrullination of arginines 212/218, 484/488/495 and 650/651 occurs at much higher rate. For example, the extent of citrullination at arginines 212/218 and 484/488/495 is significantly higher at 5 min than that at the 372 and 374 sites after 90 min (Figures 2-5a b). These observations indicate that arginines 372 and 374 are not the preferred sites of autocitrullination. Additionally, none of the observed autocitrullination sites exhibited a marked decrease in the arginine-containing parent peptide levels, which indicates that only a minor fraction of PAD4 undergoes autocitrullination, further explaining its nominal impact on the enzyme activity. Nonetheless, these results showcase how this technology provides the ability to systematically characterize the behavior of individual citrullinated isoforms of any protein – both *in vitro* and in living cells – providing a powerful new approach to understand the biology of this PTM.

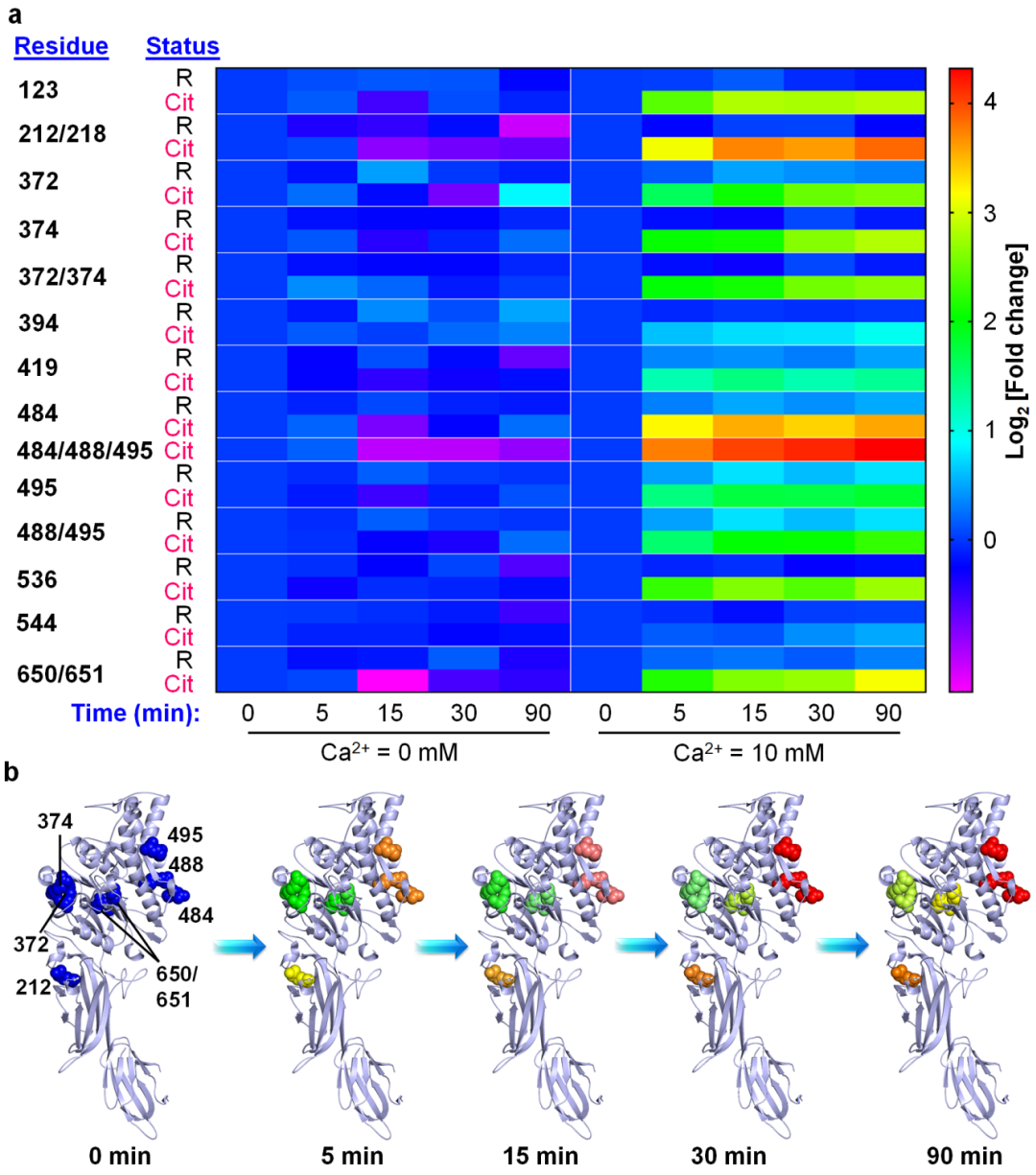


Figure 2-5. Autocitrullination sites in PAD4. a)Heat map representing the time-dependent change in peptides containing arginine or citrulline at the indicated positions (autocitrullination sites). Ca^{2+}-untreated samples are negative controls. b) Time-dependent autocitrullination at the major sites. R218 site could not be shown because of disorder in that region (PDB: 1WDA).[34] The increase in autocitrullination at these sites follows the same color code as in panel a.

2.3 Conclusion and Discussion

Herein, we report the development of a novel technique for the site-specific incorporation of citrulline into proteins in mammalian cells. Central to this technology are a photocaged citrulline (nb-Cit/**SM60**) and an *E. coli*-derived engineered leucyl-tRNA synthetase (EcLeuRS)/EctRNA$^{Leu}_{CUA}$ pair that enables the incorporation of **nb-Cit** into proteins in response to the TAG nonsense codon with high fidelity and efficiency. Subsequently, the photocage is removed with 365 nm light to generate citrulline. This technique overcomes several limitations of previous methods used to incorporate citrulline, including *in vitro* translation, post-translational mutagenesis, and *in vivo* nonsense suppression by chemically acylated tRNAs.[18, 19, 39] For example, chemically acylated tRNAs are not readily synthesized, cannot be regenerated, and consequently give poor yields. By contrast, our approach provides a highly scalable expression platform that can be readily adapted by virtually any lab to site-specifically incorporate Cit into any mammalian protein, and thereby facilitate cellular studies to understand the downstream implications of this PTM.

Specifically, we incorporated citrulline at two known autocitrullination sites, R372 and R374, in PAD4. Kinetic studies indicate that the R374Cit and R372Cit mutants are 9- and 181-fold less active than WT PAD4. Detailed studies indicate that citrullination induces local conformational changes within the active site that leads to slow reaction

between C645 and the guanidium group of the substrate, the first step in the catalytic cycle. While these results indicate that citrullination of R372 and R374 would decrease PAD4 activity, we found that autocitrullination does not impact the enzymatic activity. Quantitative proteomics studies indicate that 212/218 and 484/488/495, and not 372 and 374, are the preferred sites of citrullination. While faster autocitrullination of arginines 212/218 and 484/488/495 is likely due to their residence at the surface of PAD4, upon citrullination, they may expose deeply buried autocitrullination sites by conformational changes. Efforts are currently under way to elucidate the effect of citrullination at these major sites, particularly the 484/488/495 residues because they are present at the interface of the head-to-tail PAD4 dimer that is known to alter enzymatic activity.

Since it is well established that citrullination is critical for many physiological processes, as well as in disease pathology, this new method will provide a direct and accessible approach to understand the biology of this PTM at the molecular level. For example, histone H3 citrullination at R26 leads to the transcriptional activation of more than 200 genes in estrogen receptor-positive breast cancer cells and inhibits the methylation of the neighboring K27 residue by 30,000-fold.[15] However, the mechanism of negative crosstalk between these two PTMs remains poorly understood. Additionally, we recently showed that serine protease inhibitors (SERPINs), nicotinamide N-methyl transferase (NNMT), and pyruvate kinase isoform M2 (PKM2) are citrullinated in patients suffering from rheumatoid arthritis. Notably, the

citrullination of the SERPINs and NNMT dramatically abolishes their enzymatic activity, while citrullinated PKM2 exhibits 2-3-fold higher activity than the WT enzyme.[14] However, the underlying reasons behind such biochemical phenomenon are unclear. Finally, citrullination has been reported to impact neutrophil extracellular trap formation, pluripotency, and efficient elongation by RNA PolII but, again, the underlying mechanisms remain unclear.[1, 8] With our new enabling technology, it is now possible to incorporate citrulline on demand and mechanistically address how this PTM impacts these fundamental biological processes and pathways.

This work demonstrates the site-specific incorporation of PTM in GCE allows the dissection of function of PTM that was unable to achieved by traditional universal incorporation. After publication of the work, we further tested a newly evolved $tRNA^{Leu}$, which is significantly more efficient than previous one, for nb-Cit incorporation in PAD4. The new $tRNA^{Leu}$ can increase the overexpression of nb-Cit containing PAD4 to a comparable level to WT PAD4 without the help of eRF1-E55D.

2.4 Method:

Construction of plasmids

The previously reported pAcBac3-EcLeuTAG-EGFP-39-TAG plasmid was used to construct additional plasmids.[32] EGFP-39-TAG was replaced with WT PAD4 using SfiI restriction site to create pAcBac3-EcLeuPLRS1TAG-PAD4WT. For incorporation of nb-Cit, we introduced TAG nonsense codon at the desired sites by site-directed

mutagenesis based on the pAcBac3-EcLeuPLRS1TAG-PAD4WT plasmid. pIDTSMART eRF1 E55D was generated following a paper from the Chin lab.[40]

HEK293T cell culture and transfection

HEK293T cells were maintained at 37 °C in a humidified incubator supplemented with 5% CO_2. Cells were seeded at 9×10^6 cells per 10 cm plate 24 h before transfection. EGFP and WT PAD4 transfections were performed by incubating 10 μg plasmid DNA, 50 μL PEI (1 mg/mL; Polysciences, Warrington, PA), and 180 μL DMEM for 10 min at room temperature, followed by adding the solution dropwise to the culture medium of the cells. For nb-Cit incorporation into PAD4 at positions 372 and 374, 12 μg of PAD4 R372TAG or PAD4 R374TAG and 8 μg of pIDTSMART eRF1 E55D plasmids were incubated with a mixture of 100 μL PEI and 180 μL DMEM for 10 min at room temperature before adding to cells. Nb-Cit was added at the same time to a final concentration of 1 mM, and 2 mM sodium butyrate was added to enhance protein expression.

EGFP fluorescence analysis

EGFP fluorescence was analyzed 48 h after transfection. DMEM was exchanged with PBS and the plates were irradiated at 365 nm (120 Watt, 10 cm x 10 cm LED array; Larson Electronics) for 75 s at 4 °C to decage nb-cit. Cells were then harvested and resuspended in 600 μL CelLytic M buffer (Sigma, St. Louis, MO) with Halt Protease inhibitor Cocktail (Thermo Scientific, Waltham, MA) and Pierce Universal Nuclease for Cell Lysis (Fisher Scientific, Hampton, NH). Lysate were clarified by centrifugation

at 16,000 x *g* for 10 min and 100 µL of supernatant was transferred to a clear-bottom 96-well plate for fluorescence measurement following previously described protocol.[32] All the experiments were performed at least in duplicate.

Synthesis of nb-Cit

2-nitrobenzylamine hydrochloride salt (1 g, 5.3 mmol) was dissolved in a biphasic mixture of CH_2Cl_2 (20 mL) and sat. aq. NaHCO3 (20 mL) and cooled to at 0 ˚C with ice-salt bath. After stirring vigorously for 30 min, a solution of triphosgene (1.56 g, 5.3 mmol) in dry dichloromethane (5 ml) was added dropwise. This mixture was vigorously stirred at 0 °C for 15 min, organic layer was separated and the aq. layer was extracted with 2x10 mL CH_2Cl_2. The combined organic layer was washed with brine (10 mL), dried over anhydrous Na_2SO_4 and concentrated under reduced pressure using a rotary evaporator to give the isocyanate intermediate as an oil (740 mg) which was used directly in the next step without further purification.

To a solution of *N*-Boc-*tert*-butylester of ornithine hydrochloride (415 mg, 1.28 mmol, 1 eqv.) in CH_2Cl_2 (15 mL), Et_3N (270 µL) was added dropwise and cooled to 0 ˚C while stirring. A solution of the intermediate isocyanate (1 eqv.) in CH_2Cl_2 (2 mL) was added and the resulting mixture was stirred at room temperature for 4 h. The organic layer was washed with 2×10 mL water, followed by brine (5 mL) and dried over anhydrous Na_2SO_4. After concentrating the organic layer under reduced pressure using a rotary evaporator, the resulting crude product was purified by flash chromatography (1:1 EtOAc-hexane; 100-200 silica gel) to afford the *N*-Boc-*tert*-butyl ester of cittruline

intermediate as an oil (750 mg, 80% yield). **^{1}H-NMR (500 MHz, $CDCl_3$):** δ 7.80-7.82 (m; 1H), 7.39-7.45 (m; 2H), 7.21-7.25 (m; 1H), 6.00-6.02 (m; 1H), 5.70-5.72 (m; 1H), 5.23-5.25 (m; 1H), 4.40-4.43 (m; 2H), 3.93-4.00 (m; 1H), 2.95-3.03 (m; 2H), 1.61-1.65 (m; 1H), 1.44-1.50 (m; 1H), 1.35-1.41 (m; 2H), 1.31 (s; 9H), 1.29 (m; 9H); **^{13}C-NMR (125 MHz, $CDCl_3$):** δ 171.49, 158.53, 155.37, 147.57, 135.50, 133.38, 129.96, 127.55, 124.41, 81.45, 79.27, 77.20, 53.70, 41.15, 39.47, 27.98, 27.61, 25.92. ESI-TOF-MS (m/z): $[M+H]^+$ Calculated for $C_{22}H_{35}N_4O_7$ 466.2455, obtained 466.2428.

The intermediate was dissolved in CH_2Cl_2 (4 mL) and anhydrous trifluoroacetic acid (4 mL) was added dropwise at 4 °C and the resulting mixture was stirred at 4 °C for 4 h to have complete removal of Boc protection (judged by LC-MS). The mixture was concentrated under reduced pressure. To the resulting oil CH_2Cl_2 (~10 mL) was added and concentrated to dryness. The process was repeated several times to get rid of dissolved trifluoroacetic acid. The expected product (photocaged citrulline) was obtained as a white solid after lyophilization (435 mg, quantitative yield). **^{1}H-NMR (600 MHz, D_2O):** δ 8.03-8.05 (m; 1H), 7.63-7.66 (m; 1H), 7.44-7.47 (m; 2H), 4.78 (s; 2H), 3.68-3.72 (m; 1H), 3.96-3.98 (m; 1H), 3.18-3.22 (m; 2H), 1.76-1.90 (m; 2H), 1.55-1.71 (m; 2H); ESI-TOF-MS (m/z): $[M+H]^+$ Calculated for $C_{13}H_{19}N_4O_5$ 311.1435, obtained 311.1421.

2.5 Supplementary materials:

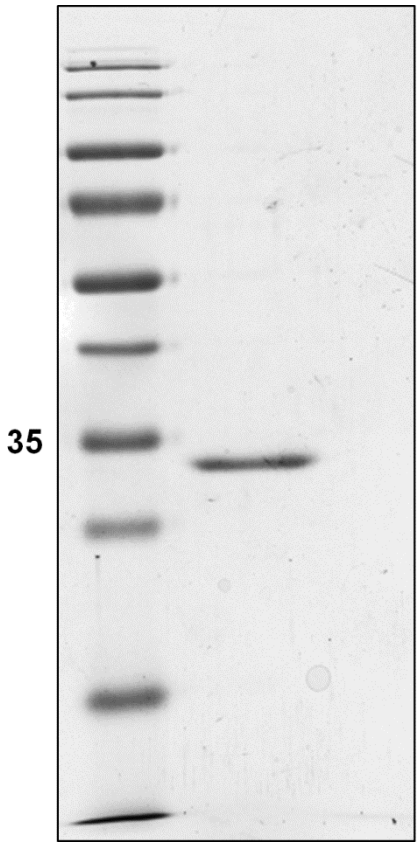


Figure 2-6. Coomassie stain of purified EGFP containing **nbcit** at 39 position.

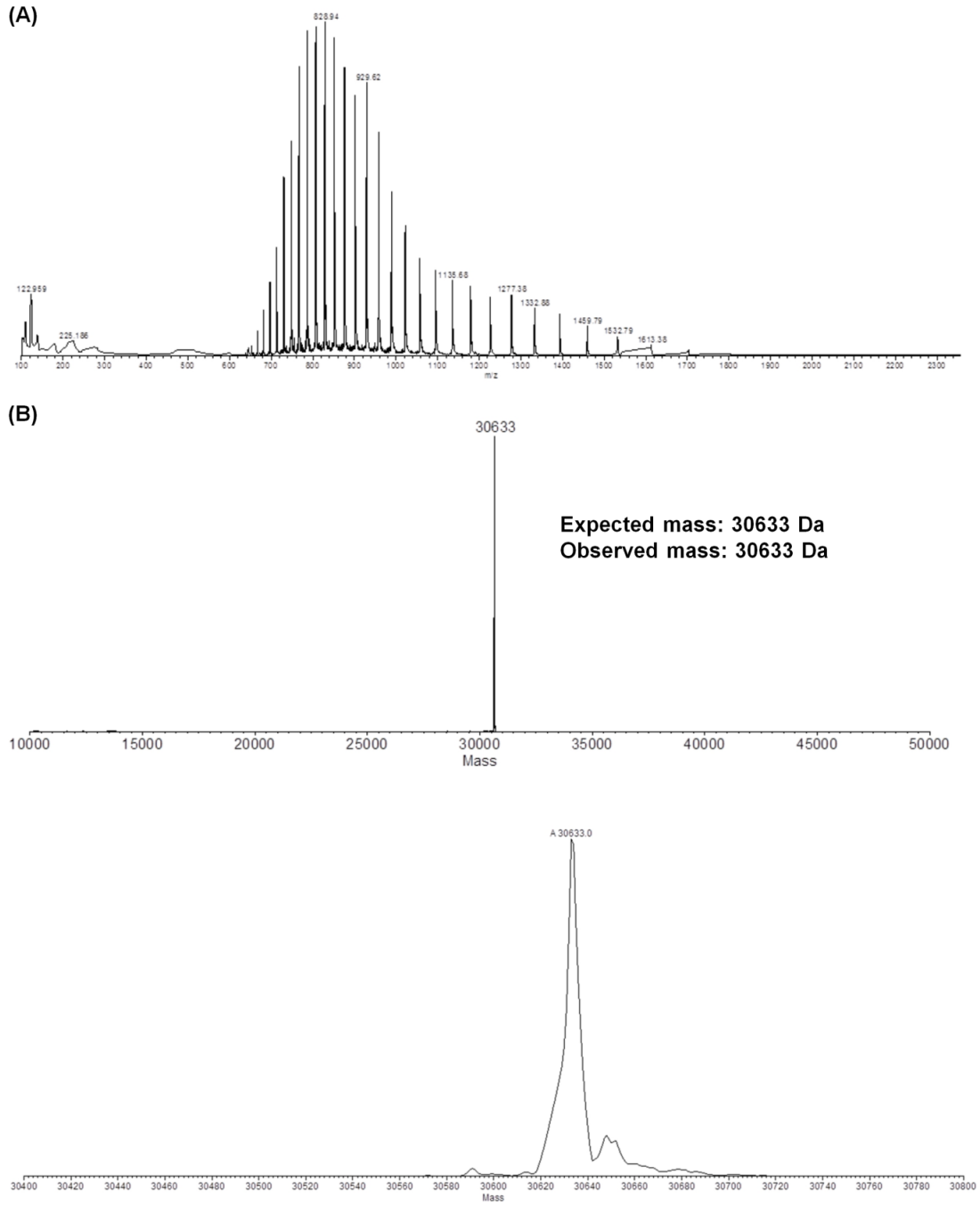


Figure 2-7. ESI mass spectrum (A) and deconvoluted spectra (B) of EGFP containing **SM60** at 39 position.

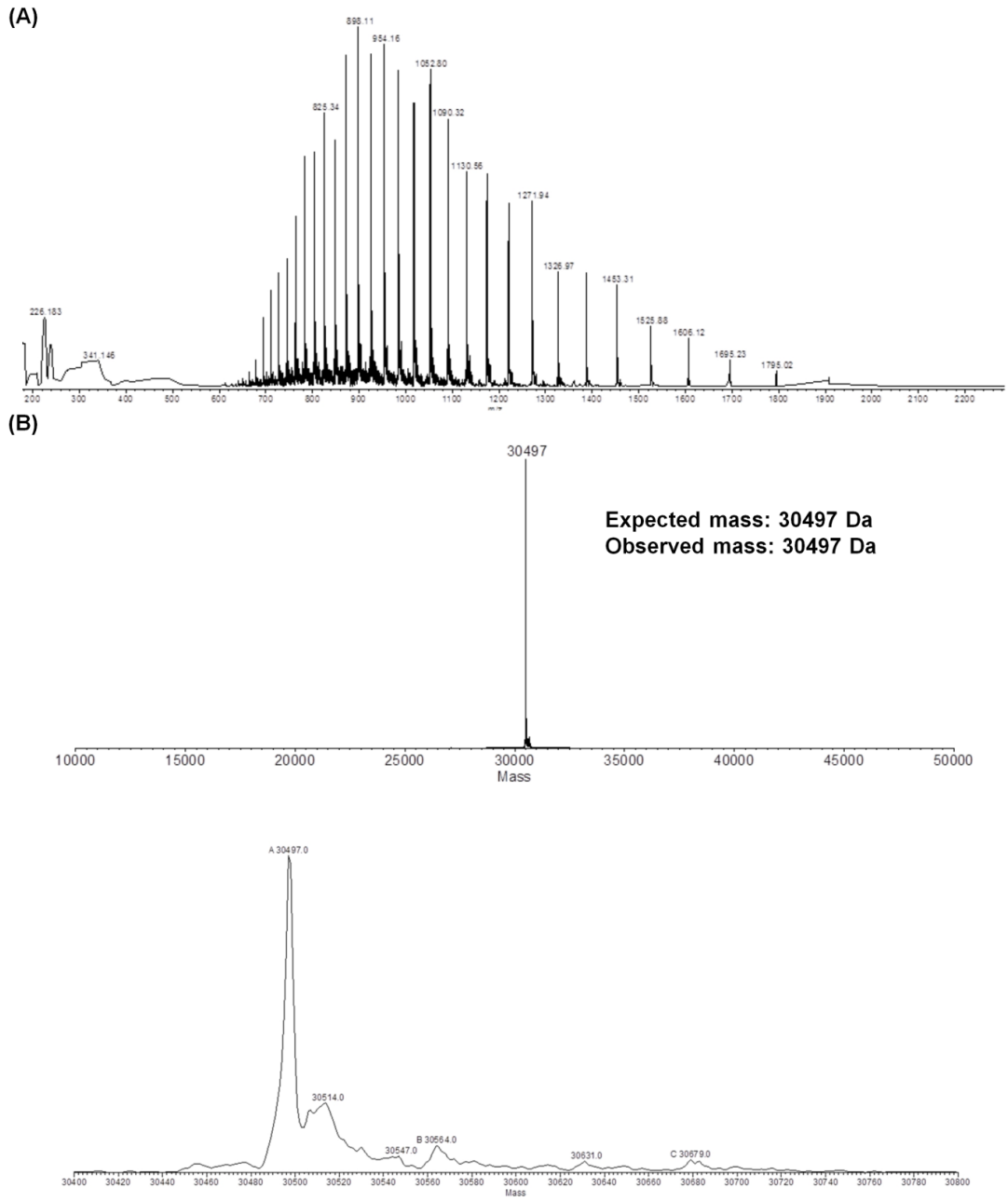
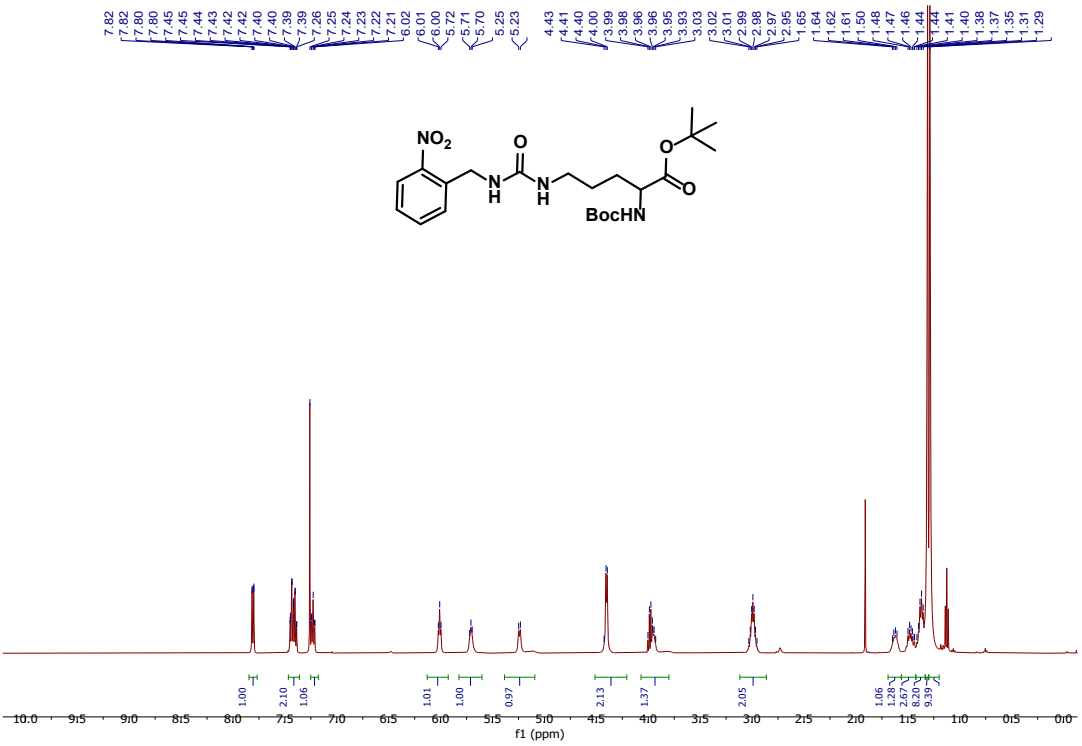


Figure 2-8. ESI mass spectrum (A) and deconvoluted spectra (B) of EGFP containing **Cit** at 39 position.


7.82
7.82
7.80
7.80
7.45
7.45
7.44
7.43
7.42
7.42
7.40
7.40
7.39
7.39
7.26
7.25
7.24
7.23
7.22
7.21
6.02
6.01
6.00
5.72
5.71
5.70
5.25
5.23
4.43
4.41
4.40
4.00
3.99
3.98
3.96
3.96
3.95
3.93
3.03
3.02
3.01
2.99
2.98
2.97
2.95
1.65
1.64
1.62
1.61
1.50
1.48
1.47
1.46
1.44
1.44
1.41
1.40
1.38
1.37
1.35
1.31
1.29
NO2
O
N
H
N
H
O
O
BocHN
1.00
2.10
1.06
1.01
1.00
0.97
2.13
1.37
2.05
1.06
1.28
2.67
8.20
9.39
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

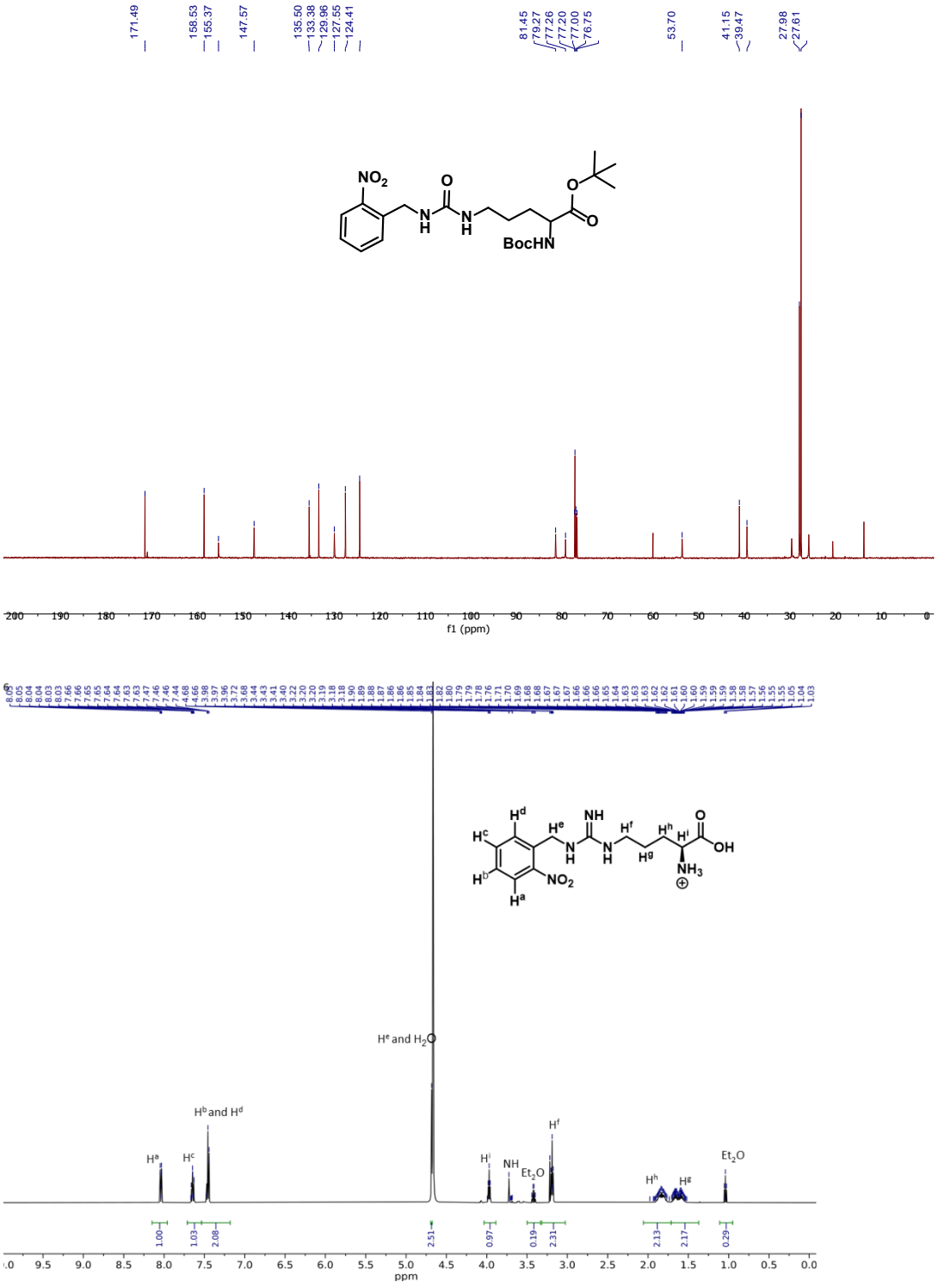


Figure 2-9. H^1-NMR and C^{13}-NMR for intermedia and product of nb-cit synthesis.

2.6 Reference:

1. Fuhrmann, J., K.W. Clancy, and P.R. Thompson, *Chemical biology of protein arginine modifications in epigenetic regulation.* Chem Rev, 2015. **115**(11): p. 5413-61.
2. Brinkmann, V., et al., *Neutrophil extracellular traps kill bacteria.* Science, 2004. **303**(5663): p. 1532-5.
3. Christophorou, M.A., et al., *Citrullination regulates pluripotency and histone H1 binding to chromatin.* Nature, 2014. **507**(7490): p. 104-8.
4. Cuthbert, G.L., et al., *Histone deimination antagonizes arginine methylation.* Cell, 2004. **118**(5): p. 545-53.
5. Kenny, E.F., et al., *Diverse stimuli engage different neutrophil extracellular trap pathways.* Elife, 2017. **6**.
6. Tanikawa, C., et al., *Regulation of histone modification and chromatin structure by the p53-PADI4 pathway.* Nat Commun, 2012. **3**: p. 676.
7. Zhang, X., et al., *Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor alpha target gene activation.* Proc Natl Acad Sci U S A, 2012. **109**(33): p. 13331-6.
8. Mondal, S. and P.R. Thompson, *Protein Arginine Deiminases (PADs): Biochemistry and Chemical Biology of Protein Citrullination.* Acc Chem Res, 2019. **52**(3): p. 818-832.
9. Slade, D.J., et al., *Protein arginine deiminase 2 binds calcium in an ordered fashion: implications for inhibitor design.* ACS Chem Biol, 2015. **10**(4): p. 1043-53.
10. Mondal, S., et al., *Halogen Bonding Increases the Potency and Isozyme Selectivity of Protein Arginine Deiminase 1 Inhibitors.* Angew Chem Int Ed Engl, 2019. **58**(36): p. 12476-12480.
11. Mondal, S., et al., *Photochemical Control of Protein Arginine Deiminase (PAD) Activity.* ACS Chem Biol, 2018. **13**(4): p. 1057-1065.
12. Wu, Z., et al., *Inhibition of PAD2 Improves Survival in a Mouse Model of Lethal LPS-Induced Endotoxic Shock.* Inflammation, 2020.
13. Bicker, K.L., et al., *Seeing citrulline: development of a phenylglyoxal-based probe to visualize protein citrullination.* J Am Chem Soc, 2012. **134**(41): p. 17015-8.
14. Tilvawala, R., et al., *The Rheumatoid Arthritis-Associated Citrullinome.* Cell Chem Biol, 2018. **25**(6): p. 691-704 e6.
15. Clancy, K.W., et al., *Citrullination/Methylation Crosstalk on Histone H3 Regulates ER-Target Gene Transcription.* ACS Chem Biol, 2017. **12**(6): p. 1691-1702.

16. Nemmara, V.V., et al., *Citrullination Inactivates Nicotinamide- N-methyltransferase.* ACS Chem Biol, 2018. **13**(9): p. 2663-2672.
17. Slack, J.L., et al., *Autodeimination of protein arginine deiminase 4 alters protein-protein interactions but not activity.* Biochemistry, 2011. **50**(19): p. 3997-4010.
18. Akahoshi, A., et al., *Site-specific incorporation of arginine analogs into proteins using arginyl-tRNA synthetase.* Biochem Biophys Res Commun, 2011. **414**(3): p. 625-30.
19. Wright, T.H., et al., *Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity.* Science, 2016. **354**(6312).
20. Chin, J.W., *Expanding and reprogramming the genetic code.* Nature, 2017. **550**(7674): p. 53-60.
21. Dumas, A., et al., *Designing logical codon reassignment - Expanding the chemistry in biology.* Chem Sci, 2015. **6**(1): p. 50-69.
22. Italia, J.S., et al., *An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes.* Nat Chem Biol, 2017. **13**(4): p. 446-450.
23. Italia, J.S., et al., *Expanding the genetic code of mammalian cells.* Biochem Soc Trans, 2017. **45**(2): p. 555-562.
24. Young, D.D. and P.G. Schultz, *Playing with the Molecules of Life.* ACS Chem Biol, 2018. **13**(4): p. 854-870.
25. Groff, D., et al., *A genetically encoded epsilon-N-methyl lysine in mammalian cells.* Chembiochem, 2010. **11**(8): p. 1066-8.
26. Italia, J.S., et al., *Genetically encoded protein sulfation in mammalian cells.* Nat Chem Biol, 2020. **16**(4): p. 379-382.
27. Luo, X., et al., *Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria.* Nat Chem Biol, 2017. **13**(8): p. 845-849.
28. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding N(epsilon)-acetyllysine in recombinant proteins.* Nat Chem Biol, 2008. **4**(4): p. 232-4.
29. Mondal, S., et al., *Site-specific incorporation of citrulline into proteins in mammalian cells.* Nature Communications, 2021. **12**(1): p. 45.
30. Chin, J.W., et al., *An expanded eukaryotic genetic code.* Science, 2003. **301**(5635): p. 964-7.
31. Wu, N., et al., *A genetically encoded photocaged amino acid.* J Am Chem Soc, 2004. **126**(44): p. 14306-7.
32. Zheng, Y., et al., *Expanding the Scope of Single- and Double-Noncanonical Amino Acid Mutagenesis in Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases.* Biochemistry, 2018. **57**(4): p. 441-445.
33. Andrade, F., et al., *Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation.* Arthritis Rheum, 2010. **62**(6): p. 1630-40.

34. Arita, K., et al., *Structural basis for Ca(2+)-induced activation of human PAD4.* Nat Struct Mol Biol, 2004. **11**(8): p. 777-83.
35. Zheng, Y., et al., *Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells.* Chem Sci, 2017. **8**(10): p. 7211-7217.
36. Schmied, W.H., et al., *Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1.* J Am Chem Soc, 2014. **136**(44): p. 15577-83.
37. Luo, Y., et al., *Activity-based protein profiling reagents for protein arginine deiminase 4 (PAD4): synthesis and in vitro evaluation of a fluorescently labeled probe.* J Am Chem Soc, 2006. **128**(45): p. 14468-9.
38. Lewis, H.D., et al., *Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation.* Nat Chem Biol, 2015. **11**(3): p. 189-91.
39. Infield, D.T., et al., *Replacing voltage sensor arginines with citrulline provides mechanistic insight into charge versus shape.* Journal of General Physiology, 2018. **150**(7): p. 1017-1024.
40. Schmied, W.H., et al., *Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1.* Journal of the American Chemical Society, 2014. **136**(44): p. 15577-15583.

Chapter 3. Examine orthogonality and efficiency of *E. coli* tRNAs as stop codon suppressor in the mammalian cell to expand genetic code expansion toolbox

3.1 Introduction

The incorporation of noncanonical amino acids (ncAA) into mammalian cells has been proven as a powerful tool to unravel biological phenomenon. Genetic code expansion (GCE) was an especially useful method to modify protein *in vivo*. GCE utilizes engineered aminoacyl tRNA synthetase (aaRS) and its cognate mutated tRNA to suppress specific codons, which allows the site- or residue-specific incorporation of ncAA site-specifically into the protein of interest (POI). The advancement of GCE in the mammalian system in the past decade has opened avenues to incorporate bioconjugation handles, post translational modification (PTM) mimic ncAA, photo cross-linker, and more ncAA that covers different functions. Despite the potential of this technique, the ncAA toolbox is restricted by a limited number of aaRS/tRNA pairs. To broaden the scope of the ncAA toolbox, additional tRNA/aaRS pairs are needed. Beside the most popular archaeal pyrrolysyl-tRNA synthetase and tRNA (PylRS-$tRNA^{Pyl}$) pair, *E. coli* leucyl-tRNA synthetase and tRNA (EcLeuRS-$tRNA^{Leu}$) pair, *E. coli* tryptophanyl-tRNA synthetase and tRNA (EcTrpRS-$tRNA^{Trp}$) pair, and *E. coli* tyrosyl -tRNA synthetase and tRNA (EcTyrRS-$tRNA^{Tyr}$) pair were developed to facilitate the ncAA incorporation in mammalian cells.

Despite the advances in past decades, GCE in the mammalian system is restricted by the limited number of available aaRS/tRNA pairs. To address this challenge, much effort has been made through different approaches. Chin's group has

evolved a PylRS-$tRNA^{Pyl}$ pair from *M. alvus (Ma)* to which is mutually orthogonal with the more established *M. mazei/M. barkeri* PylRS-$tRNA^{Pyl}$ to achieve incorporation of multiple ncAAs simultaneously.[1, 2] To expand the structural diversity of ncAAs, Lin's group has created some chimeric aaRS/tRNA pairs, which takes advantage of the orthogonality of $tRNA^{Pyl}$, and the separable catalytical domain and tRNA binding domain of some aaRSs. They fuse catalytical domain of selected aaRS with tRNA binding domain from PylRS to generate chimeric aaRSs that charge chimeric tRNAs obtained by fusing the acceptor stem of the selected tRNA with remaining sections of $tRNA^{Pyl}$. The chimera pairs suppress TAG codon like PylRS-$tRNA^{Pyl}$, but incorporate different ncAAs based on the catalytical domain. This strategy was used to develop incorporation systems for tryptophan analogs such as 6-methyl-Trp with chimeric chPheRS- $tRNA^{chPhe}$ pairs.[3] These studies have demonstrated great potential to broaden the utility of existing GCE strategies.

Encouraged by three successful *E. coli* pairs in mammalian cells, we investigated the feasibility of developing other *E. coli* pairs in mammalian cells with the following reasons: First, the *E. coli* pairs are naturally orthogonal to all other existing pairs typically used for ncAA incorporation in eukaryotes, namely archaeal PylRS-$tRNA^{Pyl}$ pair and the three other *E. coli* pairs. Second, our group developed an ATM system which allowed us to evolve *E. coli* aaRSs in the *E. coli* system, to create variants that are functional in both *E. coli* and mammalian cells.[4]

To develop new aaRS/tRNA in the mammalian system, we first needed to tackle the tRNA orthogonality in the mammalian cells. Our focus is on orthogonality of tRNA for two reasons: First, cross-reactive aaRSs lead to global misincorporation of amino acids in proteins, which would have lethal effects. This built-in negative selection allows the elimination of cross-reactive aaRSs. However, for nonsense suppressor tRNAs, cross-reaction with a host aaRS will not significantly affect cell viability but generate significant false positive signal in the selections. Consequently, careful testing is required to weed out cross reactive tRNAs. Second, aaRSs only interact with its cognate tRNA and substrate, and aaRS mutants engineered in one host typically perform similarly in a different host. Whereas tRNA interacts more extensively to the translation machinery of the host. To evolve a tRNA fit for the mammalian system requires directed evolution in the mammalian system. Even though our lab has designed a virus-assisted directed evolution of tRNAs (VADER) system to evolve tRNA in mammalian cells, the task is still labor-intensive and time-consuming.[5] Thus, to create new *E. coli* pairs for mammalian system, we set out to identify which tRNAs are naturally orthogonal as nonsense suppressors.

In this work, we extensively screen all *E. coli* tRNAs for TGA or TAG suppression in mammalian cells. We identified several *E. coli* tRNAs that are highly active and orthogonal as TAG or TGA suppressors. We demonstrate with *E. coli* phenylalanyl-tRNA synthetase (EcPheRS) that, with known mutations and paired with

screened tRNA, 4-azidophenylalanine (AzF) can be incorporated into EGFP *in vivo* in mammalian cell.[3, 6, 7] This work identified novel tRNA-aaRS pairs that can potentially be used to incorporate new ncAAs.

3.2 Results

3.2.1 Select and Design tRNAs

We analyzed the sequences of *E coli* tRNAs for all 20 amino acids based on two parameters: codon usage frequency in the *E. coli* and tRNA structural stability.[8, 9] We reasoned that for the same amino acid, the tRNA responsible for decoding the most heavily used codon would likely be more robust and efficient. For example, for Arg, we picked the tRNA for CGT over CGG, as the former accounts for 80% usage in the *E. coli*.[8] When two or more tRNA are comparable for the first parameter, structure stability is considered, because it has been shown that more stable tRNAs typically exhibit higher efficiency as a suppressor.[9-11] To generate all mutated suppressor tRNA, we mutated the anticodon to CUA/UCA and incorporate the U33C mutation, a known beneficial mutation for enhanced nonsense suppression. [12] We also changed all the wobble base pairs in the stem regions of tRNAs to enhance their stability.[10, 11]

The $tRNA^{His}$ has a unique feature: the presence of an extra nucleotide at the 5' end. An extra 5' guanylate residue can be found in bacterial, archaeal, and eukaryotic

tRNA^{His}s, except that an adenylate residue instead was reported in snail mitochondria and an uridylate residue in bacteriophage T5 (Figure 1C). [13-15] Whether the -1G modification is the most important identity element is debatable, but the loss of -1G decreases the V_{max}/Km of EcHisRS to EctRNA^{His} by several orders of magnitude.[16, 17] The origin of the modification is rooted from different pathways across species. For *E. coli* and most bacteria, the -1G is genetically encoded. Whereas for eukaryotic cells, the -1G modification is added through a post-transcriptional modification step. [18] In the mammalian system, the -1G is modified by human tRNA^{His} guanylyltransferase (hTHG1) in an untemplated manner, but the interaction between *hTHG1* and HstRNA^{His} remains unclear due to a lack of structural characterization of their complex.[19, 20] Interestingly, the homolog ScTHG1 in the yeast, recognizes the anticodon of the SctRNA^{His} to be active.[21, 22] With this information, we reasoned that it is unlikely that the *hTGH1* will recognize and modified premature EctRNA$^{His}_{CUA}$. To address this situation, we designed two versions of the His tRNA: One with G-1 encoded in the sequence (tRNA^{HisG}), and one without it (tRNA^{His}).

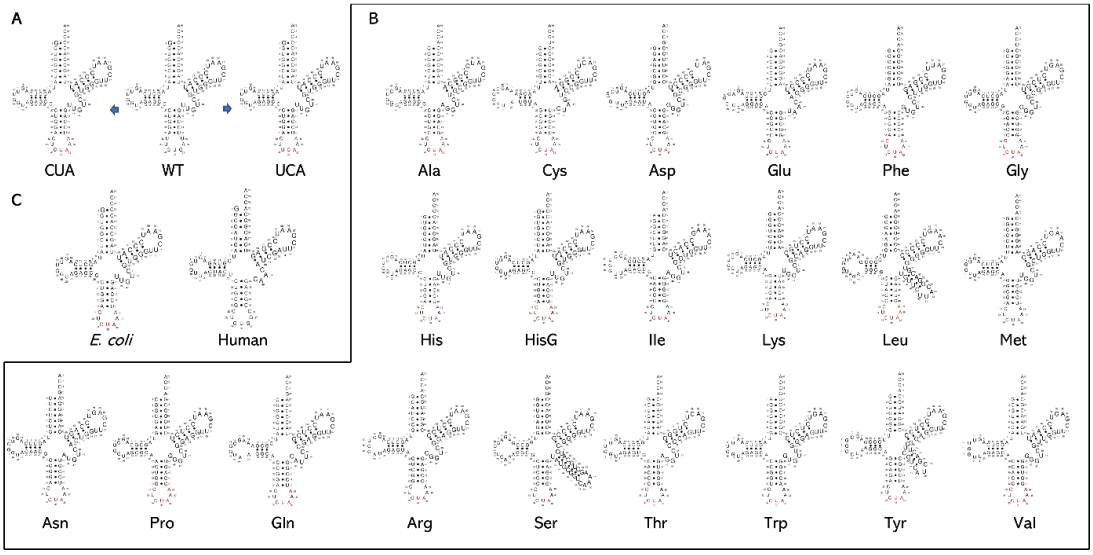


Figure 3-1. Demonstration of design and sequence of tRNAs. A) Demonstrate of mutant strategy. B) All 21 mutant tRNAs with CUA anticodon. C) Comparison of $tRNA^{His}$ from *E. coli* and human.

3.2.2 Screening active pairs.

To evaluate the activity and orthogonality of selected 21 tRNAs in the mammalian system, we designed an EGFP-based screening system. The EGFP is cloned in a PB1 construct with a TAG or TGA codon at 49 position.[23] The tRNA and aaRS were cloned into two separated pIDTsmart plasmids. (Figure 3-2A) EGFP and tRNA were co-transfected into HEK cells with and without corresponding aaRS. The cross reactive tRNA will be charged by human endogenous aaRS, which result in fluorescent signal in absence of its corresponding *E. coli* aaRS. Orthogonal and active tRNAs would yield fluorescent signal that is dependent on the presence of its corresponding *E. coli* aaRS, and the intensity of the signal would reflect its activity. EGFP with active signals are purified by Ni column and characterized by SDS-PAGE and LC/MS. (Figure 3-2C)

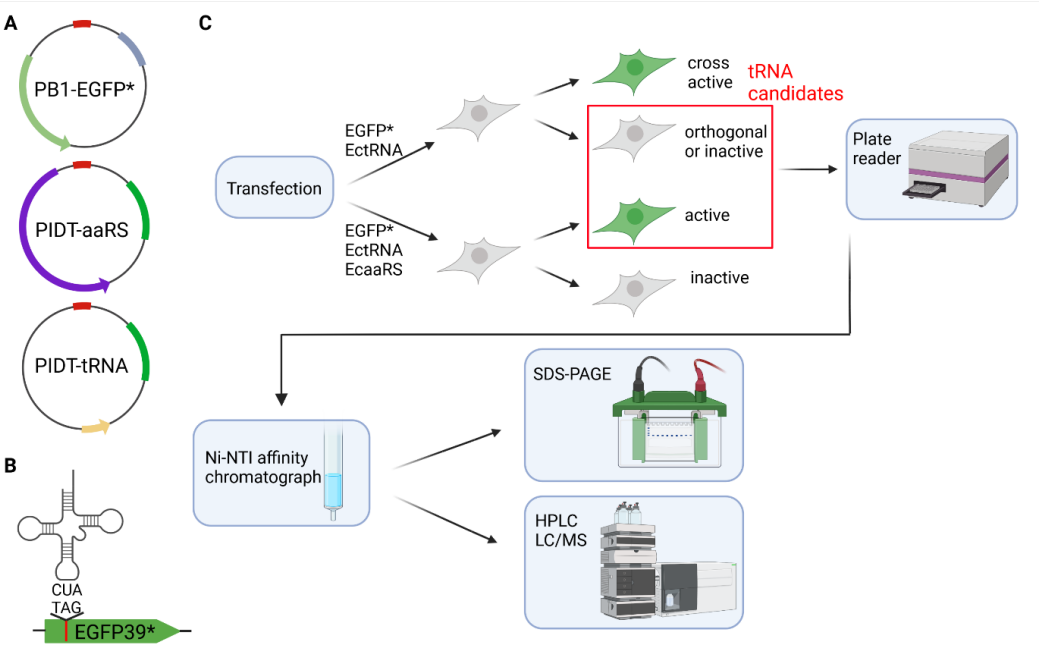


Figure 3-2. Design of screening system. A) Three plasmids system for the tRNA screening. B) EGFP Y39TAG suppression demonstration. C) tRNA orthogonality screening process.

Overall, the screening reveals that TGA suppressors are less cross reactive than TAG suppressors. From our TAG screening, several orthogonal and active *E. coli* pairs were identified. Besides $tRNA^{Leu}_{CUA}$, $tRNA^{Tyr}_{CUA}$ and $tRNA^{Trp}_{CUA}$, which were already introduced into the mammalian system, $tRNA^{HisG}_{CUA}$, $tRNA^{His}_{CUA}$, $tRNA^{Lys}_{CUA}$, $tRNA^{Ser}_{CUA}$ and $tRNA^{Phe}_{CUA}$ also demonstrated great potentials. For TGA screening, most tRNAs were found to be inactive, with $tRNA^{Arg}_{UCA}$ and $tRNA^{HisG}_{UCA}$ showing good activity but less efficient than their performance in TAG screening. (Figure 3-3,4)

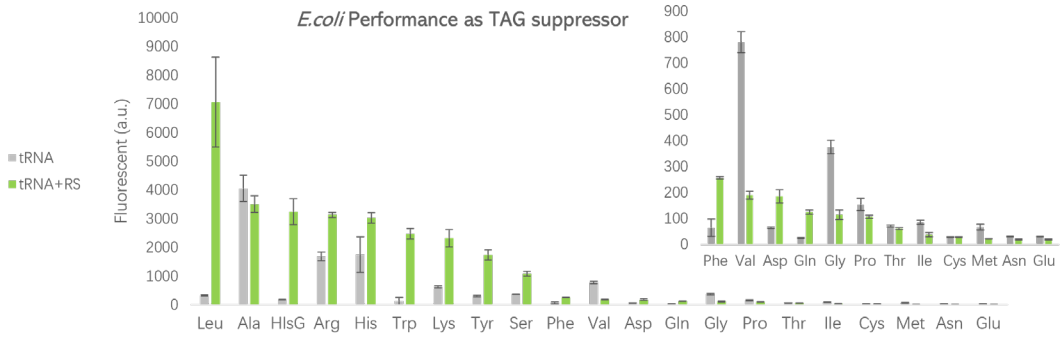


Figure 3-3. Screening result of all *E. coli* pairs (TAG) in HEK293T cells. Normalized fluorescence data for $EctRNA_{CUA}$ in HEK293T cells in absent and present of corresponding aaRS.

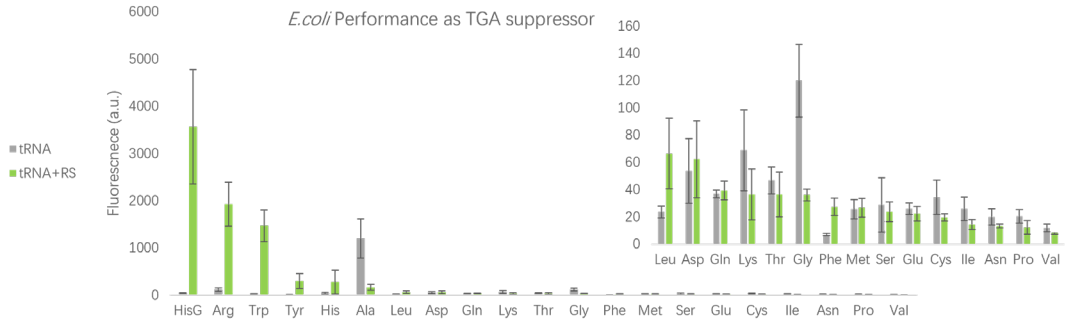


Figure 3-4. Screening result of all *E. coli* pairs (TGA) in HEK293T cells. Normalized fluorescence data for EctRNA$_{UCA}$ in HEK293T cells in absent and present of corresponding aaRS.

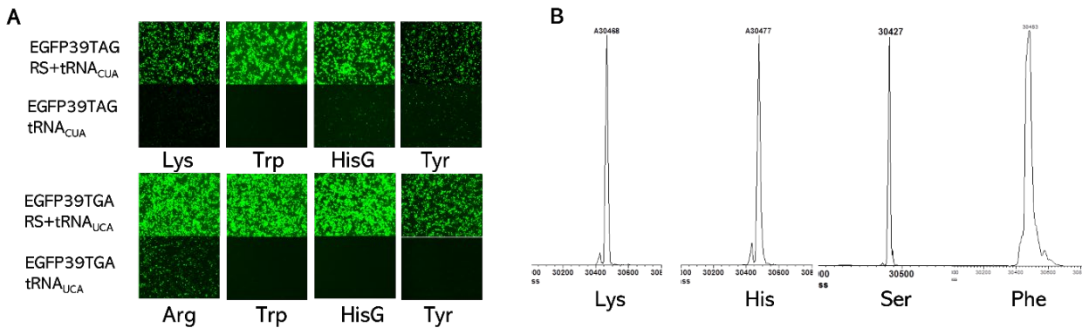


Figure 3-5. Cell picture and whole protein mass of selected active pairs. a) Cell picture under fluorescent channel in present and absent of aaRS. b) Whole protein mass of EGFP expressed from active pairs.

The contrast between tRNA$^{His}_{CUA}$ and tRNA$^{HisG}_{CUA}$ indicated that hTHG1 cannot recognize EctRNA$^{His}_{CUA}$ or unable to modify it. We reason that the outstanding orthogonality of HisGtRNA in both CUA and UCA are coming from the different identity elements used for tRNA-aaRS recognition in eukaryotic and bacterial systems. The acceptor of EctRNA^{HisG} and HstRNA^{His} are significantly different. Specifically, the

73 position is A in eukaryotes but C in *E. coli*. It is documented that A to C replacement would impair the activity of ScHisRS by four-fold. Furthermore, EcHisRS can tolerate anticodon mutations, whereas ScHisRS cannot.[24] Given that the anticodon of $tRNA^{HisG}$ was changed, along with difference at site 73, it is not surprising that the HsHisRS does not recognize $EctRNA^{HisG}$.

$EctRNA^{Arg}_{CUA}$ is cross reactive as an amber suppressor, but more orthogonal in the form of $EctRNA^{Arg}_{UCA}$. The background TAG suppressor is identified as Lys or Gln by LC/MS. Based on the literature, the mischarge may be caused by human GlnRS since 35U is a strong anticodon identity element.[25-27] This hypothesis also explained the significant reduction of cross reactive signal with $EctRNA^{Arg}_{UCA}$. When the anticodon loop switch to UCA, the middle U to C mutation reduces its affinity toward GlnRS, while enhancing the recognition by EcArgRS, as 35C is a key identity element for EcArgRS.[28]

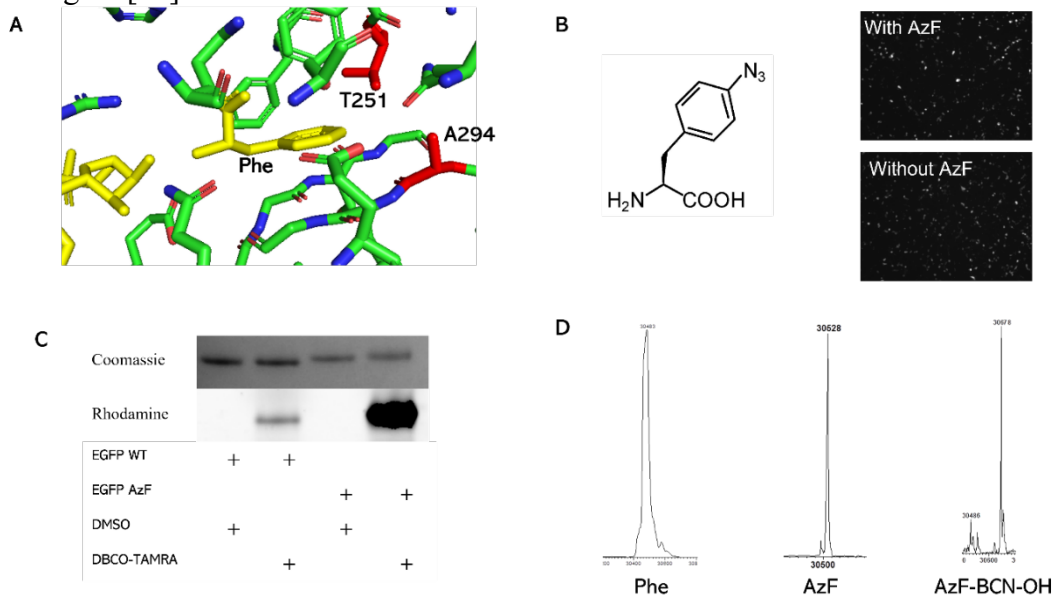


Figure 3-6. Incorporation of AzF with EcPheRS/tRNA pairs in HEK293T cells. A) Binding pocket of EcPheRS α subunits with T251 and A294 labeled in red, and Phe substrate labeled in yellow. B) AzF structure and fluorescence microscope cell image for AzF incorporation showing EGFP expression. C) Visualizing AzF incorporation in EGFP labeled by DBCO-TAMRA by SDS-PAGE in coommassie blue and Rhodamine channel. D) Characterizing AzF incorporation in EGFP with mass spectrum by comparing wild type Phe incorporation, AzF incorporation and AzF incorporation labeled by BCN-OH.

3.2.3 Incorporating AzF into mammalian cell by EcPheRS

We envisioned that the orthogonal EctRNAs can be used to incorporate ncAA if paired with a cognate engineered aaRS. To confirm this hypothesis, we searched for a known mutant of an *E. coli* aaRS that may be able to incorporate a ncAA. PheRS has a α2β2 heterotetrametric structure which is highly conserved across different species, with only mitochondrial PheRS as an exception.[29, 30] Structurally, PheRS contains two copies of each α and β subunits encoded by *pheS* and *pheT* respectively, and mutations in the binding pockets can be transplanted across species. EcPheRS was documented to incorporate *p*-azido-phenylalanine (AzF) in *E. coli* with T251G and A294G mutations previously.[31, 32] We incorporated these two mutations into the α subunit, and expressed α and β subunits from the same mRNA with a P2A sequence (Figure 3-6 A, B).

To explore the ability for *Ec*PheRS T251G A294G (EcPheRS GG) mutant to incorporate AzF in the mammalian system, we performed EGFP assay in the presence and absence of AzF. The T251G A294G mutant opens more space in the binding pocket for AzF but still accepts phenylalanine. HEK cells were transiently transfected by plasmids containing EcPheRSGG, $tRNA^{Phe}_{CUA,}$ and EGFP 39TAG. In the absence of AzF, phenylalanine is incorporated at the TAG site. In the presence of 3 mM AzF, more than 95% of protein contained AzF based on the whole protein mass spectrometry (Figure 3-6 D). To further verify the incorporation of AzF in EGFP, we performed protein-based click-chemistry with DBCO-TAMRA probe and BCN-OH. The DBCO-TAMRA labeled protein was visualized by fluorescence imaging following SDS-PAGE (Figure 3-6 C). Meanwhile, BCN-OH labeling of EGFP was quantitatively revealed by full protein mass spectrometry (Figure 3-6 D). Even though the AzF was previously incorporated in mammalian system by EcTyrRS, and chPheRS, this work proved that *E. coli* Phe pair is a suitable platform for ncAA mutagenesis in mammalian cells and that further engineering of these pairs can significantly expand the ncAA toolbox. [3, 33]

3.3 Conclusion and discussion

In conclusion, this work reports the first exhaustive evaluation of all *E. coli* tRNA/aaRS pairs nonsense suppression in mammalian cells. The screening revealed that several *E. coli* pairs can perform as efficient and orthogonal TAG and TGA

suppressors. Further engineering of these pairs can expand the ncAA toolbox and enable genetically encoding previously inaccessible important PTMs. $tRNA^{EcArg}$ was found to be cross reactive as a TAG suppressor but orthogonal as a TGA suppressor. Many important Arg PTMs have not been incorporated yet, such as mono- or di-methylated arginine. This pair is attractive for genetically encoding such PTMs. Although our lab has successfully incorporated photocaged citrulline via EcLeuRS/tRNA pair, the EcArgRS/tRNA pair could enable us to directly incorporate citrulline into proteins.[23, 34] Lysine methylations are also important PTMs that are challenging to incorporate. Orthogonal and active $tRNA^{Lys}$ in the mammalian is attractive for genetically encoding such modifications. Histidine tRNA with -1G modification outperformed most tRNA in the screening. Although some histidine PTMs are already incorporated by PylRS/tRNA pairs, a new his pair may provide higher efficiency and provide access to new structures.[35] So far, most available pairs prefer ncAAs with long bulky sides chains or aromatic side chains. SerRS/tRNA pair can enable encoding of shorter, hydrophilic ncAAs.

For tRNAs that are not active or cross reactive, it is still possible to make these useful through rational engineering or directed evolution. Using $tRNA^{Arg}$ as an example, by switching out some identity elements that match endogenous mammalian tRNA, the orthogonality and efficiency can be enhanced. Meanwhile, quadruplet codon, that extends on more nucleotide than the original codon, is also an option to improve the

active for inactive ones. [36, 37] Overall, this work lays foundation for further expansion of the ncAA toolbox in mammalian cells through developing new pairs.

3.4 Method

Plasmid generation

All tRNA synthetases were amplified directed from the T10 *E.coli* genome with listed primers. The open reading proofs are subclones between a CMV promoter and a BGH terminator by EcoRI and NheI sites. *E.coli* tRNAs are generated with long primers by a two step PCR protocol, and subcloned following a U6 promoter by AvrII and NheI.(Table) EGFP reporter plasmid was previously reported.[23] Phe α subunit GG was constructed by site specific mutagenesis based on wt Phe α subunit. Phe 4x U6 Phe tRNA TAG was made according to reported protocol.[38]

EGFP expression from HEK293T cells

HEK293T cells were cultured at 37 °C in a humidified incubator supplemented with 5% CO_2. Cells were seeded at 3.75 10^5 cells per well of 24 wells plate. 24 h before transfection. EGFP and WT PAD4 transfections were performed by incubating 0.5 μg plasmid DNA, 2.5 μL polyethylenimine (PEI) (1 mg/mL; Polysciences, Warrington, PA), and 9 μL DMEM for 10 min at room temperature, followed by adding the solution dropwise to the culture medium of the cells. For EGFP-39-TAG expression, 167 ng of

pIDTSTMART EctRNA synthetase, 167 ng pIDTSMART EctRNA plasmids and 167 ng PB1 EGFP 39TAG were incubated with a mixture of 2.5 µL PEI and 9 µL DMEM for 10 min at room temperature before adding to cells. AzF was added at the same time to a final concentration of 3 mM for AzF incorporation with PheGG mutant, and 2 mM sodium butyrate was added to enhance protein expression.

EGFP fluorescence analysis

EGFP fluorescence was analyzed 48 h after transfection. Cells were harvested and resuspended in 600 µL CelLytic M buffer (Sigma, St. Louis, MO) with Halt Protease inhibitor Cocktail (Thermo Scientific, Waltham, MA) and Pierce Universal Nuclease for Cell Lysis (Fisher Scientific, Hampton, NH). Lysates were clarified by centrifugation at 16,000 × *g* for 10 min and 100 µL of supernatant was transferred to a clear-bottom 96-well plate for fluorescence measurement. Fluorescence was measured using a Zeiss AX10 microscope (Ex. 488 nm; Em. 534 nm). All the experiments were performed at least in three duplicates.

LC-MS analysis of EGFP

LC/MS analyses of the purified EGFP were performed on Agilent Technologies 6230 TOF LC/MS with Aerns 3.6 µm WIDEEPORE XB-C8 20 LC column 100 × 4.6 mm.

Purification of EGFPs from HEK293T cells

Cells from a 10 cm plate were harvested 48 hours after transfection. For cells that overexpressed proteins containing AzF, the progress is done in a dark cold room. The cells were resuspended in 600 µL CelLytic M buffer (Sigma, St. Louis, MO) with Halt protease inhibitor cocktail (Thermo Scientific, Waltham, MA) and Pierce universal nuclease for cell lysis (Fisher Scientific, Hampton, NH). After a 20-min incubation at room temperature, 1.2 mL of equilibration buffer (20 mM Na2HPO4, 300 mM NaCl, 10 mM imidazole pH 7.4) was added. Lysate was clarified by centrifugation at 18,000 × *g* for 10 minutes at 4 °C. The clarified cell-free extract was subjected to Ni-NTA affinity chromatography using HisPur resin (Fisher Scientific, Hampton, NH) following the manufacturer's protocol.

DBCO-TAMRA labeling

1 µg EGFP 39AzF and EGFP WT were incubated with 20 µM of DBCO-TAMRA probe or DMSO control for 1 hour at room temperature in PBS buffer. The samples then prepared for SDS-PAGE electrophoresis. The SDS-PAGE was first imaged by chemdoc under Rhodamine channel, then stained by coommassie blue.

BCN-TAMRA labeling

1 µg EGFP 39AzF were incubated with 1 mM of BCN-OH probe for overnight at 4 °C temperature in PBS buffer. The samples then injected in LC/MS for mass spectrometry.

3.5 Supplementary materials

Sequence for aaRS.

All plasmid for aaRSs are based on pIDTSMART vector backbone. The aaRS are driven by CMV promoters and terminated with a Bgh terminator. The full sequence of pIDTSMART EcAlaRS will be listed below with promoter in Green color, terminator in Red color and ORF of EcAlaRS in orange. For rest of the aaRS, plasmid, only the ORF of aaRS is listed.

pIDTSMART EcAlaRS
CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATC
GTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGA
GTCTGAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGA
GACGGGCGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGGCGGC
CGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTA
CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTA
CGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACG
TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCA
AGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAAT
GGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTT
GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT
GGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA
AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA
CGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAA
CTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATA
GGGAGACCCAAGCTGGCTAGCGCCGCCACCATGAGCAAGAGCACCGCTG
AGATCCGTCAGGCGTTTCTCGACTTTTTCCATAGTAAGGGACATCAGGTAG
TTGCCAGCAGCTCCCTGGTACCCCATAACGACCCAACTTTGTTGTTTACCA
ACGCCGGGATGAACCAGTTCAAGGATGTGTTCCTTGGGCTCGACAAGCGT
AATTATTCCCGCGCTACCACTTCCCAACGCTGCGTGCGTGCGGGTGGTAA
ACACAACGACCTGGAAAACGTCGGTTACACCGCGCGTCACCATACCTTCT

TCGAAATGCTGGGCAACTTCAGCTTCGGCGACTATTTCAAACACGATGCC
ATTCAGTTTGCATGGGAACTGCTGACCAGCGAAAAATGGTTTGCCCTGCC
GAAAGAGCGTCTGTGGGTTACCGTCTATGAAAGCGACGACGAAGCCTACG
AAATCTGGGAAAAAGAAGTAGGGATCCCGCGCGAACGTATTATTCGCATC
GGCGATAACAAAGGTGCGCCATACGCATCTGACAACTTCTGGCAGATGGG
TGACACTGGTCCGTGCGGCCCGTGCACCGAAATCTTCTACGATCACGGCG
ACCACATTTGGGGGGGCCCTCCGGGAAGCCCGGAAGAAGACGGCGACCG
CTACATTGAGATCTGGAACATCGTCTTCATGCAGTTCAACCGCCAGGCCG
ATGGCACGATGGAACCGCTGCCGAAGCCGTCTGTAGATACCGGTATGGGT
CTGGAGCGTATTGCTGCGGTGCTGCAACACGTTAACTCTAACTATGACATC
GACCTGTTCCGCACGCTGATCCAGGCGGTAGCGAAAGTCACTGGCGCAAC
CGATCTGAGCAATAAATCGCTGCGCGTAATCGCTGACCACATTCGTTCTTG
TGCGTTCCTGATCGCGGATGGCGTAATGCCGTCCAATGAAAACCGTGGTT
ATGTACTGCGTCGTATCATTCGTCGCGCAGTGCGTCACGGTAATATGCTCG
GCGCGAAAGAAACCTTCTTCTACAAACTGGTTGGTCCGCTGATCGACGTT
ATGGGCTCTGCGGGTGAAGACCTGAAACGCCAGCAGGCGCAGGTTGAGC
AGGTGCTGAAGACTGAAGAAGAGCAGTTTGCTCGTACTCTGGAGCGCGGT
CTGGCGTTGCTGGATGAAGAGCTGGCAAAACTTTCTGGTGATACGCTGGA
TGGTGAAACTGCTTTCCGTCTGTACGACACCTATGGCTTCCCGGTTGACCT
GACGGCTGATGTTTGTCGTGAGCGCAACATCAAAGTTGACGAAGCTGGTT
TTGAAGCTGCAATGGAAGAGCAGCGTCGTCGCGCGCGCGAAGCCAGCGG
CTTTGGTGCCGATTACAACGCAATGATCCGTGTTGACAGTGCATCTGAATT
TAAAGGCTATGACCATCTGGAACTGAACGGCAAAGTGACTGCGCTGTTTG
TTGATGGTAAAGCGGTTGATGCCATCAATGCAGGCCAGGAAGCTGTGGTC
GTGCTGGATCAAACGCCATTCTATGCGGAATCCGGCGGTCAGGTTGGCGA
TAAAGGCGAACTGAAAGGCGCTAACTTCTCCTTTGCGGTGGAAGATACGC
AGAAATACGGCCAGGCGATTGGTCACATCGGTAAACTTGCTGCGGGTTCT
CTGAAAGTGGGCGACGCGGTGCAGGCTGATGTTGATGAGGCTCGTCGCGC
CCGTATTCGTCTGAATCACTCCGCAACGCACCTGATGCACGCTGCGCTGCG
CCAGGTTCTGGGTACTCATGTATCGCAGAAAGGTTCACTGGTTAACGACA
AGGTGCTGCGCTTCGACTTCTCACACAACGAAGCGATGAAACCAGAAGAG
ATTCGTGCGGTCGAAGACCTGGTGAACACACAGATTCGTCGCAATTTGCC
GATCGAAACCAACATCATGGATCTCGAAGCGGCGAAAGCGAAAGGTGCG
ATGGCGCTGTTCGGCGAGAAGTATGATGAGCGCGTACGCGTGCTGAGCAT
GGGCGATTTCTCTACCGAGTTGTGTGGCGGTACTCACGCCAGCCGCACTG
GTGATATTGGTCTGTTCCGCATCATCTCTGAATCGGGTACTGCTGCAGGCG
TTCGTCGTATCGAAGCGGTAACCGGAGAAGGTGCTATCGCCACCGTTCAT
GCAGACAGCGATCGCTTAAGCGAAGTCGCGCATCTGCTGAAAGGCGATAG
CAATAATCTGGCTGATAAAGTGCGCTCAGTACTGGAACGTACGCGTCAGC
TGGAAAAAGAGTTACAACAGCTTAAAGAACAAGCTGCCGCACAGGAGAG
CGCAAATCTTTCCAGTAAGGCAATTGATGTTAATGGTGTTAAGCTGTTGGT

TAGCGAGCTTAGCGGTGTTGAGCCGAAAATGTTGCGTACCATGGTTGACG
ATTTAAAAAATCAGCTGGGGTCGACAATTATCGTGCTGGCAACGGTAGTC
GAAGGTAAGGTTTCTCTGATTGCAGGCGTATCTAAGGACGTCACAGATCG
TGTGAAAGCAGGGGAACTGATTGGTATGGTCGCTCAGCAGGTGGGCGGCA
AGGGTGGTGGACGTCCTGACATGGCGCAAGCCGGTGGTACGGATGCTGCG
GCCTTACCTGCAGCGTTAGCCAGTGTGAAAGGCTGGGTCAGCGCGAAATT
GCAATAAGAATTCAACGCGTTAAGTCGACTTTAACTCGAGTCTAGAGGGC
CCGTTTAAACCCGCTGATCAGGTTGCCAGCCATCTGTTGTTTGCCCCTCCC
CCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAAT
AAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG
GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA
GCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGA
ACCCTAGGGGTGCGAGCGGATCGAGCAGTGTCGATCACTACTGGACCGCG
AGCTGTGCTGCGACcCGTGATCTTACGGCATTATACGTATGATCGGTCCAC
GATCAGCTAGATTATCTAGTCAGCTTGATGTCATAGCTGTTTCCTGAGGCT
CAATACTGACCATTTAAATCATACCTGACCTCCATAGCAGAAAGTCAAAA
GCCTCCGACCGGAGGCTTTTGACTTGATCGGCACGTAAGAGGTTCCAACT
TTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGA
GATTTTCAGGAGCTAAGGAAGCTAAAATGAGCCATATTCAACGGGAAACG
TCTTGCTTGAAGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGG
TATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCG
ATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAG
GTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGGCTAAACTGGCTG
ACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGAT
GATGCATGGTTACTCACCACTGCGATCCCAGGGAAAACAGCATTCCAGGT
ATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAG
TGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACGG
CGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTT
GGTTGGTGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAAC
AAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCTCACCGGATTCAGTC
GTCACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAA
TTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCA
GGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACA
GAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATT
GCAGTTTCACTTGATGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAA
TCACCTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCAT
AGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGATGCTCAAGTCAGAG
GTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA
GCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT
CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTA
GGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACG

AACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTG
AGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT
AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAA
GTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCG
CTCTGCTGAAGCCAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAG
ATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATTTTCTACCG
AAGAAAGGCCCA

EcArgRS
ATGAATATTCAGGCTCTTCTCTCAGAAAAAGTCCGTCAGGCCATGATTGC
GGCAGGCGCGCCTGCGGATTGCGAACCGCAGGTTCGTCAGTCAGCAAAAG
TTCAGTTCGGCGACTATCAGGCTAACGGCATGATGGCAGTTGCTAAAAAA
CTGGGTATGGCACCGCGACAATTAGCAGAGCAGGTGCTGACTCATCTGGA
TCTTAACGGTATCGCCAGCAAAGTTGAGATCGCCGGTCCAGGCTTTATCA
ACATTTTCCTTGATCCGGCATTCCTGGCTGAACATGTTCAGCAGGCGCTGG
CGTCCGATCGTCTCGGTGTTGCTACGCCAGAAAAACAGACCATTGTGGTT
GACTACTCTGCGCCAAACGTGGCGAAAGAGATGCATGTCGGTCACCTGCG
CTCTACCATTATTGGTGACGCAGCAGTGCGTACTCTGGAGTTCCTCGGTCA
CAAAGTGATTCGCGCAAACCACGTCGGCGACTGGGGCACTCAGTTCGGTA
TGCTGATTGCATGGCTGGAAAAGCAGCAGCAGGAAAACGCCGGTGAAAT
GGAGCTGGCTGACCTTGAAGGTTTCTACCGCGATGCGAAAAAGCATTACG
ATGAAGATGAAGAGTTCGCCGAGCGCGCACGTAACTACGTGGTAAAACTG
CAAAGCGGTGACGAATATTTCCGCGAGATGTGGCGCAAACTGGTCGACAT
CACCATGACGCAGAACCAGATCACCTACGATCGTCTCAACGTGACGCTGA
CCCGTGATGACGTGATGGGCGAAAGCCTCTACAACCCGATGCTGCCAGGA
ATTGTGGCGGATCTCAAAGCCAAAGGTCTGGCAGTAGAAAGCGAAGGGG
CGACCGTCGTATTCCTTGATGAGTTTAAAAACAAGGAAGGCGAACCGATG
GGCGTGATCATTCAGAAGAAAGATGGCGGCTATCTCTACACCACCACTGA
TATCGCCTGTGCGAAATATCGTTATGAAACACTGCATGCCGATCGCGTGCT
GTATTACATCGACTCCCGTCAGCATCAACACCTGATGCAGGCATGGGCGA
TCGTCCGTAAAGCAGGCTATGTACCGGAATCCGTACCGCTGGAACACCAC
ATGTTCGGCATGATGCTGGGTAAAGACGGCAAACCGTTCAAAACCCGCGC
GGGTGGTACAGTGAAACTGGCCGATCTGCTGGATGAAGCCCTGGAACGTG
CACGCCGTCTGGTGGCAGAAAAGAACCCGGATATGCCAGCCGACGAGCT
GGAAAAACTGGCTAACGCGGTTGGTATTGGTGCGGTGAAATATGCGGATC
TCTCCAAAAACCGCACCACGGACTACATCTTCGACTGGGACAACATGCTG
GCGTTTGAGGGTAATACCGCGCCATACATGCAGTATGCATACACGCGTGT
ATTGTCCGTGTTCCGTAAAGCAGAAATTGACGAAGAGCAACTGGCTGCAG
CTCCGGTTATCATCCGTGAAGATCGTGAAGCGCAACTGGCAGCTCGCCTG
CTGCAGTTTGAAGAAACCCTCACCGTGGTTGCCCGTGAAGGCACGCCGCA

TGTAATGTGTGCTTACCTGTACGATCTGGCCGGTCTGTTCTCTGGCTTCTA
CGAGCACTGCCCGATCCTCAGCGCAGAAAACGAAGAAGTGCGTAACAGC
CGTCTAAAACTGGCACAACTGACGGCGAAGACGCTGAAGCTGGGTCTGGA
TACGCTGGGTATTGAGACTGTAGAGCGTATGTAA

EcAsnRS
ATGTCACATCTCGCAGAACTGGTTGCCAGTGCGAAGGCGGCCATTAGCCA
GGCGTCAGATGTTGCCGCGTTAGATAATGTGCGCGTCGAATATTTGGGTA
AAAAAGGGCACTTAACCCTTCAGATGACGACCCTGCGTGAGCTGCCGCCA
GAAGAGCGTCCGGCAGCTGGTGCGGTTATCAACGAAGCGAAAGAGCAGG
TTCAGCAGGCGCTGAATGCGCGTAAAGCGGAACTGGAAAGCGCTGCACTG
AATGCGCGTCTGGCGGCGGAAACGATTGATGTCTCTCTGCCAGGTCGTCG
CATTGAAAACGGCGGTCTGCATCCGGTTACCCGTACCATCGACCGTATCG
AAAGTTTCTTCGGTGAGCTTGGCTTTACCGTGGCAACCGGGCCGGAAATC
GAAGACGATTATCATAACTTCGATGCTCTGAACATTCCTGGTCACCACCCG
GCGCGCGCTGACCACGACACTTTCTGGTTTGACACTACCCGCCTGCTGCGT
ACCCAGACCTCTGGCGTACAGATCCGCACCATGAAAGCCCAGCAGCCACC
GATTCGTATCATCGCGCCTGGCCGTGTTTATCGTAACGACTACGACCAGAC
TCACACGCCGATGTTCCATCAGATGGAAGGTCTGATTGTTGATACCAACA
TCAGCTTTACCAACCTGAAAGGCACGCTGCACGACTTCCTGCGTAACTTCT
TTGAGGAAGATTTGCAGATTCGCTTCCGTCCTTCCTACTTCCCGTTTACCG
AACCTTCTGCAGAAGTGGACGTCATGGGTAAAAACGGTAAATGGCTGGAA
GTGCTGGGCTGCGGGATGGTGCATCCGAACGTGTTGCGTAACGTTGGCAT
CGACCCGGAAGTTTACTCTGGTTTCGCCTTCGGGATGGGGATGGAGCGTC
TGACTATGTTGCGTTACGGCGTCACCGACCTGCGTTCATTCTTCGAAAACG
ATCTGCGTTTCCTCAAACAGTTTAAATAA

EcAspRS
ATGCGTACAGAATATTGTGGACAGCTCCGTTTGTCCCACGTGGGGCAGCA
GGTGACTCTGTGTGGTTGGGTCAACCGTCGTCGTGATCTTGGTAGCCTGAT
CTTCATCGATATGCGCGACCGCGAAGGTATCGTGCAGGTATTTTTCGATCC
GGATCGTGCGGACGCGTTAAAGCTGGCCTCTGAACTGCGTAATGAGTTCT
GCATTCAGGTCACGGGCACCGTACGTGCGCGTGACGAAAAAAATATTAAC
CGCGATATGGCGACCGGCGAAATCGAAGTGCTGGCGTCCTCGCTGACTAT
CATCAACCGCGCAGATGTTCTGCCGCTTGACTCTAACCACGTCAACACCG
AAGAAGCGCGTCTGAAATACCGCTACCTCGACCTGCGTCGTCCGGAAATG
GCTCAGCGCCTGAAAACCCGCGCTAAAATCACCAGCCTGGTGCGCCGTTT
TATGGATGACCACGGCTTCCTCGACATCGAAACTCCGATGCTGACCAAAG
CCACGCCGGAAGGCGCGCGTGACTACCTGGTGCCTTCTCGTGTGCACAAA
GGTAAATTCTACGCACTGCCGCAATCCCCGCAGTTGTTCAAACAGCTGCT
GATGATGTCCGGTTTTGACCGTTACTATCAGATCGTTAAATGCTTCCGTGA

CGAAGACCTGCGTGCTGACCGTCAGCCTGAATTTACTCAGATCGATGTGG
AAACTTCTTTCATGACCGCGCCGCAAGTGCGTGAAGTGATGGAAGCGCTG
GTGCGTCATCTGTGGCTGGAAGTGAAGGGTGTGGATCTGGGCGATTTCCC
GGTAATGACCTTTGCGGAAGCAGAACGCCGTTATGGTTCTGATAAACCGG
ATCTGCGTAACCCGATGGAACTGACTGACGTTGCTGATCTGCTGAAATCT
GTTGAGTTTGCTGTATTTGCAGGTCCGGCGAACGATCCGAAAGGTCGCGT
AGCGGCTCTGCGCGTTCCGGGCGGCGCATCGCTGACCCGTAAGCAGATCG
ACGAATACGGTAACTTCGTTAAAATCTACGGCGCGAAAGGTCTGGCTTAC
ATCAAAGTTAACGAACGCGCGAAAGGTCTGGAAGGTATCAACAGCCCGG
TAGCGAAGTTCCTTAATGCAGAAATCATCGAAGACATCCTGGATCGTACT
GCCGCGCAAGATGGCGATATGATTTTCTTCGGTGCCGACAACAAGAAAAT
TGTTGCCGACGCGATGGGTGCACTGCGCCTGAAAGTGGGTAAAGACCTTG
GTCTGACCGACGAAAGCAAATGGGCACCGCTGTGGGTTATCGACTTCCCG
ATGTTTGAAGACGACGGTGAAGGCGGCCTGACGGCAATGCACCATCCGTT
CACCTCACCGAAAGATATGACGGCTGCAGAACTGAAAGCTGCACCGGAA
AATGCGGTGGCGAACGCTTACGATATGGTCATCAATGGTTACGAAGTGGG
CGGTGGTTCAGTACGTATCCATAATGGTGATATGCAGCAGACGGTGTTTG
GTATTCTGGGTATCAACGAAGAGGAACAGCGCGAGAAATTCGGCTTCCTG
CTCGACGCTCTGAAATACGGTACTCCGCCGCACGCAGGTCTGGCATTCGG
TCTTGACCGTCTGACCATGCTGCTGACCGGCACCGACAATATCCGTGACGT
TATCGCCTTCCCGAAAACCACGGCGGCAGCGTGTCTGATGACTGAAGCAC
CGAGCTTTGCTAACCCGACTGCACTGGCTGAGCTGAGCATTCAGGTTGTG
AAGAAGGCTGAGAATAACTAA

EcCysRS
ATGCTAAAAATCTTCAATACTCTGACACGCCAAAAAGAGGAATTTAAGCC
TATTCACGCCGGGGAAGTCGGCATGTACGTGTGTGGAATCACCGTTTACG
ATCTCTGTCATATCGGTCACGGGCGTACCTTTGTTGCTTTTGACGTGGTTG
CGCGCTATCTGCGTTTCCTCGGCTATAAACTGAAGTATGTGCGCAACATTA
CCGATATCGACGACAAAATCATCAAACGCGCCAATGAAAATGGCGAAAG
CTTTGTGGCGATGGTGGATCGCATGATCGCCGAAATGCACAAAGATTTTG
ATGCTTTGAACATTCTGCGCCCGGATATGGAGCCGCGCGCGACGCACCAT
ATCGCAGAAATTATTGAACTCACTGAACAACTGATCGCCAAAGGTCACGC
TTATGTGGCGGACAACGGCGACGTGATGTTCGACGTCCCGACCGATCCAA
CTTATGGCGTGCTGTCGCGTCAGGATCTCGACCAGCTGCAGGCAGGCGCG
CGCGTTGACGTGGTCGACGACAAACGCAACCCAATGGACTTCGTTCTGTG
GAAGATGTCGAAAGAGGGCGAACCGAGCTGGCCGTCTCCGTGGGGCGCG
GGTCGTCCTGGCTGGCACATTGAATGTTCGGCAATGAACTGCAAGCAGCT
GGGTAACCACTTTGATATCCACGGCGGCGGTTCAGACCTGATGTTCCCGC
ACCACGAAAACGAAATCGCGCAGTCCACCTGTGCCCATGATGGTCAGTAT
GTGAACTACTGGATGCACTCGGGGATGGTGATGGTTGACCGCGAGAAGAT

GTCCAAATCGCTGGGTAACTTCTTTACCGTGCGCGATGTGCTGAAATACTA
CGACGCGGAAACCGTGCGTTACTTCCTGATGTCGGGCCACTATCGCAGCC
AGTTGAACTACAGCGAAGAGAACCTGAAGCAGGCGCGTGCGGCGCTGGA
GCGTCTCTACACTGCGCTGCGCGGCACAGATAAAACCGTTGCGCCTGCCG
GTGGCGAAGCGTTTGAAGCGCGCTTTATTGAAGCGATGGACGACGATTTC
AACACCCCGGAAGCCTATTCCGTACTGTTTGATATGGCGCGTGAAGTAAA
CCGTCTGAAAGCAGAAGATATGGCAGCGGCGAATGCAATGGCATCTCACC
TGCGTAAACTTTCCGCTGTATTGGGCCTGCTGGAGCAAGAACCGGAAGCG
TTCCTGCAAAGCGGCGCGCAGGCAGACGACAGCGAAGTGGCTGAGATTG
AAGCGTTAATTCAACAGCGTCTGGATGCCCGTAAAGCGAAAGACTGGGCG
GCGGCGGATGCGGCGCGTGATCGTCTTAACGAGATGGGGATCGTGCTGGA
AGATGGCCCGCAAGGGACCACCTGGCGTCGTAAGTAA

EcGlnRS
ATGAGTGAGGCAGAAGCCCGCCCGACTAACTTTATCCGTCAGATCATCGA
TGAAGATCTGGCCAGTGGTAAGCACACCACAGTACACACCCGTTTCCCGC
CGGAGCCGAATGGCTATCTGCATATTGGCCATGCGAAATCTATCTGCCTG
AACTTCGGGATCGCCCAGGACTATAAAGGCCAGTGCAACCTGCGTTTCGA
CGACACTAACCCGGTAAAAGAAGATATCGAGTATGTTGAGTCGATCAAAA
ACGACGTAGAGTGGTTAGGTTTTCACTGGTCTGGTAACGTCCGTTACTCCT
CCGATTATTTTGATCAGCTCCACGCCTATGCGATCGAACTGATCAATAAAG
GCCTGGCGTACGTTGATGAACTGACGCCGGAACAGATCCGCGAATACCGC
GGCACCCTGACGCAACCGGGTAAAAACAGCCCGTACCGCGACCGCAGCG
TTGAAGAGAACCTGGCGCTGTTCGAAAAAATGCGTGCCGGTGGTTTTGAA
GAAGGTAAAGCCTGCCTGCGTGCGAAAATCGACATGGCTTCACCGTTTAT
CGTGATGCGCGATCCGGTGCTGTACCGTATTAAATTTGCTGAACACCACC
AGACTGGCAACAAGTGGTGCATCTACCCGATGTACGACTTCACCCACTGC
ATCAGCGATGCCCTGGAAGGTATTACGCACTCTCTGTGTACGCTTGAGTTC
CAGGACAACCGTCGTCTGTACGACTGGGTACTGGACAACATCACGATTCC
TGTTCACCCGCGCCAGTATGAGTTCTCGCGCCTGAATCTGGAATACACCGT
GATGTCCAAGCGTAAGTTGAACCTGCTGGTGACCGACAAGCACGTTGAAG
GCTGGGATGACCCGCGTATGCCGACCATTTCCGGTCTGCGTCGTCGTGGTT
ACACTGCGGCTTCTATTCGTGAGTTCTGCAAACGCATCGGCGTGACCAAG
CAGGACAACACCATTGAGATGGCGTCGCTGGAATCCTGCATCCGTGAAGA
TCTCAACGAAAATGCGCCGCGCGCAATGGCGGTTATCGATCCGGTGAAAC
TGGTTATCGAAAACTATCAGGGCGAAGGCGAAATGGTTACCATGCCGAAC
CATCCGAACAAACCGGAAATGGGCAGCCGTCAGGTGCCGTTTAGCGGTGA
GATTTGGATTGATCGCGCCGATTTCCGCGAAGAAGCTAACAAGCAGTACA
AACGTCTGGTGCTGGGTAAAGAAGTGCGTCTGCGTAATGCTTATGTGATT
AAGGCAGAACGCGTCGAGAAAGATGCCGAAGGTAATATCACCACCATCTT
CTGTACTTATGACGCCGATACCTTAAGCAAAGATCCGGCAGATGGTCGTA

AAGTCAAAGGTGTTATTCACTGGGTGAGCGCGGCACATGCGCTGCCGGTT
GAAATCCGTTTGTATGACCGTCTGTTCAGCGTGCCTAACCCAGGTGCTGCG
GATGATTTCCTGTCGGTGATTAACCCGGAATCGCTGGTGATCAAACAGGG
CTTTGCTGAACCGTCGCTGAAAGATGCGGTTGCGGGTAAAGCATTCCAGT
TTGAGCGTGAAGGTTACTTCTGCCTCGATAGCCGCCATTCTACGGCGGAA
AAACCGGTATTTAACCGCACCGTTGGGCTGCGTGATACCTGGGCGAAAGT
AGGCGAGTAA

EcGluRS

ATGAAAATCAAAACTCGCTTCGCGCCAAGCCCAACAGGCTATCTGCACGT
TGGCGGCGCGCGTACTGCTCTTTACTCCTGGCTTTTTGCACGTAACCACGG
CGGTGAGTTCGTGCTGCGTATTGAAGACACCGATCTTGAGCGTTCCACGC
CGGAAGCTATCGAAGCCATTATGGATGGCATGAACTGGCTGAGCCTGGAG
TGGGATGAAGGTCCGTACTACCAGACCAAACGTTTTGATCGCTACAACGC
GGTGATCGATCAGATGCTGGAAGAGGGCACTGCTTATAAATGCTATTGCT
CTAAAGAGCGCCTGGAAGCGCTGCGCGAAGAGCAAATGGCGAAAGGTGA
GAAGCCGCGTTATGACGGTCGCTGCCGCCACAGCCATGAGCATCATGCTG
ATGATGAACCGTGTGTTGTACGTTTTGCTAACCCGCAGGAAGGTTCTGTTG
TTTTTGACGATCAGATCCGTGGTCCGATCGAGTTCAGCAACCAGGAACTG
GACGATCTTATTATCCGCCGTACCGATGGTTCCCCAACCTATAACTTCTGT
GTGGTTGTCGATGACTGGGATATGGAAATCACCCACGTTATCCGTGGCGA
AGACCATATCAACAACACGCCACGCCAGATCAACATTCTTAAGGCCCTGA
AAGCGCCGGTGCCGGTTTACGCGCACGTTTCTATGATCAATGGCGATGAC
GGTAAAAAACTGTCCAAACGTCACGGGGCAGTCAGCGTAATGCAGTATCG
TGATGACGGTTATTTGCCAGAAGCACTGCTGAACTATCTGGTGCGTCTGG
GCTGGTCCCACGGCGATCAGGAAATCTTCACTCGTGAAGAGATGATCAAA
TACTTCACTTTGAATGCCGTCAGCAAATCTGCCAGTGCGTTCAACACCGAC
AAGCTGCTGTGGCTGAACCATCACTACATTAACGCGCTGCCGCCGGAGTA
TGTTGCTACTCACTTACAGTGGCACATTGAGCAGGAAAATATCGATACCC
GTAACGGCCCGCAGCTGGCTGATCTGGTGAAACTGCTGGGCGAACGCTGC
AAGACGCTGAAAGAGATGGCACAGAGCTGCCGTTATTTCTACGAAGATTT
TGCTGAGTTCGATGCCGACGCCGCGAAAAAACATCTGCGTCCGGTAGCGC
GTCAGCCGCTGGAAGTGGTTCGTGACAAACTGGCCGCGATTACTGACTGG
ACCGCTGAAAACGTTCATCACGCTATTCAGGCGACGGCGGATGAGCTGGA
AGTGGGTATGGGTAAAGTTGGTATGCCGCTGCGTGTCGCCGTAACCGGTG
CGGGGCAGTCTCCAGCACTGGATGTTACCGTTCACGCAATTGGTAAGACC
CGCAGTATCGAGCGTATCAACAAAGCGCTGGATTTTATTGCTGAACGCGA
AAATCAGCAGTAA

EcGlyRS α

ATGCAAAAGTTTGATACCAGGACCTTCCAGGGCTTGATCCTGACCTTACA
GGATTACTGGGCTCGCCAGGGCTGCACCATTGTTCAACCATTGGACATGG
AAGTCGGCGCGGGAACCTCTCACCCAATGACCTGTCTGCGCGAGCTGGGG
CCAGAACCGATGGCGGCTGCTTATGTTCAGCCTTCTCGTCGCCCGACCGAT
GGTCGCTACGGCGAAAACCCCAACCGTTTACAGCACTACTATCAGTTCCA
GGTGGTCATTAAGCCATCGCCGGACAATATTCAGGAGCTGTACCTCGGTT
CTCTGAAAGAGCTGGGCATGGACCCGACTATTCACGACATCCGTTTCGTG
GAAGATAACTGGGAAAACCCGACGCTGGGTGCCTGGGGACTGGGCTGGG
AAGTGTGGCTGAACGGCATGGAAGTGACGCAGTTCACTTACTTCCAGCAG
GTTGGTGGTCTGGAGTGTAAACCGGTTACCGGCGAGATCACCTACGGTCT
GGAACGTCTGGCCATGTACATTCAGGGCGTAGACAGCGTTTACGACCTGG
TCTGGAGCGACGGCCCGCTGGGTAAAACCACCTACGGCGACGTGTTCCAT
CAGAACGAAGTGGAGCAGTCCACTTACAACTTCGAATACGCGGATGTGGA
CTTCCTGTTCACCTGCTTCGAGCAGTACGAGAAAGAAGCGCAGCAGCTGC
TGGCGCTGGAAAATCCGCTGCCGCTGCCAGCCTACGAGCGTATTCTGAAA
GCCGCCCACAGCTTCAACCTGCTGGATGCGCGTAAAGCCATCTCCGTCAC
CGAGCGTCAGCGCTACATTCTGCGCATTCGCACCCTGACCAAAGCAGTGG
CAGAAGCATACTACGCTTCCCGTGAAGCCCTCGGCTTCCCGATGTGCAAC
AAAGATAAGTAA

EcGlyRS β
ATGTCTGAGAAAACTTTTCTGGTGGAAATCGGCACTGAAGAGCTGCCACC
AAAAGCACTGCGCAGCCTGGCTGAGTCCTTTGCTGCGAACTTTACTGCGG
AGCTGGATAACGCTGGCCTCGCACACGGCACCGTTCAATGGTTTGCTGCT
CCGCGTCGTCTGGCGCTGAAAGTAGCTAACCTGGCGGAAGCGCAACCGGA
TCGTGAAATCGAAAAACGCGGCCCGGCGATTGCCCAGGCGTTCGACGCTG
AAGGCAAACCGAGCAAAGCGGCAGAAGGTTGGGCGCGTGGTTGCGGTAT
TACCGTTGACCAGGCTGAGCGTCTGACTACCGATAAAGGCGAATGGCTGC
TGTATCGCGCCCATGTGAAGGGCGAAAGCACCGAAGCACTGCTGCCGAAT
ATGGTTGCGACTTCTCTGGCGAAACTGCCGATCCCGAAACTGATGCGTTG
GGGCGCAAGCGACGTGCACTTCGTGCGTCCGGTGCACACCGTGACCCTGC
TGCTGGGCGACAAAGTCATTCCGGCAACCATTCTGGGCATTCAGTCCGAT
CGCGTGATTCGCGGCCACCGCTTTATGGGCGAGCCGGAATTCACCATCGA
TAACGCCGATCAGTATCCGGAAATTCTGCGTGAGCGTGGGAAAGTCATCG
CCGATTACGAAGAACGTAAGGCGAAGATTAAAGCCGATGCCGAAGAAGC
AGCGCGTAAGATTGGCGGTAACGCTGACTTAAGCGAAAGCCTGCTGGAAG
AAGTGGCTTCGCTGGTGGAGTGGCCGGTCGTTCTGACCGCAAAATTCGAA
GAGAAATTCCTCGCGGTGCCGGCTGAAGCGCTGGTTTACACCATGAAAGG
TGACCAGAAATACTTCCCGGTGTATGCGAACGACGGCAAACTGCTGCCGA
ACTTTATCTTCGTTGCCAACATCGAATCGAAAGATCCGCAGCAGATTATCT
CCGGTAACGAGAAAGTCGTTCGTCCGCGTCTGGCGGATGCCGAGTTCTTC

TTCAACACCGACCGTAAAAAACGTCTTGAAGATAACCTGCCGCGCCTGCA
AACCGTGTTGTTCCAGCAACAGTTGGGGACGCTGCGCGACAAAACTGACC
GCATCCAGGCGCTGGCTGGCTGGATTGCTGAACAGATTGGCGCTGACGTT
AACCACGCTACCCGTGCGGGTCTGCTGTCTAAGTGCGACCTGATGACCAA
CATGGTCTTCGAGTTCACCGACACCCAGGGCGTTATGGGGATGCACTATG
CGCGTCACGATGGCGAAGCGGAAGATGTCGCGGTGGCGCTGAATGAGCA
GTATCAGCCGCGTTTTGCTGGTGATGACCTGCCGTCCAACCCAGTAGCTTG
TGCGCTGGCGATTGCTGACAAGATGGATACCCTGGCGGGTATCTTCGGTA
TCGGTCAGCATCCGAAAGGCGACAAAGACCCGTTTGCGCTGCGTCGTGCC
GCGCTTGGCGTGCTGCGAATTATCGTTGAGAAGAACCTCAACCTTGATCT
GCAAACGCTGACCGAAGAAGCGGTGCGTCTGTATGGCGATAAGCTGACTA
ATGCCAACGTAGTTGATGATGTTATCGACTTTATGCTCGGTCGCTTCCGCG
CCTGGTATCAGGACGAAGGTTATACCGTTGACACCATCCAGGCGGTACTG
GCGCGTCGTCCGACTCGTCCGGCTGATTTCGATGCCCGTATGAAAGCGGT
ATCGCATTTCCGTACCCTGGATGCAGCTGCTGCACTGGCGGCGGCGAACA
AACGTGTATCTAACATTCTGGCGAAATCTGACGAAGTGCTGAGCGACCGC
GTGAATGCCTCTACCCTGAAAGAGCCGGAAGAAATTAAACTGGCGATGCA
GGTTGTGGTGCTACGTGACAAGCTGGAGCCGTACTTTACGGAAGGTCGTT
ACCAGGATGCGCTGGTCGAACTGGCTGAGCTGCGTGAACCGGTTGATGCT
TTCTTCGATAAAGTGATGGTCATGGTTGATGACAAAGAATTGCGTATCAA
CCGTCTGACCATGCTGGAGAAACTGCGCGAACTGTTCCTGCGCGTTGCGG
ATATTTCGCTGTTGCAATAA

EcHisRS

ATGGCAAAAAACATTCAAGCCATTCGCGGCATGAACGATTACCTGCCTGG
CGAAACGGCCATCTGGCAGCGCATTGAAGGCACACTGAAAAACGTGCTCG
GCAGCTACGGTTACAGTGAAATCCGCTTGCCGATTGTAGAGCAGACCCCG
CTATTCAAACGTGCGATTGGTGAAGTCACCGACGTGGTTGAAAAAGAGAT
GTACACCTTTGAGGATCGCAATGGCGACAGCCTGACTCTGCGCCCTGAAG
GGACGGCGGGCTGTGTACGCGCCGGCATCGAGCATGGTCTTCTGTACAAT
CAGGAACAGCGTCTGTGGTATATCGGGCCGATGTTCCGTCACGAGCGTCC
GCAGAAAGGGCGTTATCGTCAGTTCCATCAGTTGGGCTGCGAAGTTTTCG
GTCTGCAAGGTCCGGATATCGACGCTGAACTGATTATGCTCACTGCCCGCT
GGTGGCGCGCGCTGGGTATTTCCGAGCACGTAACTCTTGAGCTGAACTCT
ATCGGTTCGCTGGAAGCACGCGCCAATTACCGCGATGCGCTGGTGGCATT
CCTTGAGCAGCATAAAGAAAAGCTGGACGAAGACTGCAAACGCCGCATG
TACACTAACCCGCTGCGCGTGCTGGATTCAAAAAATCCGGAAGTGCAGGC
GCTTCTCAACGACGCTCCGGCATTAGGTGACTATCTGGACGAGGAATCTC
GTGAGCATTTTGCCGGTCTGTGCAAACTGCTGGAGAGCGCGGGGATCGCT
TACACCGTAAACCAGCGTCTGGTGCGTGGTCTGGATTACTACAACCGTAC
CGTTTTCGAGTGGGTGACTAACAGTCTCGGCTCCCAGGGCACCGTGTGTG

CAGGCGGTCGTTATGACGGTCTTGTGGAACAACTGGGCGGTCGTGCAACA
CCGGCTGTCGGTTTTGCTATGGGCCTCGAACGTCTTGTATTGTTAGTACAG
GCCGTTAATCCGGAATTTAAAGCCGATCCTGTTGTCGATATATACCTGGTG
GCTTCAGGTGCTGATACACAATCTGCGGCTATGGCATTAGCTGAGCGTCT
GCGTGATGAATTACCGGGCGTGAAATTGATGACCAACCACGGCGGCGGCA
ACTTTAAGAAACAGTTTGCCCGTGCTGATAAATGGGGTGCCCGCGTTGCT
GTGGTGCTGGGTGAGTCTGAAGTGGCTAACGGCACAGCAGTAGTGAAGG
ATTTGCGCTCTGGTGAGCAAACGGCAGTTGCGCAGGATAGCGTAGCCGCG
CATTTGCGCACGTTACTGGGTTAA

EcIleRS

ATGAGTGACTATAAATCAACCCTGAATTTGCCGGAAACAGGGTTCCCGAT
GCGTGGCGATCTCGCCAAGCGCGAACCCGGAATGCTGGCGCGTTGGACTG
ATGATGATCTGTACGGCATCATCCGTGCGGCTAAAAAAGGCAAAAAAACC
TTCATTCTGCATGATGGCCCTCCTTATGCGAATGGCAGCATTCATATTGGT
CACTCGGTTAACAAGATTCTGAAAGACATTATCGTGAAGTCCAAAGGGCT
TTCCGGTTATGACTCGCCGTATGTGCCTGGCTGGGACTGCCACGGTCTGCC
GATCGAGCTGAAAGTCGAGCAAGAATACGGTAAGCCGGGTGAGAAATTC
ACCGCCGCCGAGTTCCGCGCCAAGTGCCGCGAATACGCGGCGACCCAGGT
TGACGGTCAACGCAAAGACTTTATCCGTCTGGGCGTGCTGGGCGACTGGT
CGCACCCGTACCTGACCATGGACTTCAAAACTGAAGCCAACATCATCCGC
GCGCTGGGCAAAATCATCGGCAACGGTCACCTGCACAAAGGCGCGAAGC
CAGTTCACTGGTGCGTTGACTGCCGTTCTGCGCTGGCGGAAGCGGAAGTT
GAGTATTACGACAAAACTTCTCCGTCCATCGACGTTGCTTTCCAGGCAGTC
GATCAGGATGCACTGAAAGCAAAATTTGCCGTAAGCAACGTTAACGGCCC
AATCTCGCTGGTAATCTGGACCACCACGCCGTGGACTCTGCCTGCCAACC
GCGCAATCTCTATTGCACCAGATTTCGACTATGCGCTGGTGCAGATCGAC
GGTCAGGCCGTGATTCTGGCGAAAGATCTGGTTGAAAGCGTAATGCAGCG
TATCGGCGTGACCGATTACACCATTCTCGGCACGGTAAAAGGTGCGGAGC
TTGAGCTGCTGCGCTTTACCCATCCGTTTATGGGCTTCGACGTTCCGGCAA
TCCTCGGCGATCACGTTACCCTGGATGCCGGTACCGGTGCCGTTCACACCG
CGCCTGGCCACGGCCCGGACGACTATGTGATCGGTCAGAAATACGGCCTG
GAAACCGCTAACCCGGTTGGCCCGGACGGCACTTATCTGCCGGGCACTTA
TCCGACGCTGGATGGCGTGAACGTCTTCAAAGCGAACGACATCGTCGTTG
CGCTGCTGCAGGAAAAAGGCGCGCTGCTGCACGTTGAGAAAATGCAGCA
CAGCTATCCGTGCTGCTGGCGTCACAAAACGCCGATCATCTTCCGCGCGA
CGCCGCAGTGGTTCGTCAGCATGGATCAGAAAGGTCTGCGTGCGCAGTCA
CTGAAAGAGATCAAAGGCGTGCAGTGGATCCCGGACTGGGGCCAGGCGC
GTATCGAGTCGATGGTTGCTAACCGTCCTGACTGGTGTATCTCCCGTCAGC
GCACCTGGGGTGTACCGATGTCACTGTTCGTGCACAAAGACACGGAAGAG
CTGCATCCGCGTACCCTTGAACTGATGGAAGAAGTGGCAAAACGCGTTGA

AGTCGATGGCATCCAGGCGTGGTGGGATCTCGATGCGAAAGAGATCCTCG
GCGACGAAGCTGATCAGTACGTGAAAGTGCCGGACACATTGGATGTATGG
TTTGACTCCGGATCTACCCACTCTTCTGTTGTTGACGTGCGTCCGGAATTT
GCCGGTCACGCAGCGGACATGTATCTGGAAGGTTCTGACCAACACCGCGG
CTGGTTCATGTCTTCCCTAATGATCTCCACCGCGATGAAGGGTAAAGCGCC
GTATCGTCAGGTACTGACCCACGGCTTTACCGTGGATGGTCAGGGCCGCA
AGATGTCTAAATCCATCGGCAATACCGTTTCGCCGCAGGATGTGATGAAC
AAACTGGGCGCGGATATTCTGCGTCTGTGGGTGGCATCAACCGACTACAC
CGGTGAAATGGCCGTTTCTGACGAGATCCTGAAACGTGCTGCCGATAGCT
ATCGTCGTATCCGTAACACCGCGCGCTTCCTGCTGGCAAACCTGAACGGTT
TTGATCCAGCAAAAGATATGGTGAAACCGGAAGAGATGGTGGTACTGGAT
CGCTGGGCCGTAGGTTGTGCGAAAGCGGCACAGGAAGACATCCTCAAGG
CGTACGAAGCATACGATTTCCACGAAGTGGTACAGCGTCTGATGCGCTTC
TGCTCCGTTGAGATGGGTTCCTTCTACCTCGACATCATCAAAGACCGTCAG
TACACCGCCAAAGCGGACAGTGTGGCGCGTCGTAGCTGCCAGACTGCGCT
ATATCACATCGCAGAAGCGCTGGTGCGCTGGATGGCACCAATCCTCTCCT
TCACCGCTGATGAAGTGTGGGGCTACCTGCCGGGCGAACGTGAAAAATAC
GTCTTCACCGGTGAGTGGTACGAAGGCCTGTTTGGCCTGGCAGACAGTGA
AGCGATGAACGATGCGTTCTGGGACGAGCTGTTGAAAGTGCGTGGCGAAG
TGAACAAAGTCATTGAGCAAGCGCGTGCCGACAAGAAAGTGGGTGGCTC
GCTGGAAGCGGCAGTAACCTTGTATGCAGAACCGGAACTGTCGGCGAAAC
TGACCGCGCTGGGCGATGAATTACGATTTGTCCTGTTGACCTCCGGCGCTA
CCGTTGCAGACTATAACGACGCACCTGCTGATGCTCAGCAGAGCGAAGTA
CTCAAAGGGCTGAAAGTCGCGTTGAGTAAAGCCGAAGGTGAGAAGTGCC
CACGCTGCTGGCACTACACCCAGGATGTCGGCAAGGTGGCGGAACACGCA
GAAATCTGCGGCCGCTGTGTCAGCAACGTCGCCGGTGACGGTGAAAAACG
TAAGTTTGCCTAA

EcLeuRS

ATGCAAGAGCAATACCGCCCGGAAGAGATAGAATCCAAAGTACAGCTTC
ATTGGGATGAGAAGCGCACATTTGAAGTAACCGAAGACGAGAGCAAAGA
GAAGTATTACTGCCTGTCTATGCTTCCCTATCCTTCTGGTCGACTACACAT
GGGCCACGTACGTAACTACACCATCGGTGACGTGATCGCCCGCTACCAGC
GTATGCTGGGCAAAAACGTCCTGCAGCCGATCGGCTGGGACGCGTTTGGT
CTGCCTGCGGAAGGCGCGGCGGTGAAAAACAACACCGCTCCGGCACCGT
GGACGTACGACAACATCGCGTATATGAAAAACCAGCTCAAAATGCTGGGC
TTTGGTTATGACTGGAGCCGCGAGCTGGCAACCTGTACGCCGGAATACTA
CCGTTGGGAACAGAAATTCTTCACCGAGCTGTATAAAAAAGGCCTGGTAT
ATAAGAAGACTTCTGCGGTCAACTGGTGCCCGAACGACCAGACCGTACTG
GCGAACGAACAAGTTATCGACGGCTGCTGCTGGCGCTGCGATACCAAAGT
TGAACGTAAAGAGATCCCGCAGTGGTTTATCAAAATCACTGCTTACGCTG

ACGAGCTGCTCAACGATCTGGATAAACTGGATCACTGGCCAGACACCGTT
AAAACCATGCAGCGTAACTGGATCGGTCGTTCCGAAGGCGTGGAGATCAC
CTTCAACGTTAACGACTATGACAACACGCTGACCGTTTACACTACCCGCCC
GGACACCTTTATGGGTTGTACCTACCTGGCGGTAGCTGCGGGTCATCCGCT
GGCGCAGAAAGCGGCGGAAAATAATCCTGAACTGGCGGCCTTTATTGACG
AATGCCGTAACACCAAAGTTGCCGAAGCTGAAATGGCGACGATGGAGAA
AAAAGGCGTCGATACTGGCTTTAAAGCGGTTCACCCATTAACGGGCGAAG
AAATTCCCGTTTGGGCAGCAAACTTCGTATTGATGGAGTACGGCACGGGC
GCAGTTATGGCGGTACCGGGGCACGACCAGCGCGACTACGAGTTTGCCTC
TAAATACGGCCTGAACATCAAACCGGTTATCCTGGCAGCTGACGGCTCTG
AGCCAGATCTTTCTCAGCAAGCCCTGACTGAAAAAGGCGTGCTGTTCAAC
TCTGGCGAGTTCAACGGTCTTGACCATGAAGCGGCCTTCAACGCCATCGC
CGATAAACTGACTGCGATGGGCGTTGGCGAGCGTAAAGTGAACTACCGCC
TGCGCGACTGGGGTGTTTCCCGTCAGCGTTACTGGGGCGCGCCGATTCCG
ATGGTGACGCTGGAAGACGGTACCGTAATGCCGACCCCGGACGACCAGCT
GCCGGTGATCCTGCCGGAAGATGTGGTAATGGACGGCATTACCAGCCCGA
TTAAAGCAGATCCGGAGTGGGCGAAAACTACCGTTAACGGTATGCCAGCA
CTGCGTGAAACCGACACTTTCGACACCTTTATGGAGTCCTCCTGGTACTAT
GCGCGCTACACTTGCCCGCAGTACAAAGAAGGTATGCTGGATTCCGAAGC
GGCTAACTACTGGCTGCCGGTGGATATCTACATTGGTGGTATTGAACACG
CCATTATGCACCTGCTCTACTTCCGCTTCTTCCACAAACTGATGCGTGATG
CAGGCATGGTGAACTCTGACGAACCAGCGAAACAGTTGCTGTGTCAGGGT
ATGGTGCTGGCAGATGCCTTCTACTATGTTGGCGAAAACGGCGAACGTAA
CTGGGTTTCCCCGGTTGATGCTATCGTTGAACGTGACGAGAAAGGCCGTA
TCGTGAAAGCGAAAGATGCGGCAGGCCATGAACTGGTTTATACCGGCATG
AGCAAAATGTCCAAGTCGAAGAACAACGGTATCGACCCGCAGGTGATGG
TTGAACGTTACGGCGCGGACACCGTTCGTCTGTTTATGATGTTTGCTTCTC
CGGCTGATATGACTCTCGAATGGCAGGAATCCGGTGTGGAAGGGGCTAAC
CGCTTCCTGAAACGTGTCTGGAAACTGGTTTACGAGCACACAGCAAAAGG
TGATGTTGCGGCACTGAACGTTGATGCGCTGACTGAAAATCAGAAAGCGC
TGCGTCGCGATGTGCATAAAACGATCGCTAAAGTGACCGATGATATCGGC
CGTCGTCAGACCTTCAACACCGCAATTGCGGCGATTATGGAGCTGATGAA
CAAACTGGCGAAAGCACCAACCGATGGCGAGCAGGATCGCGCTCTGATG
CAGGAAGCACTGCTGGCCGTTGTCCGTATGCTTAACCCGTTCACCCCGCAC
ATCTGCTTCACGCTGTGGCAGGAACTGAAAGGCGAAGGCGATATCGACAA
CGCGCCGTGGCCGGTTGCTGACGAAAAAGCGATGGTGGAAGACTCCACGC
TGGTCGTGGTGCAGGTTAACGGTAAAGTCCGTGCCAAAATCACCGTTCCG
GTGGACGCAACGGAAGAACAGGTTCGCGAACGTGCTGGCCAGGAACATC
TGGTAGCAAAATATCTTGATGGCGTTACTGTACGTAAAGTGATTTACGTAC
CAGGTAAACTCCTCAATCTGGTCGTTGGCTAA

EcLysRS

ATGTCTGAACAACACGCACAGGGCGCTGACGCGGTAGTCGATCTTAACAA
TGAACTGAAAACGCGTCGTGAGAAGCTGGCGAACCTGCGCGAGCAGGGG
ATTGCCTTCCCGAACGATTTCCGTCGCGATCATACCTCTGACCAATTGCAC
GCAGAATTCGACGGCAAAGAGAACGAAGAACTGGAAGCGCTGAACATCG
AAGTCGCCGTTGCTGGCCGCATGATGACCCGTCGTATTATGGGTAAAGCG
TCTTTCGTTACCCTGCAGGACGTTGGCGGTCGCATTCAGCTGTACGTTGCC
CGTGACGATCTCCCGGAAGGCGTTTATAACGAGCAGTTCAAAAAATGGGA
CCTCGGCGACATCCTCGGCGCGAAAGGTAAGCTGTTCAAAACCAAAACCG
GCGAACTGTCTATCCACTGCACCGAGTTGCGTCTGCTGACCAAAGCACTG
CGTCCGCTGCCGGATAAATTCCACGGCTTGCAGGATCAGGAAGCGCGCTA
TCGTCAGCGTTATCTCGATCTCATCTCCAACGATGAATCCCGCAACACCTT
TAAAGTGCGCTCGCAGATCCTCTCTGGTATTCGCCAGTTCATGGTGAACCG
CGGCTTTATGGAAGTTGAAACGCCGATGATGCAGGTGATCCCTGGCGGTG
CCGCTGCGCGTCCGTTTATCACCCACCATAACGCGCTGGATCTCGACATGT
ACCTGCGTATCGCGCCGGAACTGTACCTCAAGCGTCTGGTGGTTGGTGGC
TTCGAGCGTGTATTCGAAATCAACCGTAACTTCCGTAACGAAGGTATTTCC
GTACGTCATAACCCAGAGTTCACCATGATGGAACTCTACATGGCTTACGC
AGATTACAAAGATCTGATCGAGCTGACCGAATCGCTGTTCCGTACTCTGG
CACAGGATATTCTCGGTAAGACGGAAGTGACCTACGGCGACGTGACGCTG
GACTTCGGTAAACCGTTCGAAAAACTGACCATGCGTGAAGCGATCAAGAA
ATATCGCCCGGAAACCGACATGGCGGATCTGGACAACTTCGACTCTGCGA
AAGCAATTGCTGAATCTATCGGCATCCACGTTGAGAAGAGCTGGGGTCTG
GGCCGTATCGTTACCGAGATCTTCGAAGAAGTGGCAGAAGCACATCTGAT
TCAGCCGACCTTCATTACTGAATATCCGGCAGAAGTTTCTCCGCTGGCGCG
TCGTAACGACGTTAACCCGGAAATCACAGACCGCTTTGAGTTCTTCATTGG
TGGTCGTGAAATCGGTAACGGCTTTAGCGAGCTGAATGACGCGGAAGATC
AGGCGCAACGCTTCCTGGATCAGGTTGCCGCGAAAGACGCAGGTGACGAC
GAAGCGATGTTCTACGATGAAGATTACGTCACCGCACTGGAACATGGCTT
ACCGCCGACAGCAGGTCTGGGAATTGGTATCGACCGTATGGTAATGCTGT
TCACCAACAGCCATACCATCCGCGACGTTATTCTGTTCCCGGCGATGCGTC
CGGTAAAATAA

EcMetRS

ATGACTCAAGTCGCGAAGAAAATTCTGGTGACGTGCGCACTGCCGTACGC
TAACGGCTCAATCCACCTCGGCCATATGCTGGAGCACATCCAGGCTGATG
TCTGGGTCCGTTACCAGCGAATGCGCGGCCACGAGGTCAACTTCATCTGC
GCCGACGATGCCCACGGTACACCGATCATGCTGAAAGCTCAGCAGCTTGG
TATCACCCCGGAGCAGATGATTGGCGAAATGAGTCAGGAGCATCAGACTG
ATTTCGCAGGCTTTAACATCAGCTATGACAACTATCACTCGACGCACAGC
GAAGAGAACCGCCAGTTGTCAGAACTTATCTACTCTCGCCTGAAAGAAAA

CGGTTTTATTAAAAACCGCACCATCTCTCAGCTGTACGATCCGGAAAAAG
GCATGTTCCTGCCGGACCGTTTTGTGAAAGGCACCTGCCCGAAATGTAAA
TCCCCGGATCAATACGGCGATAACTGCGAAGTCTGCGGCGCGACCTACAG
CCCGACTGAACTGATCGAGCCGAAATCGGTGGTTTCTGGCGCTACGCCGG
TAATGCGTGATTCTGAACACTTCTTCTTTGATCTGCCCTCTTTCAGCGAAA
TGTTGCAGGCATGGACCCGCAGCGGTGCGTTGCAGGAGCAGGTGGCAAAT
AAAATGCAGGAGTGGTTTGAATCTGGCCTGCAACAGTGGGATATCTCCCG
CGACGCCCCTTACTTCGGTTTTGAAATTCCGAACGCGCCGGGCAAATATTT
CTACGTCTGGCTGGACGCACCGATTGGCTACATGGGTTCTTTCAAGAATCT
GTGCGACAAGCGCGGCGACAGCGTAAGCTTCGATGAATACTGGAAGAAA
GACTCCACCGCCGAGCTGTACCACTTCATCGGTAAAGATATTGTTTACTTC
CACAGCCTGTTCTGGCCTGCCATGCTGGAAGGCAGCAACTTCCGCAAGCC
GTCCAACCTGTTTGTTCATGGCTATGTGACGGTGAACGGCGCAAAGATGT
CCAAGTCTCGCGGCACCTTTATTAAAGCCAGCACCTGGCTGAATCATTTTG
ACGCAGACAGCCTGCGTTACTACTACACTGCGAAACTCTCTTCGCGCATTG
ATGATATCGATCTCAACCTGGAAGATTTCGTTCAGCGTGTGAATGCCGAT
ATCGTTAACAAAGTGGTTAACCTGGCCTCCCGTAATGCGGGCTTTATCAAC
AAGCGTTTTGACGGCGTGCTGGCAAGCGAACTGGCTGACCCGCAGTTGTA
CAAAACCTTCACTGATGCCGCTGAAGTGATTGGTGAAGCGTGGGAAAGCC
GTGAATTTGGTAAAGCCGTGCGCGAAATCATGGCGCTGGCTGATCTGGCT
AACCGCTATGTCGATGAACAGGCTCCGTGGGTGGTGGCGAAACAGGAAG
GCCGCGATGCCGACCTGCAGGCAATTTGCTCAATGGGCATCAACCTGTTC
CGCGTGCTGATGACTTACCTGAAGCCGGTACTGCCGAAACTGACCGAGCG
TGCAGAAGCATTCCTCAATACGGAACTGACCTGGGATGGTATCCAGCAAC
CGCTGCTGGGCCACAAAGTGAATCCGTTCAAGGCGCTGTATAACCGCATC
GATATGAGGCAGGTTGAAGCACTGGTGGAAGCCTCTAAAGAAGAAGTAA
AAGCCGCTGCCGCGCCGGTAACTGGCCCGCTGGCAGATGATCCGATTCAG
GAAACCATCACCTTTGACGACTTCGCTAAAGTTGACCTGCGCGTGGCGCT
GATTGAAAACGCAGAGTTTGTTGAAGGTTCTGACAAACTGCTGCGCCTGA
CGCTGGATCTCGGCGGTGAAAAACGCAATGTCTTCTCCGGTATTCGTTCTG
CTTACCCGGATCCGCAGGCACTGATTGGTCGTCACACCATTATGGTGGCTA
ACCTGGCACCACGTAAAATGCGCTTCGGTATCTCTGAAGGCATGGTGATG
GCTGCCGGTCCTGGCGGGAAAGATATTTTCCTGCTAAGCCCGGATGCCGG
TGCTAAACCGGGTCATCAGGTGAAATAA

EcPheRS α

ATGTCACATCTCGCAGAACTGGTTGCCAGTGCGAAGGCGGCCATTAGCCA
GGCGTCAGATGTTGCCGCGTTAGATAATGTGCGCGTCGAATATTTGGGTA
AAAAAGGGCACTTAACCCTTCAGATGACGACCCTGCGTGAGCTGCCGCCA
GAAGAGCGTCCGGCAGCTGGTGCGGTTATCAACGAAGCGAAAGAGCAGG
TTCAGCAGGCGCTGAATGCGCGTAAAGCGGAACTGGAAAGCGCTGCACTG

AATGCGCGTCTGGCGGCGGAAACGATTGATGTCTCTCTGCCAGGTCGTCG
CATTGAAAACGGCGGTCTGCATCCGGTTACCCGTACCATCGACCGTATCG
AAAGTTTCTTCGGTGAGCTTGGCTTTACCGTGGCAACCGGGCCGGAAATC
GAAGACGATTATCATAACTTCGATGCTCTGAACATTCCTGGTCACCACCCG
GCGCGCGCTGACCACGACACTTTCTGGTTTGACACTACCCGCCTGCTGCGT
ACCCAGACCTCTGGCGTACAGATCCGCACCATGAAAGCCCAGCAGCCACC
GATTCGTATCATCGCGCCTGGCCGTGTTTATCGTAACGACTACGACCAGAC
TCACACGCCGATGTTCCATCAGATGGAAGGTCTGATTGTTGATACCAACA
TCAGCTTTACCAACCTGAAAGGCACGCTGCACGACTTCCTGCGTAACTTCT
TTGAGGAAGATTTGCAGATTCGCTTCCGTCCTTCCTACTTCCCGTTTACCG
AACCTTCTGCAGAAGTGGACGTCATGGGTAAAAACGGTAAATGGCTGGAA
GTGCTGGGCTGCGGGATGGTGCATCCGAACGTGTTGCGTAACGTTGGCAT
CGACCCGGAAGTTTACTCTGGTTTCGCCTTCGGGATGGGGATGGAGCGTC
TGACTATGTTGCGTTACGGCGTCACCGACCTGCGTTCATTCTTCGAAAACG
ATCTGCGTTTCCTCAAACAGTTTAAATAA

EcPheRS β
ATGAAATTCAGTGAACTGTGGTTACGCGAATGGGTGAACCCGGCGATTGA
TAGCGATGCGCTGGCAAATCAAATCACTATGGCGGGCCTGGAAGTTGACG
GTGTAGAACCGGTTGCCGGCAGCTTCCACGGCGTGGTCGTTGGTGAAGTG
GTTGAGTGTGCGCAGCATCCGAACGCTGACAAACTGCGTGTGACAAAAGT
GAATGTCGGCGGCGATCGCCTGCTGGACATCGTCTGCGGTGCGCCAAACT
GCCGTCAGGGCCTGCGTGTAGCGGTAGCGACCATTGGTGCTGTTCTGCCG
GGTGATTTCAAAATTAAAGCGGCGAAACTGCGTGGCGAACCGTCTGAAGG
GATGCTGTGCTCCTTCTCTGAACTGGGCATTTCTGACGATCACAGCGGCAT
TATCGAACTGCCTGCGGATGCGCCGATTGGCACCGATATCCGTGAATACC
TGAAACTTGATGACAACACCATCGAAATCAGCGTGACGCCAAACCGTGCC
GACTGCTTAGGCATCATTGGTGTTGCGCGTGACGTTGCCGTGCTGAACCA
GCTGCCGCTGGTTCAACCGGAAATCGTTCCGGTTGGTGCGACCATCGACG
ACACGCTGCCGATTACAGTCGAAGCGCCGGAAGCCTGCCCGCGTTATCTT
GGCCGTGTGGTAAAAGGCATTAACGTTAAAGCGCCAACTCCGCTGTGGAT
GAAAGAAAAACTGCGTCGTTGCGGGATCCGTTCTATCGATGCAGTTGTTG
ACGTCACCAACTATGTGCTGCTCGAACTGGGCCAGCCGATGCACGCTTTC
GATAAAGATCGCATTGAAGGCGGCATTGTGGTGCGGATGGCGAAAGAGG
GCGAAACGCTGGTGCTGCTCGACGGTACTGAAGCGAAGCTGAATGCTGAC
ACTCTGGTCATCGCCGACCACAACAAGGCGCTGGCGATGGGCGGCATCTT
CGGTGGCGAACACTCTGGCGTGAATGACGAAACACAAAACGTGCTGCTGG
AATGCGCGTTCTTTAGCCCGCTGTCTATCACCGGTCGTGCTCGTCGTCATG
GCCTGCATACCGATGCGTCTCACCGTTATGAGCGTGGCGTTGATCCGGCA
CTGCAGCACAAAGCGATGGAACGTGCGACCCGTCTGCTGATCGACATCTG
CGGTGGTGAGGCTGGCCCGGTAATTGATATCACCAACGAAGCAACGCTGC

CGAAGCGTGCAACCATCACTCTACGTCGTAGCAAACTGGATCGCCTGATC
GGCCATCATATTGCGGATGAGCAGGTAACTGACATTCTGCGTCGTCTCGG
CTGCGAAGTGACCGAAGGCAAAGACGAATGGCAGGCAGTTGCGCCGAGC
TGGCGTTTCGATATGGAGATTGAAGAAGATCTGGTTGAAGAAGTCGCGCG
TGTTTACGGCTACAACAACATCCCGGATGAGCCGGTACAGGCAAGCCTGA
TTATGGGTACTCACCGTGAAGCTGACCTGTCGCTCAAGCGCGTGAAAACG
CTGCTCAACGACAAAGGCTATCAGGAAGTGATCACCTACAGCTTCGTTGA
TCCGAAAGTGCAGCAGATGATCCATCCAGGCGTTGAAGCCTTACTGCTGC
CAAGCCCGATCTCTGTTGAAATGTCAGCAATGCGTCTTTCTCTGTGGACTG
GCCTGCTGGCAACCGTGGTGTACAACCAGAACCGTCAGCAGAACCGTGTG
CGCATTTTCGAAAGCGGTCTGCGTTTCGTACCAGATACTCAGGCACCGTTG
GGCATTCGTCAGGATCTGATGTTAGCCGGTGTGATTTGCGGTAACCGTTAC
GAAGAGCACTGGAACCTGGCAAAAGAGACCGTTGATTTCTATGATTTGAA
AGGCGATCTTGAATCCGTTCTCGACCTGACCGGTAAACTGAATGAGGTTG
AGTTCCGTGCAGAAGCGAATCCGGCACTGCATCCGGGGCAATCCGCAGCG
ATTTATCTGAAAGGTGAACGTATTGGTTTTGTTGGGGTTGTTCATCCTGAA
CTGGAACGTAAACTGGATCTTAACGGTCGCACTCTGGTGTTCGAACTGGA
GTGGAACAAGCTCGCAGACCGCGTGGTGCCTCAGGCGCGCGAGATTTCTC
GCTTCCCGGCGAACCGTCGTGACATCGCGGTGGTGGTCGCAGAAAACGTT
CCCGCAGCGGATATTTTATCCGAATGTAAGAAAGTTGGCGTAAATCAGGT
AGTTGGCGTAAACTTATTTGACGTGTACCGCGGTAAGGGTGTTGCGGAGG
GGTATAAGAGCCTCGCCATAAGCCTGATCCTGCAAGATACCAGCCGTACA
CTCGAAGAAGAGGAGATTGCCGCTACCGTCGCCAAATGTGTAGAGGCATT
AAAAGAGCGATTCCAGGCATCATTGAGGGATTTA

EcProRS

ATGCGTACTAGCCAATACCTGCTCTCCACTCTCAAGGAGACACCTGCCGA
CGCCGAGGTGATCAGCCATCAGCTGATGCTGCGCGCCGGGATGATCCGCA
AGCTGGCCTCCGGGTTATATACCTGGCTGCCGACCGGCGTGCGCGTTCTG
AAAAAAGTCGAAAACATCGTGCGTGAAGAGATGAACAACGCCGGTGCGA
TCGAGGTGTCGATGCCGGTGGTTCAGCCAGCCGATTTGTGGCAAGAGAGT
GGTCGTTGGGAACAGTACGGTCCGGAACTGCTGCGTTTTGTTGACCGTGG
CGAGCGTCCGTTCGTACTCGGCCCAACTCATGAAGAAGTTATCACTGACC
TGATTCGTAACGAGCTTAGCTCTTACAAACAGCTGCCGCTGAACTTCTATC
AGATCCAGACCAAGTTCCGCGACGAAGTGCGTCCGCGTTTCGGCGTCATG
CGTTCCCGCGAATTCCTGATGAAAGATGCTTACTCTTTCCATACTTCTCAG
GAATCCCTGCAGGAAACCTACGATGCAATGTATGCGGCCTACAGCAAAAT
CTTCAGCCGCATGGGGCTGGATTTCCGCGCCGTACAAGCCGACACCGGTT
CTATCGGCGGCAGCGCCTCTCACGAATTCCAGGTGCTGGCGCAGAGCGGT
GAAGACGATGTGGTCTTCTCCGACACCTCTGACTATGCAGCGAACATTGA
ACTGGCAGAAGCTATCGCGCCGAAAGAACCGCGCGCTGCTGCTACCCAGG

AAATGACGCTGGTTGATACGCCGAACGCGAAAACCATCGCGGAACTGGTT
GAACAGTTCAATCTGCCGATTGAGAAAACGGTTAAGACTCTGCTGGTTAA
AGCGGTTGAAGGCAGCAGCTTCCCGCAGGTTGCGCTGCTGGTGCGCGGTG
ATCACGAGCTGAACGAAGTTAAAGCAGAAAAACTGCCGCAGGTTGCAAG
CCCGCTGACTTTCGCGACCGAAGAAGAAATTCGTGCCGTGGTTAAAGCCG
GTCCGGGTTCACTGGGTCCGGTAAACATGCCGATTCCGGTGGTGATTGAC
CGTACCGTTGCGGCGATGAGTGATTTCGCTGCTGGTGCTAACATCGATGGT
AAACACTACTTCGGCATCAACTGGGATCGCGATGTCGCTACCCCGGAAGT
TGCAGATATCCGTAACGTGGTGGCTGGCGATCCAAGCCCGGATGGCCAGG
GTAGGCTGCTGATCAAACGTGGTATCGAAGTTGGTCACATCTTCCAGCTG
GGTACCAAGTACTCCGAAGCACTGAAAGCCTCCGTACAGGGTGAAGATGG
CCGTAACCAAATCCTGACGATGGGTTGCTACGGTATCGGGGTAACGCGTG
TGGTAGCTGCGGCGATTGAGCAGAACTACGACGAACGAGGCATCGTATGG
CCTGACGCTATCGCGCCGTTCCAGGTGGCGATTCTGCCGATGAACATGCA
CAAATCCTTCCGCGTACAAGAGCTTGCTGAGAAACTGTACAGCGAACTGC
GTGCACAAGGTATCGAAGTGCTGCTGGATGACCGCAAAGAGCGTCCGGGC
GTGATGTTTGCTGATATGGAACTGATCGGTATTCCGCACACTATTGTGCTG
GGCGACCGTAACCTCGACAACGACGATATCGAATATAAATATCGTCGCAA
CGGCGAGAAACAGTTAATTAAGACTGGTGACATCGTCGAATATCTGGTGA
AACAGATTAAAGGCTAA

EcSerRS
ATGCTCGATCCCAATCTGCTGCGTAATGAGCCAGACGCAGTCGCTGAAAA
ACTGGCACGCCGGGGCTTTAAGCTGGATGTAGATAAGCTGGGCGCTCTTG
AAGAGCGTCGTAAAGTATTGCAGGTCAAAACGGAAAACCTGCAAGCGGA
GCGTAACTCCCGATCGAAATCCATTGGCCAGGCGAAAGCGCGCGGGGAA
GATATCGAGCCTTTACGTCTGGAAGTGAACAAACTGGGCGAAGAGCTGGA
TGCAGCAAAAGCCGAGCTGGATGCTTTACAGGCTGAAATTCGCGATATCG
CGCTGACCATCCCTAACCTGCCTGCAGATGAAGTGCCGGTAGGTAAAGAC
GAAAATGACAACGTTGAAGTCAGCCGCTGGGGTACCCCGCGTGAGTTTGA
CTTTGAAGTTCGTGACCACGTGACGCTGGGTGAAATGCACTCTGGCCTCG
ACTTTGCAGCTGCAGTTAAGCTGACTGGTTCCCGCTTTGTGGTAATGAAAG
GGCAGATTGCTCGCATGCACCGCGCACTGTCGCAGTTTATGCTGGATCTGC
ATACCGAACAGCATGGCTACAGTGAGAACTATGTTCCGTACCTGGTTAAC
CAGGACACGCTGTACGGTACGGGTCAACTGCCGAAATTTGCTGGCGATCT
GTTCCATACTCGTCCGCTGGAAGAAGAAGCAGACACCAGTAACTATGCGC
TGATCCCAACGGCAGAAGTTCCGCTGACTAACCTGGTGCGCGGTGAAATC
ATCGATGAAGATGATCTGCCAATTAAGATGACCGCCCACACCCCATGCTT
CCGTTCTGAAGCCGGTTCATATGGTCGTGACACCCGTGGTCTGATCCGTAT
GCACCAGTTCGACAAAGTTGAAATGGTGCAGATCGTGCGCCCAGAAGACT
CAATGGCGGCGCTGGAAGAGATGACTGGTCATGCAGAAAAAGTCCTGCA

GTTGCTGGGCCTGCCGTACCGTAAAATCATCCTTTGCACTGGCGACATGG
GCTTTGGCGCTTGCAAAACTTACGACCTGGAAGTATGGATCCCGGCACAG
AACACCTACCGTGAGATCTCTTCCTGCTCCAACGTTTGGGATTTCCAGGCA
CGTCGTATGCAGGCACGTTGCCGCAGCAAGTCGGACAAGAAAACCCGTCT
GGTTCATACCCTGAACGGTTCTGGTCTGGCTGTTGGTCGTACGCTGGTTGC
AGTAATGGAAAACTATCAGCAGGCTGATGGTCGTATTGAAGTACCAGAAG
TTCTGCGTCCGTATATGAACGGACTGGAATATATTGGCTAA

EcThrRS
ATGCCTGTTATAACTCTTCCTGATGGCAGCCAACGCCATTACGATCACGCT
GTAAGCCCCATGGATGTTGCGCTGGACATTGGTCCAGGTCTGGCGAAAGC
CTGTATCGCAGGGCGCGTTAATGGCGAACTGGTTGATGCTTGCGATCTGA
TTGAAAACGACGCACAACTGTCGATCATTACCGCCAAAGACGAAGAAGGT
CTGGAGATCATTCGTCACTCCTGTGCGCACCTGTTAGGGCACGCGATTAA
ACAACTTTGGCCGCATACCAAAATGGCAATCGGCCCGGTTATTGACAACG
GTTTTTATTACGACGTTGATCTTGACCGCACGTTAACCCAGGAAGATGTCG
AAGCACTCGAGAAGCGGATGCATGAGCTTGCTGAGAAAAACTACGACGT
CATTAAGAAGAAAGTCAGCTGGCACGAAGCGCGTGAAACTTTCGCCAACC
GTGGGGAGAGCTACAAAGTCTCCATTCTTGACGAAAACATCGCCCATGAT
GACAAGCCAGGTCTGTACTTCCATGAAGAATATGTCGATATGTGCCGCGG
TCCGCACGTACCGAACATGCGTTTCTGCCATCATTTCAAACTAATGAAAAC
GGCAGGGGCTTACTGGCGTGGCGACAGCAACAACAAAATGTTGCAACGT
ATTTACGGTACGGCGTGGGCAGACAAAAAAGCACTTAACGCTTACCTGCA
GCGCCTGGAAGAAGCCGCGAAACGCGACCACCGTAAAATCGGTAAACAG
CTCGACCTGTACCATATGCAGGAAGAAGCGCCGGGTATGGTATTCTGGCA
CAACGACGGCTGGACCATCTTCCGTGAACTGGAAGTGTTTGTTCGTTCTAA
ACTGAAAGAGTACCAGTATCAGGAAGTTAAAGGTCCGTTCATGATGGACC
GTGTCCTGTGGGAAAAAACCGGTCACTGGGACAACTACAAAGATGCAATG
TTCACCACATCTTCTGAGAACCGTGAATACTGCATTAAGCCGATGAACTG
CCCGGGTCACGTACAAATTTTCAACCAGGGGCTGAAGTCTTATCGCGATC
TGCCGCTGCGTATGGCCGAGTTTGGTAGCTGCCACCGTAACGAGCCGTCA
GGTTCGCTGCATGGCCTGATGCGCGTGCGTGGATTTACCCAGGATGACGC
GCATATCTTCTGTACTGAAGAACAAATTCGCGATGAAGTTAACGGATGTA
TCCGTTTAGTCTATGATATGTACAGCACTTTTGGCTTCGAGAAGATCGTCG
TCAAACTCTCCACTCGTCCTGAAAAACGTATTGGCAGCGACGAAATGTGG
GATCGTGCTGAGGCGGACCTGGCGGTTGCGCTGGAAGAAAACAACATCCC
GTTTGAATATCAACTGGGTGAAGGCGCTTTCTACGGTCCGAAAATTGAATT
TACCCTGTATGACTGCCTCGATCGTGCATGGCAGTGCGGTACAGTACAGC
TGGACTTCTCTTTGCCGTCTCGTCTGAGCGCTTCTTATGTAGGCGAAGACA
ATGAACGTAAAGTACCGGTAATGATTCACCGCGCAATTCTGGGGTCGATG
GAACGTTTCATCGGTATCCTGACCGAAGAGTTCGCTGGTTTCTTCCCGACC

TGGCTTGCGCCGGTTCAGGTTGTTATCATGAATATTACCGATTCACAGTCT
GAATACGTTAACGAATTGACGCAAAAACTATCAAATGCGGGCATTCGTGT
TAAAGCAGACTTGAGAAATGAGAAGATTGGCTTTAAAATCCGCGAGCACA
CTTTGCGTCGCGTCCCATATATGCTGGTCTGTGGTGATAAAGAGGTGGAAT
CAGGCAAAGTTGCCGTTCGCACCCGCCGTGGTAAAGACCTGGGAAGCATG
GACGTAAATGAAGTGATCGAGAAGCTGCAACAAGAGATTCGCAGCCGCA
GTCTTAAACAATTGGAGGAATAA

EcTrpRS
ATGACTAAGCCCATCGTTTTTAGTGGCGCACAGCCCTCAGGTGAATTGAC
CATTGGTAACTACATGGGTGCGCTGCGTCAGTGGGTAAACATGCAGGATG
ACTACCATTGCATTTACTGTATCGTTGACCAACACGCGATCACCGTGCGCC
AGGATGCACAGAAGCTGCGTAAAGCGACGCTGGATACGCTGGCCTTGTAT
CTGGCTTGTGGTATCGATCCTGAGAAAAGCACCATTTTTGTTCAGTCCCAC
GTGCCGGAACATGCACAGTTAGGCTGGGCACTGAACTGCTATACCTACTT
CGGCGAACTGAGTCGCATGACGCAGTTTAAAGATAAATCTGCGCGTTATG
CCGAGAACATCAACGCTGGTCTGTTTGACTATCCGGTGCTGATGGCAGCG
GACATCCTGCTGTATCAAACTAATCTGGTACCGGTGGGTGAAGACCAGAA
ACAGCACCTCGAACTGAGCCGCGATATTGCCCAGCGTTTCAACGCGCTGT
ATGGCGAGATCTTTAAGGTGCCGGAGCCGTTTATTCCGAAATCTGGCGCG
CGCGTAATGTCGCTGCTGGAGCCGACCAAGAAGATGTCCAAGTCTGACGA
TAATCGCAATAACGTTATCGGCCTGCTGGAAGATCCGAAATCGGTAGTGA
AGAAAATCAAACGTGCGGTCACTGACTCCGACGAGCCGCCGGTAGTTCGC
TACGATGTGCAGAACAAAGCGGGCGTTTCCAACCTGTTGGATATCCTTTC
AGCGGTAACGGGCCAGAGCATCCCAGAACTGGAAAAACAGTTCGAAGGC
AAGATGTATGGTCATCTGAAAGGTGAAGTGGCTGATGCCGTTTCCGGTAT
GCTGACTGAATTGCAGGAACGCTATCACCGTTTCCGCAACGATGAAGCCT
TCCTGCAACAGGTGATGAAAGATGGCGCGGAAAAAGCCAGCGCGCACGC
TTCCCGTACGCTAAAAGCGGTGTACGAAGCGATTGGTTTTGTGGCGAAGC
CGTAA

EcTyrRS
ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTAGC
CCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCG
ATCGCGCTCTATTGCGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGGG
CATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAAG
CCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTT
CAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGG
GTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGG
AGAAAACTCTGCTATCGCGGCGAACAACTATGACTGGTTCGGCAATATGA
ATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAGA

TGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGG
GATTTCGTTCACTGAGTTTTCCTACAACCTGTTGCAGGGTTATGACTTCGC
CTGTCTGAACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCTGACC
AGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTGCATCAG
AATCAGGTGTTTGGCCTGACCGTTCCGCTGATCACTAAAGCAGATGGCAC
CAAATTTGGTAAAACTGAAGGCGGCGCAGTCTGGTTGGATCCGAAGAAAA
CCAGCCCGTACAAATTCTACCAGTTCTGGATCAACACTGCGGATGCCGAC
GTTTACCGCTTCCTGAAGTTCTTCACCTTTATGAGCATTGAAGAGATCAAC
GCCCTGGAAGAAGAAGATAAAAACAGCGGTAAAGCACCGCGCGCCCAGT
ATGTACTGGCGGAGCAGGTGACTCGTCTGGTTCACGGTGAAGAAGGTTTA
CAGGCGGCAAAACGTATTACCGAATGCCTGTTCAGCGGTTCTTTGAGTGC
GCTGAGTGAAGCGGACTTCGAACAGCTGGCGCAGGACGGCGTACCGATG
GTTGAGATGGAAAAGGGCGCAGACCTGATGCAGGCACTGGTCGATTCTGA
ACTGCAACCTTCCCGTGGTCAGGCACGTAAAACTATCGCCTCCAATGCCA
TCACCATTAACGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAA
GAAGATCGTCTGTTTGGTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAAT
TACTGTCTGATTTGCTGGAAATAA

EcValRS
ATGGAAAAGACATATAACCCACAAGATATCGAACAGCCGCTTTACGAGCA
CTGGGAAAAGCAGGGCTACTTTAAGCCTAATGGCGATGAAAGCCAGGAA
AGTTTCTGCATCATGATCCCGCCGCCGAACGTCACCGGCAGTTTGCATATG
GGTCACGCCTTCCAGCAAACCATCATGGATACCATGATCCGCTATCAGCG
CATGCAGGGCAAAAACACCCTGTGGCAGGTCGGTACTGACCACGCCGGG
ATCGCTACCCAGATGGTCGTTGAGCGCAAGATTGCCGCAGAAGAAGGTAA
AACCCGTCACGACTACGGCCGCGAAGCTTTCATCGACAAAATCTGGGAAT
GGAAAGCGGAATCTGGCGGCACCATTACCCGTCAGATGCGCCGTCTCGGC
AACTCCGTCGACTGGGAGCGTGAACGCTTCACCATGGACGAAGGCCTGTC
CAATGCGGTGAAAGAAGTTTTCGTTCGTCTGTATAAAGAAGACCTGATTT
ACCGTGGCAAACGCCTGGTAAACTGGGATCCGAAACTGCGCACCGCTATC
TCTGACCTGGAAGTCGAAAACCGCGAATCGAAAGGTTCGATGTGGCACAT
CCGCTATCCGCTGGCTGACGGTGCGAAAACCGCAGACGGTAAAGATTATC
TGGTGGTCGCGACTACCCGTCCAGAAACCCTGCTGGGCGATACTGGCGTA
GCCGTTAACCCGGAAGATCCGCGTTACAAAGATCTGATTGGCAAATATGT
CATTCTGCCGCTGGTTAACCGTCGTATTCCGATCGTTGGCGACGAACACGC
CGACATGGAAAAAGGCACCGGCTGCGTGAAAATCACTCCGGCGCACGAC
TTTAACGACTATGAAGTGGGTAAACGTCACGCCCTGCCGATGATCAACAT
CCTGACCTTTGACGGCGATATCCGTGAAAGCGCCCAGGTGTTCGATACCA
AAGGTAACGAATCTGACGTTTATTCCAGCGAAATCCCTGCAGAGTTCCAG
AAACTGGAGCGTTTTGCTGCACGTAAAGCAGTCGTTGCCGCAGTTGACGC
GCTTGGCCTGCTGGAAGAAATTAAACCGCACGACCTGACCGTTCCTTACG

GCGACCGTGGCGGCGTAGTTATCGAACCAATGCTGACCGACCAGTGGTAC
GTGCGTGCCGATGTCCTGGCGAAACCGGCGGTTGAAGCGGTTGAGAACGG
CGACATTCAGTTCGTACCGAAGCAGTACGAAAACATGTACTTCTCCTGGA
TGCGCGATATTCAGGACTGGTGTATCTCTCGTCAGTTGTGGTGGGGTCACC
GTATCCCGGCATGGTATGACGAAGCGGGTAACGTTTATGTTGGCCGCAAC
GAAGACGAAGTGCGTAAAGAAAATAACCTCGGTGCTGATGTTGTCCTGCG
TCAGGACGAAGACGTTCTCGATACCTGGTTCTCTTCTGCGCTGTGGACCTT
CTCTACCCTTGGCTGGCCGGAAAATACCGACGCCCTGCGTCAGTTCCACCC
AACCAGCGTGATGGTATCTGGTTTCGACATCATTTTCTTCTGGATTGCCCG
CATGATCATGATGACCATGCACTTCATCAAAGATGAAAATGGCAAACCGC
AGGTGCCGTTCCACACCGTTTACATGACCGGCCTGATTCGTGATGACGAA
GGCCAGAAGATGTCCAAATCCAAGGGTAACGTTATCGACCCACTGGATAT
GGTTGACGGTATTTCGCTGCCAGAACTGCTGGAAAAACGTACCGGCAATA
TGATGCAGCCGCAGCTGGCGGACAAAATCCGTAAGCGCACCGAGAAGCA
GTTCCCGAACGGTATTGAGCCGCACGGTACTGACGCGCTGCGCTTCACCC
TGGCGGCGCTGGCGTCTACCGGTCGTGACATCAACTGGGATATGAAGCGT
CTGGAAGGTTACCGTAACTTCTGTAACAAGCTGTGGAACGCCAGCCGCTT
TGTGCTGATGAACACAGAAGGTCAGGATTGCGGCTTCAACGGCGGCGAAA
TGACGCTGTCGCTGGCGGACCGCTGGATTCTGGCGGAGTTCAACCAGACC
ATCAAAGCGTACCGCGAAGCGCTGGACAGCTTCCGCTTCGATATCGCCGC
AGGCATTCTGTATGAGTTCACCTGGAACCAGTTCTGTGACTGGTATCTCGA
GCTGACCAAGCCGGTAATGAACGGTGGCACCGAAGCAGAACTGCGCGGT
ACTCGCCATACGCTGGTGACTGTACTGGAAGGTCTGCTGCGCCTCGCGCA
TCCGATCATTCCGTTCATCACCGAAACCATCTGGCAGCGTGTGAAAGTACT
TTGCGGTATCACTGCCGACACCATCATGCTGCAGCCGTTCCCGCAGTACG
ATGCATCTCAGGTTGATGAAGCCGCACTGGCCGACACCGAATGGCTGAAA
CAGGCGATCGTTGCGGTACGTAACATCCGTGCAGAAATGAACATCGCGCC
GGGCAAACCGCTGGAGCTGCTGCTGCGTGGTTGCAGCGCGGATGCAGAAC
GTCGCGTAAATGAAAACCGTGGCTTCCTGCAAACCCTGGCGCGTCTGGAA
AGTATCACCGTGCTGCCTGCCGATGACAAAGGTCCGGTTTCCGTTACGAA
GATCATCGACGGTGCAGAGCTGCTGATCCCGATGGCTGGCCTCATCAACA
AAGAAGATGAGCTGGCGCGTCTGGCGAAAGAAGTGGCGAAGATTGAAGG
TGAAATCAGCCGTATCGAGAACAAACTGGCGAACGAAGGCTTTGTCGCCC
GCGCACCGGAAGCGGTCATCGCGAAAGAGCGTGAGAAGCTGGAAGGCTA
TGCGGAAGCGAAAGCGAAACTGATTGAACAGCAGGCTGTTATCGCCGCGC
TGTAA

Plasmid for tRNA.

All plasmid for tRNA are based on pIDTSMART vector backbone. The tRNA are driven by U6 promoter. The full sequence of pIDTSMART EcAlatRNATAG will be listed below with promoter in Green color, and ORF of $tRNA^{EcAla}_{CUA}$ in orange. For rest of the tRNA plasmids, only the tRNA sequence are listed in Table 1 and 2 for TAG and TGA suppressor respectively.

pIDTSMART $EcAlatRNA_{CUA}$

CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATC
GTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGA
GTCTGAGGCTCGTCCTGAATGATATGCGaCCGCCGGAGGGTTGCGTTTGAG
ACGGGCGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGCAAAAAA
TGGTGGAGCTATGCGGGATCGAACCGCAGACCTCCTGCTTTAGAGGCAGG
CGCTCTCCCAGCTGAGCTATAGCCCCGGTGTTTCGTCCTTTCCACAAGATA
TATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATA
GTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTA
CTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTA
ATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGA
ATCATGGGAAATAGGCCCTCTTCCTGCCCGAcCTAGGGGTGCGAGCGGAT
CGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACcCGTGAT
CTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTC
AGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCA
TACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTG
ACTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGAT
CACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAG
CTAAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTA
AATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAA
TGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATG
CGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTT
ACAGATGAGATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCC
GACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCAC
TGCGATCCCAGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATT
CAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATT
CGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCGCGTATTTCGTCTCGC
TCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGATTTTG
ATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCAT

AAACTCTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCA
CTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTT
GGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAA
CTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATA
TGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCACTTGATGCTCGA
TGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCTCACCTTCGGGT
GGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAG
GTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGA
CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA
GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGC
TACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG
TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAG
GATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

tRNA (CUA)	anticodon usage	Sequence
Ala	59%	GGGGCTATAGCTCAGCTGGGAGAGCGCCTGC**CTCTAAA**GCAGGAGGTCTGCGGTTCGATCCCGCATAGCTCCACCA
Arg	80%	GCATCCGTAGCTCAGCTGGATAGAGTACTCGG**CTCTAAA**CCGAGCGGTCGGAGGTTCGAATCCTCCCGGATGCACCA
Asn	100%	TCCTCTGTAGTTCAGTCGGTAGAACGGCGGA**CTCTAAA**TCCGTATGTCACTGGTTCGAGTCCAGTCAGAGGAGCCA
Asp	100%	GGAGCGGTAGTTCAGTCGGTTAGAATACCTGC**CTCTAAA**GCAGGGGGTCGCGGGTTCGAGTCCCGTCCGTTCCGCCA
Cys	100%	GGCGCGTTAACAAAGCGGTTATGTAGCGGA**CTCTAAA**TCCGTCTAGTCCGGTTCGACTCCGGAACGCGCCTCCA
Gln	70%	TGGGGTATCGCCAAGCGGTAAGGCACCGGA**CTCTAAA**TCCGGCATTCCGAGGTTCGAATCCTCGTACCCCAGCCA

Glu	100%	GTCCCCTTCGTCTAGAGGCCCAGGACACCGCC**CTCT AAA**GGCGGTAACAGGGGTTCGAATCCCCTAGGGGA CGCCA
Gly	75%	GCGGGAATAGCTCAGTTGGTAGAGCACGACC**CTCT AAA**GGTCGGGGTCGCGAGTTCGAGTCTCGTTTCCCG CTCCA
His	100%	GTGGCTATAGCTCAGTTGGTAGAGCCCTGGA**CTCTA AA**TCCAGTTGTCGTGGGTTCGAATCCCATTAGCCAC CCCA
HisG	100%	GGTGGCTATAGCTCAGTTGGTAGAGCCCTGGA**CTCT AAA**TCCAGTTGTCGTGGGTTCGAATCCCATTAGCCA CCCCA
Ile	93%	AGGCTTGTAGCTCAGGTGGTTAGAGCGCACCC**CTCT AAA**GGGTGAGGTCGGTGGTTCAAGTCCACTCAGGC CTACCA
Leu	46%	GCGAAGGTGGCGGAATTGGTAGACGCGCTAGC**CTC TAAA**GTTAGTGTTCTTACGGACGTGGGGGTTCAAGT CCCCCCCCTCGCACCA
Lys	100%	GGGTCGTTAGCTCAGTTGGTAGAGCAGTTGA**CTCTA AA**TCAATTGGTCGCAGGTTCGAATCCTGCACGACCC ACCA
Met	100%	GGCCCTTTAGCTCAGTGGTTAGAGCAGGCGA**CTCT AAA**TCGCTTGGTCGCTGGTTCAAGTCCAGCAAGGGC CACCA
Phe	100%	GCCCGGATAGCTCAGTCGGTAGAGCAGGGGA**CTCT AAA**TCCCCGTGTCCTTGGTTCGATTCCGAGTCCGGG CACCA
Pro	28%	CGGCGAGTAGCGCAGCTTGGTAGCGCAACTGG**CTC TAAA**CCAGTGGGTCGGAGGTTCGAATCCTCTCTCGC CGACCA
Ser	33%	GGTGAGGTGGCCGAGAGGCTGAAGGCGCTCCC**CTC TAAA**GGGAGTATGCGGTCAAAAGCTGCATCCGGGG TTCGAATCCCCGCCTCACCGCCA
Thr	63%	GCTGATATGGCTCAGTTGGTAGAGCGCACCC**CTCTA AA**GGGTGAGGTCCCCAGTTCGACTCTGGGTATCAGC ACCA
Trp	100%	AGGGGCGTAGTTCAATTGGTAGAGCACCGGT**CTCT AAA**ACCGGGTGTTGGGAGTTCGAGTCTCTCCGCCCC TGCCA

Tyr	100%	GGTGGGGTTCCCGAGCGGCCAAAGGGAGCAGA**CTC TAAA**TCTGCCGTCATCGACTTCGAAGGTTCGAATCC TTCCCCCACCACCA
Val	65%	GCGTTCATAGCTCAGTTGGTTAGAGCACCACC**CTCT AAA**GGTGGGGGTCGTTGGTTCGAGTCCAATTGAAC GCACCA

Table 1. EctRNA$_{CUA}$ sequence used in this chapter, with anticodon loop highlighted.

The anticodon usage in bacteria were used as to select the tRNA.

tRNA (UCA)	anticodon usage	Sequence
Ala	59%	GGGGCTATAGCTCAGCTGGGAGAGCGCCTGC**CTTC AAA**GCAGGAGGTCTGCGGTTCGATCCCGCATAGCTC CACCA
Arg	80%	GCATCCGTAGCTCAGCTGGATAGAGTACTCGG**CTTC AAA**CCGAGCGGTCGGAGGTTCGAATCCTCCCGGAT GCACCA
Asn	100%	TCCTCTGTAGTTCAGTCGGTAGAACGGCGGA**CTTCA AA**TCCGTATGTCACTGGTTCGAGTCCAGTCAGAGGA GCCA
Asp	100%	GGAGCGGTAGTTCAGTCGGTTAGAATACCTGC**CTTC AAA**GCAGGGGGTCGCGGGTTCGAGTCCCGTCCGTTC CGCCA
Cys	100%	GGCGCGTTAACAAAGCGGTTATGTAGCGGA**CTTCA AA**TCCGTCTAGTCCGGTTCGACTCCGGAACGCGCCT CCA
Gln	70%	TGGGGTATCGCCAAGCGGTAAGGCACCGGA**CTTCA AA**TCCGGCATTCCGAGGTTCGAATCCTCGTACCCCA GCCA
Glu	100%	GTCCCCTTCGTCTAGAGGCCCAGGACACCGCC**CTTC AAA**GGCGGTAACAGGGGTTCGAATCCCCTAGGGGA CGCCA
Gly	75%	GCGGGAATAGCTCAGTTGGTAGAGCACGACC**CTTC AAA**GGTCGGGGTCGCGAGTTCGAGTCTCGTTTCCCG CTCCA
His	100%	GTGGCTATAGCTCAGTTGGTAGAGCCCTGGA**CTTCA AA**TCCAGTTGTCGTGGGTTCGAATCCCATTAGCCAC CCCA

HisG	100%	GGTGGCTATAGCTCAGTTGGTAGAGCCCTGGA**CTTC AAA**TCCAGTTGTCGTGGGTTCGAATCCCATTAGCCA CCCCA
Ile	93%	AGGCTTGTAGCTCAGGTGGTTAGAGCGCACCC**CTTC AAA**GGGTGAGGTCGGTGGTTCAAGTCCACTCAGGC CTACCA
Leu	46%	GCGAAGGTGGCGGAATTGGTAGACGCGCTAGC**CTT CAAA**GTTAGTGTTCTTACGGACGTGGGGGTTCAAGT CCCCCCCCTCGCACCA
Lys	100%	GGGTCGTTAGCTCAGTTGGTAGAGCAGTTGA**CTTCA AA**TCAATTGGTCGCAGGTTCGAATCCTGCACGACCC ACCA
Met	100%	GGCCCTTTAGCTCAGTGGTTAGAGCAGGCGA**CTTC AAA**TCGCTTGGTCGCTGGTTCAAGTCCAGCAAGGGC CACCA
Phe	100%	GCCCGGATAGCTCAGTCGGTAGAGCAGGGGA**CTTC AAA**TCCCCGTGTCCTTGGTTCGATTCCGAGTCCGGG CACCA
Pro	28%	CGGCGAGTAGCGCAGCTTGGTAGCGCAACTGG**CTT CAAA**CCAGTGGGTCGGAGGTTCGAATCCTCTCTCGC CGACCA
Ser	33%	GGTGAGGTGGCCGAGAGGCTGAAGGCGCTCCC**CTT CAAA**GGGAGTATGCGGTCAAAAGCTGCATCCGGGG TTCGAATCCCCGCCTCACCGCCA
Thr	63%	GCTGATATGGCTCAGTTGGTAGAGCGCACCC**CTTCA AA**GGGTGAGGTCCCCAGTTCGACTCTGGGTATCAGC ACCA
Trp	100%	AGGGGCGTAGTTCAATTGGTAGAGCACCGGT**CTTC AAA**ACCGGGTGTTGGGAGTTCGAGTCTCTCCGCCCC TGCCA
Tyr	100%	GGTGGGGTTCCCGAGCGGCCAAAGGGAGCAGA**CTT CAAA**TCTGCCGTCATCGACTTCGAAGGTTCGAATCC TTCCCCCACCACCA
Val	65%	GCGTTCATAGCTCAGTTGGTTAGAGCACCACC**CTTC AAA**GGTGGGGGTCGTTGGTTCGAGTCCAATTGAAC GCACCA

Table 2. EctRNA$_{UCA}$ sequence used in this chapter, with anticodon loop highlighted.

The anticodon usage in bacteria were used as to select the tRNA.

Whole protein mass spectrum data.

EGFP protein generated by active tRNA/RS pairs were purified and characterized by LC/MS spectrum. (Table 3 and 4) And all the individual deconvoluted peak is listed below. (Figure 3-5)

Pairs	**Expected**	**Observed**	
		TAG	TGA
Ala	30413	30411	
Arg	30498	30496	30496
Asp	30456	30455	
Gln	30469	30466	
His	30477	30477	
HisG	30477	30477	30475
Leu	30454	30452	
Lys	30469	30466	
Phe	30489	30486	
Ser	30428	30427	
Trp	30527	30524	30494
Tyr	30504	30504	30503

Table 3. Summary of selected EGFP generated by active aaRS/tRNA with both TAG and TGA suppressor.

tRNA TAG	**Observed** TAG	Incorporated AA
Ala	30410	Ala
Arg	30467	Gln
Asp	30496	Arg

Table 4. Summary of selected EGFP generated by cross-active tRNA with TAG suppressor.

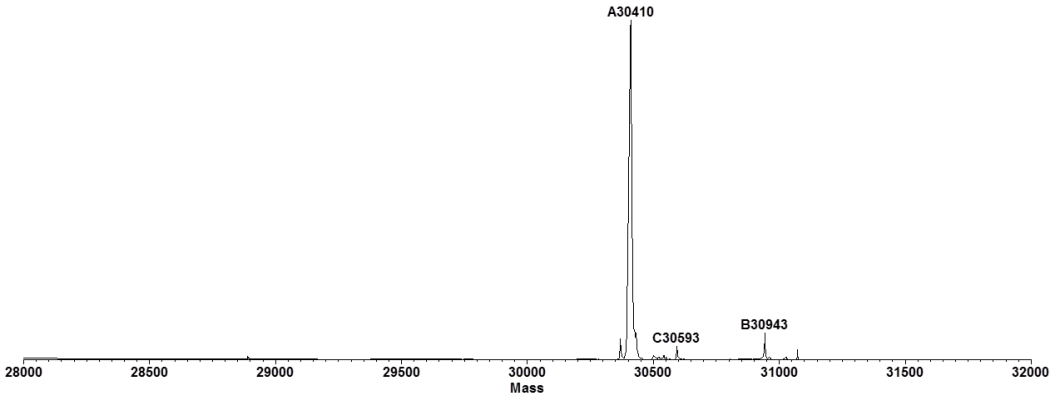


Figure 3-7. a-1) Deconvoluted whole protein mass with EGFP + $tRNA^{EcAla}_{CUA.}$

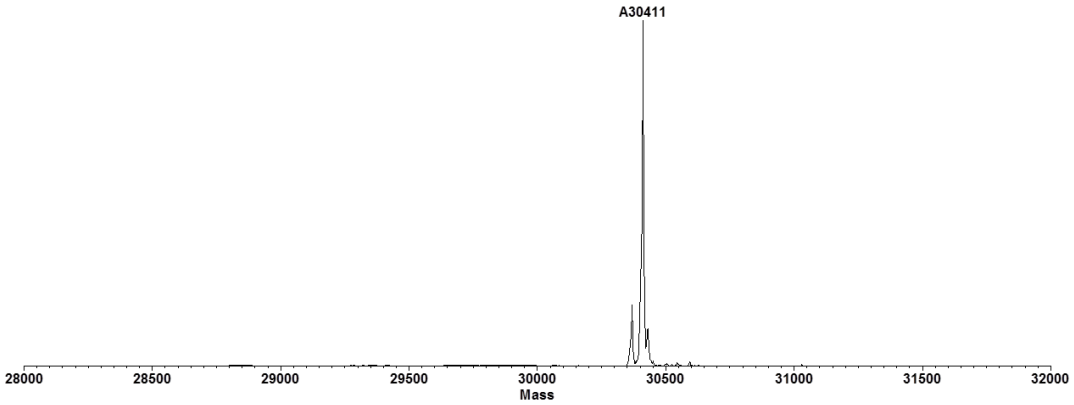


Figure 3-7. a-2) Deconvoluted whole protein mass with EGFP + $tRNA^{EcAla}_{CUA}$ +EcAlaRS.

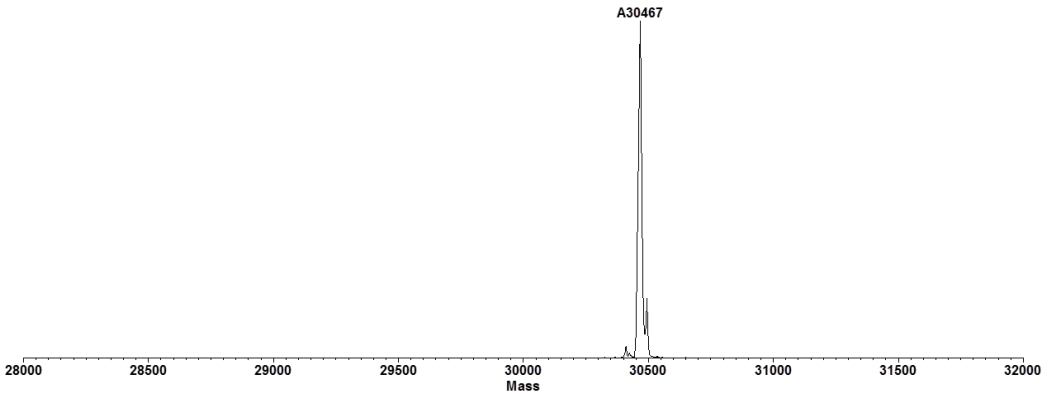


Figure 3-7. b-1) Deconvoluted whole protein mass with EGFP + $tRNA^{EcArg}_{CUA.}$

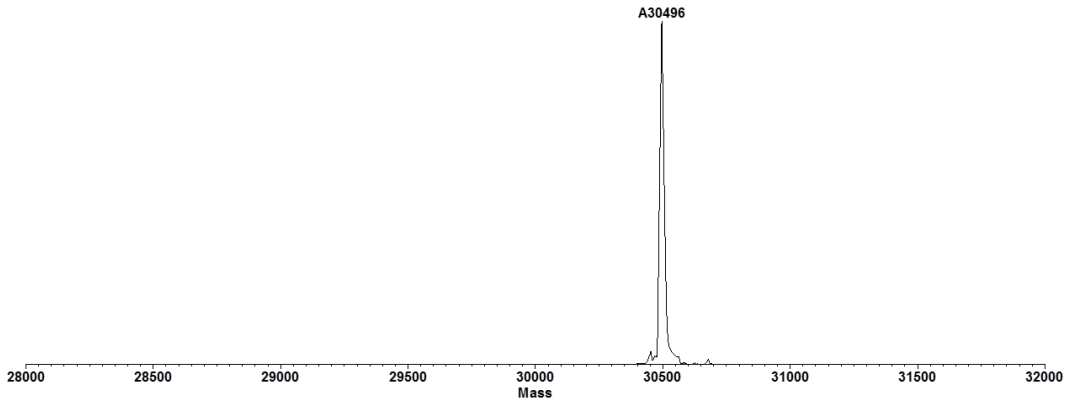


Figure 3-7. b-2) Deconvoluted whole protein mass with EGFP + $tRNA^{EcArg}_{CUA.}$ +EcArgRS.

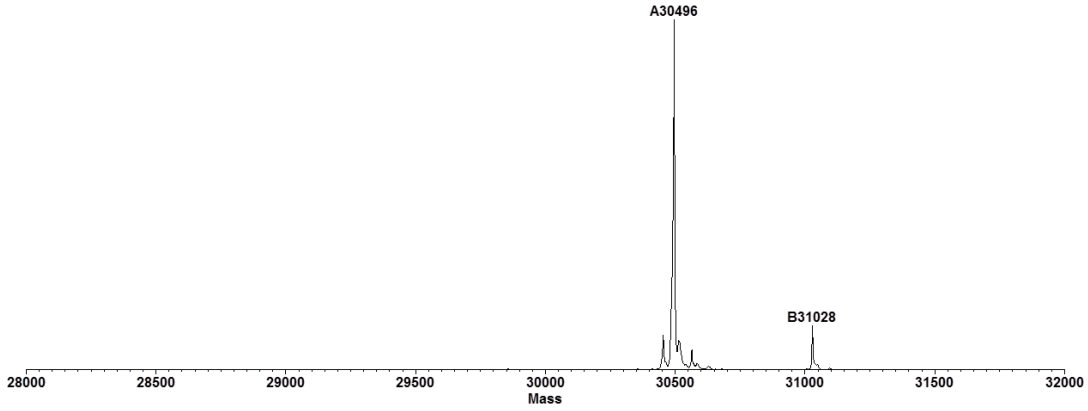


Figure 3-7. b-3) Deconvoluted whole protein mass with EGFP + $tRNA^{EcArg}_{UCA}$ +EcArgRS.

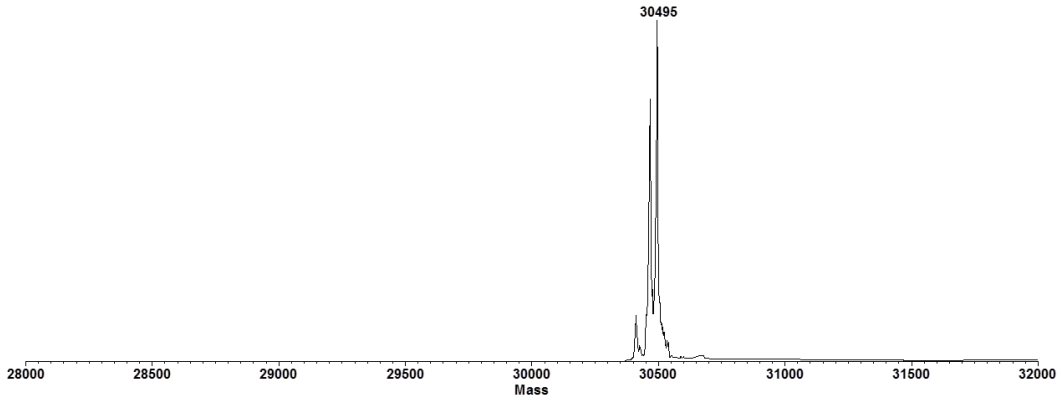


Figure 3-7. c-1) Deconvoluted whole protein mass with EGFP + $tRNA^{EcAsp}_{CUA.}$

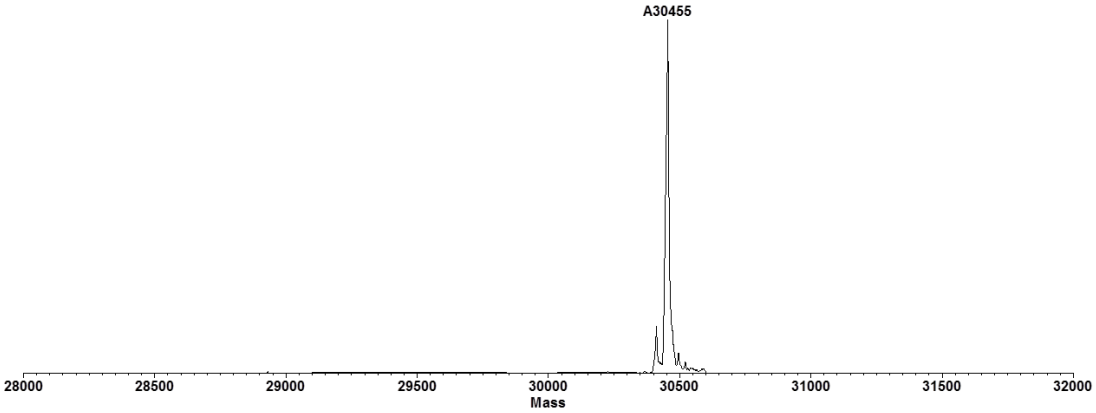


Figure 3-7. c-2) Deconvoluted whole protein mass with EGFP + $tRNA^{EcAsp}_{CUA}$ +EcAspRS.

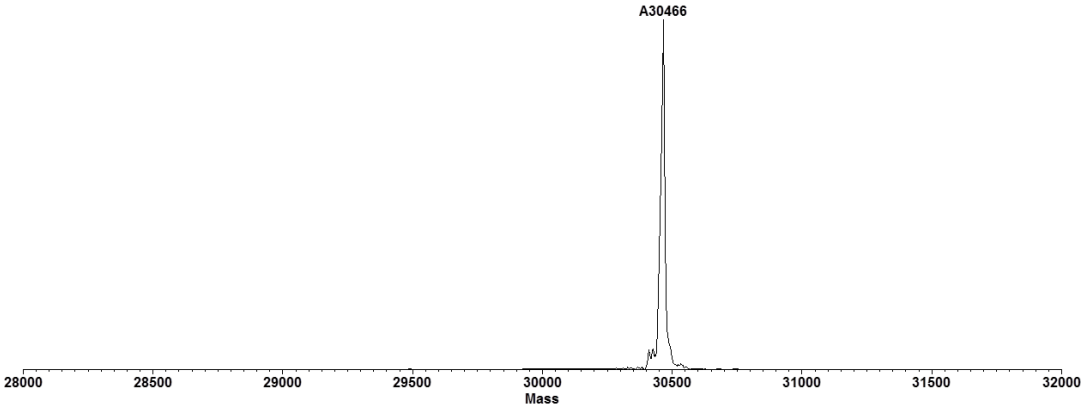


Figure 3-7. d) Deconvoluted whole protein mass with EGFP + $tRNA^{EcGln}_{CUA}$ +EcGlnRS.

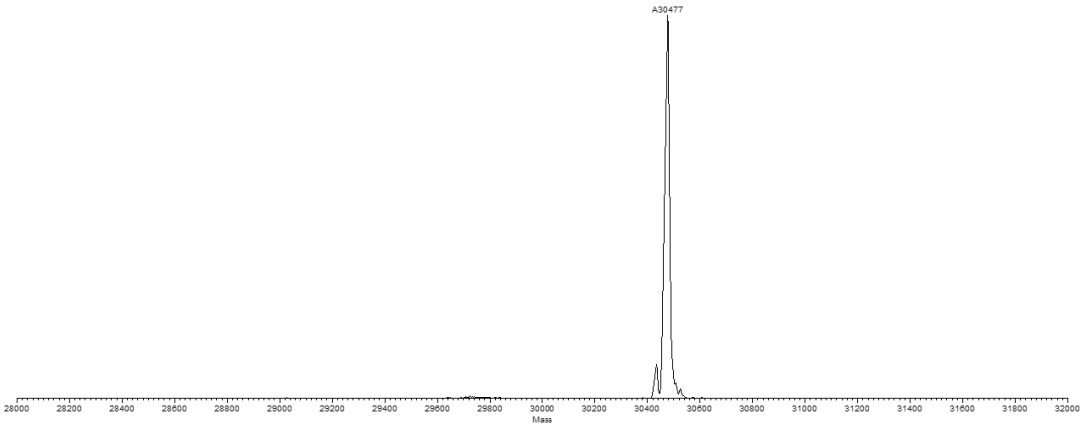


Figure 3-7. e) Deconvoluted whole protein mass with EGFP + $tRNA^{EcHis}_{CUA}$ +EcHisRS.

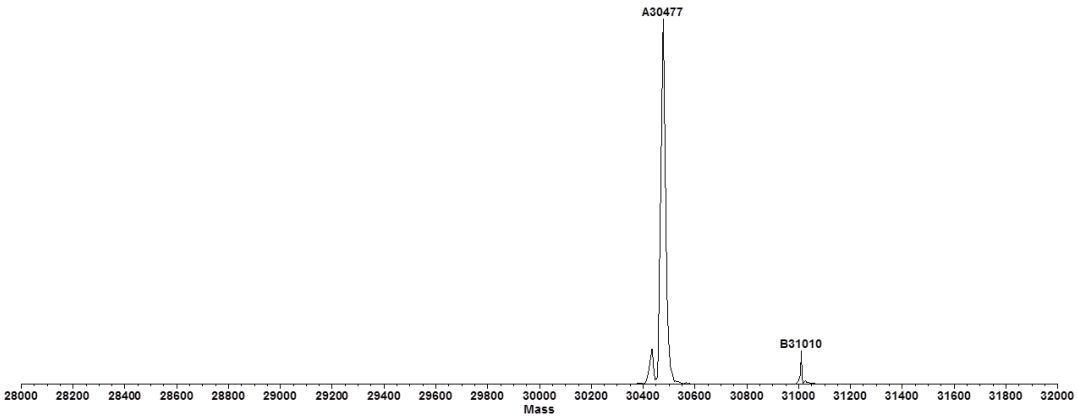


Figure 3-7. F-1) Deconvoluted whole protein mass with EGFP + $tRNA^{EcHisG}_{CUA}$ +EcHisRS.

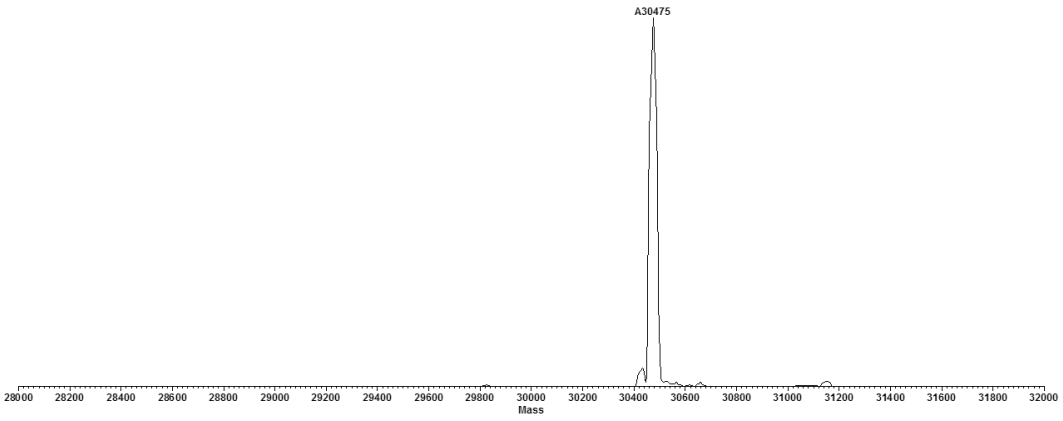


Figure 3-7. F-2) Deconvoluted whole protein mass with EGFP + $tRNA^{EcHisG}_{UCA}$ +EcHisRS.

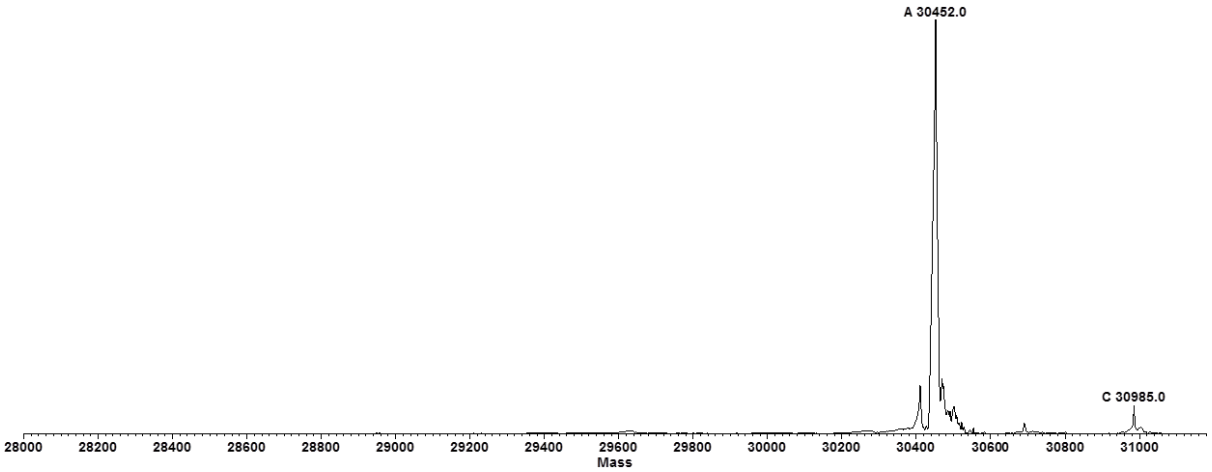


Figure 3-7. g) Deconvoluted whole protein mass with EGFP + $tRNA^{EcLeu}_{CUA}$ +EcLeuRS.

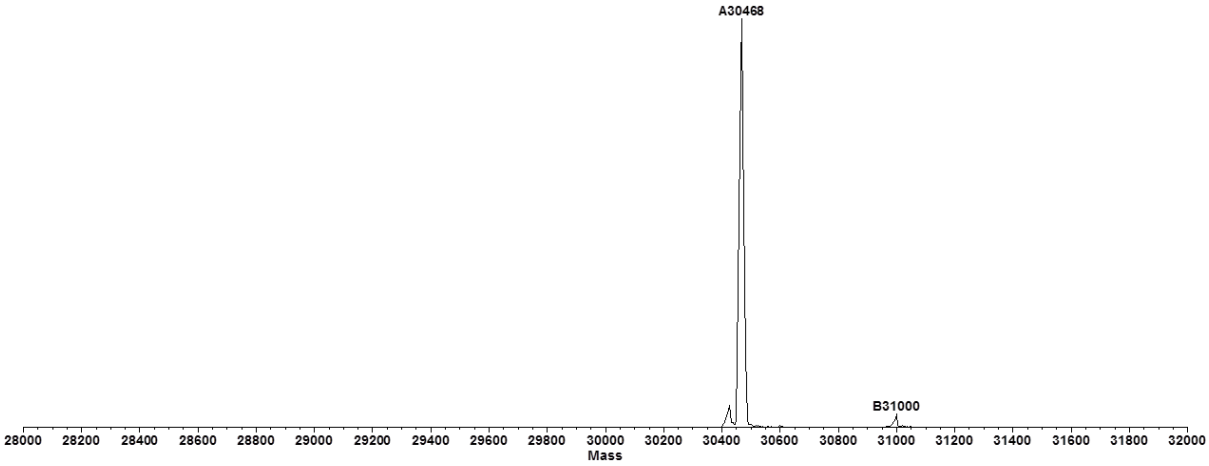


Figure 3-7. h) Deconvoluted whole protein mass with EGFP + $tRNA^{EcLys}_{CUA}$ +EcLysRS.

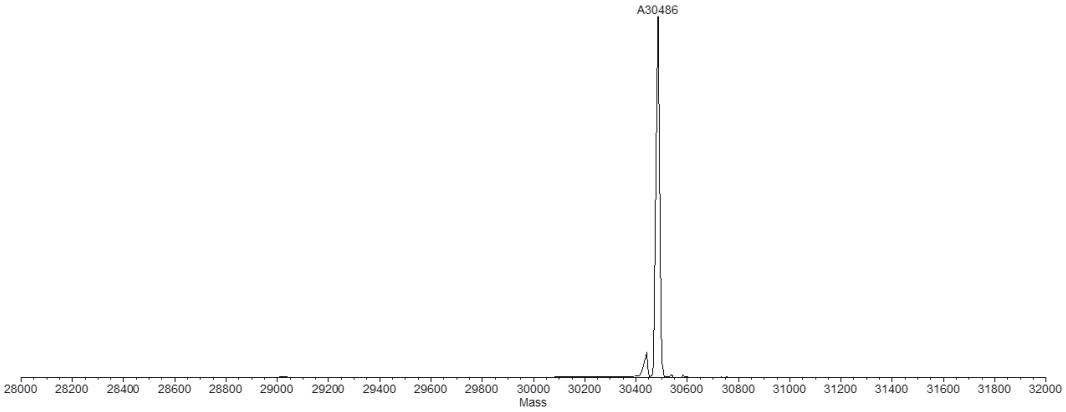


Figure 3-7. i) Deconvoluted whole protein mass with EGFP + $tRNA^{EcPhe}_{CUA}$

+EcPheRS.

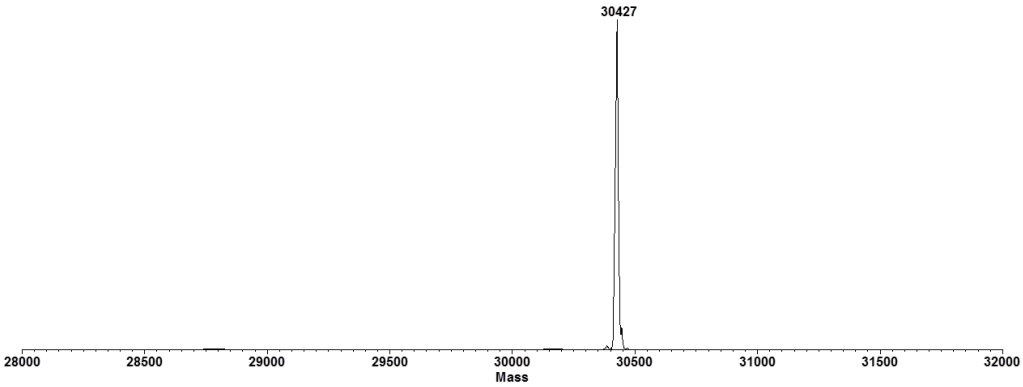


Figure 3-7. j) Deconvoluted whole protein mass with EGFP + $tRNA^{EcSer}_{CUA}$

+EcSerRS.

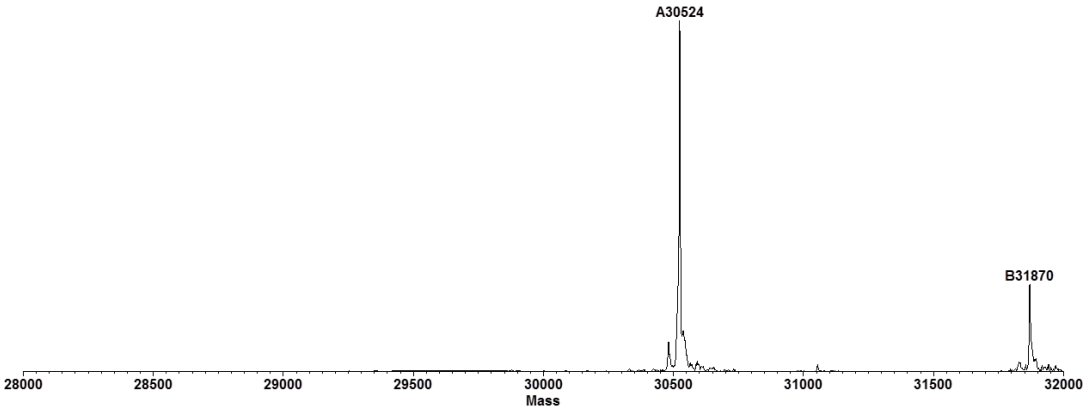


Figure 3-7. k) Deconvoluted whole protein mass with EGFP + $tRNA^{EcTrp}_{CUA}$

+EcTrpRS.

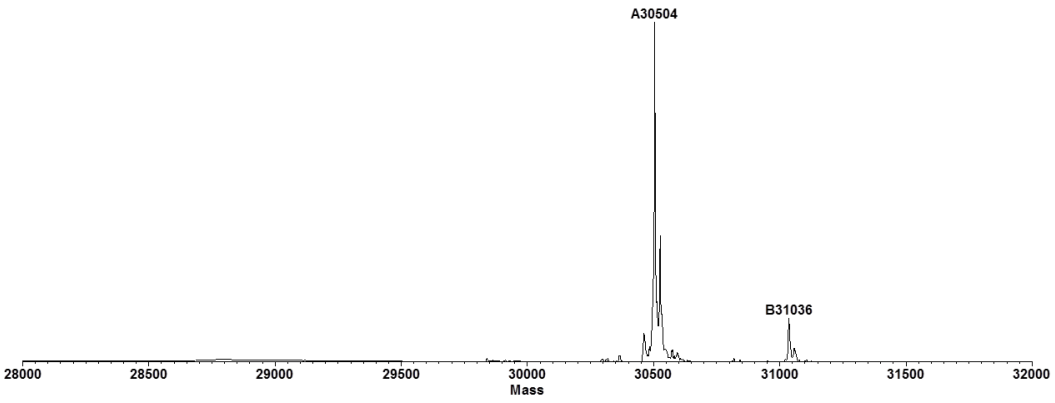


Figure 3-7. l-1) Deconvoluted whole protein mass with EGFP + $tRNA^{EcTyr}_{CUA}$ +EcTyrRS.

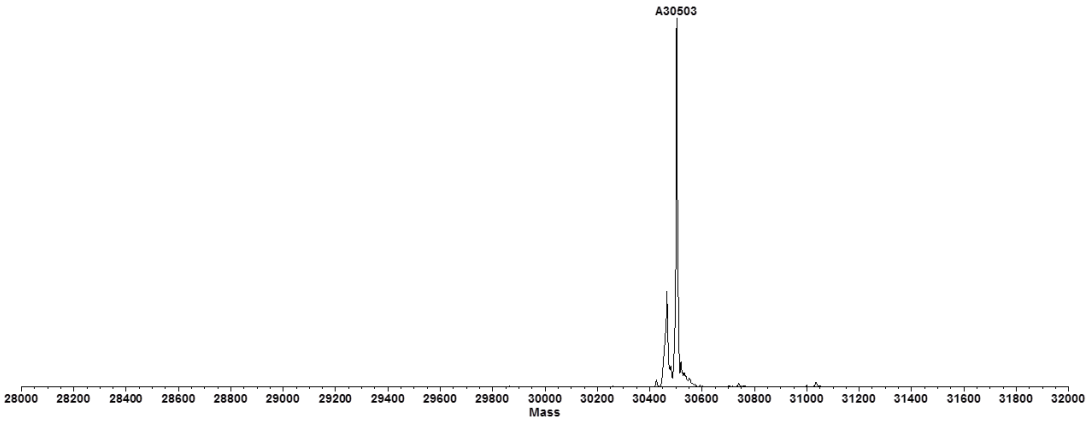


Figure 3-7. l-2) Deconvoluted whole protein mass with EGFP + $tRNA^{EcTyr}_{UCA}$ +EcTyrRS.

Whole protein mass for AzF containing EGFP .

AzF containing EGFP generated by EcPheRS T251G A294G with $tRNA^{Phe}_{CUA}$ were purified and characterized by LC/MS spectrum. (Figure 3-8)

Sequence of EcPheRS α T251G A294G:

ATGTCACATCTCGCAGAACTGGTTGCCAGTGCGAAGGCGGCCATTAGCCA
GGCGTCAGATGTTGCCGCGTTAGATAATGTGCGCGTCGAATATTTGGGTA
AAAAAGGGCACTTAACCCTTCAGATGACGACCCTGCGTGAGCTGCCGCCA
GAAGAGCGTCCGGCAGCTGGTGCGGTTATCAACGAAGCGAAAGAGCAGG
TTCAGCAGGCGCTGAATGCGCGTAAAGCGGAACTGGAAAGCGCTGCACTG
AATGCGCGTCTGGCGGCGGAAACGATTGATGTCTCTCTGCCAGGTCGTCG
CATTGAAAACGGCGGTCTGCATCCGGTTACCCGTACCATCGACCGTATCG
AAAGTTTCTTCGGTGAGCTTGGCTTTACCGTGGCAACCGGGCCGGAAATC
GAAGACGATTATCATAACTTCGATGCTCTGAACATTCCTGGTCACCACCCG
GCGCGCGCTGACCACGACACTTTCTGGTTTGACACTACCCGCCTGCTGCGT
ACCCAGACCTCTGGCGTACAGATCCGCACCATGAAAGCCCAGCAGCCACC

GATTCGTATCATCGCGCCTGGCCGTGTTTATCGTAACGACTACGACCAGAC
TCACACGCCGATGTTCCATCAGATGGAAGGTCTGATTGTTGATACCAACA
TCAGCTTTACCAACCTGAAAGGCACGCTGCACGACTTCCTGCGTAACTTCT
TTGAGGAAGATTTGCAGATTCGCTTCCGTCCTTCCTACTTCCCGTTTGGCG
AACCTTCTGCAGAAGTGGACGTCATGGGTAAAAACGGTAAATGGCTGGAA
GTGCTGGGCTGCGGGATGGTGCATCCGAACGTGTTGCGTAACGTTGGCAT
CGACCCGGAAGTTTACTCTGGTTTCGGCTTCGGGATGGGGATGGAGCGTC
TGACTATGTTGCGTTACGGCGTCACCGACCTGCGTTCATTCTTCGAAAACG
ATCTGCGTTTCCTCAAACAGTTTAAATAA

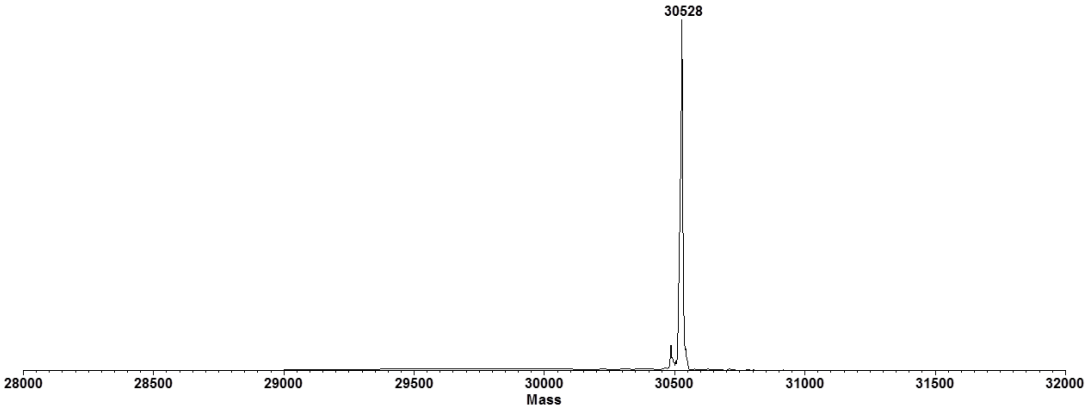


Figure 3-8. Whole protein mass for AzF containing EGFP.

Click chemistry with EGFP based AzF.

Click chemistry was performed with EGFP based AzF with DBCO-4PEG-TAMRA and BCN-OH. The conjugation was characterized by fluorescent SDS-PAGE and whole protein mass respectively.

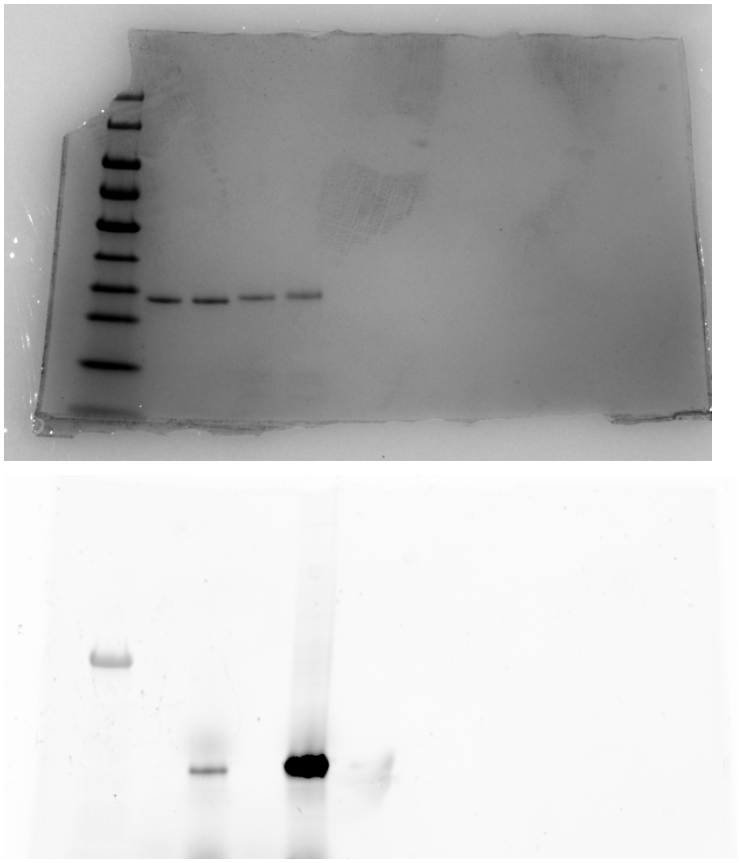


Figure 3-9 Unmodified gel SDS-PAGE picture with Coomassie stain and same gel under fluorescent channel. The lane 1 is protein molecular weight ladder, with from bottom to top in order of: 10, 15, 25, 35,45, 50, 70, 110, 180, and 250 kD. The EGFP protein bands is migrated between 25 and 35kD. The fluorescent signal in bottom picture is from 70kD band. From Lane 2 to Lane 5 are EGFPWT+DMSO, EGFPWT+DBCO-TAMRA, EGFP39AzF+DMSO and EGFP39AxF+DBCO-TAMRA.

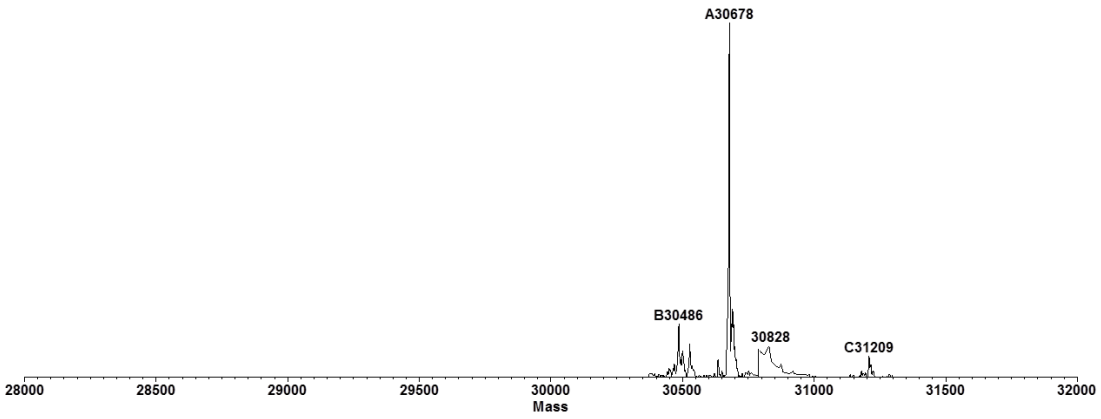


Figure 3-10.Whole protein mass of EGFP39AzF conjugated with BCN-OH. The 30678 peak is agreed with expected mass for conjugated product. The 30486 peak is coming from Phe incorporated by this PheRS T251G A294G.

3.6 Reference:

1. Dunkelmann, D.L., et al., *Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids.* Nat Chem, 2020. **12**(6): p. 535-544.
2. Willis, J.C.W. and J.W. Chin, *Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs.* Nature Chemistry, 2018. **10**(8): p. 831-837.
3. Ding, W., et al., *Chimeric design of pyrrolysyl-tRNA synthetase/tRNA pairs and canonical synthetase/tRNA pairs for genetic code expansion.* Nat Commun, 2020. **11**(1): p. 3154.
4. Italia, J.S., et al., *An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes.* Nature Chemical Biology, 2017. **13**(4): p. 446-450.
5. Kelemen, R.E., et al., *Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells.* bioRxiv, 2022: p. 2022.01.21.477302.
6. Kwon, I. and D.A. Tirrell, *Site-Specific Incorporation of Tryptophan Analogues into Recombinant Proteins in Bacterial Cells.* Journal of the American Chemical Society, 2007. **129**(34): p. 10431-10437.
7. Kwon, I. and S.I. Lim, *Tailoring the Substrate Specificity of Yeast Phenylalanyl-tRNA Synthetase toward a Phenylalanine Analog Using Multiple-Site-Specific Incorporation.* ACS Synthetic Biology, 2015. **4**(5): p. 634-643.

8. Ikemura, T., *Codon usage and tRNA content in unicellular and multicellular organisms.* Mol Biol Evol, 1985. **2**(1): p. 13-34.
9. Chan, P.P., et al., *tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes.* Nucleic Acids Res, 2021. **49**(16): p. 9077-9096.
10. Schmied, W.H., et al., *Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1.* Journal of the American Chemical Society, 2014. **136**(44): p. 15577-15583.
11. Anderson, J.C. and P.G. Schultz, *Adaptation of an Orthogonal Archaeal Leucyl-tRNA and Synthetase Pair for Four-base, Amber, and Opal Suppression.* Biochemistry, 2003. **42**(32): p. 9598-9608.
12. Chatterjee, A., et al., *A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli.* Biochemistry, 2013. **52**(10): p. 1828-1837.
13. Yoon, Y.G. and M.D. Koob, *Transformation of isolated mammalian mitochondria by bacterial conjugation.* Nucleic Acids Research, 2005. **33**(16): p. e139-e139.
14. Sprinzl, M., et al., *Compilation of tRNA sequences.* Nucleic acids research, 1980. **8**(1): p. r1-r22.
15. Yokobori, S. and S. Pääbo, *Transfer RNA editing in land snail mitochondria.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(22): p. 10432-10435.
16. Fromant, M., P. Plateau, and S. Blanquet, *Function of the Extra 5'-Phosphate Carried by Histidine tRNA.* Biochemistry, 2000. **39**(14): p. 4062-4067.
17. Yan, W. and C. Francklyn, *Cytosine 73 is a discriminator nucleotide in vivo for histidyl-tRNA in Escherichia coli.* Journal of Biological Chemistry, 1994. **269**(13): p. 10022-10027.
18. Cooley, L., B. Appel, and D. Söll, *Post-transcriptional nucleotide addition is responsible for the formation of the 5' terminus of histidine tRNA.* Proceedings of the National Academy of Sciences, 1982. **79**(21): p. 6475-6479.
19. Gu, W., et al., *Depletion of Saccharomyces cerevisiae tRNA(His) guanylyltransferase Thg1p leads to uncharged tRNAHis with additional m(5)C.* Molecular and cellular biology, 2005. **25**(18): p. 8191-8201.
20. Hyde Samantha, J., et al., *tRNAHis guanylyltransferase (THG1), a unique 3'-5' nucleotidyl transferase, shares unexpected structural homology with canonical 5'-3' DNA polymerases.* Proceedings of the National Academy of Sciences, 2010. **107**(47): p. 20305-20310.
21. Jackman, J.E. and E.M. Phizicky, *Identification of Critical Residues for G−1 Addition and Substrate Recognition by tRNAHis Guanylyltransferase.* Biochemistry, 2008. **47**(16): p. 4817-4825.

22. Jackman, J.E. and E.M. Phizicky, *tRNAHis guanylyltransferase adds G-1 to the 5' end of tRNAHis by recognition of the anticodon, one of several features unexpectedly shared with tRNA synthetases.* RNA (New York, N.Y.), 2006. **12**(6): p. 1007-1014.
23. Zheng, Y., et al., *Expanding the Scope of Single- and Double-Noncanonical Amino Acid Mutagenesis in Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases.* Biochemistry, 2018. **57**(4): p. 441-445.
24. Nameki, N., et al., *Identity elements of Saccharomyces cerevisiae tRNA(His).* Nucleic acids research, 1995. **23**(3): p. 389-394.
25. Schulman, L.H. and H. Pelka, *In vitro conversion of a methionine to a glutamine-acceptor tRNA.* Biochemistry, 1985. **24**(25): p. 7309-14.
26. Italia, J.S., et al., *Genetically encoded protein sulfation in mammalian cells.* Nat Chem Biol, 2020. **16**(4): p. 379-382.
27. Varshney, U., C.P. Lee, and U.L. RajBhandary, *Direct analysis of aminoacylation levels of tRNAs in vivo. Application to studying recognition of Escherichia coli initiator tRNA mutants by glutaminyl-tRNA synthetase.* J Biol Chem, 1991. **266**(36): p. 24712-8.
28. Shimada, A., et al., *Structural and mutational studies of the recognition of the arginine tRNA-specific major identity element, A20, by arginyl-tRNA synthetase.* Proceedings of the National Academy of Sciences, 2001. **98**(24): p. 13537-13542.
29. Mermershtain, I., et al., *Idiosyncrasy and identity in the prokaryotic phe-system: Crystal structure of E. coli phenylalanyl-tRNA synthetase complexed with phenylalanine and AMP.* Protein Science, 2011. **20**(1): p. 160-167.
30. Klipcan, L., et al., *The tRNA-Induced Conformational Activation of Human Mitochondrial Phenylalanyl-tRNA Synthetase.* Structure, 2008. **16**(7): p. 1095-1104.
31. Kirshenbaum, K., I.S. Carrico, and D.A. Tirrell, *Biosynthesis of Proteins Incorporating a Versatile Set of Phenylalanine Analogues.* ChemBioChem, 2002. **3**(2-3): p. 235-237.
32. Datta, D., et al., *A Designed Phenylalanyl-tRNA Synthetase Variant Allows Efficient in Vivo Incorporation of Aryl Ketone Functionality into Proteins.* Journal of the American Chemical Society, 2002. **124**(20): p. 5652-5653.
33. Italia, J.S., et al., *Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes.* Cell Chem Biol, 2018. **25**(10): p. 1304-1312.e5.
34. Mondal, S., et al., *Site-specific incorporation of citrulline into proteins in mammalian cells.* Nature Communications, 2021. **12**(1): p. 45.
35. Xiao, H., et al., *Genetic Incorporation of Histidine Derivatives Using an Engineered Pyrrolysyl-tRNA Synthetase.* ACS Chemical Biology, 2014. **9**(5): p. 1092-1096.

36. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon.* Proceedings of the National Academy of Sciences, 2004. **101**(20): p. 7566-7571.
37. Chatterjee, A., et al., *A bacterial strain with a unique quadruplet codon specifying non-native amino acids.* Chembiochem, 2014. **15**(12): p. 1782-6.
38. Zheng, Y., et al., *Virus-Enabled Optimization and Delivery of the Genetic Machinery for Efficient Unnatural Amino Acid Mutagenesis in Mammalian Cells and Tissues.* ACS Synthetic Biology, 2017. **6**(1): p. 13-18.

Chapter 4. Creating a Pre-ATMK for screening orthogonal $tRNA^{Lys}$/LysRS pairs in *E. coli*

4.1 Introduction

The acetylation and methylation of lysine residues on histones was first discovered about five decades ago, revealing the interplay between acetylation and methylation in gene regulation [1] The regulatory roles of post-translational modifications (PTM) are not just limited to dynamic gene regulation, but also extended to protein expression regulation. For instance, the discovery of SUMOylation of lysine residues revealed a novel protein degradation process.[2] Subsequently, with increasing attention on lysine PTMs and advancements in mass spectrometry, more than 100,000 sites of Lys modifications in over 10,000 proteins have been mapped.[3] PTMs such as propionylation (KPr), butylation (Kbu), crotonylation (Kcro), malonylation (Kmal), succinylation (Ksucc), glutarylation (Kglu), β-hydroxybutylation (KBhb), 2-hydroxyisobutyrylation (Khib), lactylation (Klac) and benzoylation (KBz) have been reported recently, all with different levels of understanding of their function and regulation.[4-6]

To dissect the biological function of post-translationally modified lysine residues (PTM-Lys), as discussed in chapter 1, obtaining a homogenous sample of the protein of interest containing the PTM is critical. Directly extracting the modified protein from cells normally results in a sample containing heterogeneous modification of the desired site, as well as mixed modification patterns on other sites. To address this challenge, chemical synthesis of small peptides which contain the modified residue, ncAA

mimicking, GCE and other methods have been used. Specifically, for lysine PTMs, chemical synthesis has been used to generate Tau with K280 acetylation; a succinyl lysine analog was generated by selective thiol-ene reaction with cysteine to study Ksucc at H2BK34; acetyl lysine was incorporated into Ran within mammalian cells using the PylRS/tRNA pair to study the function of acetylation of K71. [7-10] Despite extensive effort, some important lysine-PTMs remain challenging to incorporate directly. For instance, methylated lysine PTMs can be only incorporated by solid phase peptide synthesis, which limits the protein targets by size and folding accuracy.[11-13] Dimethyl-lysine has been incorporated using the PylRS/tRNA pair. However, since the binding pocket of PylRS does not uptake dimethyl-lysine, the incorporation could only be achieved through incorporation of a precursor of dimethyl-lysine, which then require several steps of reaction workup.[14] The advantage of using GCE technology is the ability to produce site-specifically modified, natively folded protein in large quantities regardless of protein size. [15] Therefore, this indirect incorporation limits the utility of the GCE technology. There is a pressing need to develop platforms which can provide access to currently unavailable lysine modifications, to fully exploit the potential of GCE in PTM studies.

With the previously obtained knowledge described in Chapter 3, we concluded that the *E. coli* $tRNA^{Lys}$/aaRS has the potential to be evolved for use in mammalian systems. In this chapter, I explore the possibility of designing an ATMK strain for evolving the EcLysRS to incorporate lysine-PTMs and other ncAAs in mammalian cells.

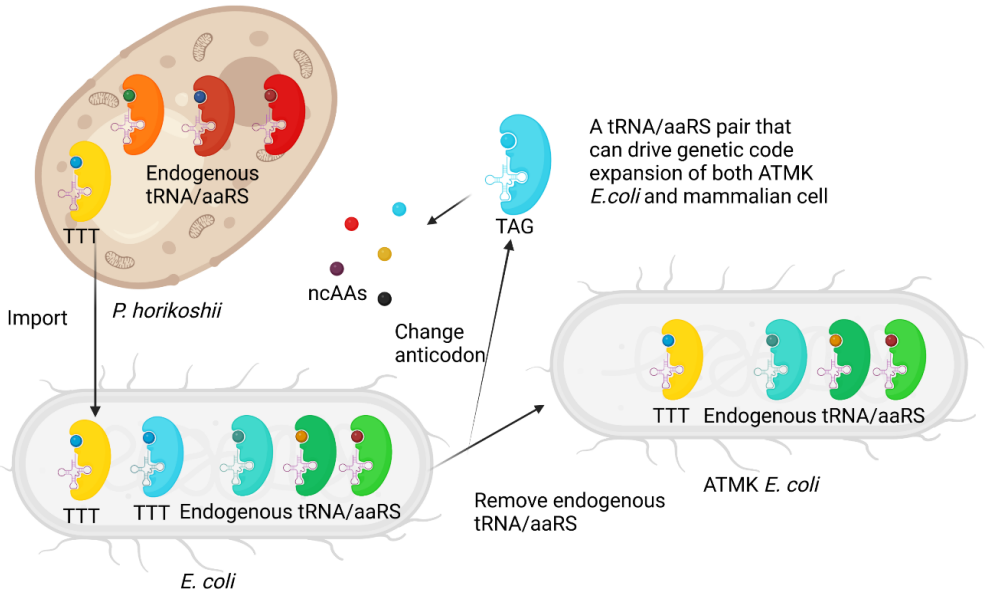


Figure 4-1. Design of ATMK for directed evolution of EcArgRS.

4.2 Result

4.2.1 Blueprint for ATMK strain

The core principle behind building an altered translation machinery lysine strain (ATMK) is to liberate the EcLysRS/tRNA pair in *E. coli* by replacing it with an orthogonal pair, therefore granting EcLysRS/tRNA orthogonality in ATMK for subsequent directed evolution. Specifically, on the genomic level, the ATMK strain would not contain any EcLysRS and $tRNA^{Lys}$. Meanwhile, an imported pair which can

functionally replace EcLysRS/tRNA is introduced via an inducible plasmid. (Figure 4-1) There are several critical steps involved in constructing an ATMK strain. The first step is building an *E. coli* strain which can be used to test candidate pairs for functional replacement, named pre-ATMK. This strain is constructed by knocking out the endogenous *E. coli* LysRS, which disables the pre-ATMK strain's ability to suppress lysine codons. Thus, pre-ATMK is unable to synthesize proteins unless the imported pair compensates for the missing function. To first propagate this strain, the *E. coli* LysRS is re-introduced to the cells on an inducible plasmid, to ensure the strain's viability without an imported pair. The second step for construction of an ATMK strain is to test each potential imported pair in the pre-ATMK strain. The imported pair must be orthogonal in *E. coli* and actively compensate for the loss of EcLysRS. Once these two challenges are overcome, subsequently the endogenous $tRNA^{Lys}$s can be depleted to make the ATMK strain.

Creation of the ATMK strain is the first step for creating the new pair, however, there remains potential challenges in evolving *E. coli* LysRS in this engineered strain. To evolve LysRS in *E. coli*, the corresponding tRNA must suppress an unassigned codon, commonly TAG or TGA. In some cases, this mutated tRNA loses orthogonality if the critical identity elements of the tRNA are altered. This issue can be resolved by

switching to a different codon or occupying the cross-reactive aaRS by upregulating the corresponding tRNA.[16, 17]

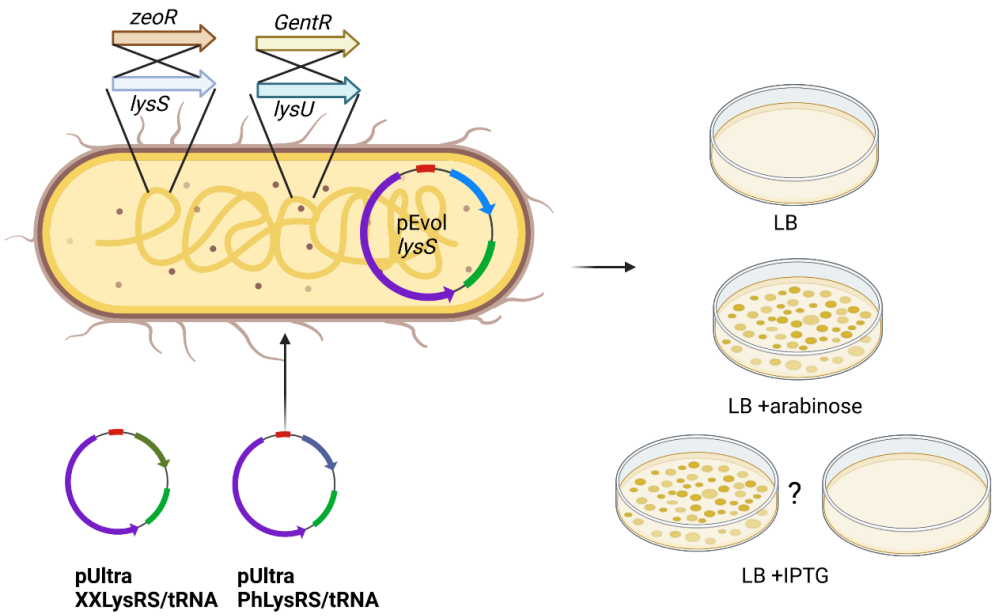


Figure 4-2. Design of Pre-ATMK for screening orthogonal LysRS/tRNA pair for replacing loss of EcArgRS function.

4.2.2 Knocking out LysRS

To construct the pre-ATMK strain, we replaced LysRS with an antibiotic resistance gene. Unlike the aaRSs for other amino acids, there are two LysRS genes in *E. coli* encoded by *lysS* and *lysU*. The two aaRSs share 88% homology, indicating the shared origin and function of these aaRSs. While *lysS* is constitutively expressed, *lysU* is induced by heat stress.[18] To fully remove LysRS function, *lysS* and *lysU* were sequentially replaced by ZeoR and GentR, respectively. The genome editing was enabled by λ-Red recombination, which was carried out via introduction of a pKD46

plasmid.[19] The arabinose-inducible pKD46 plasmid is rejected by the cell when the incubation temperature is above 30 ºC, allowing a conditional turn on and removal of the λ-Red system upon completion of genome editing. We performed colony PCR with primers located upstream and within the ORF of the resistance gene to verify the successful recombination and location of genome editing. (Figure 4-3) Upon the knockout of *lysS* and *lysU*, the survival of the strain is dependent on inducible expression of LysRS introduced by a pEvol plasmid.

The purpose of the pre-ATMK strain is to serve as a screening platform for potential imported pairs. For successful screening, the viability of the engineered cell should be tightly dependent on the function of imported pairs. To test this, we performed a mock screening by observing the growth of the pre-ATMK strain with or without inducing expression of EcLysRS. Unsurprisingly, knockout of *lysU* does not affect the viability of the strain. Likewise, knocking out only *lysS* does not lead to lethality, implying that *lysU* compensates for the loss of *lysS*. However, when both *lysU* and *lysS* are knocked out, the pre-ATMK strain's viability is dependent on the induction of the pEvol plasmid. Thus, the pre-ATMK strain functions as expected and can be used as a screening platform for imported pairs.

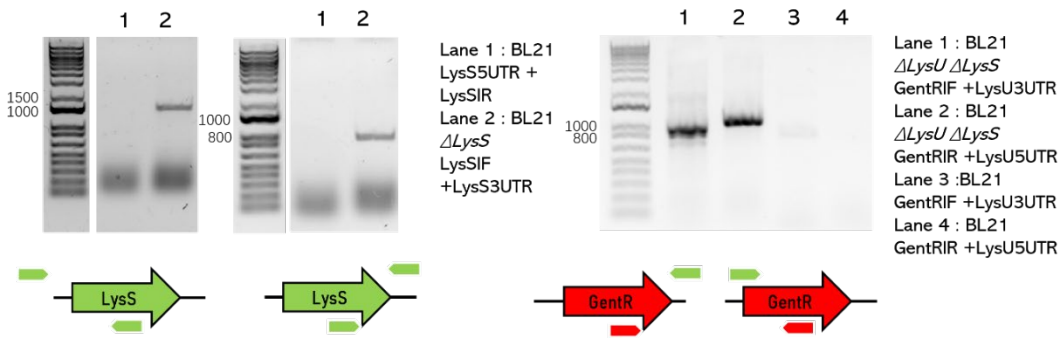


Figure 4-3. Colony PCR product on agarose gel. The band size matches expectation indicates success replace target gene with resistant gene.

4.2.3 Mock screening with EcLysRS in Pre-ATMK strain

The screening of pre-ATMK is designed to be a viability test by plating cells on selective media. The pre-ATMK strain lacks the ability to synthesize proteins without expression of a functional LysRS introduced by a plasmid. The positive control EcLysRS encoded by *lysS* is driven by an *araB* promoter, induced by arabinose, on a pEvol plasmid. Meanwhile, the LysRS being tested, EcLysRS in this mock screening, is driven by a *tacI* promoter, induced by IPTG, on a pUltra plasmid. By supplying arabinose or IPTG in the media, the positive control aaRS or testing aaRS are expressed, respectively. The mock screening results demonstrated that the pre-ATMK strain grows on agar plates in an arabinose-dependent manner and the growth intensity has a positive correlation with arabinose concentration. In addition, the pre-ATMK strain grows with addition of IPTG in the absence of arabinose, indicating successful mock screening. (Figure 4)

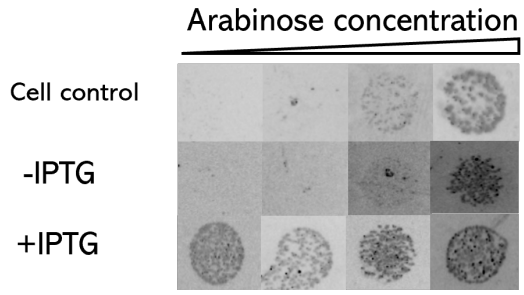


Figure 4-4. Mock screening with pUltra EcLys. Cell control is Pre-ATMK carries only pEvol EcLysRS. The testing strain is Pre-ATMK carries pEvol EcLysRS and pUltra EcLysRS. IPTG concentration is 1 mM and the arabinose concentration is 0%, 0.02%, 0.2% and 2%.

4.2.4 Imported LysRS selection

Given limited resources, we cannot test all the possible aaRS/tRNA pairs from archaeal and eukaryotic organisms to identify a LysRS/tRNA pair that is orthogonal to *E. coli.* Inspired by utilization of the MjTyr pair in the *E. coli* system, we first looked for a reported suppressor in *E. coli* that is based on the LysRS/tRNA pair, although with a modified anticodon loop the orthogonal suppressor pair is not guaranteed to be orthogonal when used as a wild type pair. In this chapter, we investigated *P. horikoshii* tRNA and aaRS which was previously reported to be orthogonal in *E. coli* with a quadruplet codon.[20]

Besides searching for documented orthogonal tRNAs, we can also utilize bioinformatic tools. Utilizing advancements in the bioinformatics field can avoid repetitive labor by greatly narrowing down the candidate pairs to be tested. To address

the issue of tRNA/aaRS orthogonality in *E. coli*, Chin's group designed a bioinformatic tool named tREX. The chemical space of tRNAs is relatively limited due to conserved structures and building blocks, as most tRNAs are distinguished by important identity elements. The tREX system generates a score based on the identical elements and significant structural differences between candidate and endogenous *E. coli* tRNAs. These computationally calculated scores are used to yield candidates which are likely orthogonal for subsequent wet lab variation. [21] For our aaRS of interest, LysRS, unfortunately tREX was not able to yield any candidate tRNAs.

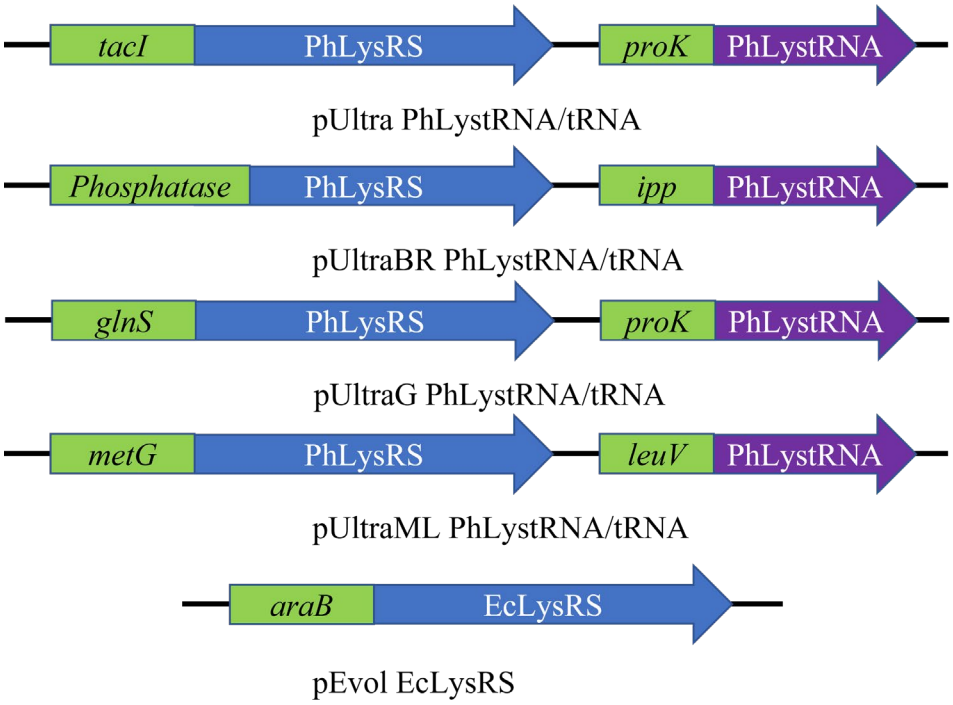


Figure 4-5. Key component of pEvol and pUltra plasmid used in this chapter. Positive control versions of pUltra with EcLysRS contains no tRNA cassette.

4.2.5 Test PhtRNA^{Lys}/aaRS

The *P. horikoshii* tRNA is proposed and reported to be orthogonal in *E. coli* due to the conserved A73 in tRNA^{Lys} in prokaryotes versus G73 in archaea. [20, 22] The PhLysRS is reported to be toxic when constitutively expressed in *E. coli*, but the toxicity can be minimized by truncation of the PhLysRS after residue S357.[20] We generated two versions of a pUltra plasmid containing either full length PhLysRS/tRNA$_{GAA}$ or ΔPhLysRS/tRNA$_{GAA}$. The PhLysRS in the pUltra plasmid is driven by a *tacI* promoter, which is induced by IPTG, whereas the EcLysRS in pEvol is driven by an *araB* promoter which is induced by arabinose. The strain was screened by plating on LB with addition of arabinose or IPTG as a positive control or for testing, respectively. The results showed that the strain grows up when only arabinose is added to the media but fail to grow up when induced by either IPTG alone or IPTG and arabinose. (Figure 4-6A) Growth in the presence of arabinose but death when both arabinose and IPTG are added indicates that the ΔPhLysRS is toxic to *E. coli* host cells. To verify this hypothesis, pUltra ΔPhLysRS/tRNA$_{GAA}$ was transformed into wild type T10 *E. coli* cells and plated on LB containing IPTG. These results agreed with the observation in pre-ATMK, that the addition of IPTG at varying concentrations results

in cell death. (Figure 4-6B).

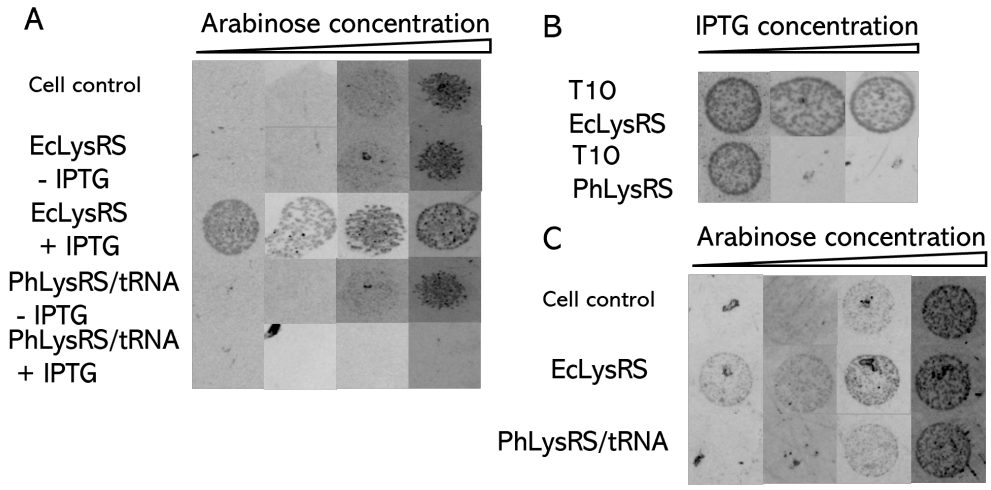


Figure 4-6. A) Testing PhLysRS/tRNA pair in Pre-ATMK strain. Arabinose concentration is 0%, 0.02%, 0.2%, and 2%. IPTG concentration is 1 mM. B) Testing toxicity in PhLysRS/tRNA pair in T10 with different concentration of IPTG :0 mM, 0.1 mM and 1 mM. C) Testing PhLysRS/tRNA in pUltraG construct.

To reduce the toxicity of the ΔPhLysRS/tRNA pair, we rationalized that a lower expression level of ΔPhLysRS may minimize the toxicity. Thus, we created a new pUltra construct, switching the strong *tacI* promoter to a constitutive weak promoter, namely *glnS*. The new plasmid is named pUltraG and was first tested with EcLysRS. Unfortunately, the *glnS* promoter is weak, that even with EcLysRS, pUltraG cannot barely compensate for the loss of function of genomic LysRS (Figure 4-6C) Therefore, ΔPhLysRS/tRNA was not tested using the pUltraG plasmid. We also tested two other promoters, namely *metG* promoter in pUltraML and ribulose phosphate promoter in pUltraBR. Unfortunately, neither of these two constitutive promoters is strong enough to express enough protein to compensate the lost function even if with WT EcLysRS.

4.3 Conclusion and discussion

In this chapter, we explored the possibility of building an ATMK strain as a platform for EcLysRS directed evolution. The motivation of this project was rooted in the need for more GCE tools to use in mammalian systems. The EcLysRS/tRNA pair is a feasible candidate due to the orthogonality of the tRNA in mammalian cells, small number of assigned codons and tRNA copies, and potential to access novel ncAAs. The pre-ATMK strain was successfully generated to screen orthogonal imported pairs and ΔPhlysRS was a reasonable candidate, due to its previous success as a quadruplet codon suppressor in *E. coli*. However, ΔPhlysRS paired with its wild type anti-codon tRNA seemed to amplify the toxicity in *E. coli,* which resulted in an unsuccessful attempt. However, in this work, we only tried one potential pair from archaea and checked bioinformatics database from tREX. It is possible that a broader search for a foreign lysyl pair can reveal a candidate that passes the screening in the pre-ATMK strain.

The ATM system is a platform with the advantage of evolving an aaRS/tRNA pair that works in both mammalian cells and *E. coli* cells, yet it is not the only possible path to evolving the EcLysRS pair for use in mammalian systems. For example, the EcLeuRS was first evolved in yeast. Besides directly evolve EcLysRS in the eukaryotic system, the Terrier group has developed a system to evolve an *E. coli* pair in *E. coli* without solving the orthogonality issue.[23, 24] Furthermore, building a directed evolution platform in a mammalian system for universal ncAA candidates can be

challenging, but designing a selection scheme for specific ncAAs can be achieved by modifying the previously developed VADER system.[25]

4.4 Method

Gene knock-out by Lambda Red recombination

BL21(DE3) was transformed with pKD46 which containing Lambda red system. To prepare recombination ready competent cell. 20 mL of the pKD46 containing BL21(DE3) was cultured in 30 ºC to OD_{600} reach 0.2 and the lambda Red system is induced by addition of 0.02% arabinose. When OD_{600} reaches 0.55 the cell was washed by ice-cold water by centrifugation for 3 times and then resuspended in 50 μL of ice-cold DDI water.

To remove the *E. coli* lysyl–tRNA synthetase (*lysS*) from this strain, the gene encoding zeocin resistance (*ShBle*) driven by the EM-7 promoter and the CYC1 transcription terminator was PCR amplified using primers lysS-ZeoR-F and lysS-ZeoR-R to generate the PCR product *lysS*::ZeoR. 300 ng of the *lysS*::ZeoR PCR cassette was transformed in the prepared pKD46 induced competent cell, and the resulting strains were plated on LB–Agar plates supplemented with Zeocin. The resulting colonies were screened via colony PCR using lysS5UTR, ZeoRIR, lysS3UTR and ZeoRIF, as well as sequencing these colony PCR products. This strain was named BL21(DE3) *ΔlysS*.

To remove the *E. coli* lysyl–tRNA synthetase (*lysU*) from BL21(DE3), the pEvol EcLysRS was transformed in BL21(DE3) *ΔlysS* to compensate the lost function of both EcLysRS. The BL21 (DE3) *ΔlysS* containing pKD46 and pEvol EcLysRS were prepared into competent cell as described above. The *lysU::*GentR PCR cassette was amplified using lysU-Gent-F and lysU.Gent-R. 300 ng of *lysU*::GentR PCR cassette was transformed into prepared BL21(DE3) *ΔlysS*. Resulting gentamycin-resistant colonies were screened for the desired recombination using colony PCR primers LysU5UTR, GentRIR, GentRIF and LysU3UTR, as well as sequencing of the PCR product. The resulting strain was named Pre-ATMK.

Dot plating on selective plate.

Colony was picked and inoculated in LB with 2% arabinose and proper antibiotics and cultured overnight at 37 ºC with 250RPM as seeding culture. The seeding culture was washed with PBS by centrifugation. The cell density was measured by optical density at 600 nm (OD_{600}), and then diluted to 0.1 with PBS. The sample was serial diluted from OD_{600} 10^{-1} to 10^{-7}. 10 μl of each sample from 10^{-3} to10^{-7} was doted on the plate with selective conditions. The selective plates were incubated in the 37 ºC incubator overnight, then picture of plate was taken and recorded.

4.5 Supplement information

Sequence of $tRNA^{PhLys}_{UCC}$

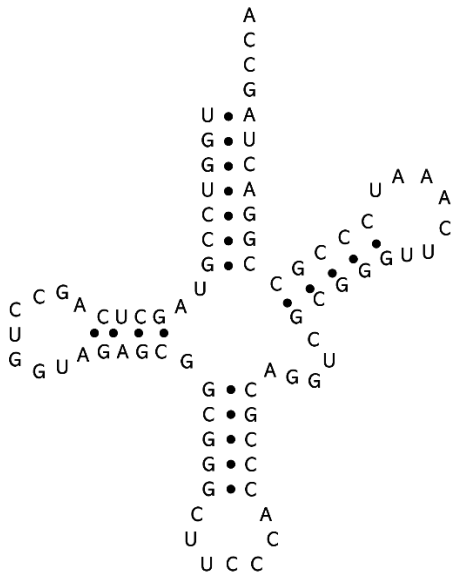


Figure 4-7. Sequence of $tRNA^{PhLys}_{UCC}$ used in this work.

Sequence of PhLysRS and ΔPhLysRS

Full length PhLysRS:

ATGGTACATTGGGCTGATTATATTGCGGATAAGATCATTCGTGAACGCGG
AGAGAAAGAGAAATACGTTGTAGAAAGTGGTATCACTCCGTCTGGTTATG
TACATGTGGGCAATTTCCGTGAGTTATTCACAGCATACATTGTCGGACACG
CGTTACGCGACAAAGGATATGAAGTTCGCCACATTCACATGTGGGATGAC
TATGATCGCTTTCGCAAAGTACCGCGTAATGTCCCGCAAGAGTGGAAAGA
CTACTTAGGCATGCCCATCTCTGAGGTGCCTGACCCTTGGGGGTGCCACG
AGTCTTATGCAGAGCATTTTATGCGTAAATTCGAGGAAGAGGTAGAAAAA
CTGGGCATTGAGGTCGACTTCTTATATGCTTCCGAATTGTACAAGCGCGGT
GAATACTCGGAGGAGATCCGCCTTGCATTTGAAAAGCGTGACAAGATTAT
GGAGATCCTGAATAAATATCGTGAAATCGCCAAACAACCGCCATTGCCAG
AGAACTGGTGGCCAGCCATGGTCTACTGTCCGGAACACCGCCGCGAAGCG
GAAATTATTGAATGGGACGGTGGATGGAAAGTCAAATACAAGTGTCCTGA
AGGCCATGAAGGCTGGGTTGATATTCGTTCGGGCAACGTAAAACTGCGTT
GGCGTGTTGATTGGCCGATGCGTTGGTCACATTTCGGCGTAGATTTTGAGC
CTGCGGGCAAGGACCACCTTGTAGCTGGATCTTCATATGATACTGGAAAG
GAAATCATTAAGGAGGTATACGGGAAGGAAGCTCCACTGAGCCTTATGTA
CGAATTTGTAGGAATCAAAGGACAAAAGGGTAAGATGTCAGGGTCTAAA
GGTAATGTTATTCTTTTGTCGGATCTGTATGAAGTACTGGAGCCAGGGCTT
GTCCGCTTTATCTATGCGCGTCACCGCCCGAACAAGGAGATTAAAATTGA

TTTGGGTTTGGGTATCTTGAATTTGTACGACGAGTTCGATAAAGTAGAGCG
CATCTACTTTGGAGTTGAGGGAGGTAAGGGAGACGACGAGGAACTGCGTC
GTACATATGAATTATCCATGCCAAAGAAACCAGAACGCCTTGTAGCACAA
GCGCCTTTTCGTTTCTTAGCAGTACTGGTGCAGCTGCCCCACTTGACTGAA
GAGGATATCATCAACGTATTGATCAAGCAAGGGCATATTCCACGTGATCT
TAGTAAGGAGGACGTAGAGCGTGTTAAACTTCGCATTAACTTGGCTCGCA
ACTGGGTCAAGAAATATGCACCAGAAGATGTTAAGTTCAGCATTTTAGAA
AAACCTCCCGAAGTTGAGGTCTCCGAAGACGTACGTGAAGCAATGAACGA
AGTAGCCGAGTGGTTGGAAAACCACGAAGAGTTTTCCGTAGAAGAGTTCA
ACAATATTCTGTTTGAGGTCGCGAAGCGTCGCGGAATTTCCAGTCGTGAG
TGGTTTAGCACGCTTTACCGTCTTTTTATCGGGAAAGAACGCGGACCTCGT
CTGGCATCGTTTCTTGCGTCTCTTGATCGTTCGTTCGTAATTAAACGCCTGC
GCCTGGAAGGCTAA

ΔPhLysRS:
ATGGTTCATTGGGCCGATTATATTGCTGATAAAATAATTAGAGAGAGGGG
GGAGAAGGAGAAGTACGTTGTTGAGAGTGGAATAACGCCAAGTGGTTAC
GTTCACGTTGGGAACTTTAGGGAGCTTTTTACAGCTTATATTGTGGGCCAT
GCCCTAAGGGATAAGGGGTATGAGGTTAGGCACATCCACATGTGGGATGA
TTATGATAGATTTAGGAAGGTTCCAAGGAACGTTCCCCAGGAATGGAAAG
ATTACCTGGGAATGCCCATTAGTGAAGTTCCTGATCCCTGGGGATGCCAT
GAGAGTTATGCTGAACACTTCATGAGAAAGTTCGAGGAGGAGGTAGAAA
AATTAGGGATCGAAGTTGACTTTCTTTATGCGAGTGAACTCTACAAGAGA
GGGGAATATTCTGAGGAGATAAGGTTAGCCTTTGAGAAAAGGGATAAGAT
AATGGAGATACTAAACAAGTATAGGGAAATTGCGAAACAACCTCCCCTTC
CAGAGAACTGGTGGCCCGCAATGGTTTACTGCCCTGAGCATAGGAGGGAA
GCAGAGATCATTGAATGGGATGGGGGCTGGAAGGTTAAGTATAAGTGCCC
CGAAGGTCACGAGGGATGGGTTGATATAAGGAGTGGGAACGTGAAACTG
AGGTGGCGTGTTGATTGGCCCATGCGTTGGTCTCACTTTGGCGTTGACTTC
GAACCTGCTGGAAAGGATCATCTTGTGGCTGGTTCAAGCTACGATACGGG
AAAGGAGATTATAAAGGAAGTTTATGGAAAGGAAGCTCCGTTATCTTTAA
TGTATGAGTTTGTTGGAATTAAGGGGCAGAAGGGGAAGATGAGTGGTAGT
AAGGGAAATGTTATTTTACTCAGCGATCTGTATGAGGTTCTTGAGCCAGGT
CTCGTTAGATTTATCTACGCTCGGCATAGGCCAAACAAGGAGATAAAGAT
AGATCTAGGTCTTGGCATTCTAAACCTCTACGATGAGTTCGATAAAGTTGA
GAGAATATACTTCGGGGTTGAGGGTGGTAAAGGTGATGATGAAGAATTAA
GGAGGACTTACGAGCTTTCATAA

Plate Picture for dot plating

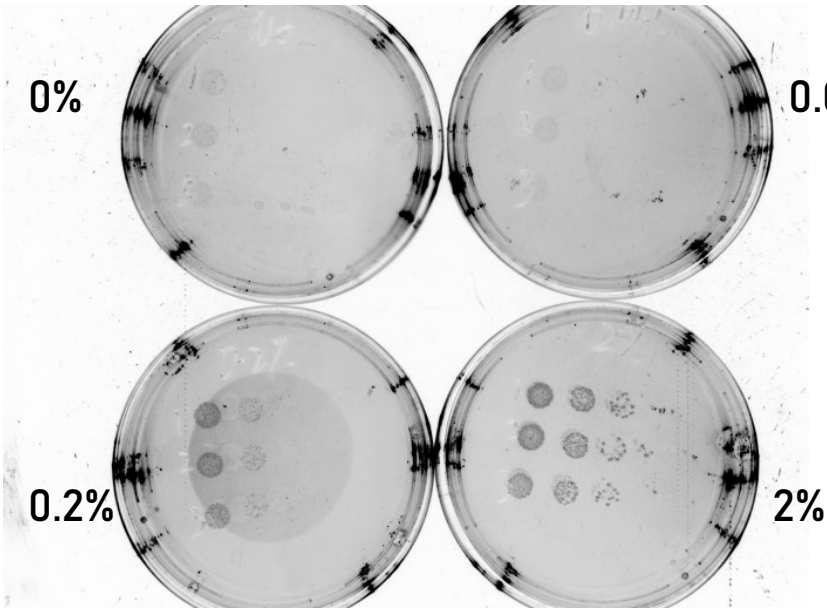


Figure 4-8. Pre-ATMK arabinose dependence test. 3 repeats of Pre-ATMK growth on plate with different concentration of arabinose.

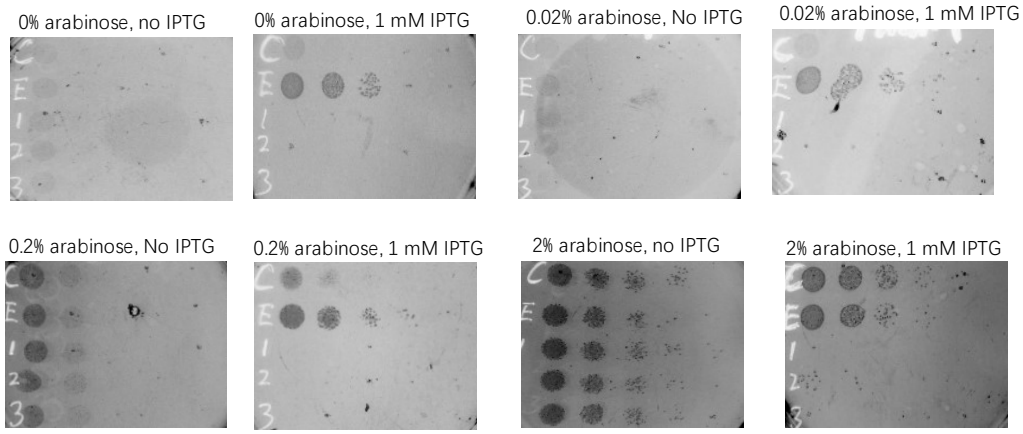


Figure 4-9. Unmodified picture of figure 4-6 A). The second most concentrated spot is presenting in the Figure 4-6 A)

4.6 Reference

1. Ali, I., et al., *Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics.* Chemical Reviews, 2018. **118**(3): p. 1216-1252.
2. Brooks, C.L. and W. Gu, *p53 Ubiquitination: Mdm2 and Beyond.* Molecular Cell, 2006. **21**(3): p. 307-315.
3. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Research, 2012. **40**(D1): p. D261-D270.
4. Zhang, D., et al., *Metabolic regulation of gene expression by histone lactylation.* Nature, 2019. **574**(7779): p. 575-580.
5. Huang, H., et al., *Lysine benzoylation is a histone mark regulated by SIRT2.* Nature Communications, 2018. **9**(1): p. 3374.
6. Huang, H., et al., *Quantitative Proteomic Analysis of Histone Modifications.* Chemical Reviews, 2015. **115**(6): p. 2376-2418.
7. Haj-Yahya, M. and H.A. Lashuel, *Protein Semisynthesis Provides Access to Tau Disease-Associated Post-translational Modifications (PTMs) and Paves the Way to Deciphering the Tau PTM Code in Health and Diseased States.* Journal of the American Chemical Society, 2018. **140**(21): p. 6611-6621.
8. Jing, Y., et al., *Site-Specific Installation of Succinyl Lysine Analog into Histones Reveals the Effect of H2BK34 Succinylation on Nucleosome Dynamics.* Cell Chemical Biology, 2018. **25**(2): p. 166-174.e7.
9. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding Nε-acetyllysine in recombinant proteins.* Nature Chemical Biology, 2008. **4**(4): p. 232-234.
10. de Boor, S., et al., *Small GTP-binding protein Ran is regulated by posttranslational lysine acetylation.* Proceedings of the National Academy of Sciences, 2015. **112**(28): p. E3679-E3688.
11. Worden, E.J., et al., *Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L.* Cell, 2019. **176**(6): p. 1490-1501.e12.
12. Hsu, P.L., et al., *Crystal Structure of the COMPASS H3K4 Methyltransferase Catalytic Module.* Cell, 2018. **174**(5): p. 1106-1116.e9.
13. Bilokapic, S. and M. Halic, *Nucleosome and ubiquitin position Set2 to methylate H3K36.* Nature Communications, 2019. **10**(1): p. 3795.
14. Wang, Z.A., et al., *A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation.* Angewandte Chemie International Edition, 2017. **56**(1): p. 212-216.

15. Mukai, T., et al., *Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases.* Biochem Biophys Res Commun, 2008. **371**(4): p. 818-22.
16. Italia, J.S., et al., *Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes.* Cell Chem Biol, 2018. **25**(10): p. 1304-1312.e5.
17. Italia, J.S., et al., *An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes.* Nature Chemical Biology, 2017. **13**(4): p. 446-450.
18. Lévêque, F., et al., *Homology of lysS and lysU, the two Escherichia coli genes encoding distinct lysyl-tRNA synthetase species.* Nucleic Acids Res, 1990. **18**(2): p. 305-12.
19. Mosberg, J.A., M.J. Lajoie, and G.M. Church, *Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate.* Genetics, 2010. **186**(3): p. 791-9.
20. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon.* Proceedings of the National Academy of Sciences, 2004. **101**(20): p. 7566-7571.
21. Cervettini, D., et al., *Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs.* Nat Biotechnol, 2020. **38**(8): p. 989-999.
22. Ibba, M., et al., *Substrate recognition by class I lysyl-tRNA synthetases: a molecular basis for gene displacement.* Proc Natl Acad Sci U S A, 1999. **96**(2): p. 418-23.
23. Yoo, T.H. and D.A. Tirrell, *High-Throughput Screening for Methionyl-tRNA Synthetases That Enable Residue-Specific Incorporation of Noncanonical Amino Acids into Recombinant Proteins in Bacterial Cells.* Angewandte Chemie International Edition, 2007. **46**(28): p. 5340-5343.
24. Chin Jason, W., et al., *An Expanded Eukaryotic Genetic Code.* Science, 2003. **301**(5635): p. 964-967.
25. Kelemen, R.E., et al., *Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells.* bioRxiv, 2022: p. 2022.01.21.477302.

Chapter 5. Creating a Pre-ATMF strain to screening orthogonal PheRS/tRNA pairs in *E. coli.*

5.1 Introduction.

Tyrosine, tryptophan, and phenylalanine are three aromatic amino acids that play unique rules in biological systems. These three amino acids have structurally bulky and chemically hydrophobic side chains, which enable π-π interactions with other biomolecules. Differing from the other two, the hydroxyl group of tyrosine offers a reactive site for redox chemistry which may contribute to its critical role in protein functions.[1] Tyrosine can be oxidized to a tyrosyl radical and serve as a critical residue for electron transfer in the catalytic center of several metalloproteins [2, 3] Tyrosine can also be exposed on the protein surface and undergo post-translational modifications (PTMs) such as phosphorylation, sulfation, and ubiquitylation, or can alter the enzyme activity, promote protein-protein interaction, change the protein chemical property, or direct protein degradation.[4-9] Tryptophan can also be post-translationally modified, but mostly to form enzyme cofactors such as chromophore maturation of green fluorescent protein (GFP) or tryptophan tryptophylquinone of MADH.[10, 11] Unlike tyrosine and tryptophan, less PTMs have been reported for phenylalanine. In fact, phenylalanine is widely used as a tyrosine analog to study tyrosine function.[5]

Due to the unique function and structure of the aforementioned aromatic amino acids, the ability to site-specifically incorporate tyrosine and phenylalanine analogs into proteins would be a beneficial addition to the chemical toolbox to probe the function of tyrosine in proteins. Using phenylalanyl-tRNA synthetase (PheRS) to incorporate

noncanonical aromatic amino acids can be dated back to 1998, where Rolf Furter incorporated *p*-fluoro-phenylalanine into *E. coli* with a PheRS/tRNA pair from *Saccharomyces cerevisiae* (ScPheRS).[12] However, the utilization of PheRS to incorporate tyrosine and phenylalanine analogs was soon overshadowed by the discovery and application of the *M. jannaschii* tyrosyl RS (MjTyrRS)/tRNA and pyrrolysylRS (PylRS)/tRNA pairs which exhibit a more plastic binding pocket with better substrate specificity.[13] Even though extensive studies have been done to expand the substrate scope of ScPheRS to more phenylalanine, tyrosine, and tryptophan analogs, it is rarely used to incorporate noncanonical amino acids (ncAAs) in *E. coli* because of the superiority of MjTyrRS and PylRS.[14-17] However, as a major heavy lifter in *E. coli*, MjTyrRS is not compatible in mammalian systems, and the *E. coli* tyrosyl-tRNA synthetase (EcTyrRS) and tryptophanyl-tRNA synthetase (EcTrpRS) cannot fill up the substrate variety vacancy left by MjTyrRS despite extensive efforts.[18-21] Building upon my work described in Chapter 3, it was demonstrated that the EcPheRS/tRNA pair can be used to incorporate ncAAs into proteins in mammalian cells. In this chapter, I describe the development of an altered translational machinery phenylalanine (ATMF) strain to further evolve EcPheRS to expand its substrate scope.

Cytoplasmic PheRS has a highly conserved $\alpha_2\beta_2$ heterotetrametric structure, making it one of the largest and most complex enzymes of the aaRS family.[22] On the other hand, mitochondrial PheRS has a simple structure consisting of only a tRNA binding domain and a catalytic domain. (Figure 5-1) [23] Surprisingly, bacterial PheRS

shares considerable structural similarity with the cytoplasmic PheRS instead of the mitochondrial PheRS, even though most mitochondrial aaRS are evolutionarily derived from bacteria. The catalytic domains across all types of PheRS are highly conserved, whereas the anticodon-binding domains (ABDs) and the tRNA binding module at the N-terminus of the α subunit differs between bacterial and cytoplasmic PheRS. [24] In bacteria, the ABD and tRNA binding module at the N-terminus of the α subunit have identity elements for EctRNA^{Phe} and define its orthogonality in the mammalian system. Meanwhile, the anticodon recognition of EctRNA^{Phe} is detrimental to efficiency of EctRNA$^{Phe}_{CUA}$ and devastating to EctRNA$^{Phe}_{UCA}$ activity.[25] Therefore, evolving the EcPheRS for better substrate specificity is as important as evolving EcPheRS for anti-codon compatibility.

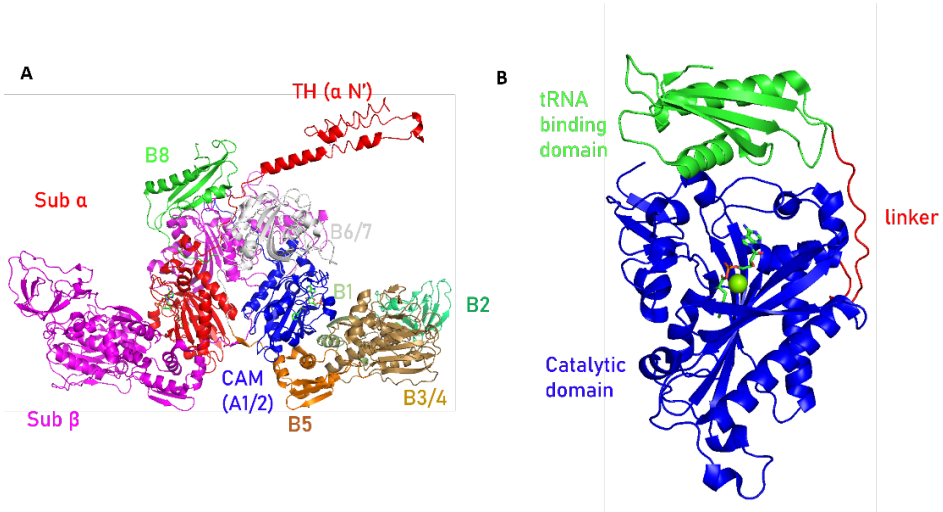


Figure 5-1. Crystal strucutre of EcPheRS and Human mit PheRS. A) Crystal structure of EcPheRS (PDB:3teg). Red and pink color represent α and β subunits respectively. Other colors represent different domain of protein. B) Crystal structure of H.mitPheRS.

Blue and green color represents homolog of H.mitPheRS domains with EcPheRS domains.

There are two GCE pairs based on PheRS that work in the mammalian system: the EcPheRS/tRNA pair discussed in Chapter 3, and the chimeric PheRS/tRNA pair by the group of Lin.[26] In Chapter 3, two known mutations were made in EcPheRS to incorporate azido-phenylalanine (AzF) into mammalian cells, proving the feasibility and potential of this pair. Therefore, the focus of this chapter is to build a phenylanalyl ATM (ATMF) strain to evolve EcPheRS for improved substrate specificity. The chimera PheRS/tRNA pair is made up of a fusion aaRS and a fusion tRNA. Both PylRS and PheRS are categorized as class IIc aaRS, which are made up of a catalytic domain, editing domain, and sometimes an ABD. In the case of the human mitochondrial phenylalanyl-tRNA synthetase (hmPheRS), it is a minimal aaRS that only contains a catalytic domain and an ABD that are linked with a short structureless peptide linker.[24, 27, 28] The design of the chimera PheRS has a fused catalytic domain of hmPheRS with a PylRS ABD, fused with a short flexible peptide linker. Meanwhile, to pair with the chimeric PheRS, the chimeric tRNA fused accepter stem of $hmtRNA^{Phe}$ with other parts from $tRNA^{Pyl}$. This chimeric PheRS (chPheRS) leverages the orthogonality of PylRS/tRNA pairs, as well the substrate profile of PheRS.[26] The chimeric design is intriguing because it bypasses the orthogonality issue and the new pair theoretically can be used in both bacterial and mammalian systems. However, the drawbacks of this system are that it is not compatible with the PylRS/tRNA system, and

the suppression efficiency is limiting. In this chapter, I focus on building an ATMF strain to evolve and test the chPheRS/tRNA pair in mammalian systems.

5.2 Result

5.3.1 Knocking out PheRS to construct Pre-ATMF

The core design for the ATMF strain is similar to ATMK, which was discussed in Chapter 4. Briefly, the ATMF strain would enable directed evolution of the liberated EcPheRS/tRNA pair in the ATMF *E. coli* strain. The liberation of EcPheRS/tRNA is achieved by knocking out the genomic copy, accompanied by the introduction of an orthogonal PheRS/tRNA pair to ensure cell viability. To test orthogonality of the imported PheRS in the *E. coli* strain, a Pre-ATMF strain is needed. The Pre-ATMF strain has endogenous PheRS knocked out and re-introduced on an arabinose-inducible pEvol plasmid to make the strain dependent on arabinose. The imported PheRS/tRNA pair is introduced into Pre-ATMF via an IPTG-inducible pUltra plasmid. Therefore, the ATMF strain can be propagated in the presence of arabinose, and the orthogonal and active imported pair enables the strain to survive in the absence of arabinose but in the presence of IPTG. (Figure 5-2)

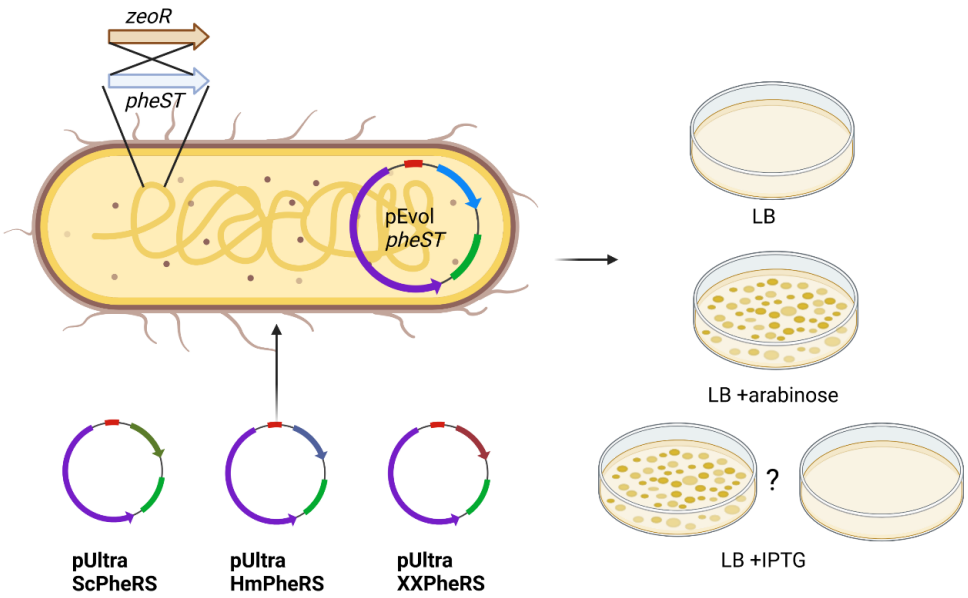


Figure 5-2. Design of Pre-ATMF for screening orthogonal PheRS/tRNA pair for replacing loss of EcPheRS function.

EcPheRS contains two subunits encoded by *pheS* and *pheT*. These two genes are linked by several nucleotides and exist in the same operon, *pheST*. To knock out the *pheST* operon, a temperaure-sensitive pKD46 plasmid harboring the lambda Red recombination system and an arabinose-inducible pEvol plasmid harboring the *E. coli pheST* operon were incorporated into the B95 cell line. The lambda Red system on the pKD46 plasmid can be induced with arabinose and can be discarded from the organism with growth temperatures above 30 °C. The pEvol plasmid encodes EcPheRS to ensure cell viability after knocking out the endogenous copy. For phenotypically knocked out screening and strain maintenance, the genomic *pheST* operon is replaced by a zeocin resistance (*zeoR*) gene. Recombination primers are designed with 45-nucleotide long

homology to the *pheST* locus to completely remove it from the genome and avoid chances of miss-targeting.

The knockout is verified genotypically and phenotypically. Colony PCR is performed to genotypically examine the replacement of the *pheST* operon with *zeoR* and primers is located upstream and downstream of the *pheST* operon. Phenotypically, the viability of Pre-ATMF is dependent on the presence of arabinose, which induces the *pheST* operon on the pEvol plasmid. To evaluate the phenotype, the Pre-ATMF strain is plated on plates consisting of gradient concentrations of arabinose. (Figure 5-3)

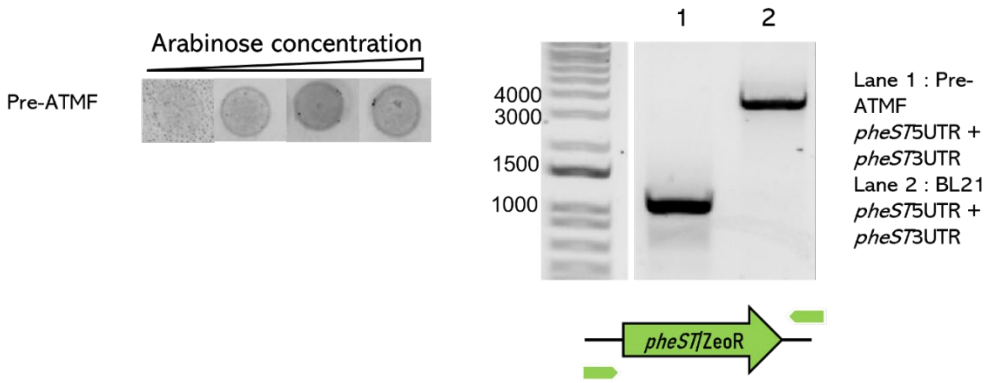


Figure 5-3. Verification of Pre-ATMF strain. Arabinose concentration are 0%, 0.02%, 0.2% and 2%. The colony PCR are performed with primers annealing up and down stream of *pheST*. The products are expected to be 1200 bp for ZeoR and 3700 bp for *pheST,* the result matches with expectation.

5.2.2 Imported PheRS candidate.

In theory, to identify an orthogonal imported pair, all PheRS/tRNA pairs from all domains of life should be tested. However, given the limited time and resources, some parameters are rationally designed and screened. First, aaRS/tRNA pairs from different domains of life tend to be orthogonal to each other, for example the MjTyrRS/tRNA pair in *E. coli* and EcTyrRS/tRNA pair in mammalian systems. Second, tRNAs with different identity elements are likely to be orthogonal to each other's corresponding aaRS. Third, a pair that is evolved to be a suppressor tRNA in the host is likely to be orthogonal as well in Phe sense codon. The first parameter allows us to narrow down the potential candidates to the whole eukaryotic and archaeal domain, yet this candidate pool remains large in number. For the second parameter, with advances in the bioinformatics field, the lab of Jason Chin recently designed a tRNA scoring system, named tREX, to identify tRNAs that are orthogonal to *E. coli* aaRS/tRNA pairs. Unfortunately, tREX did not yield an orthogonal PheRS/tRNA pair from their database.[29] For the third parameter, we focus on the yeast PheRS/tRNA pair which has been engineered as a GCE tool in the *E. coli* host.[12, 17]

5.2.3 Building pUltra ScPheRS/tRNA

The pUltra plasmid contains a *tacI* promoter controlling ScPheRS expression, and an *lpp* promoter constitutively expressing $SctRNA^{Phe}$. The α and β subunit of ScPheRS

is encoded by *FRS1* and *FRS2,* respectively. Unlike *pheS* and *pheT* in *E. coli*, *FRS1* and *FRS2* are not under the same operon. To build them in one ORF under the same promoter, *FRS2* and *FRS1* are fused together, similarly to their homology, the *pheST* operon as previously reported.(Figure 5-4) [30] The cloning work is designed to be sequential, first incorporating the *FRS2-FRS1* cluster, then incorporating SctRNA^{Phe}. The subclone of ScPheRS was straightforward, yet the subclone of tRNA led to no clones. The subclone of tRNA was repeated, with more plating of the recovered culture from transformation. The second attempt yielded three clones with different mutations in the ScPheRS gene, which indicate the SctRNA^{Phe}-WT is toxic in *E. coli.*

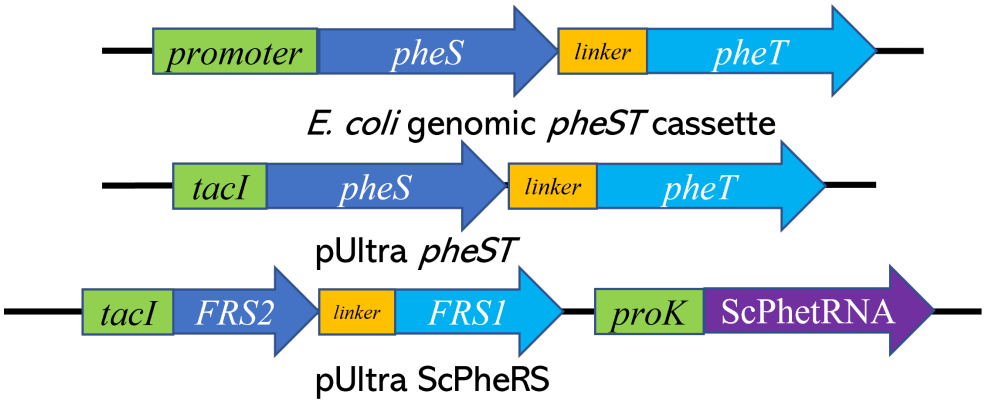


Figure 5-4. Key components of vectors involve in this chapter. *FRS2* and *FRS1* is linked by the same linker sequence as *pheS* and *pheT* linked in the *E. coli* genome.

5.2.4 Testing chPheRS/chPhetRNA

The chPheRS/chtRNA^{Phe} pair was generated by the group of Lin and further evolved to enlarge the substrate scope.[26, 31] The chimeric pair takes advantage of the orthogonality of tRNA^{Pyl} which makes it accessible for both bacterial and mammalian

systems. However, the reported study is focused on the bacterial system and offered limited information on the efficiency in the mammalian system. We then reconstructed the reported chPheRS and chtRNA^{Phe} in two pIDTsmart plasmids, where chPheRS is driven by a CMV promoter and chtRNA^{Phe} is driven by a U6 promoter. To evaluate the efficiency of chPheRS/chtRNA^{Phe} pair in the mammalian system, a PB1 EGFP39TAG reporter plasmid and pIDTsmart tRNA$^{chPhe}_{CUA}$ was co-transfected with or without pIDTsmart chPheRS. The result showed that the pair has poor activity in the mammalian system. Due to a different plasmid format used in our testing and the reported work, it is possible that after optimization, this pair could weakly suppress TAG codons in the mammalian system as the authors demonstrated with DOPARS, an evolved DOPA-specific chPheRS.[31]

5.3 Conclusion and discussion

The EcPheRS/tRNA pair has the potential to expand the current mammalian ncAA scope with directed evolution. The proof of concept is demonstrated in Chapter 3 by incorporating AzF into EGFP in HEK293T cells with a known EcPheRS mutant. In this chapter, I attempted to accomplish this goal by building an ATMF strain as a platform for the directed evolution of EcPheRS. However, the strain cannot be built unless an imported PheRS/tRNA pair that is orthogonal to the *E. coli* system can be found. The ScPheRS/tRNA pair was chosen as the testing candidate because it was engineered as an amber suppressor in *E. coli*, which indicates a certain level of orthogonality.

Unfortunately, the wild type $tRNA^{ScPhe}$ appeared toxic in *E. coli* which does not leave space to modify. At this point, we do not have another good candidate to test in the Pre-ATMF strain. The next step would be to screen eukaryotic PheRS/tRNA pairs using bioinformatic tools to identify candidates that do not share the same identity elements with EcPheRS/tRNA. Chin's work is on this direction and within their database, no orthogonal and active tRNA was identified.[29] We would carefully monitor new literature for potential candidates to push this method forward.

Other than the ATMF system, there are other directed evolution platforms that can be used to evolve EcPheRS/tRNA. For instance, the EcLeuRS/tRNA pair was evolved in yeast. Thus, the EcPheRS/tRNA can be tested for orthogonality in yeast for possible directed evolution. Even though, directed evolution in the mammalian system directly is more time consuming and more sophisticated, it is possible to design a mammalian-based directed evolution system based on the VADER system.[32] Besides the EcPheRS/tRNA pair, the hmPheRS/tRNA pair is also a good starting point due to its simple structure and has already been evolved to incorporate DOPA in mammalian cells.[31] The catalytic domain of hmPheRS is highly homologous with EcPheRS, therefore, the mutant evolved in one may be shared in the other. Evolving chPheRS in *E. coli* and then transplanting the found mutations in EcPheRS for mammalian applications is another way to go.

5.4 Methods

Gene knock-out by Lambda Red recombination

The pEvol EcPheRS and pKD46 were transformed in BL21(DE3) to compensate the lost function of EcPheRS and introduce Lambda red system. To prepare recombination ready competent cell. 20 mL of the pKD46 and pEvol containing BL21(DE3) was cultured in 30 ºC to OD_{600} reach 0.2 and the lambda Red system is induced by addition of 0.02% arabinose. When OD_{600} reaches 0.55 the cell was washed by ice-cold water by centrifugation for 3 times and then resuspended in 50 μL of ice-cold DDI water.

To remove the *E. coli* phenylalanyl–tRNA synthetase (*pheST*) from this strain, the gene encoding zeocin resistance (*ShBle)* driven by the EM-7 promoter and the CYC1 transcription terminator was PCR amplified using primers pheST-ZeoR-F and pheST-ZeoR-R to generate the PCR product *pheST*::ZeoR. 300 ng of the *pheST*::ZeoR PCR cassette was transformed in the prepared lambda Red system induced competent cell, and the resulting strains were plated on LB–Agar plates supplemented with Zeocin and arabinose. The resulting colonies were screened via colony PCR using *pheST*5UTR, ZeoRIR, *pheST*3UTR and ZeoRIF, as well as sequencing these colony PCR products. This strain was named Pre-ATMF.

Dot plating on selective plate.

Colony was picked and inoculated in LB with 2% arabinose and proper antibiotics and cultured overnight at 37 ºC with 250RPM as seeding culture. The seeding culture was washed with PBS by centrifugation. The cell density was measured by optical density at 600 nm (OD_{600}), and then diluted to 0.1 with PBS. The sample was serial diluted from OD_{600} 10^{-1} to 10^{-7}. 10 µl of each sample from 10^{-3} to10^{-7} was doted on the plate with selective conditions. The selective plates were incubated in the 37 ºC incubator overnight, then picture of plate was taken and recorded.

5.5 Supplementary materials

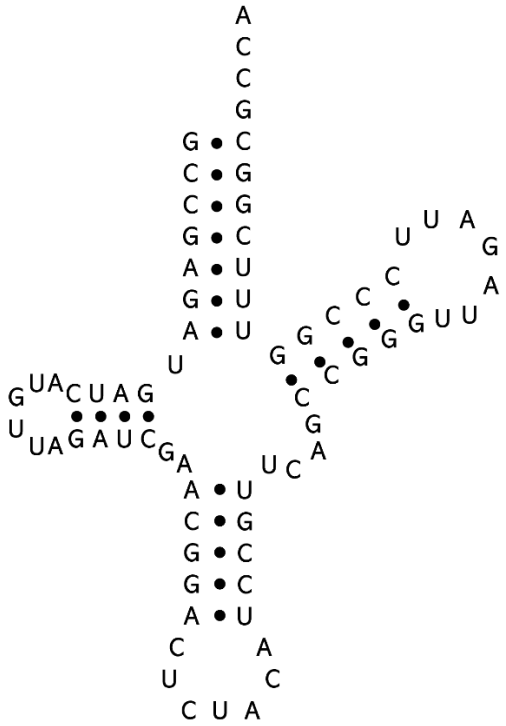


Figure 5-5. Chimera $tRNA^{Phe}_{CUA}$ sequence.

Chimera PheRS sequence.

GSlinker (red); PylRS binding domain (green); HmPheRS catalytic domain (blue);

ATGGATAAGAAGCCGCTGGATGTTCTGATCTCTGCGACCGGTCTGTGG
ATGTCCCGTACCGGCACGCTGCACAAGATCAAGCACTATGAGATTTCT
CGTTCTAAAATCTACATCGAAATGGCGTGTGGTGACCATCTGGTTGTG
AACAACTCTCGTTCTTGTCGTCCCGCACGTGCATTCCGTTATCATAAAT
ACCGTAAAACCTGCAAACGTTGTCGTGTTTCTGACGAAGATATCAACA
ACTTCCTGACCCGTTCTACCGAAGGCAAAACCTCTGTTAAAGTTAAAG
TTGTTTCTGAGCCGAAAGTGAAAAAAGCGATGCCGAAATCTGTTTCTC
GTGCGCCGAAACCGCTGGAAAATCCGGTTTCTGCGAAAGCGTCTACC
GACACCTCTCGTTCTGTTCCGTCTCCGGCGAAATCTACCCCGAACTCTC
CGGTTCCGACCTCTGCAAGCGCCCCAGCTCTGACTAAATCCCAGACGG
ACCGTCTGGAGGTGCTGCTGAACCCAAAGGATGAAATCTCTCTGAAC
AGCGGCAAGCCTTTCCGTGAGCTGGAAAGCGAGCTGCTGTCTCGTCGT
AAAAAGGATCTGCAACAGATCTACGCTGAGGAACGCGAGGGAGGCG
GTGGGAGTGGGGGCGGTGGAAGCGGAGGTGGCGGGAGTGGAGGAGG
AGGAAGTCAGGCCTGGGGATCGAGGCCTCCTGCAGCAGAGTGTGCCA
CCCAAAGAGCTCCAGGCAGTGTGGTGGAGCTGCTGGGCAAATCCTAC
CCTCAGGACGACCACAGCAACCTCACCCGGAAGGTCCTCACCAGAGT
TGGCAGGAACCTGCACAACCAGCAGCATCACCCTCTGTGGCTGATCA
AGGAGAGGGTGAAGGAGCACTTCTACAAGCAGTATGTGGGCCGCTTT
GGGACCCCGTTGTTCTCGGTCTACGACAACCTTTCTCCAGTGGTCACG
ACCTGGCAGAACTTTGACAGCCTGCTCATCCCAGCTGATCACCCCAGC
AGGAAGAAGGGGGACAACTATTACCTGAATCGGACTCACATGCTGAG
AGCGCACACGTCTGCACACCAGTGGGACTTGCTGCACGCGGGACTGG
ATGCCTTCCTGGTGGTGGGTGATGTCTACAGGCGTGACCAGATCGACT
CCCAGCACTACCCTATTTTCCACCAGCTGGAGGCCGTGCGGCTCTTCT
CCAAGCATGAGTTATTTGCTGGTATAAAGGATGGAGAAAGCCTGCAG
CTCTTTGAACAAAGTTCTCGCTCTGCGCATAAACAAGAGACACACACC
ATGGAGGCCGTGAAGCTTGTAGAGTTTGATCTTAAGCAAACGCTTACC
AGGCTCATGGCACATCTTTTTGGAGATGAGCTGGAGATAAGATGGGT
AGACTGCTACTTCCCTTTTACACATCCTTCCTTTGAGATGGAGATCAAC
TTTCATGGAGAATGGCTGGAAGTTCTTGGCTGCGGGGTGATGGAACA
ACAACTGGTCAATTCAGCTGGTGCTCAAGACCGAATCGGCTGGGCTTT
TGGCCTAGGATTAGAAAGGCTAGCCATGATCCTCTACGACATCCCTGA
TATCCGTCTCTTCTGGTGTGAGGACGAGCGCTTCCTGAAGCAGTTCTG
TGTATCCAACATTAATCAGAAGGTGAAGTTTCAGCCTCTTAGCAAATA
A

5.6 Reference

1. Glover, S.D., et al., *Photochemical Tyrosine Oxidation in the Structurally Well-Defined α3Y Protein: Proton-Coupled Electron Transfer and a Long-Lived Tyrosine Radical.* Journal of the American Chemical Society, 2014. **136**(40): p. 14039-14051.
2. Minnihan, E.C., D.G. Nocera, and J. Stubbe, *Reversible, Long-Range Radical Transfer in E. coli Class Ia Ribonucleotide Reductase.* Accounts of Chemical Research, 2013. **46**(11): p. 2524-2535.
3. Zhu, G., et al., *Dissecting the Mechanism of the Nonheme Iron Endoperoxidase FtmOx1 Using Substrate Analogues.* JACS Au, 2022.
4. Stipanuk, M.H., et al., *Cysteine dioxygenase: a robust system for regulation of cellular cysteine levels.* Amino Acids, 2009. **37**(1): p. 55-63.
5. Chen, L., et al., *Use of a Tyrosine Analogue To Modulate the Two Activities of a Nonheme Iron Enzyme OvoA in Ovothiol Biosynthesis, Cysteine Oxidation versus Oxidative C–S Bond Formation.* Journal of the American Chemical Society, 2018. **140**(13): p. 4604-4612.
6. Mo, X., et al., *Tyrosine phosphorylation tunes chemical and thermal sensitivity of TRPV2 ion channel.* eLife, 2022. **11**: p. e78301.
7. Kehoe, J.W. and C.R. Bertozzi, *Tyrosine sulfation: a modulator of extracellular protein–protein interactions.* Chemistry & Biology, 2000. **7**(3): p. R57-R61.
8. Yang, Y.S., et al., *Tyrosine sulfation as a protein post-translational modification.* Molecules, 2015. **20**(2): p. 2138-64.
9. Geimonen, E., et al., *Tyrosine residues direct the ubiquitination and degradation of the NY-1 hantavirus G1 cytoplasmic tail.* J Virol, 2003. **77**(20): p. 10760-868.
10. Barondeau David, P., et al., *Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures.* Proceedings of the National Academy of Sciences, 2003. **100**(21): p. 12111-12116.
11. Davidson, V.L. and C.M. Wilmot, *Posttranslational biosynthesis of the protein-derived cofactor tryptophan tryptophylquinone.* Annu Rev Biochem, 2013. **82**: p. 531-50.
12. Furter, R., *Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli.* Protein Sci, 1998. **7**(2): p. 419-26.
13. Krahn, N., et al., *Engineering aminoacyl-tRNA synthetases for use in synthetic biology.* Enzymes, 2020. **48**: p. 351-395.

14. Datta, D., et al., *A Designed Phenylalanyl-tRNA Synthetase Variant Allows Efficient in Vivo Incorporation of Aryl Ketone Functionality into Proteins.* Journal of the American Chemical Society, 2002. **124**(20): p. 5652-5653.
15. Kwon, I., P. Wang, and D.A. Tirrell, *Design of a Bacterial Host for Site-Specific Incorporation of p-Bromophenylalanine into Recombinant Proteins.* Journal of the American Chemical Society, 2006. **128**(36): p. 11778-11783.
16. Kwon, I. and D.A. Tirrell, *Site-Specific Incorporation of Tryptophan Analogues into Recombinant Proteins in Bacterial Cells.* Journal of the American Chemical Society, 2007. **129**(34): p. 10431-10437.
17. Kwon, I. and S.I. Lim, *Tailoring the Substrate Specificity of Yeast Phenylalanyl-tRNA Synthetase toward a Phenylalanine Analog Using Multiple-Site-Specific Incorporation.* ACS Synthetic Biology, 2015. **4**(5): p. 634-643.
18. Grasso, K.T., et al., *A Facile Platform to Engineer Escherichia coli Tyrosyl-tRNA Synthetase Adds New Chemistries to the Eukaryotic Genetic Code, Including a Phosphotyrosine Mimic.* ACS Central Science, 2022. **8**(4): p. 483-492.
19. Italia, J.S., et al., *Genetically encoded protein sulfation in mammalian cells.* Nat Chem Biol, 2020. **16**(4): p. 379-382.
20. Italia, J.S., et al., *Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes.* Cell Chem Biol, 2018. **25**(10): p. 1304-1312.e5.
21. Italia, J.S., et al., *An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes.* Nature Chemical Biology, 2017. **13**(4): p. 446-450.
22. Moor, N., et al., *Prokaryotic and eukaryotic tetrameric phenylalanyl-tRNA synthetases display conservation of the binding mode of the tRNA(Phe) CCA end.* Biochemistry, 2003. **42**(36): p. 10697-708.
23. Klipcan, L., et al., *The tRNA-Induced Conformational Activation of Human Mitochondrial Phenylalanyl-tRNA Synthetase.* Structure, 2008. **16**(7): p. 1095-1104.
24. Klipcan, L., et al., *Crystal structure of human mitochondrial PheRS complexed with tRNA(Phe) in the active "open" state.* J Mol Biol, 2012. **415**(3): p. 527-37.
25. Mermershtain, I., et al., *Idiosyncrasy and identity in the prokaryotic phe-system: Crystal structure of E. coli phenylalanyl-tRNA synthetase complexed with phenylalanine and AMP.* Protein Science, 2011. **20**(1): p. 160-167.
26. Ding, W., et al., *Chimeric design of pyrrolysyl-tRNA synthetase/tRNA pairs and canonical synthetase/tRNA pairs for genetic code expansion.* Nat Commun, 2020. **11**(1): p. 3154.

27. Smith, T.F. and H. Hartman, *The evolution of Class II Aminoacyl-tRNA synthetases and the first code.* FEBS Letters, 2015. **589**(23): p. 3499-3507.
28. Kavran Jennifer, M., et al., *Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation.* Proceedings of the National Academy of Sciences, 2007. **104**(27): p. 11268-11273.
29. Cervettini, D., et al., *Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs.* Nat Biotechnol, 2020. **38**(8): p. 989-999.
30. Sanni, A., et al., *Construction of a FRS1-FRS2 operon encoding the structural genes for the alpha and beta subunits of yeast phenylalanyl-tRNA synthetase and its use in deletion analysis.* Nucleic Acids Res, 1990. **18**(8): p. 2087-92.
31. Zhao, H., et al., *Directed-evolution of translation system for efficient unnatural amino acids incorporation and generalizable synthetic auxotroph construction.* Nature Communications, 2021. **12**(1): p. 7039.
32. Kelemen, R.E., et al., *Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells.* bioRxiv, 2022: p. 2022.01.21.477302.

Chapter 6. Developing platform for directed evolution of EcArgRS as GCE tool in mammalian cells.

6.1 Introduction

Arginine, as one of the three positive charged amino acids, plays several critical roles in protein functions. It serves as an essential element of serine/arginine (SR)-rich splicing factors, functions critically in nuclear localization sequences for viral proteins Tas and Res, and presents in the common protein binding arginine−glycine−aspartate (RGD) motif, as well as playing a role in improving protein stability to prevent protein aggregation.[1-4] In addition to its diverse functions, arginine is among one of the most post-translationally modified amino acid in the forms of monomethylarginine (MMA), symmetric dimethylarginine (SDMA), asymmetric dimethylarginine (ADMA), citrulline, phosphoarginine, and more. [5] (Figure 6-1) Within these PTMs, methylation and citrullination are of interest among researchers due to their presence in histone proteins. Arginine methylation is believed to function in pre-mRNA splicing, DNA damage signaling, mRNA translation, cell signaling, and cell fate decision, and citrullination is a negative regulator of the methylation.[6]

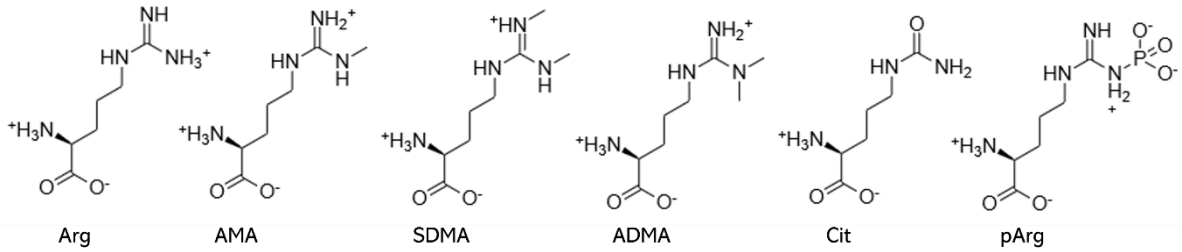


Figure 6-1. Arginine based PTMs.

Currently, to obtain a homogenous methylated sample, histone protein is incubated *in vitro* with PRMTs or CARMs, which are methyltransferases responsible for arginine methylation. [7] However, the crosstalk between different PTMs in histones and the different functions of various PTM sites cannot be investigated through this approach. With this *in vitro* enzymatic methylation method, the sample cannot reassemble protein in physiological condition with other nearby PTMs, and the location of the methylation is hardly controlled. The *in vitro* translation system and solid-phase synthesis of protein allow site-specific installation of PTMs, but restoration of the physiological condition and nearby PTMs remain challenging. [8] Aside from arginine PTMs in histone protein, citrullination is found in many other proteins and is believed to be involved with protein degradation, NETosis formations, and protein translocation, which are mostly seen in mammalian systems and deeply discussed in chapter 2. Given the limited approach to dissection arginine PTM, there is an urgent need to develop a method to site specifically incorporate arginine PTM and obtain protein samples in their native environments.

For the site-specific study of arginine and its PTMs in their native environment, GCE is a perfect tool. As for arginine citrullination, we accomplished *in vivo* incorporation with a nitrobenzyl citrulline by EcLeuRS. This photocaged citrulline incorporation served as a powerful tool to reveal the biological function of citrullination on 372 and 374 in PAD4.[9] With this advancement, we envision a more direct incorporation of citrulline, while extending this strategy to incorporate other arginine PTMs. Inspired by a report on utilizing yeast ArgRS to incorporate citrulline with *in*

vitro translation system method, and the orthogonality of arginine $tRNA_{CUA}$ in the mammalian system, we believe EcArgRS/tRNA has the potential to fulfill this task. [8] Therefore, the next step is to design a scheme to engineer or evolve EcArgRS to accept more substrates.

To evolve the EcArgRS/tRNA pair into a GCE tool in the mammalian system, we considered three available platforms including *E. coli*, *S. cerevisiae* (yeast), and mammalian cells. The *E. coli* system has the advantage of high transformation efficiency, library maintenance, easy recovery, quick doubling time, and already existing directed evolution schemes. However, our candidate, ArgRS, originated from *E. coli*, and thus faces an orthogonality challenge in the *E. coli* host. To overcome cross-reactivity with the endogenous wild type ArgRS, there are two options: replace the endogenous ArgRS with an orthogonal pair or conditionally disable the endogenous ArgRS by removing its substrate. The altered translational machinery (ATM) system is based on the first path, which involves replacing the endogenous aaRS/tRNAs with imported counterparts to liberate the target aaRS/tRNAs for evolution and application in the strain. [10] Specifically, for ArgRS/tRNA, the ATMR would be challenging to construct due to the comparatively large number of anticodons assigned to arginine and the large number of $EctRNA^{Arg}$ genes responsible for decoding them. Specifically, seven endogenous Arg-tRNAs need to be knocked out, and at least three orthogonal tRNAs must be imported and optimized for proper expression levels, assuming an orthogonal ArgRS/tRNAs pair is available. [11] Alternatively, the endogenous wild

type ArgRS is disabled by depleting its substrate arginine. The goal of directed evolution of ArgRS is to create a variant with a higher affinity for its substrate(s) than arginine. This difference between WT and mutant ArgRS would create a conditional orthogonality in the absence of arginine and in the presence of an ncAA; the library would only be functional under this condition. The second evolutionary strategy has been applied to EcMetRS by the Terrier group and proved to be feasible with substantial optimizations.[12] In this chapter, we adopt this design for EcArgRS evolution platform.

The yeast system has similar characteristics as the *E. coli* system. It falls short of *E. coli* on transformation efficiency, doubling time, and ease of library recovery. However, as the simplest eukaryotic system, it very likely orthogonal to aaRS/tRNA pairs originating from *E. coli*, and it also partially resembles mammalian cell environments. The yeast-based directed evolution scheme was designed by the Schultz group and utilized to evolve EcLeuRS.[13] The scheme depends on GAL4, a transcription factor, and controlled expression of essential genes. To conditionally control the GAL4, two stop codons are incorporated into the GAL4, and only when these two stop codons are suppressed is the GAL4 successfully expressed, resulting in expression of downstream genes. Based on the downstream gene, positive or negative selection can be performed. Although the yeast system is well characterized, the orthogonality of EcArgRS in the yeast system is questionable, even if it is highly orthogonal in the mammalian system.

The mammalian system is the destination for the incorporation of arginine PTMs. Therefore, directed evolution in mammalian cells would have incomparable advantages for orthogonality and compatibility. However, the mammalian cell will not maintain and amplify the imported plasmid. This trait of mammalian cells limits library size and recovery of library members, and it therefore hinders development of an appropriate directed evolution scheme. Our lab developed an AAV-based directed evolution scheme in the mammalian system to evolve tRNAs. [14] AAV compensates for the limitations of the library plasmids in mammalian cells, in that its ability to enter and replicate within cells aids delivery of library members, and its ability to be recovered, intact, from mammalian cells enables recovery of library members. It is possible to tune this AAV-based system for aaRS evolution in the mammalian system. However, due to time limits and the extensive optimization and development necessary to enable aaRS evolution in this system, this was not the path we took in this chapter. Due to the lack of a suitable directed evolution platform, we turned our focus to enzyme engineering through rational design using established structural information.

Overall, in this chapter, we focused on the development of a directed evolution platform for EcArgRS evolution with the final goal of expanding the GCE toolbox for mammalian systems. Based on the reported structure, we rationally designed several mutants and tested the incorporating efficiency of arginine PTMs.

6.2 Result

6.2.1 Structure inspection and mutant design of EcArgRS.

Rational design of the EcArgRS binding pocket to accept other substrates than arginine generally follows one method: shorten or neutralize the side chains of amino acids to allow access of larger or uncharged ncAAs. In the binding pocket of the EcArgRS, D118, D317, Y313 and R324 coordinate the binding of substrate.(Figure 6-2) Changing D118 and D317 abolishes the binding property; meanwhile, changes weaken the binding in the case of Y313 and R324.[15] Among these, the D317 is likely to form ionic bonds or salt bridges with guanidine group of the side chain of arginine, dictating the possible lengths and charges of substrate side chains. Considering citrulline is a deaminated arginine, which is charge-neutral, we hypothesized that mutating D317 to permit neutral and shorter side chains on substrates would allow for citrulline and methyl arginine uptake by the mutant RS. With this hypothesis, we mutated D317 to A, C, G, N, S, P, or T via site-directed mutagenesis and tested their ability to incorporate arginine PTMs in the HEK293T cell line.

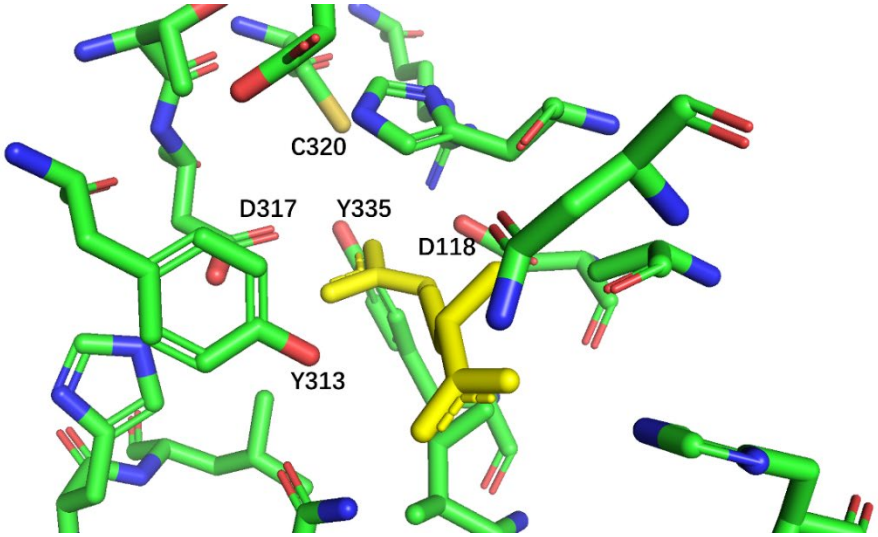


Figure 6-2. Crystal structure of EcArgRS. The Arg substrate is highlighted in yellow and other nearby residues are labeled out. (PBD:4oby)

6.2.2 Characterize D317 mutant of ArgRS

To test D317 mutants with various ncAAs of interest, we assessed stop codon suppression in an EGFP construct in the absence or presence of a desired ncAA. In chapter 3, we demonstrated that EctRNA^{Arg} containing the UCA anticodon is orthogonal with low background. Therefore, the EcArgRS mutant, EctRNA$_{UCA}$, and EGFP 39TGA were co-transfected in HEK293T in the absence or presence of citrulline. Citrulline was tested first due to its structural similarity to the native substrate arginine, and because with a minor spacing change of one amino acid in the binding pocket, we hypothesize that citrulline has a better chance to be accepted. An active mutant will accept citrulline as a substrate, resulting in suppression of TGA in EGFP and fluorescent cells. All seven mutations showed a detrimental effect on ArgRS activity,

which agrees with the reported essentiality of D317. Among the seven mutants, D317S showed the largest change between presence and absence of citrulline. Interestingly, with WT ArgRS, the addition of citrulline induced expression of EGFP 39TGA, which could be potentially explained by the fact that citrulline is a substrate for arginine biosynthesis. Thus, the change of fluorescence induced by addition of citrulline needs to be evaluated by comparing change seen with the ArgRS mutants to that seen with the WT. D317A, D317G and D317S mutants were tested in comparison with WT in the presence of 1 mM, 5 mM and 10 mM citrullines. At 5 mM, D317S showed a 2-fold increase in signal whereas WT showed a 1.5-fold increase. (Figure 6-3) We hypothesis that this difference potentially resulted from the uptake of citrulline by D317S. Therefore, an alternative method to verify citrulline incorporation is needed.

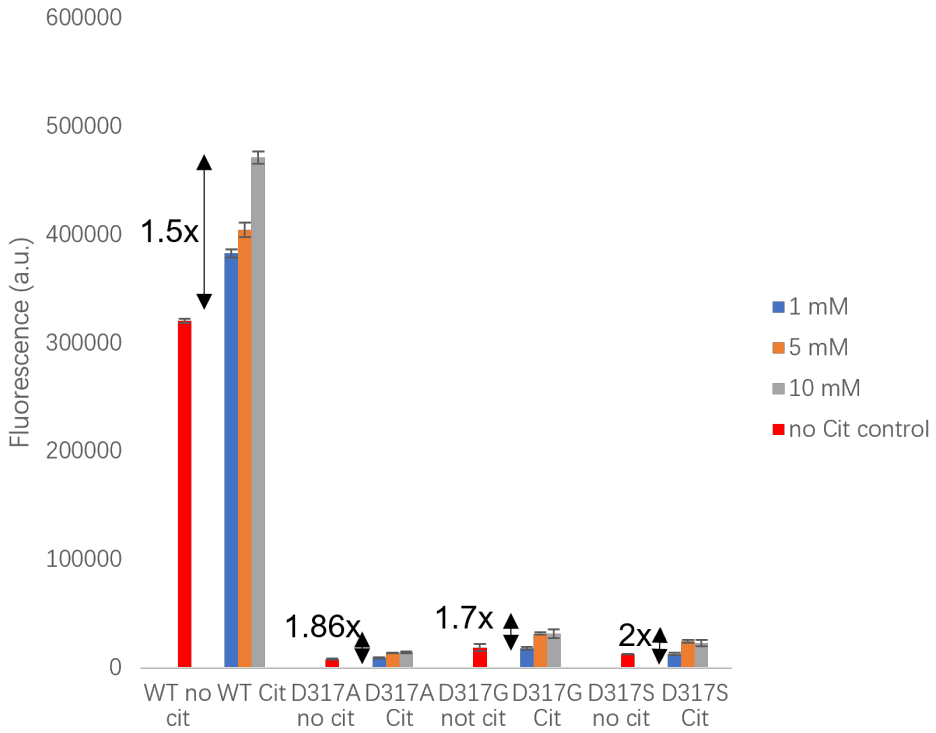


Figure 6-3. Fluorescence data of selected ArgRS mutants compare with WT ArgRS in present and absent of Cit.

6.2.3 Determine citrulline incorporation

The common method to determine ncAA incorporation is mass spectrometry. Particularly, for EGFP, whole protein mass from LC/MS is the preferred method due to the small amount of sample required and simple process. However, citrulline has a 0.98 da mass difference with arginine, and therefore the incorporation of citrulline can hardly be distinguished from arginine by whole protein mass. Trypsin digestion combined with tandem mass or high-resolution mass spectrometry can distinguish arginine and citrulline; however, the D317S mutant lowered incorporation efficiency by 10-fold, which made sample accumulation difficult. Other determination methods such as use of a citrulline-specific probe or citrulline-specific antibody can qualitatively detect citrulline but cannot quantitively determine the percentage of incorporation.

To overcome the detection challenges, we proposed a protease-based citrulline detection method in order to distinguish between citrulline and arginine with sub-microgram levels of protein sample. In this approach, the arginine codon is replaced with a stop codon to incorporate citrulline or arginine in a protease cleavage site that requires arginine. The Factor Xa protease (Fxa) cleaves after an arginine residue in an Ile-Glu/Asp-Gly-Arg sequence or less preferred Gly-Arg sequence.[16] We argue that switching Arg to Cit in the cleavage site prevents the recognition of Fxa and results in

a full length reporter protein. Based on this idea, we designed a reporter gene by fusing SUMO and EGFP with a modified Fxa cleavage site linker. The modified Fxa cleavage site sequence has the arginine codon replaced by TGA. (Figure 6-4A) The reporter gene is co-transfected with EctRNA$^{Arg}_{UCA}$ and ArgRS WT or D317S mutant in the presence of citrulline. Meanwhile, a reporter protein with the WT Fxa sequence is used as a positive control and trial test sample for protocol optimization. The assay was performed with protein samples from SUMO-EGFP with the WT Fxa sequence, SUMO-EGFP with the modified sequence suppressed by WT ArgRS and SUMO-EGFP with the modified sequence suppressed by D317S in the presence of citrulline. (Figure 6-4) All three samples remain full length in the absent of Fxa protease and were cleaved into two fragments in the presence of Fxa protease. Under the assumption that the Fxa protease cannot recognize citrulline, this result indicates that the D317S mutant installed arginine at the cleavage site, triggering Fxa protease cleavage. (Figure 6-4B) The assumption that the Fxa protease cannot recognize citrulline can be verified by incorporating citrulline at the site, and we generated the citrulline-containing sample by using EcLeuRS. Up to writing of this thesis, this experiment has not yet been performed due to time limits.

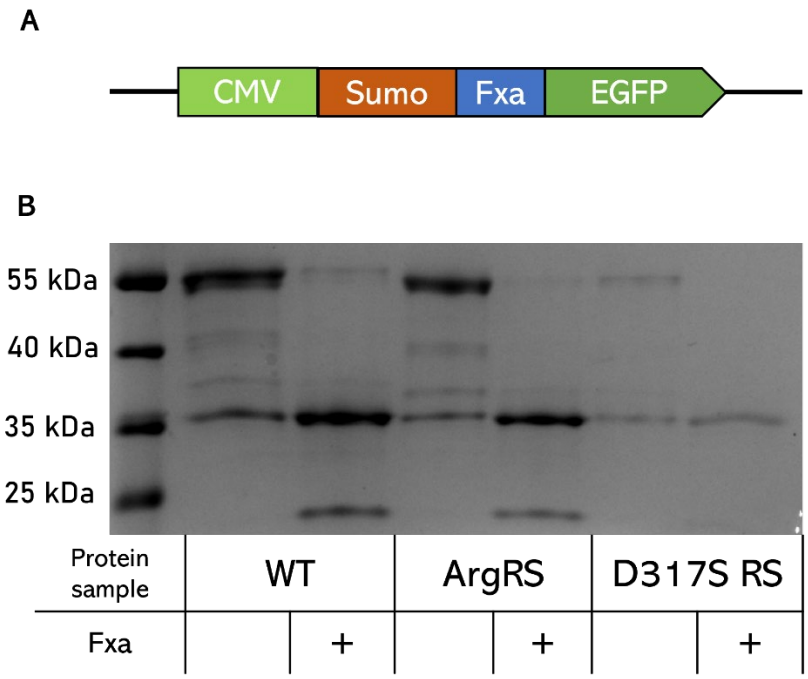


Figure 6-4. A) Construct design of Fxa based citrulline detection method. The Fxa recogniztion sequence is IEGR or IEG(TGA), flanked by 4 nts GS linkers. B) SDS-PAGE of Fxa assay with different protein sample. WT: WT IEGR sequence without any GCE method involved. ArgRS: IEG(TGA) sequence, with TGA suppressed by WT ArgRS/tRNA$_{UCA}$ pair with cellur Arg. D317SRS: IEG (TGA) sequence, with TGA suppressed by D317SRS/tRNA$_{UCA}$ piar with 5 mM Cit.

The MMA and SDMA were also tested with 317S ArgRS with no positive signal detected. In summary, this attempted single site mutagenesis did not yield an active mutant that can expand the chemistry space for ncAA incorporation.

6.2.4 Testing ArgRS pair orthogonality in the yeast.

The success of directed evolution of EcLeuRS in yeast demonstrated the feasibility of 1) yeast as a platform and 2) the GAL4-based selection scheme.[13] ArgRS shares

similar tRNA and codon numbers in *E. coli* with LeuRS and thus faces similar challenges for evolution in *E. coli* on a platform such as ATM strain creation.[11] (Table 1) The high number of tRNAs and codons shared by both of these systems would not only require extensive work to knock out all the tRNAs, but also adds another layer of difficulty to the introduction and balance of orthogonal imported tRNAs. Use of the yeast-based platform offered a potential alternative path to success: the biggest challenge presented is that it requires tRNA orthogonality in yeast.

Isotype	tRNA Count by Anticodon						Total
Arg	ACG 4	GCG	CCG 1	TCG	CCT 1	TCT 1	7
Leu	AAG	GAG 1	CAG 4	TAG 1	CAA 1	TAA 1	8

Table 6-1. *E. coli* tRNA counts by anticodon for Arg and Leu.

We tested the orthogonality of EcArgRS in yeast using a yEGFP-based fluorescence assay. EcArgRS and $tRNA_{CUA/UCA}$ were cloned into a pESC plasmid with ADH1 and SUP4 promoters, respectively. Also, pESC plasmids containing only EcArgRS or $tRNA_{CUA/UCA}$ are generated as control plasmids. The yEGFP TAG/TGA constructs were on pGAD plasmids, driven by the ADH1 promoter. To perform the assay, the reporter pGAD plasmid and pESC plasmid containing full or partial GCE elements were co-transformed into yeast, and the resulting cell fluorescence was

measured to determine suppression efficiency. Subsequently, yEGFP was purified on a Ni-column and characterized by whole protein mass via LC/MS.

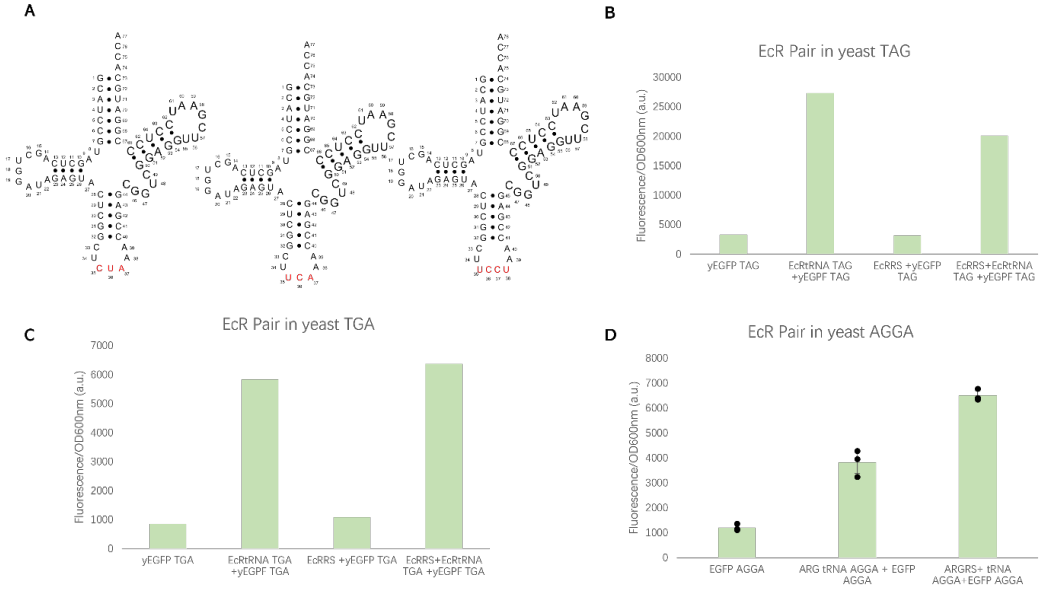
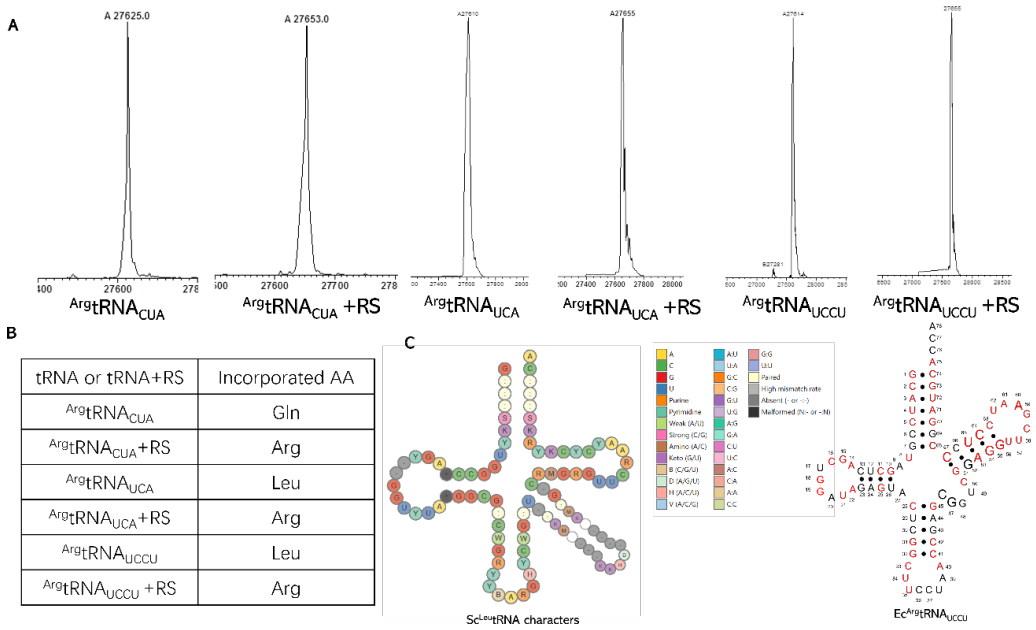


Figure 6-5. A) tRNA^{EcArg} teseted in yeast. BCD)Fluorescence readout for tRNAs with different anticodon. WT yEGFP is around 200,000 F/OD (a.u.).

Both tRNA$_{CUA}$ and tRNA$_{UCA}$ showed high levels of cross-reactivity. With ArgRS and EctRNA$^{Arg}_{CUA}$ or EctRNA$^{Arg}_{UCA}$, arginine was installed at the stop codon in the yEGFP. However, with only EctRNA$^{Arg}_{CUA}$ or EctRNA$^{Arg}_{UCA}$ in the absence of ArgRS, the amino acids installed in the yEGFP are lysine/glutamine or leucine, respectively. (Figure 6-6AB) For the sample from EctRNA$^{Arg}_{CUA}$, we reason that glutamine was incorporated instead of lysine, which was previously reported. The leucine incorporation with EctRNA$^{Arg}_{UCA}$ likely resulted from its high similarity to SctRNA^{Leu}. [17] (Figure 6-6C) An AGGA-suppressing quadruplet codon tRNA was made with the aim of lowering the cross-reactivity seen with yeast endogenous aaRSs. Unfortunately,

the UCCU anticodon still contained identity elements for SctRNA^{Leu}, and ScLeuRS remains cross-reactive with this EctRNA^{Arg} variant.



tRNA or tRNA+RS	Incorporated AA
^{Arg}tRNA$_{CUA}$	Gln
^{Arg}tRNA$_{CUA}$+RS	Arg
^{Arg}tRNA$_{UCA}$	Leu
^{Arg}tRNA$_{UCA}$+RS	Arg
^{Arg}tRNA$_{UCCU}$	Leu
^{Arg}tRNA$_{UCCU}$ +RS	Arg

Figure 6-6. A) Mass spectrometry data for yEGFP TAG/TGA/AGGA suppressed by different components of EcArgRS/tRNA pairs. B) Summarized mass spectrometry data with corresponding incorporated amino acids. C) Comparison between SctRNA^{Leu} with EctRNA^{Arg}. The matching elements are labeled in red on EctRNA^{Arg}.

6.2.5 Yeast-based selection.

The pilot experiment using yEGFP expression indicated that the observed cross-reactivity may render this platform unsuitable to perform selection. We decided to test the pair in a more stringent selection system to verify the result. The yeast selection system contains two key elements: 1) a selective strain of *S. cerevisiae* that is

conditionally auxotrophic to Leu, His, Trp and uracil acid, and 2) a pGAD plasmid harboring stop codon-containing genes which encode GAL4 to control the expression of genes involved in the biosynthesis of histidine and uracil acid, namely *HIS3* and *URA3*, respectively.[13] This system enables two levels of positive selection. The first level is qualitative selection in the form of indicating whether candidates are active or inactive; the second level is a quantitative selection, allowing adjustment of the stringency of the selection. Suppressing the stop codon and enabling expression of functional GAL4 protein leads to expression of *HIS3* and *URA3* and allows the selected strain to grow on the agar plate without histidine and uracil acid supplement. This phenomenon yields a basic understanding of the activity of the target aaRS and tRNA pair. In addition, the promoter of *HIS3* is negatively regulated by a competitive inhibitor 3-aminotriazole (3-AT), which allows fine tuning of the stringency of the selection. By creating a gradient concentration of 3-AT, the difference between aaRS/tRNA variant activities is revealed. The negative selection is based on the ability of Orotidine 5'-phosphate decarboxylase (ODCase), which is encoded by GAL4 triggered *URA3*, to convert 5-FOA into the toxic compound. On a plate with 5-FOA, without a complete set of orthogonal aaRS/tRNA pair, the strain should be able to survive. However, presence of cross-reactive tRNA or aaRS leads to lethal toxicity.

To test $tRNA^{Arg}_{UCCU}$ orthogonality in the yeast with selection system, tRNA only, aaRS only, and tRNA with aaRS are separately co-transformed with pGAD-GAL4(2xAGGA) and plated on selective agar plates. The conditions include SD-Leu-

Trp as cell control, SD-Leu-Trp-His as positive control, SD-Leu-Trp-His + 20 mM 3-AT as another positive control, and SD-Leu-Trp + 0.1% 5-FOA as negative control. The results from the positive and negative selections came back in disagreement. For the positive selection with 3-AT, the cells expressing the tRNA only or the tRNA and aaRS do not exhibit growth defects, indicating a successful suppression of AGGA codon. Unexpectedly, the cells containing EcArgRS only did not survive. This result agreed with the yEGFP assay results suggesting that $EctRNA^{Arg}$ is cross-reactive with the endogenous aaRS. However, on the 5-FOA-containing negative selection plate, not only did the RS only plate grow as expected, but the tRNA only plate also grew with a comparable intensity. The disagreement between negative and positive selection indicated that one of the selection systems does not function properly. To tackle this inconsistency, we performed both positive and negative selections with different stringency. The growth pattern was expected to change with the concentration of 3-AT and 5-FOA if the selection system is functional. As expected, as concentration of 3-AT increases, fewer cells grow on the corresponding plates. This observation suggests that the positive selection works properly since 3-AT is a competitive inhibitor with GAL4. However, with increasing concentrations of 5-FOA on the negative selection plates, no difference in cell growth can be detected, which suggests that the negative selection is dysfunctional.

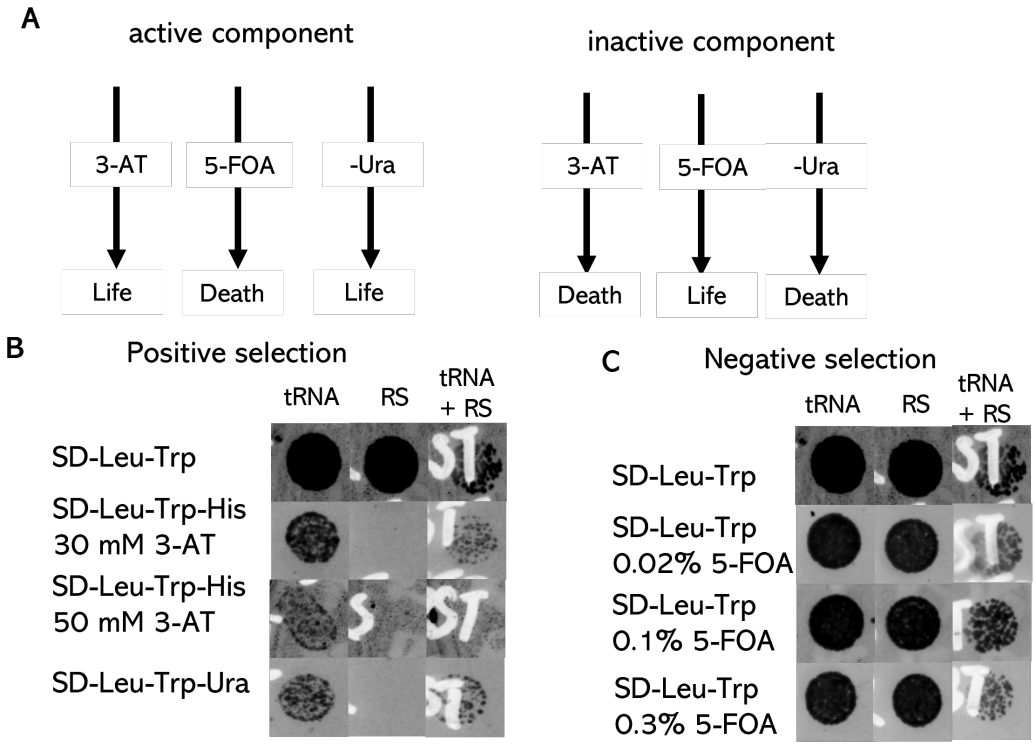


Figure 6-7. A) Expectation of selection system. B) Positive selection result. tRNA stands for strain received a plasmid contain EctRNA$^{Arg}_{UCCU}$. RS stands for strain received a plasmid contrain EcArgRS. tRNA+RS stands for strain received a plasmid contain EcArgRS/tRNA$_{UCCU}$ Pair. SD-Leu-Trp is cell quality control. All components are expected to live on the plates as the data shown. C) Negative selection result.

Combining the yEGFP assay and the positive selection, we conclude that the EctRNA^{Arg} is cross-reactive to yeast-endogenous aaRSs depends on different anticodons. With the CUA anticodon targeting TAG codons, the tRNA is mischarged by GlnRS due to strong anticodon recognition, which is also observed in mammalian cells. With the UCA or UCCU anticodon targeting TGA or AGGA codons, the tRNA is mischarged by LeuRS due to high similarity between tRNA^{Arg} and tRNA^{Leu} sequences. Currently, ongoing work aims to optimize this selection by increasing the

number of stop codons in the GAL4 gene. In conclusion, the yeast selection system is not suitable for evolution of EcArgRS due to the cross-reactivity of EctRNA^{Arg}.

6.2.6 FACS-based directed evolution system in *E. coli*

As evolving ArgRS in yeast is not feasible due to a lack of tRNA orthogonality, we seek to evolve ArgRS in *E. coli*. However, we do not have an EcArgRS-orthogonal *E. coli* strain available, and building an ATMR would be challenging in many aspects. The Terrier group has developed a FACS-based directed evolution system without creating a fully orthogonal strain, which they used to evolve EcMetRS.[12, 18] In this system, the wild type tRNA^{Arg} decodes conventional Arg codons and does not need to be modified. The endogenous ArgRS and the mutant ArgRS co-exist in the host cell and their activities are distinguished by timely supply of amino acid substrates. When the environmental arginine is replaced by ncAA, the mutant ArgRS charges ncAAs to globally replace arginine in the newly synthesized proteins. The reporter plasmid encoding sGFP is induced after the swap of arginine to ncAA, and it can only be expressed if the mutant ArgRS is able to uptake the ncAA as a substrate. Therefore, the fluorescent signal in the cells corresponds to the activity of the mutant ArgRS, which allows desired cells to be sorted through FACS.

To achieve this selective environment, the ncAA is added at a given time, whereas arginine needs to be present for cell culture and removed when the ncAA is introduced. *E. coli* acquire arginine from both the environment and biosynthesis. To deplete the

supply of arginine for *E. coli,* we need to change the media recipe and disturb the arginine biosynthesis pathway to create an arginine auxotroph. Meanwhile, the GFP reporter needs to contain several non-essential arginine sites as the incorporation sites. To operate the selection, the reporter gene is located on an inducible plasmid and co-transformed into the auxotroph strain with mutant EcArgRS. The strain is cultured in an arginine-containing media to a desired induction confluency, then the media is switched to an arginine-free, ncAA-containing media and GFP is induced for activity testing. Fluorescent signal intensity represents the activity of the mutant ArgRS, and when paired with FACS, active candidates can be selected and characterized.

6.2.7 Generation of an arginine auxotroph by knocking out *argH*.

The arginine biosynthetic pathway in *E. coli* consists of eight genes that convert glutamate to arginine, namely *argA* to *argH*. [19] For this specific directed evolution purpose, we require either *argG* or *argH* to be knocked out, as citrulline is a substrate of the pathway upstream of *argG*. We first looked for an existing arginine auxotroph strain in the Coli Genetic Stock Center (CGSC), which has auxotroph strains for all 20 amino acids. The only arginine auxotroph in the CGSC has *argA* deleted, which does not match our needs. Therefore, we decided to knock out *argH,* which encodes argininosuccinate lyase, to create an arginine auxotrophic *E. coli*. The *argH* is replaced with *ZeoR* by lambda red homologous recombination. The strain was genotypically verified by colony PCR using primers annealing within the open reading frame of *ZeoR*,

and primers located up- and downstream of *argH*. More importantly, the phenotypic performance of the strain was verified by spot plating on synthetic agar plates supplemented with arginine, citrulline, or neither substrate. The genotypic and phenotypic testing demonstrated that knocking out *argH* resulted in an arginine-dependent, but not citrulline-dependent, auxotroph. (Figure 6-8B) Meanwhile, because the selection scheme required the strain to be viable after arginine depletion, the strain was characterized by a growth curve experiment with varying concentrations of arginine. The result showed that the log phase of cell growth was stalled after depletion of arginine, but the strain maintained some viability, as indicated by slight increases in cell density after that. This result implied that in the absence of arginine input, the cell can salvage old proteins to recycle arginine for new proteins.

(Figure 6-8C) The ability to recycle cellular arginine will lead to a basal level of false positive signal that cannot be avoided. This basal signal serves as a selection threshold, such that only very active candidates producing very high signals can be chosen.

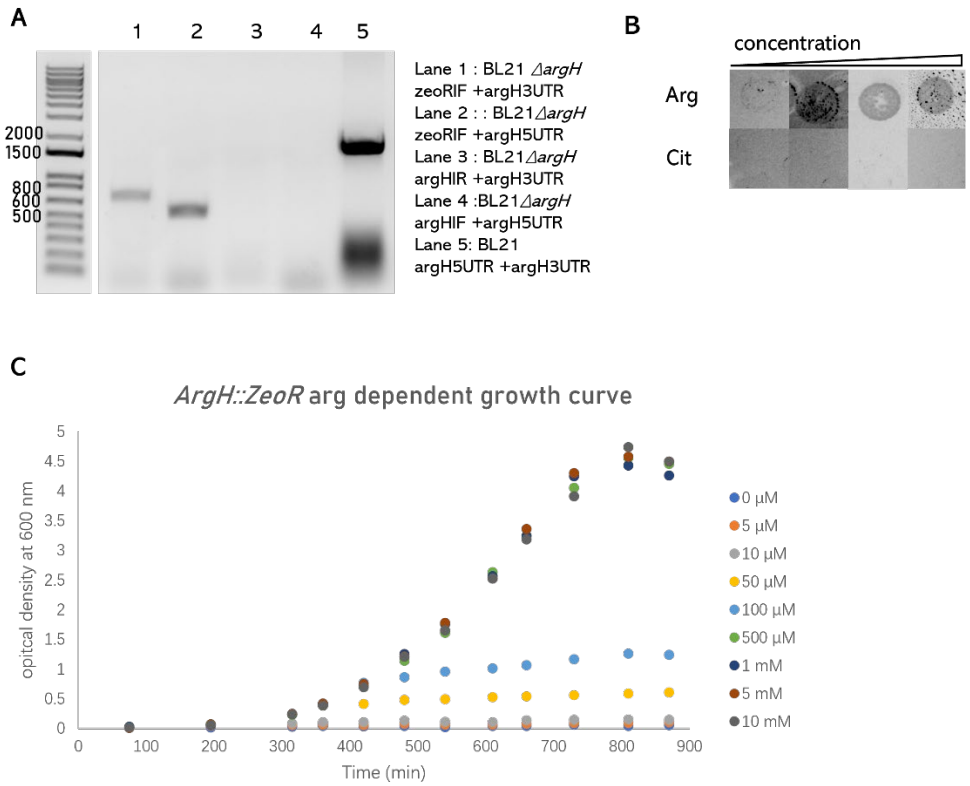


Figure 6-8. A) Colony PCR product on agarose gel. The band size matches expectation indicates success replace target gene with resistant gene. B) Verification of *argH* knock out strain. The concentration gradient for both Arg and Cit are 0 mM, 0.1 mM, 1mM, and 10 mM. The strain only growth in present of Arg instead of Cit. C) Arg dependent growth curve of BL21 (DE3) *ΔargH.* Cells density plateaued once Arg is depleted in the media.

6.2.8 Building a split-GFP system for selection

In Terrier's work, the GFP reporter was tailored for the selection scheme: they tested the essentiality of all methionine sites and relocated them to structurally irrelevant locations. [12] By the same principle, the GFP reporter for the ArgRS

selection needs to contain at least a few nonessential arginine sites. By inspecting the structure of GFP and other fluorescent proteins, such as mCherry and YGFP, it has been found that position R96 is conserved among such proteins, and it is essential for fluorophore maturation.[20] (Figure 6-9) To bypass this essential arginine, we decided to utilize the split GFP (sGFP) system that was first reported by the Waldo group.[21] The GFP protein contains a barrel-shaped shell formed by 11 beta sheets and a fluorophore core in the middle. Removing a beta sheet leads to non-fluorescent protein. However, the beta sheet expressed separately in the form of a short peptide can spontaneously fuse back to the GFP to fill the gap and restore the fluorescence. Using this property, the Waldo group disconnected the number 11 beta sheet (β11) from the rest of GFP (β1-10) and engineered both fragments for better solubility. Despite the efforts, β11 remained only moderately soluble and often fused with other proteins as a C-terminal tag to boost solubility. (Figure 6-10)

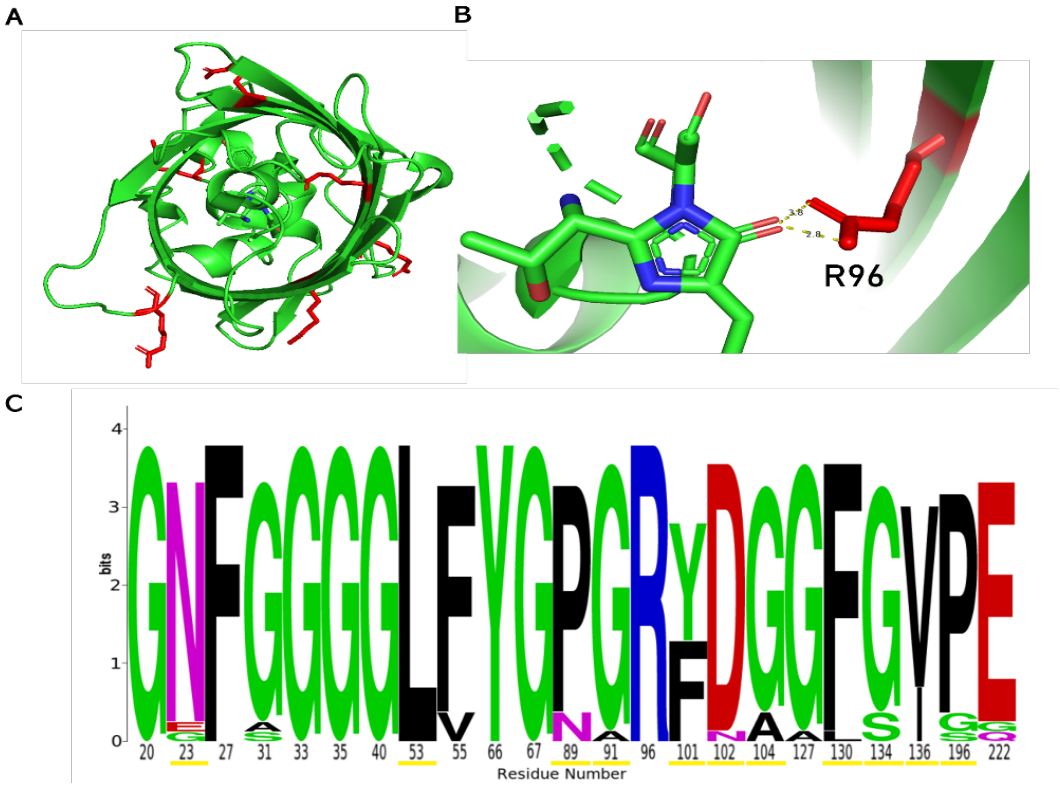


Figure 6-9. A) sfGFP structure with all Arg residues labeled red. B) SfGFP structure zooms in R96 residue. C) R96 conservation cross fluorescence proteins.[22]

We inspected the sequence of the two fragments of sGFP to locate all arginine locations. R96 is in the β1-10 segments with three other arginines, and β11 contains one arginine codon in the middle. This distribution of arginines was beneficial for our aims: we sought to make the β1-10 piece in the presence of arginine and the β11 piece in the presence of ncAA. To optimize their solubility and increase the arginine numbers on the reporter (to reduce the false positive signal), the β11 was fused with a SUMO tag. After testing three fusion formats for β11 and SUMO, we selected an N-terminal SUMO fused to β11 with an eight-amino acid GS linker. This SUMO-β11 construct

contains total six arginine sites and can fuse back with β1-10 when expressed sequentially or simultaneously.

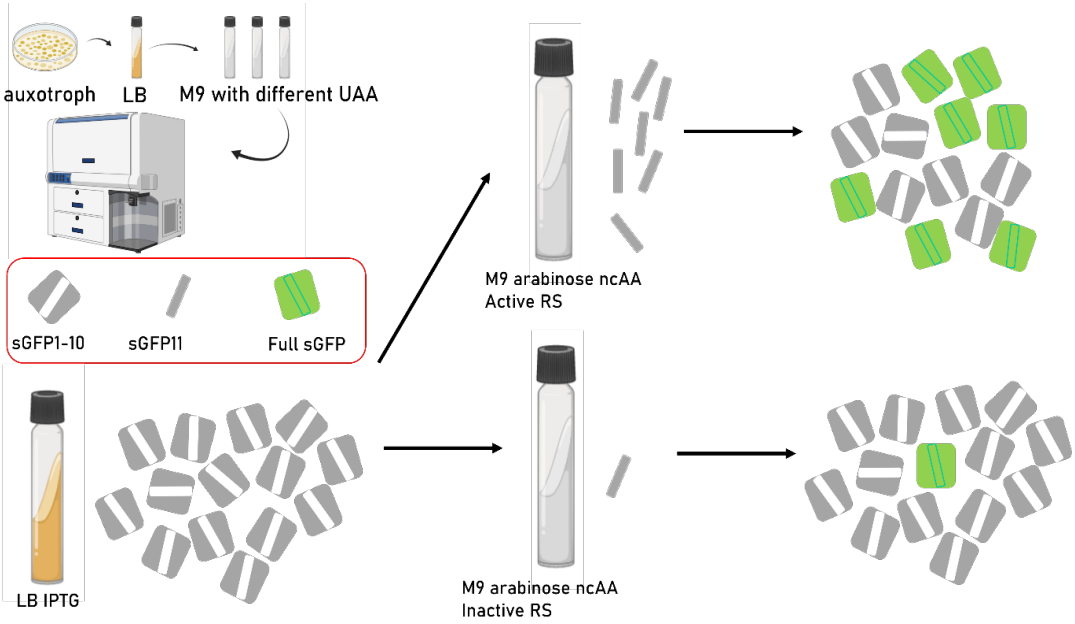


Figure 6-10. Principle of splitting GFP and selection.

6.2.9 Characterizing the selection system with WT EcArgRS and arginine.

We tested the feasibility of this selection strategy using a mock selection method. Usually, a mock selection would be performed with a known mutant that is able to charge an ncAA. However, we did not have such a known pair for this purpose. Therefore, we decided to characterize the method with WT EcArgRS and arginine by controlling the supplementation of substrate. Specifically, β1-10 and SUMO-β11 are harbored by two inducible plasmids, respectively: IPTG-induced pUltra and arabinose-induced pEvol. The two plasmids were transformed into the arginine auxotroph and

induced sequentially. In the first stage, the strain was incubated in arginine-containing media and the pUltra was induced to express and accumulate β1-10 to ensure the R96 was installed. In the second stage, the media was swapped to arginine-free media, and the strain was allowed to fully consume free cellular arginine. At this stage, arginine was added into the media and the pEvol plasmid was induced to express SUMO-β11 for 7 hours. At the end of the third stage, cells were harvested, and cell fluorescence normalized to OD was measured. Our aim was to distinguish between the fluorescence seen with and without arginine at this stage. There are several key elements necessary for this protocol to work: first, in stage 1, β1-10 is expected to be present in great excess comparing to SUMO-β11, which is expressed at stage 3. The amount of β1-10 determines the dynamic range of the signal, which should not be saturated by SUMO-β11. Second, stage 2 needs to be long enough to ensure the depletion of free cellular arginine in order to limit the false positive signal. Third, temperature must remain constant, as the poor solubility of β11 means that its accessible concentration is sensitive to temperature during expression and incubation.

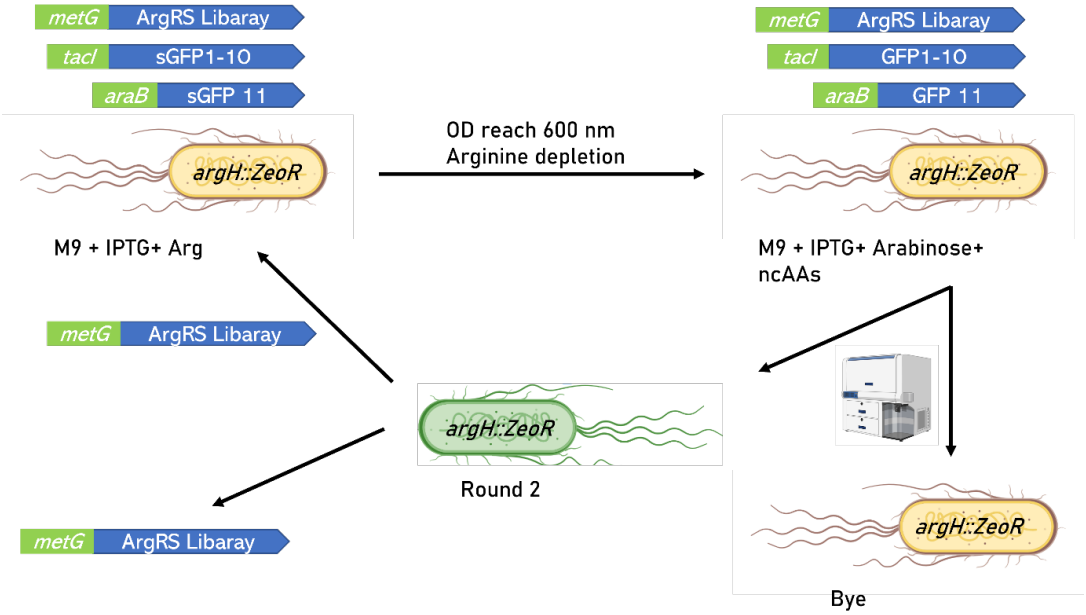


Figure 6-11. Selection scheme of FACS based selection with spliting GFP.

To test this idea and optimize the protocol, several conditions were tested. We started by modifying the conventional protein overexpression protocol, in which protein expression is induced at OD_{600} 0.6 and post-induction culture is incubated at 30 ºC for 12 hours. One slight modification from this protocol was the replacement of LB by M9 minimal media containing arginine. When the culture reached OD_{600} 0.2, the pUltra plasmid was induced by IPTG for expression of β1-10, and when the culture reached OD_{600} 0.6, the culture was washed by arginine-free M9 media three times and then incubates at 37 ºC for one hour to consume the cellular arginine. After incubation, the culture was induced by arabinose to express SUMO-β11, and arginine was added back in the selected samples. The post-induction culture was incubated at 30 ºC for 12 hours before harvest. The fluorescence was normalized by OD_{600} to represent the protein yield per cell. Several comparison samples were set with induction variation

and arginine addition. For this trial, the signal from the desired positive sample was no different from the un-induced cell control, and it was likely caused by short expression time for β1-10. In the later trails, aspects such as media selection and expression time for the first stage, incubation time for the second stage, and the incubation temperature were adjusted to optimize the yield. At this point in time, we have obtained a working condition. First, the cells were cultured in LB media and β1-10 expression was induced at OD_{600} 0.6, after which there was continued culture for 4 hours. In the second stage, the culture was washed three time and inoculated at OD_{600} 0.4 into arginine-free M9 media. The incubation time was set at four hours to ensure the depletion of cellar arginine. And for the third stage, the incubation temperature was set to room temperature or lower. With this working condition, a two-fold difference between positive sample and control can be observed. Due to time constraints, no more tests have been performed beyond this condition, but it is believed that these conditions are far from optimized since the overall signal is significantly lower than that of split GFP formed from co-expression of β1-10 and SUMO-β11 in LB media.

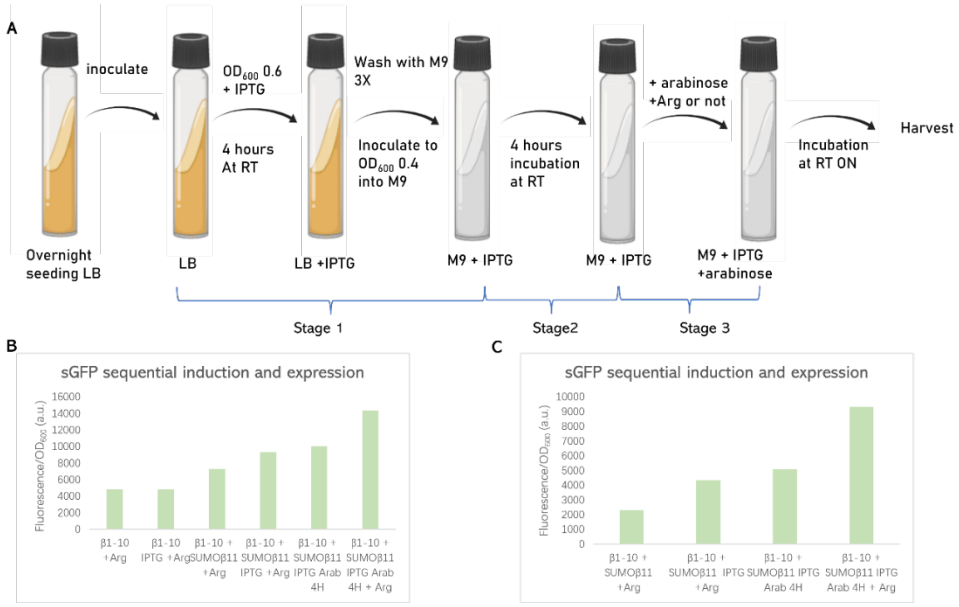


Figure 6-12. A) The latest functional procedure for mock selection. B) Fluorescence signal of strain with different conditions. C) Fluorescence data subtracts cell background.

6.3 Conclusion and discussion.

The EcArgRS/tRNA pair has great potential to expand the GCE toolbox and tackle the community's urgent need for studies of arginine PTMs such as methylated arginine and citrulline. In chapter 3, we discovered that EctRNA$^{Arg}_{UCA}$ is nearly orthogonal in mammalian cells and can be used as a TGA suppressor in conjunction with a mutant EcArgRS. In this chapter, we sought to rationally engineer EcArgRS, and to create a directed evolution platform for evolving EcArgRS in yeast or *E. coli.*

The rational design of the catalytical pocket was based on the structure of EcArgRS. in which D317 is responsible for binding the guanidine group of arginine. [23] Therefore, it is reasonable to hypothesize that shortening and neutralizing the side chain

in position 317 could enable binding of citrulline and methylarginines. However, it is also documented that mutation at D317 abolished the binding activity toward arginine. Thus, mutating D317 could be detrimental for substrate binding, or it could be beneficial for ncAAs while orthogonally excluding arginine. The results from our experiments showed that mutating D317 was indeed detrimental for substrate binding, and there was no evidence that ncAAs were incorporated into protein using our D317 mutants. Although this was a failed attempt to engineer a functional mutant, a method to verify citrulline incorporation based on an arginine-specific protease was designed and tested. With a little more polish through addition of a citrulline-containing control, this method can be a quick and efficient tool for verifying citrulline incorporation with small amounts of protein samples.

In this chapter, we did not explore the possibility of designing a mammalian cell-based directed evolution strategy due to time limits and the greater appeal of other, more accessible approaches. However, it is possible to design such a platform in the future. The incorporation of AAV into mammalian selection schemes could enable feasible directed evolution schemes by allowing for large library coverage, easy library recovery and high throughput selections.[14] Meanwhile, the AAV capsid can serve as the selection pressure: if active ArgRS can incorporate ncAAs into the capsid, AAV containing active ArgRS library members can be selectively retrieved. However, this strategy depends on identifying ncAAs that can be incorporated successfully into the AAV capsid, and thus could be challenging to apply.

The attempt to use the yeast-based platform developed by the Schultz group failed due to the tRNA's cross-reactivity with endogenous aaRS. Efforts were made to see if strategic anticodon selection could result in increased orthogonality, but unless we can modify or evolve the tRNA significantly, it has not currently demonstrated any potential to be orthogonal. Because the ultimate destination of this pair is mammalian systems, evolving the tRNA to be orthogonal in the yeast is unnecessary. Furthermore, we have shown that the 5-FOA based negative selection is not effective and needs to be optimized.

The FACS-based split GFP directed evolution platform in *E. coli* for arginine is being developed and requires significant optimization. Currently, we have achieved the ability to discriminate between cells containing a mock "active" ArgRS and the control group. Thus, we envision that this strategy could prove useful for selecting desired ArgRS with further development. The shortcoming of the platform is that execution of the experiment is rather complicated, and background is significant. The difference between positive and control groups in our mock selection is 2-fold compared to baseline, but whether this difference is significant enough for selection remains unknown, especially since the signal generated by mutant ArgRS in a real selection is likely to be significantly lower. This question could be addressed by an improved mock selection that uses a known mutant ArgRS with its corresponding ncAA as the mock "active" ArgRS. However, no such mutant currently exists. Alternatively, it is possible build a TAG-containing SUMO-β11 and use PylRS with ncAA as the mock "active"

aaRS and introduce a EctRNA$^{Arg}_{CUA}$ to mimic the background. In this chapter, we have explored three different potential selection systems for evolving EcArgRS to incorporate arginine PTMs. However, none of the tested platforms yielded the desired EcArgRS. The FACS-based system has demonstrated great potential and should be further explored. For future experiments, it is also possible to use sfGFP as the reporter gene with R96K mutation--although R96 is highly conserved, it is also documented that R96K can restore fluorescence.[20]

6.4 Method:

EGFP expression from HEK293T cells

HEK293T cells were cultured at 37 °C in a humidified incubator supplemented with 5% CO2. Cells were seeded at 3.75 10^5 cells per well of 24 wells plate. 24 h before transfection. EGFP and WT PAD4 transfections were performed by incubating 0.5 µg plasmid DNA, 2.5 µL polyethylenimine (PEI) (1 mg/mL; Polysciences, Warrington, PA), and 9 µL DMEM for 10 min at room temperature, followed by adding the solution dropwise to the culture medium of the cells. For EGFP-39-TAG expression, 167 ng of pIDTSTMART EctRNA^{Arg}-synthetase, 167 ng pIDTSMART EctArgRNA$_{UCA}$ plasmids and 167 ng PB1 EGFP 39TAG were incubated with a mixture of 2.5 µL PEI and 9 µL DMEM for 10 min at room temperature before adding to cells. Citrilline was added at the same time to a final concentration of 1mM, 3 mM or 5 mM for citrulline

incorporation with EcArgRS mutants, and 2 mM sodium butyrate was added to enhance protein expression.

EGFP fluorescence analysis

EGFP fluorescence was analyzed 48 h after transfection. Cells were harvested and resuspended in 600 µL CelLytic M buffer (Sigma, St. Louis, MO) with Halt Protease inhibitor Cocktail (Thermo Scientific, Waltham, MA) and Pierce Universal Nuclease for Cell Lysis (Fisher Scientific, Hampton, NH). Lysates were clarified by centrifugation at 16,000 × *g* for 10 min and 100 µL of supernatant was transferred to a clear-bottom 96-well plate for fluorescence measurement. Fluorescence was measured using a Zeiss AX10 microscope (Ex. 488 nm; Em. 534 nm). All the experiments were performed at least in three duplicates.

LC-MS analysis of EGFP

LC/MS analyses of the purified EGFP were performed on Agilent Technologies 6230 TOF LC/MS with Aerns 3.6 µm WIDEEPORE XB-C8 20 LC column 100 × 4.6 mm.

Dot plating on selective plate.

Colony was picked and inoculated in LB with 1mM arginine and proper antibiotics and cultured overnight at 37 ºC with 250RPM as seeding culture. The seeding culture was washed with PBS by centrifugation for three times. The cell density was measured by optical density at 600 nm (OD_{600}), and then diluted to 0.1 with PBS. The sample was

serial diluted from OD_{600} 10^{-1} to 10^{-7}. 10 µl of each sample from 10^{-3} to10^{-7} was doted on the plate with selective conditions. The selective plates were incubated in the 37 ºC incubator overnight, then picture of plate was taken and recorded.

yGFP fluorescence analysis

Colony was picked and inoculated in SD-Leu-Trp media and incubated at 250RPM, 30 ºC for 24 hours as seeding. The seeding culture was inoculated into 10 ml of SD-Leu-Trp media in 1:100 ratio and cultured in same condition for 48 hours. Then the culture was harvested by centrifugation at 1000xg followed by resuspend in PBS buffer. The cell density was measured by OD_{600} and diluted to 0.1 for measurement. The sample was measured with OD600 and fluorescent channel with a setting following setting: Ex. 488 nm; Em. 534 nm. The data was normalized by using divide fluorescent reading by OD_{600}.

GAL4 based selection

The GAL4 based selection was following protocol described in Schultz group paper with different plasmids combination.[13]

6.5 Supplementary Materials

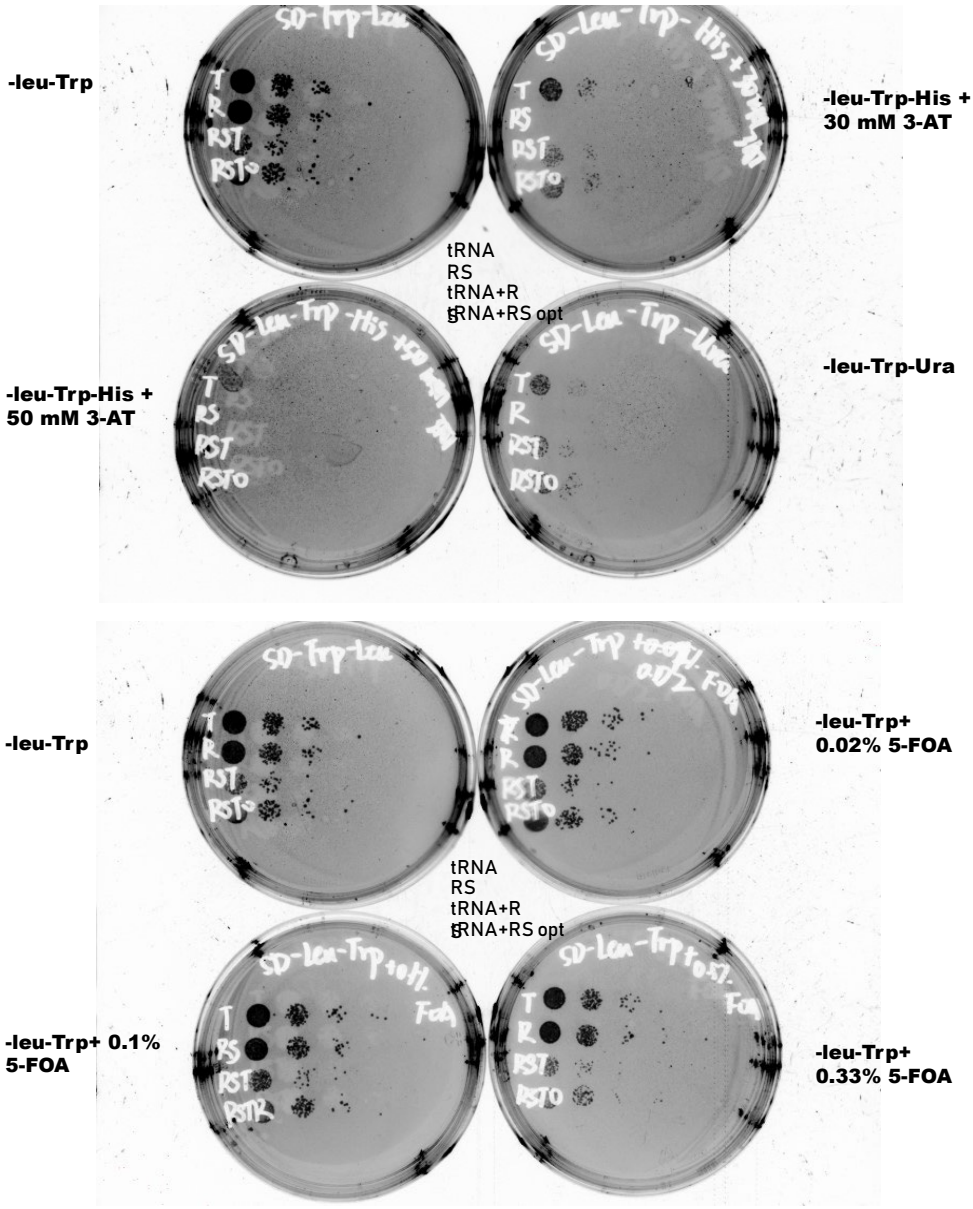


Figure 6-13. Unmodified plate picture of Figure 6-7.

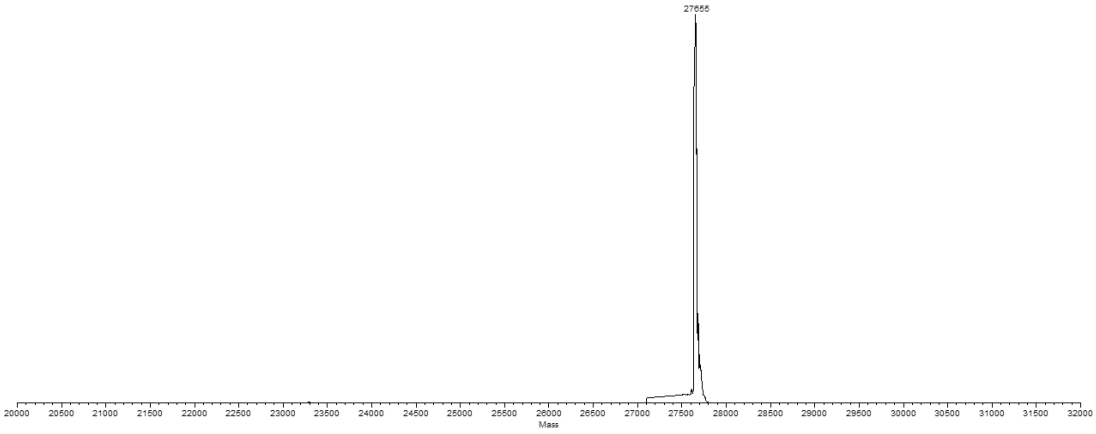


Figure 6-14 a-1). yEGFP from tRNA$^{scArg}_{UCCU}$ + EcArgRS+ yEGFP39AGGA.

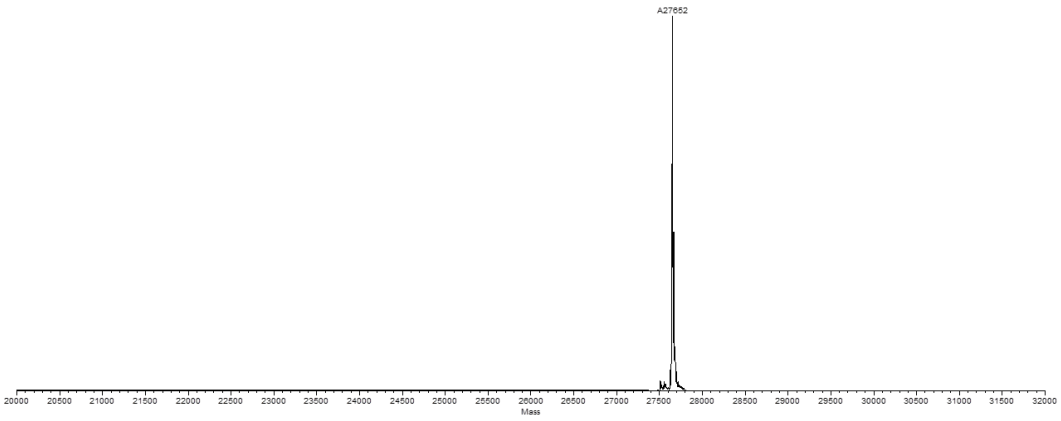


Figure 6-14. a-2)yEGFP from tRNA$^{scArg}_{UCCU}$ + yEGFP39AGGA.

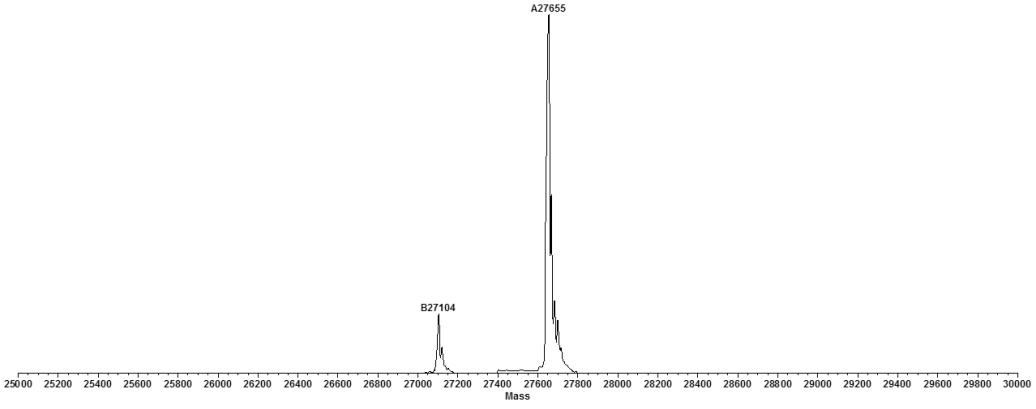


Figure 6-14. b-1)yEGFP from $tRNA^{scArg}_{UCA}$ + EcArgRS+ yEGFP39TGA

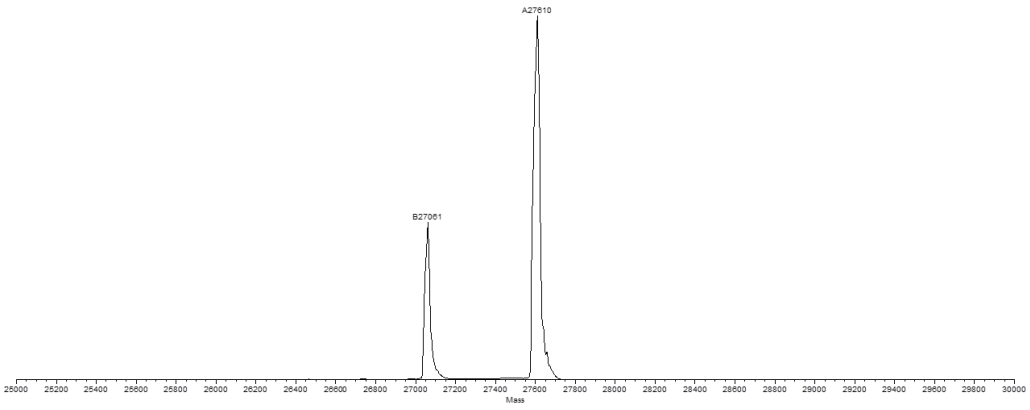


Figure 6-14. b-2)yEGFP from $tRNA^{scArg}_{UCA}$ + yEGFP39TGA

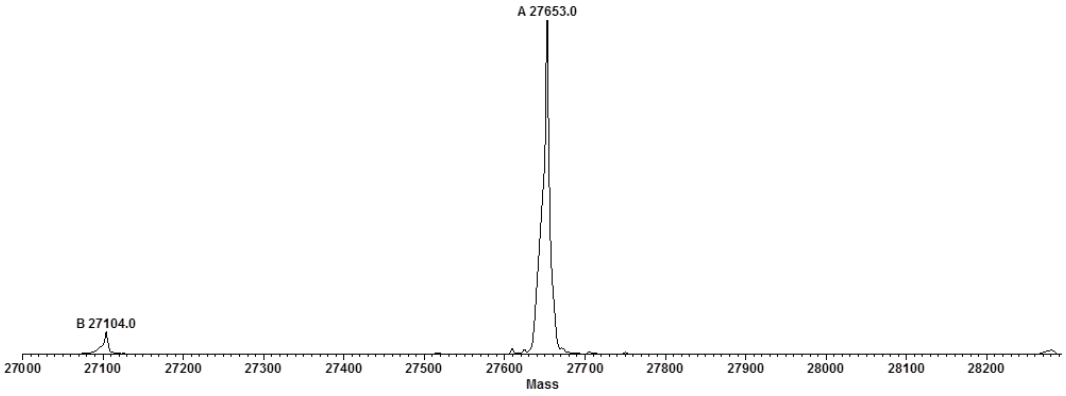


Figure 6-14. C-1)yEGFP from $tRNA^{scArg}_{CUA}$ + EcArgRS+ yEGFP39TAG

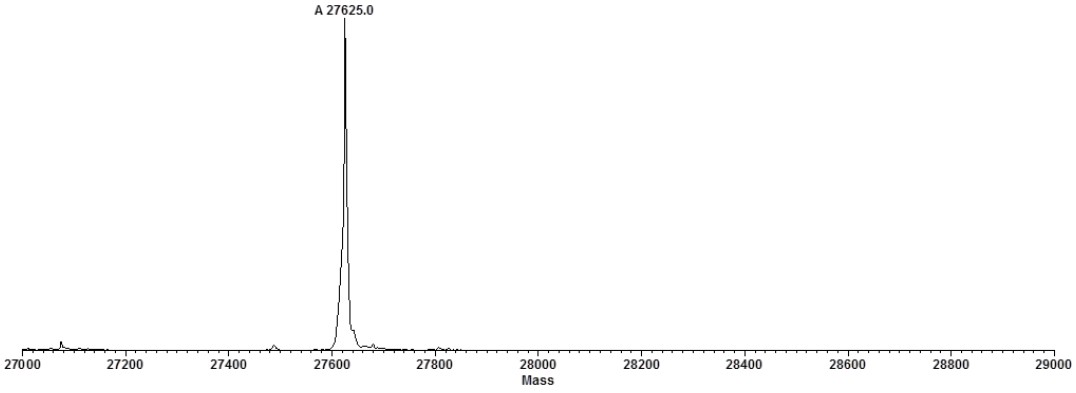


Figure 6-14. C-2)yEGFP from $tRNA^{scArg}_{CUA}$ + yEGFP39TAG

6.6 Reference

1. Graveley, B.R. and T. Maniatis, *Arginine/Serine-Rich Domains of SR Proteins Can Function as Activators of Pre-mRNA Splicing.* Molecular Cell, 1998. **1**(5): p. 765-771.
2. Truant, R. and R. Cullen Bryan, *The Arginine-Rich Domains Present in Human Immunodeficiency Virus Type 1 Tat and Rev Function as Direct Importin β-Dependent Nuclear Localization Signals.* Molecular and Cellular Biology, 1999. **19**(2): p. 1210-1217.
3. Xuan, J.W., et al., *Site-directed mutagenesis of the arginine-glycine-aspartic acid sequence in osteopontin destroys cell adhesion and migration functions.* J Cell Biochem, 1995. **57**(4): p. 680-90.
4. Baynes, B.M., D.I.C. Wang, and B.L. Trout, *Role of Arginine in the Stabilization of Proteins against Aggregation.* Biochemistry, 2005. **44**(12): p. 4919-4925.
5. Slade, D.J., et al., *Chemical and biological methods to detect post-translational modifications of arginine.* Biopolymers, 2014. **101**(2): p. 133-43.
6. Blanc, R.S. and S. Richard, *Arginine Methylation: The Coming of Age.* Mol Cell, 2017. **65**(1): p. 8-24.
7. Di Lorenzo, A. and M.T. Bedford, *Histone arginine methylation.* FEBS Letters, 2011. **585**(13): p. 2024-2031.
8. Akahoshi, A., et al., *Site-specific incorporation of arginine analogs into proteins using arginyl-tRNA synthetase.* Biochemical and Biophysical Research Communications, 2011. **414**(3): p. 625-630.
9. Mondal, S., et al., *Site-specific incorporation of citrulline into proteins in mammalian cells.* Nature Communications, 2021. **12**(1): p. 45.
10. Italia, J.S., et al., *An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes.* Nature Chemical Biology, 2017. **13**(4): p. 446-450.
11. Chan, P.P., et al., *tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes.* Nucleic Acids Res, 2021. **49**(16): p. 9077-9096.
12. Yoo, T.H. and D.A. Tirrell, *High-Throughput Screening for Methionyl-tRNA Synthetases That Enable Residue-Specific Incorporation of Noncanonical Amino Acids into Recombinant Proteins in Bacterial Cells.* Angewandte Chemie International Edition, 2007. **46**(28): p. 5340-5343.
13. Chin Jason, W., et al., *An Expanded Eukaryotic Genetic Code.* Science, 2003. **301**(5635): p. 964-967.
14. Kelemen, R.E., et al., *Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells.* bioRxiv, 2022: p. 2022.01.21.477302.

15. Bi, K., et al., *Crystal structure of E. coli arginyl-tRNA synthetase and ligand binding studies revealed key residues in arginine recognition.* Protein Cell, 2014. **5**(2): p. 151-9.
16. Eaton, D., H. Rodriguez, and G.A. Vehar, *Proteolytic processing of human factor VIII. Correlation of specific cleavages by thrombin, factor Xa, and activated protein C with activation and inactivation of factor VIII coagulant activity.* Biochemistry, 1986. **25**(2): p. 505-12.
17. Lin, B.Y., P.P. Chan, and T.M. Lowe, *tRNAviz: explore and visualize tRNA sequence features.* Nucleic Acids Research, 2019. **47**(W1): p. W542-W547.
18. Truong, F., et al., *Two-strain, cell-selective protein labeling in mixed bacterial cultures.* J Am Chem Soc, 2012. **134**(20): p. 8551-6.
19. Gorini, L., W. Gundersen, and M. Burger, *Genetics of regulation of enzyme synthesis in the arginine biosynthetic pathway of Escherichia coli.* Cold Spring Harb Symp Quant Biol, 1961. **26**: p. 173-82.
20. Wood, T.I., et al., *Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis.* Biochemistry, 2005. **44**(49): p. 16211-20.
21. Cabantous, S., T.C. Terwilliger, and G.S. Waldo, *Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.* Nature Biotechnology, 2005. **23**(1): p. 102-107.
22. Zimmer, M.H., et al., *Structural Consequences of Chromophore Formation and Exploration of Conserved Lid Residues amongst Naturally Occurring Fluorescent Proteins.* Chem Phys, 2014. **429**: p. 5-11.
23. Bi, K., et al., *Crystal structure of E. coli arginyl-tRNA synthetase and ligand binding studies revealed key residues in arginine recognition.* Protein & Cell, 2014. **5**(2): p. 151-159.

Chapter 7. Unpublished data and small story

7.1 Double suppression with pSer and Para-benzoyl phenylalanine to trap pSer reader

7.2.1 Brief introduction

Phosphorylation is one of the most abundant PTMs, and among all other targets, serine is the most phosphorylated residue. The GCE method to incorporate pSer was first developed by Söll's lab based on a unique aminoacyl–tRNA synthetase devoted to Sep-tRNA^{Cys} in *Methanogenic archaea*. Instead of a dedicated CysRS, *M. archaea* utilize a SepRS to charge pSer on tRNA^{Cys}, and then form cystine by enzymatic reaction. (Figure 7-1)

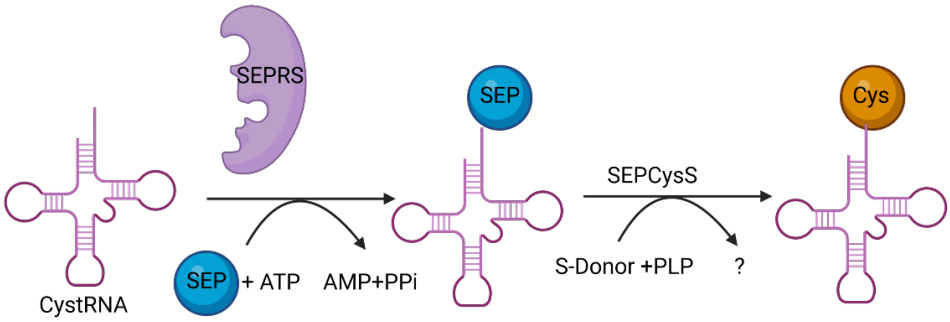


Figure 7-1. Scheme of tRNA^{Cys} formation in *Methanogenic archaea*.

Jason Chin's lab further evolved the tRNA and SEPRS to enhance the activity and specificity. A dedicated EF-Tu for SEP was also evolved to assist the suppression.[1-3] With the advance on pSer incorporation, the Rinehart lab has demonstrated a technique to screen substrates of known pSer recognition domain. Rinehart's method could efficiently enlarge the substrate pool of the known domain yet would not be an efficient

method to search for known or unknown pSer recognition domains for a specific substrate. [4] Rinehart's method could efficiently enlarge the substrate pool of the known domain yet would not be an efficient method to search for known or even unknown pSer recognition domains for a specific substrate.

We envisioned an approach to screen pSer reader protein by combining pSer incorporation with a photo-crosslinker incorporation. Para-benzoyl phenylalanine (BpA) was first incorporated by the Schultz lab with *Methanococcus jannaschii* tyrosyl-tRNA synthetase (MjTyrRS) mutant.[5] It serves as a photo-crosslinker to form a covalent band with a nearby residue upon light radiation. In Dr. Yunan Zheng's work, pBpA was utilized to trap bromo domain which recognizes nearby acetyl lysine (AcK) PTM, demonstrating that it could efficiently crosslink to partner proteins for further studies.[6] In this work, we first proved concept by incorporating both pSer and BpA into sfGFP, and then tested possibility of applying this technique 14-3-3ζ, a natural pSer containing protein.[7]

7.2.2 Result and discussion

We subcloned pSerRS/tRNA$_{CUA}$ pair into a pUltra plasmid, to pair with a pEvol plasmid that harbors MjBPA/tRNA$_{UCA}$ pair, and a pET22b plasmid that harbors the reporter gene sfGFP. The pSerRS/tRNA$_{CUA}$ and MjBPA/tRNA$_{UCA}$ pair are first tested individually with sfGFP 151TAG and sfGFP TGA respectively. Both are active and equivalent for about 20 percent of WT sfGFP overexpression and are characterized by

LC/MS (Figure 7-2 and 7-4). Moving forward, we incorporate both BPA and pSer in sfGFP 3TGA 151TAG to test the efficiency of double suppression. Even though both pairs efficiently suppress the stop codon, when introduced and induced together, the suppression efficiency is lower than 1% relative to WT sfGFP. This less than 1% efficiency on sfGFP nearly deny the possibility to perform double suppression of BPA and pSer on other proteins.

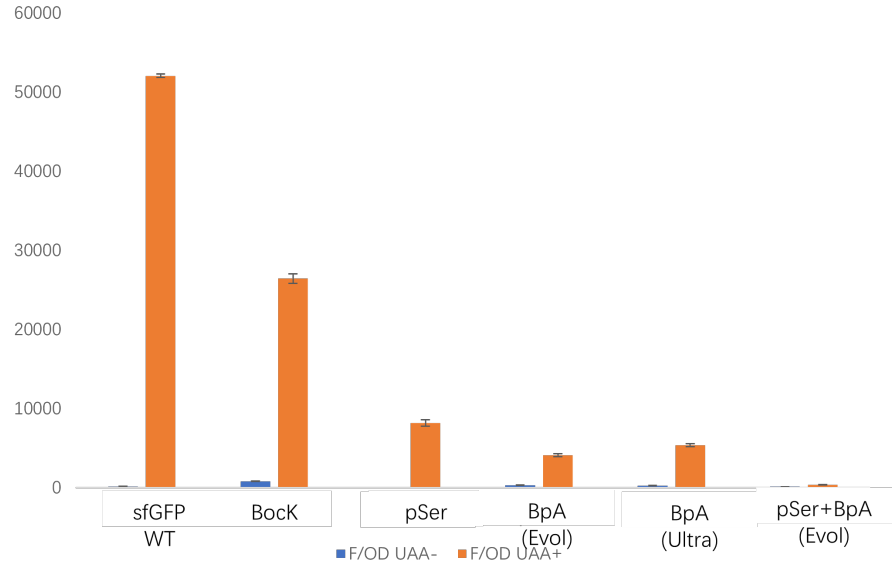


Figure 7-2. Normalized fluorescence data indicating incorporation effiency of different substrates with different construct. BocK is incorporated by MmPylRS/tRNA pair as a control. MjBPARS/tRNA pair shows comparble efficiency expressed by either pUltra or pEvol.

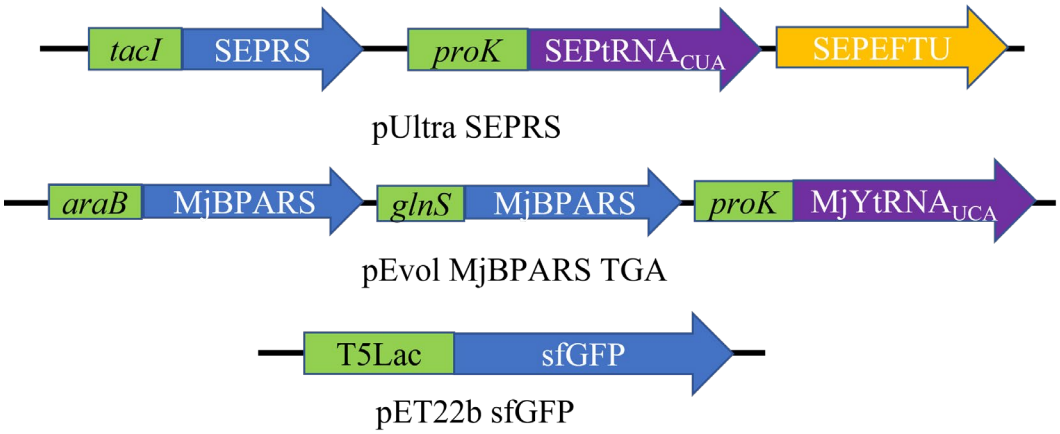


Figure 7-3. Key elements of pUltra, pEvol and pET22b vectors.

To optimize the incorporation efficiency, we focus on two aspects: facilitating suppression by removing releasing factor 1 and decreasing number of TAG in *E. coli* genome; and increase cellular pSer concentration by knocking out *serB* gene. Two *E. coli* strains were generated to boost TAG suppression efficiency, namely C321 and B95. The C321 removed all TAG codon which makes suppressor $tRNA_{CUA}$ specific to artificial TAG site that installed for ncAA incorporation.[8] However, the benefits come with growth defects and T5 promoter specificity.[9] To overcome these short coming and maintain the advantage for TAG suppression, B95 is generated by removing 95 TAG codon from BL21 strain. The suppression enhancement from B95 is comparable with C321 meanwhile share all the strengths of BL21, which are T7 promoter compatibility, normal doubling time, and optimized incubation temperate.[10] Beside TAG codon removal, both strains removed RF1, which compete with mutant suppressor tRNA. We compared the performance of B95 and C321 regard of BPA, pSer

double suppression, and the B95 outperformed C321 by increasing the efficiency for two-folds to 3.3%. Therefore, B95 is chosen as working strain for *serB* knock out.

The *serB* encodes phosphoserine phosphatase, which catalyzes the last step in serine biosynthesis by convert phosphoserine to serine. By knocking out *serB* in the B95 strain, the cellar phosphoserine is accumulated to enhance the pSer incorporation. The B95 *ΔA ΔserB::zeoR* strain is created by using lambda red system and verified by colony PCR. The suppression efficiency in the deletion of *serB* is two-folds than it in B95 *ΔA,* which indicates increasing substrate concentration is beneficial for incorporation, but it is not a game changer. The most interesting benefit from using B95 strain instead of BL21 strain is the improvement of the purity of the double suppression product. The double suppression of pSer and BPA as 25% of product containing pSer and Trp. By changing to B95 based strain, the purity of the product increase to more than 90% It is interesting because the tryptophan is incorporated by MjBPA pair, which is suppressing TGA codon, whereas the B95 strain supposedly enhance the TAG suppression. I tested single suppression with MjBPA pair, and no tryptophan incorporation was observed. (Figure 7-3)

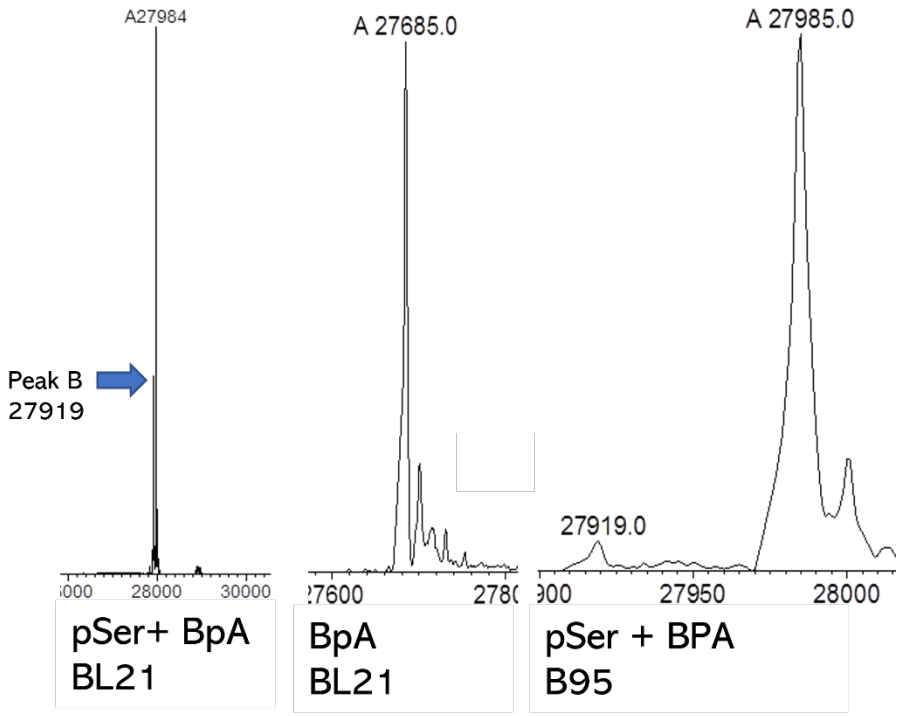


Figure 7-4. Mass spectrometry data of whole protein mass on sfGFP 3BPA 151pSer and sfGFP 3BPA. Proteins are expressed and purified from different strain as labeled.

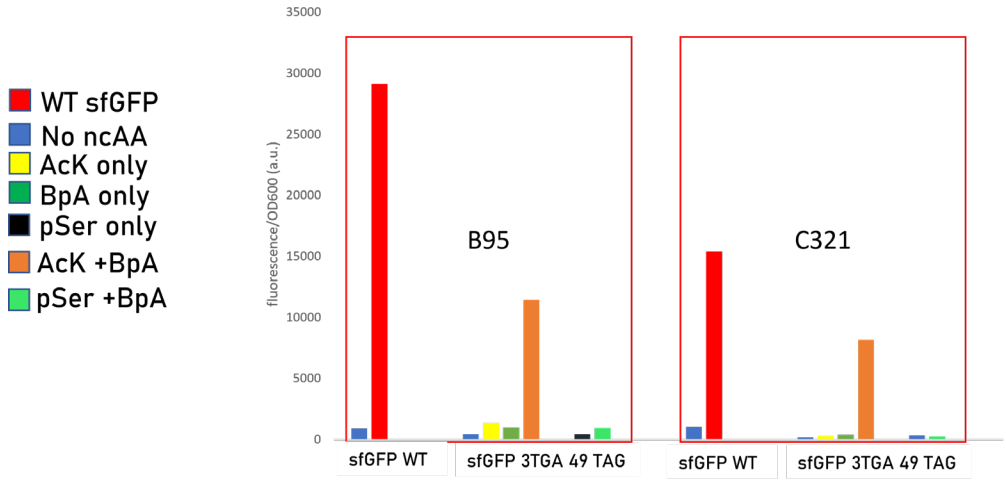


Figure 7-5. Double suppressions with different pairs, substrates, and strain to compare the efficiency.

I switched promoters for SEPRS and MjBPARS, hoping that altering the expression pattern may change the result. The pEvol plasmid contains two expression cassettes with *araB* and *glnS* promoters, and it is possible that the *glnS* promoter could help slowly accumulate SEPRS which result in boosting the ncAA containing protein yield. However, switching promoters did not improve the yield.

I also compared double incorporation of AcK and BPA, as Dr. Zheng published, with pSer and BPA in the same conditions. We first tested incorporation efficiency for each individual ncAA, AcK, BPA and pSer, and results are comparable. However, when performing dual incorporation, the efficiency of the AcK and BPA dual remain as single suppression of either ncAA; the efficiency of the pSer and BPA dual are barely detectable. In conclusion, this SEPRS/tRNA does not compatible with MjTyrRS/tRNA which unknown reason.

7.2 Incorporation of novel ncAAs with EcleuRS in mammalian cell

7.2.1 Brief introduction

The cellular signaling pathways mostly operate through interaction between biomolecules, such as protein-protein interaction, protein DNA interaction and DNA-RNA interaction and more. Strong interaction partner can be identified by technique

such as pulldown and immunoprecipitation. However, mapping and tacking transient interaction is still a challenge that calls for more tools.

Using GCE method to incorporate bio-orthogonal handle containing ncAAs, such as click chemistry handle or photo-crosslinker to trap the transient and weak protein to molecular interaction is proved to be a superior method.[11] In mammalian system, the technique is especially useful because of the complicated native environment of cells. The current available long chain azide containing ncAA is incorporated by PylRS/tRNA pair which makes it limited in application due to most of PTM ncAA is also incorporated by same pair.[12] In this work we tested two long chain azide containing ncAA (1 and 2) which was synthesized by the Woo lab (Harvard University) with our EcLeuRS based system, and demonstrated that these two are very efficient substrate.

I also tested two ncAAs (3 and 4) which is synthesized by Dr. Roy and they are active substrate for our EcLeuRS based system.

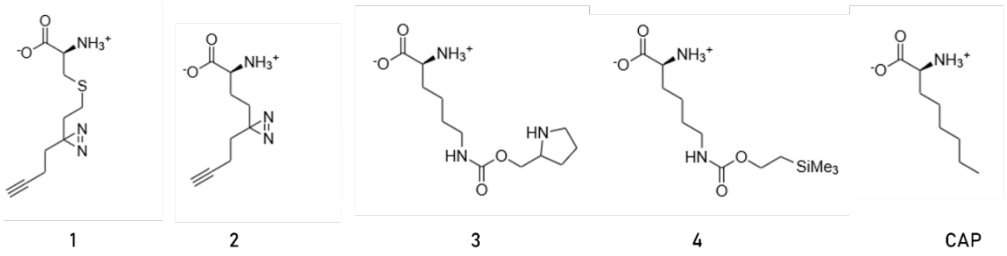


Figure 7-6. ncAAs involved in this section.

7.2.2 Result and discussion

Four ncAAs are tested to incorporate in the mammalian system with EcLeuRS/tRNA pair which Dr Zheng engineered.[13] 1 and 2 are azide containing ncAA could be beneficial for trapping PTM triggered protein-protein interactions pairing with PylRS/tRNA to incorporate PTMs. 1 and 2 are incorporated in EGFP in mammalian cells with a high efficiency, that comparable to a commonly used ncAA CAP, by both LeuRS1 and LeuRS2. The ncAA containing EGFP are purified to verify by LC/MS, and the result match with expectation. (Figure 7-6) The high incorporation efficiency of these two ncAAs by LeuRS may enable double or even triple suppression to study more complex situation such as PTM triggered PPI or even how different PTMs crosstalk affects PPI in physiological environment.

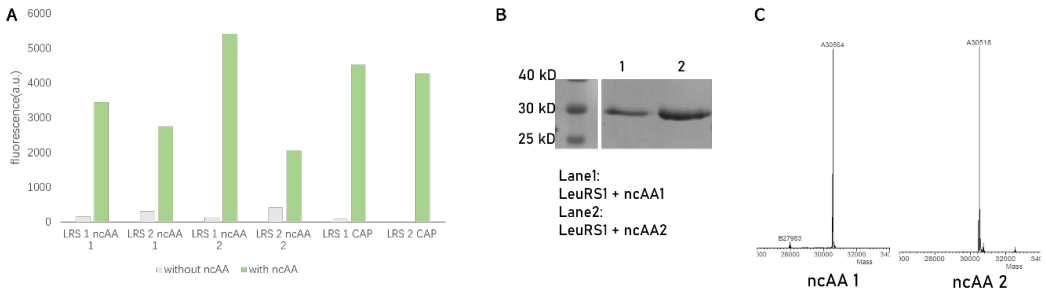


Figure 7-7. A) normalized fluorescent data for EGFP TAG expression with different EcLeuRS mutant in present and absent of ncAA. B) Purified ncAA1 or ncAA2 containing EGFP. C) Mass spectrometry for whole protein mass of ncAA containing EGFP.

Additionally, ncAAs 3 and 4 were tested in a similar manner, but the EGFP reporter expression was found to be similar in the absence and the presence of these ncAAs. In

conclusion, 3 and 4 are not suitable substrates for EcLeuRS. *O*-nitrobenzoyl arginine was also tested with EcLeuRS in a similar manner, but it failed to incorporate despite its structural similarity with *o*-nitrobenzoyl citrulline, which is accepted by EcLeuRS.[14]

7.3 Method

EGFP fluorescence analysis

EGFP fluorescence was analyzed 48 h after transfection. Cells were then harvested and resuspended in 600 μL CelLytic M buffer (Sigma, St. Louis, MO) with Halt Protease inhibitor Cocktail (Thermo Scientific, Waltham, MA) and Pierce Universal Nuclease for Cell Lysis (Fisher Scientific, Hampton, NH). Lysates were clarified by centrifugation at 16,000 × *g* for 10 min and 100 μL of supernatant was transferred to a clear-bottom 96-well plate for fluorescence measurement. Fluorescence was measured using a Zeiss AX10 microscope (Ex. 488 nm; Em. 534 nm). All the experiments were performed at least in three duplicates.

sfGFP fluorescence analysis

Colony was picked and inoculated in LB media and incubated at 250RPM, 37 ºC for 16 hours as seeding. The seeding culture was inoculated into 10 ml of LB media in 1:100 ratio and cultured in same condition until OD_{600} reaches 0.6. Then the culture was induced by addition of 1mM IPTG or 0.02% arabinose depends on the construct. The culture was then incubated at 30 ºC 250RPM for 10 hours. The sample then was

harvested by centrifugation at 5000 xg followed by resuspended in PBS. The cell density was measured at OD_{600} and diluted to 0.1. The sample was measured with OD_{600} and fluorescent channel with a setting following setting: Ex. 488 nm; Em. 534 nm. The data was normalized by using divide fluorescent reading by OD_{600}.

LC-MS analysis of GFP

LC/MS analyses of the purified GFP were performed on Agilent Technologies 6230 TOF LC/MS with Aerns 3.6 μm WIDEEPORE XB-C8 20 LC column 100 × 4.6 mm.

7.4 Reference

1. Sauerwald, A., et al., *RNA-Dependent Cysteine Biosynthesis in Archaea.* Science, 2005. **307**(5717): p. 1969-1972.
2. Rogerson, D.T., et al., *Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog.* Nat Chem Biol, 2015. **11**(7): p. 496-503.
3. Park, H.S., et al., *Expanding the genetic code of Escherichia coli with phosphoserine.* Science, 2011. **333**(6046): p. 1151-4.
4. Barber, K.W., et al., *Encoding human serine phosphopeptides in bacteria for proteome-wide identification of phosphorylation-dependent interactions.* Nature Biotechnology, 2018. **36**(7): p. 638-644.
5. Chin, J.W., et al., *Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli.* Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11020-4.
6. Zheng, Y., et al., *Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis.* ACS Chem Biol, 2018. **13**(5): p. 1137-1141.
7. Pennington, K.L., et al., *The dynamic and stress-adaptive signaling hub of 14-3-3: emerging mechanisms of regulation and context-dependent protein–protein interactions.* Oncogene, 2018. **37**(42): p. 5587-5604.
8. Wannier, T.M., et al., *Adaptive evolution of genomically recoded Escherichia coli.* Proceedings of the National Academy of Sciences, 2018. **115**(12): p. 3090-3095.
9. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* Science, 2013. **342**(6156): p. 357-60.

10. Mukai, T., et al., *Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon.* Sci Rep, 2015. **5**: p. 9699.
11. Wang, S., A.O. Osgood, and A. Chatterjee, *Uncovering post-translational modification-associated protein-protein interactions.* Curr Opin Struct Biol, 2022. **74**: p. 102352.
12. Hoffmann, J.-E., et al., *A Bifunctional Noncanonical Amino Acid: Synthesis, Expression, and Residue-Specific Proteome-wide Incorporation.* Biochemistry, 2018. **57**(31): p. 4747-4752.
13. Zheng, Y., et al., *Expanding the Scope of Single- and Double-Noncanonical Amino Acid Mutagenesis in Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases.* Biochemistry, 2018. **57**(4): p. 441-445.
14. Mondal, S., et al., *Site-specific incorporation of citrulline into proteins in mammalian cells.* Nature Communications, 2021. **12**(1): p. 45.

Appendix I

Recombination cassettes

Chapter 4 Recombination Cassettes

Sequences of various DNA elements:

The following sequences are the dsDNA PCR products that were electroporated for recombination. Primers used are listed in MM and primer list. Important features are mentioned prior to the sequence with color code in parenthesis.

lysU::Gent^{R} PCR cassettes:

Gent^{R} (green);

CTCTTTATCTACGCTAAATTGAAAGCTGGATTTAGAGGAACCAAACCCTA
ATTACGCACACCGTGGAAACGGATGAAGGCACGAACCCAGTTGACATAA
GCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAA
CTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGG
TTTTCATGGCTTGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATC
CAAGCAGCAAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCA
GCAACGATGTTACGCAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCC
TAAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTCGCACATGTAGGC
TCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTTGATCTTTTCGGTC
GTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCCGAT
TACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTC
GACCAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAGGTTT
GAGCAGCCGCGTAGTGAGATCTATATCTATGATCTCGCAGTCTCCGGCGA
GCACCGGAGGCAGGGCATTGCCACCGCGCTCATCAATCTCCTCAAGCATG
AGGCCAACGCGCTTGGTGCTTATGTGATCTACGTGCAAGCAGATTACGGT

GACGATCCCGCAGTGGCTCTCTATACAAAGTTGGGCATACGGGAAGAAGT
GATGCACTTTGATATCGACCCAAGTACCGCCACCTAACAATTCGTTCAAG
CCGAGATCGGCTTCCCGGATTTCACTTTAATGAACGAAGCAGTCAGGCGA
CTGCTTCGTTCAT

lysS::Zeo^{R} *PCR* cassettes

EM-7 promoter (red); Zeo^{R} (green); CYC1 terminator (blue);

CAATTTATCGAAGCAAGTTTGAAAGCAGGGTTATGAGGAACCAACGGTGT
TGACAATTAATCATCGGCATAGTATATTGGCATAGTATAATACGACAAGG
TGAGGAACTAAACCATGGCCAAGCTGACCAGTGCCGTTCCGGTGCTCACC
GCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTT
CTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACG
TGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACC
CTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTC
GGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCG
AGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGC
CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTCCGAC
GGCGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTC
GATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACA
TCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCC
TATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAA
ATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGA
AAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAG
CTGCATTACGTTATGCTCACAACCCCGGCAAATGTCGGGGTTTTTTT

Chapter 5 Recombination Cassettes

pheST::Zoe^{R} PCR cassettes

EM-7 promoter (red); Zeo^{R} (green); CYC1 terminator (blue);

TAAGTCCAACGAACCAGTGTCACCACTGACACAATGAGGAAAACCGGTGT
TGACAATTAATCATCGGCATAGTATATTGGCATAGTATAATACGACAAGG
TGAGGAACTAAACCATGGCCAAGCTGACCAGTGCCGTTCCGGTGCTCACC
GCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTT
CTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACG
TGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACC
CTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTC
GGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCG

AGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGC
CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTCCGAC
GGCGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTC
GATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACA
TCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCC
TATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAA
ATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGA
AAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAG
CTACCTATGGCGCTTACAAAAGCTGAAATGTCAGAATATCTGTTTGA

Chapter 6 Recombination Cassettes

argH::Zoe^{R} PCR cassettes

EM-7 promoter (red); Zeo^{R} (green); CYC1 terminator (blue);

GGCGCAGCTTTCCGGCATTGAATTTCAAAATAAGGAAACAGAGTTGGTGT
TGACAATTAATCATCGGCATAGTATATTGGCATAGTATAATACGACAAGG
TGAGGAACTAAACCATGGCCAAGCTGACCAGTGCCGTTCCGGTGCTCACC
GCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTT
CTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACG
TGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACC
CTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTC
GGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCG
AGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGC
CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTCCGAC
GGCGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTC
GATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACA
TCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCC
TATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAA
ATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGA
AAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAG
CTGAACATTTATATGTATAAATTTGAGCCTGGCTTATCGCCGGGCTT

Chapter 7 Recombination Cassettes

SerB::Zoe^{R} PCR cassettes
EM-7 promoter (red); Zeo^{R} (green); CYC1 terminator (blue)
CCCTGCATCAGCGCCACTAACAAGGCAAGACAGAACAGGACAATCGGTG
TTGACAATTAATCATCGGCATAGTATATtGGCATAGTATAATACGACAAGG
TGAGGAACTAAACCATGGCCAAGCTGACCAGTGCCGTTCCGGTGCTCACC

GCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTT
CTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACG
TGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACC
CTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTC
GGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCG
AGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGC
CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTCCGAC
GGCGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTC
GATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACA
TCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCC
TATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAA
ATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGA
AAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAG
CTACCATCACCGAGGTGCGTCTTGCTGCGTCGCCAATGGTGGCGCGT

SerC::Zoe^{R} PCR cassettes

EM-7 promoter (red); Zeo^{R} (green); CYC1 terminator (blue)

ATCGATTGACCGCGGGTTAATAGCAACGCAACGTGGTGAGGGGAAGGTGT
TGACAATTAATCATCGGCATAGTATATtGGCATAGTATAATACGACAAGGT
GAGGAACTAAACCATGGCCAAGCTGACCAGTGCCGTTCCGGTGCTCACCG
CGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTC
TCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGT
GACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCC
TGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCG
GAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGA
GATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCC
GGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTCCGACG
GCGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTCG
ATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACAT
CCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCT
ATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAA
TTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGAA
AACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAGCT
TGCCGAAATTTTGCTTAATCCCCACAGCCAGCCTGTGGGGTTTTT

Appendix II

Oligos List

Chapter 2

PAD4WT-Forward:
ATTATTAGAATTGGCCAAGGAGGCCACCATGGACTACAAGGACGACGAC
GACAAG
PAD4WT-Reverse:
ATTATTAGAATTCGGCCTTAGAGGCCTCAGTGGTGGTGGTGGTGGTGGTG
GTGGTGGTGGTGGTGGGGCACCATGTTCCACCACTTGAA
PAD4 R372TAG Inner Forward:
GTCTTCGACTCTCCTTAGAACTAGGGCCTGAAG
PAD4 R372TAG Inner Reverse:
CTTCAGGCCCTAGTTCTAAGGAGAGTCGAAGAC
PAD4 R374TAG Inner Forward:
GTCTTCGACTCTCCTAGGAACTAGGGCCTGAAG
PAD4 R374TAG Inner Reverse:
CTTCAGGCCCTAGTTCCTAGGAGAGTCGAAGAC

Chapter 3

pIDSMART Forward: CCCGTGTAAAACGACGGCCAGTTTATC
pIDSMART Reverse: GCCGTAAGATCACGGGTCGCAGC
Ala tRNA iR:
GGAGAGCGCCTGCCTCTAAAGCAGGAGGTCTGCGGTTCGATCCCGCATAG
CTCCATTTTTTGCTAGCGGATCGACGAGAGC
Ala tRNA iF
GAACCGCAGACCTCCTGCTTTAGAGGCAGGCGCTCTCCCAGCTGAGCTAT
AGCCCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC
Gly tRNA iR
GTAGAGCACGACCCTCTAAAGGTCGGGGTCGCGAGTTCGAGTCTCGTTTC
CCGCTTTTTTTGCTAGCGGATCGACGAGAGC
Gly tRNA iF
GAACTCGCGACCCCGACCTTTAGAGGGTCGTGCTCTACCAACTGAGCTAT
TCCCGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC
Pro tRNA iR
GTAGCGCAACTGGCTCTAAACCAGTGGGTCGGAGGTTCGAATCCTCCCTC
GCCGATTTTTTGCTAGCGGATCGACGAGAGCAGCGCG
Pro tRNA iF
GAACCTCCGACCCACTGGTTTAGAGCCAGTTGCGCTACCAAGCTGCGCTA
CTCGCCGGGTGTTTCGTCCTTTCCAC
Thr tRNA iR
GTAGAGCGCACCCCTCTAAAGGGTGAGGTCCCCAGTTCGACTCTGGGTAT
CAGCATTTTTTGCTAGCGGATCGACGAGAGC

Thr tRNA iF

GAACTGGGGACCTCACCCTTTAGAGGGGTGCGCTCTACCAACTGAGCCAT
ATCAGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

Val tRNA iR

GTTAGAGCACCACCCTCTAAAGGTGGGGGTCGTTGGTTCGAGTCCAACTG
AACGCATTTTTTGCTAGCGGATCGACGAGAGC

Val tRNA iF

GAACCAACGACCCCCACCTTTAGAGGGTGGTGCTCTAACCAACTGAGCTA
TGAACGCGGTGTTTCGTCCTTTCCACAAGATATATAAAG

Phe tRNA iR

GTAGAGCAGGGGACTCTAAATCCCCGTGTCCTTGGTTCGATTCCGAGTCC
GGGCATTTTTTGCTAGCGGATCGACGAGAGC

Phe tRNA iF

CGAACCAAGGACACGGGGATTTAGAGTCCCCTGCTCTACCGACTGAGCTA
TCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAG

Asn tRNA iR

GTAGAACGGCGGACTCTAAATCCGTATGTCACTGGTTCGAGTCCAGTCAG
AGGAGTTTTTTGCTAGCGGATCGACGAGAGC

Asn tRNA iF

GAACCAGTGACATACGGATTTAGAGTCCGCCGTTCTACCGACTGAACTAC
AGAGGAGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

Asp tRNA iR

GTTAGAATACCTGCCTCTAAAGCAGGGGGTCGCGGGTTCGAGTCCCGCCC
GTTCCGTTTTTTGCTAGCGGATCGACGAGAGC

Asp tRNA iF

GAACCCGCGACCCCCTGCTTTAGAGGCAGGTATTCTAACCGACTGAACTA
CCGCTCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

Glu tRNA iR

CCCAGGACACCGCCCTCTAAAGGCGGTAACAGGGGTTCGAATCCCCTAGG
GGACGTTTTTTGCTAGCGGATCGACGAGAGC

Glu tRNA iF

GAACCCCTGTTACCGCCTTTAGAGGGCGGTGTCCTGGGCCTCTAGACGAA
GGGGACGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

His tRNA iR

GTAGAGCCCTGGACTCTAAATCCAGTTGTCGTGGGTTCGAATCCCATTAGC
CACCTTTTTTGCTAGCGGATCGACGAGAGC

His tRNA iF

GAACCCACGACAACTGGATTTAGAGTCCAGGGCTCTACCAACTGAGCTAT
AGCCACGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

His tRNA iF G1

GAACCCACGACAACTGGATTTAGAGTCCAGGGCTCTACCAACTGAGCTAT
AGCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

Gln tRNA iR

GTAAGGCACCGGACTCTAAATCCGGCATTCCGAGGTTCGAATCCTCGTAC
CCCAGTTTTTTGCTAGCGGATCGACGAGAGC

Gln tRNA iF

GAACCTCGGAATGCCGGATTTAGAGTCCGGTGCCTTACCGCTTGGCGATA
CCCCAGGTGTTTCGTCCTTTCCACAAGATATATAAAGCC

Cys tRNA iR

GTTATGTAGCGGACTCTAAATCCGCCTAGTCCGGTTCGACTCCGGAACGC
GCCTTTTTTTGCTAGCGGATCGACGAGAGC

Cys tRNA iF

CGAACCGGACTAGGCGGATTTAGAGTCCGCTACATAACCGCTTTGTTAAC
GCGCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCC

Ser tRNA iR

CTGAAGGCGCTCCCCTCTAAAGGGAGTATGCGGTCAAAAGCTGCATCCGG
GGTTCGAATCCCCGCCTCACCGTTTTTTGCTAGCGGATCGACGAGAGC

Ser tRNA iF

CGAACCCCGGATGCAGCTTTTGACCGCATACTCCCTTTAGAGGGGAGCGC
CTTCAGCCTCTCGGCCACCTCACCGGTGTTTCGTCCTTTCCACAAGATATA
TAAAGCCAAGAAATC

Arg tRNA iR

GATAGAGTACTCGGCTCTAAACCGAGCGGTCGGAGGTTCGAATCCTCCCG
GATGCATTTTTTGCTAGCGGATCGACGAGAGC

Arg tRNA iF

GAACCTCCGACCGCTCGGTTTAGAGCCGAGTACTCTATCCAGCTGAGCTA
CGGATGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

Leu tRNA iR

GTAGACGCGCTAGCCTCTAAAGTTAGTGTTCTTACGGACGTGGGGGTTCA
AGTCCCCCCCCCTCGCATTTTTTGCTAGCGGATCGACGAGAGC

Leu tRNA iF

CGTCCGTAAGAACACTAACTTTAGAGGCTAGCGCGTCTACCAATTCCGCC
ACCTTCGCGGTGTTTCGTCCTTTCCACAAGATATATAAAG

Ile tRNA iR

GTTAGAGCGCACCCCTCTAAAGGGTGAGGTCGGTGGTTCAAGTCCACCCA
GGCCTATTTTTTGCTAGCGGATCGACGAGAGC

Ile tRNA iF

GAACCACCGACCTCACCCTTTAGAGGGGTGCGCTCTAACCACCTGAGCTA
CAAGCCTGGTGTTTCGTCCTTTCCACAAGATATATAAAGC

Met tRNA iR

GTTAGAGCAGGCGACTCTAAATCGCCTGGTCGCTGGTTCAAGTCCAGCAA
GGGCCATTTTTTGCTAGCGGATCGACGAGAGC

Met tRNA iF

GAACCAGCGACCAGGCGATTTAGAGTCGCCTGCTCTAACCACTGAGCTAA
AGGGCCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCC

Lys tRNA IR

GGTAGAGCAGTTGACTCTAAATCAATTGGTCGCAGGTTCGAATCCTGCAC
GACCCACCATTTTTTGCTAGCGGATCGACGAGAGC

Lys tRNA IF

GAACCTGCGACCAATTGATTTAGAGTCAACTGCTCTACCAACTGAGCTAA
CGACCCGGTGTTTCGTCCTTTCCACAAGATATATAAAG

Trp tRNA iR

CACCGGTCTCTAAAACCGGGTGTTGGGAGTTCGAGTCTCTCCGCCCCTGCC
ATTTTTTGCTAGCGGATCGACGAGAG

Trp tRNA iF

CAACACCCGGTTTTAGAGACCGGTGCTCTACCAATTGAACTACGCCCCTG
GTGTTTCGTCCTTTCCACAAGATATATAAAG

Tyr tRNA iR

GGAGCAGACTCTAAATCTGCCGTCATCGACTTCGAAGGTTCGAATCCTTCC
CCCACCACCATTTTTTGCTAGCGGATCGACGAGAGC

Tyr tRNA iF

GAAGTCGATGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAA
CCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCC

AlaRS NheI-F

AATAATAGGCTAGCGCCGCCACCATGAGCAAGAGCACCGCTGAGATC

AlaRS EcoRI-R

AATAATAGTTGAATTCTTATTGCAATTTCGCGCTGACCCAGCC

GlyRSa NheI-F

AATAATAGGCTAGCGCCGCCACCATGTCTGAGAAAACTTTTCTGGTGGAA
ATCGGC

GlyRSa SalI-R

AATAATAAAGTCGACTTATTGCAACAGCGAAATATCCGCAACGC

ProRSa NheI-F

AATAATAGGCTAGCGCCGCCACCATGCGTACTAGCCAATACCTGCTCTCC
AC

ProRSa SalI-R

AATAATAAAAGTCGACTTAGCCTTTAATCTGTTTCACCAGATATTCGACGA
TG

ThrRS NheI-F

AATAATAGGCTAGCGCCGCCACCATGCCTGTTATAACTCTTCCTGATGGCA
GCC
ThrRS EcoRI-R
AATAATAGACGAATTCTTATTCCTCCAATTGTTTAAGACTGCGGCTGC
ValRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGGAAAAGACATATAACCCACAAGA
TATCGAACAGCC
ValRS EcoRI-R
AATAATAGACGAATTCTTACAGCGCGGCGATAACAGCC
PheRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGTCACATCTCGCAGAACTGGTTGCC
PheRS EcoRI-R
AATAATAGACGAATTCTTATTTAAACTGTTTGAGGAAACGCAGATCGTTTT
CG
AsnRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGTCACATCTCGCAGAACTGGTTGCC
AsnRS EcoRI-R
AATAATAGACGAATTCTTATTTAAACTGTTTGAGGAAACGCAGATCGTTTT
CGAAG
AspRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGCGTACAGAATATTGTGGACAGCTC
CG
AspRS EcoRI-R
AATAATAGACGAATTCTTAGTTATTCTCAGCCTTCTTCACAACCTGAATGC
GluRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGAAAATCAAAACTCGCTTCGCGCCA
AG
GluRS EcoRI-R
AATAATAGACGAATTCTTACTGCTGATTTTCGCGTTCAGCAATAAAATCCA
G
HisRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGGCAAAAAACATTCAAGCCATTCGC
GG
HisRS EcoRI-R
GACGAATTCTTAACCCAGTAACGTGCGCAAATGCG
GlnRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGAGTGAGGCAGAAGCCCGC
GlnRS EcoRI-R
AATAATAGACGAATTCTTACTCGCCTACTTTCGCCCAGGTATC
CysRS NheI-F

AATAATAGGCTAGCGCCGCCACCATGCTAAAAATCTTCAATACTCTGACA
CGCCAAAAAGAG
CysRS EcoRI-R
AATAATAGACGAATTCTTACTTACGACGCCAGGTGGTCCC
SerRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGCTCGATCCCAATCTGCTGCGTAAT
G
SerRS EcoRI-R
AATAATAGACGAATTCTTAGCCAATATATTCCAGTCCGTTCATATACGGAC
G
ArgRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGAATATTCAGGCTCTTCTCTCAGAA
AAAGTCCGTC
ArgRS EcoRI-R
AATAATAGACGAATTCTTACATACGCTCTACAGTCTCAATACCCAGCG
IleRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGAGTGACTATAAATCAACCCTGAAT
TTGCCGGAAAC
IleRS EcoRI-R
AATAATAGACGAATTCTTAGGCAAACTTACGTTTTTCACCGTCACCG
MetRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGACTCAAGTCGCGAAGAAAATTCTG
GTGAC
MetRS EcoRI-R
AATAATACGACGAATTCTTATTTCACCTGATGACCCGGTTTAGCACC
TrpRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGACTAAGCCCATCGTTTTTAGTGGC
GC
TrpRS EcoRI-R
AATAATAGTTGAATTCTTACGGCTTCGCCACAAAACCAATCGCTTC
TyrRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGGCAAGCAGTAACTTGATTAAACAA
TTG
TyrRS EcoRI-R
AATAATAGACGAATTCTTATTTCCAGCAAATCAGACAGTAATTCTTTTTAC
CGC
LeuRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGCAAGAGCAATACCGCCCGG
LeuRS EcoRI-R
AATAATAGACGAATTCTTAGCCAACGACCAGATTGAGGAGTTTACCTG
GlyRSb NheI-F

AATAATAGGCTAGCGCCGCCACCATGCAAAAGTTTGATACCAGGACCTTC
CAGG
GlyRSb SalI-R
AATAATAAAGTCGACTTACTTATCTTTGTTGCACATCGGGAAGCCG
PheRSb NheI-F
AATAATAGGCTAGCGCCGCCACCATGAAATTCAGTGAACTGTGGTTACGC
GAATGG
PheRSb EcoRI-R
AATAATAGACGAATTCTAAATCCCTCAATGATGCCTGGAATCGCTC
General AvrII-F
AATAATACCTAGGTGCGCTGCTTCGCGATG
General AvrII-R
AATAATACCTAGGGATCCGCTAGGGTTCTTTCCGCCTCAGAAG
LysRS SalI-R
AATAATAGTCGACTTATTTTACCGGACGCATCGCCGGG
LysRS NheI-F
AATAATAGGCTAGCGCCGCCACCATGTCTGAACAACACGCACAGGGCG
Lys tRNA IR
GGTAGAGCAGTTGACTCTAAATCAATTGGTCGCAGGTTCGAATCCTGCAC
GACCCACCATTTTTTGCTAGCGGATCGACGAGAGC
Lys tRNA IF
GAACCTGCGACCAATTGATTTAGAGTCAACTGCTCTACCAACTGAGCTAA
CGACCCGGTGTTTCGTCCTTTCCACAAGATATATAAAG

Chapter 4

LysS ZeoR F:
CAATTTATCGAAGCAAGTTTGAAAGCAGGGTTATGAGGAACCAACGGTGTTGAC
AATTAATCATCGGCATAGTATATtGGCATAGTATAA
LysS ZeoR R:
AAAAAAACCCCGACATTTGCCGGGGTTGTGAGCATAACGTAATGCAGCTTGCAA
ATTAAAGCCTTCGAGCGTCCCAAAACCTTCTCAAGC
ZeoR IR
GAACAGGGTCACGTCGTCCCGGACC
ZeoR IF
CATCAGCGCGGTCCAGGACCAGGTG
LysU Gent F
CTCTTTATCTACGCTAAATTGAAAGCTGGATTTAGAGGAACCAAACCCTAATTAC
GCACACCGTGGAAACGGATGAAGGCACGAACCCAG
LysU Gent R
TGAACGAAGCAGTCGCCTGACTGCTTCGTTCATTAAAGTGAAATCCGGGAAGCCG
ATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACT

GentR IF
CTCTTGATCTTTTCGGTCGTGAGTTC
GentR IR
GTTGGGAGTAGGTGGCTACGTCTCC
LysU 5UTR
TATGCTGGCAACAGGCGCGGTCGCC
LysU 3UTR
GTTGTTTCTTTTGATTATCCATCA
LysS 5UTR
GGAAGATAACAAATCCGACATCGGCTGGGGCAGCCAG
LysS 3UTR
GCAGTGATCCGCCGTTGCCGTTATGAGCTTGGC

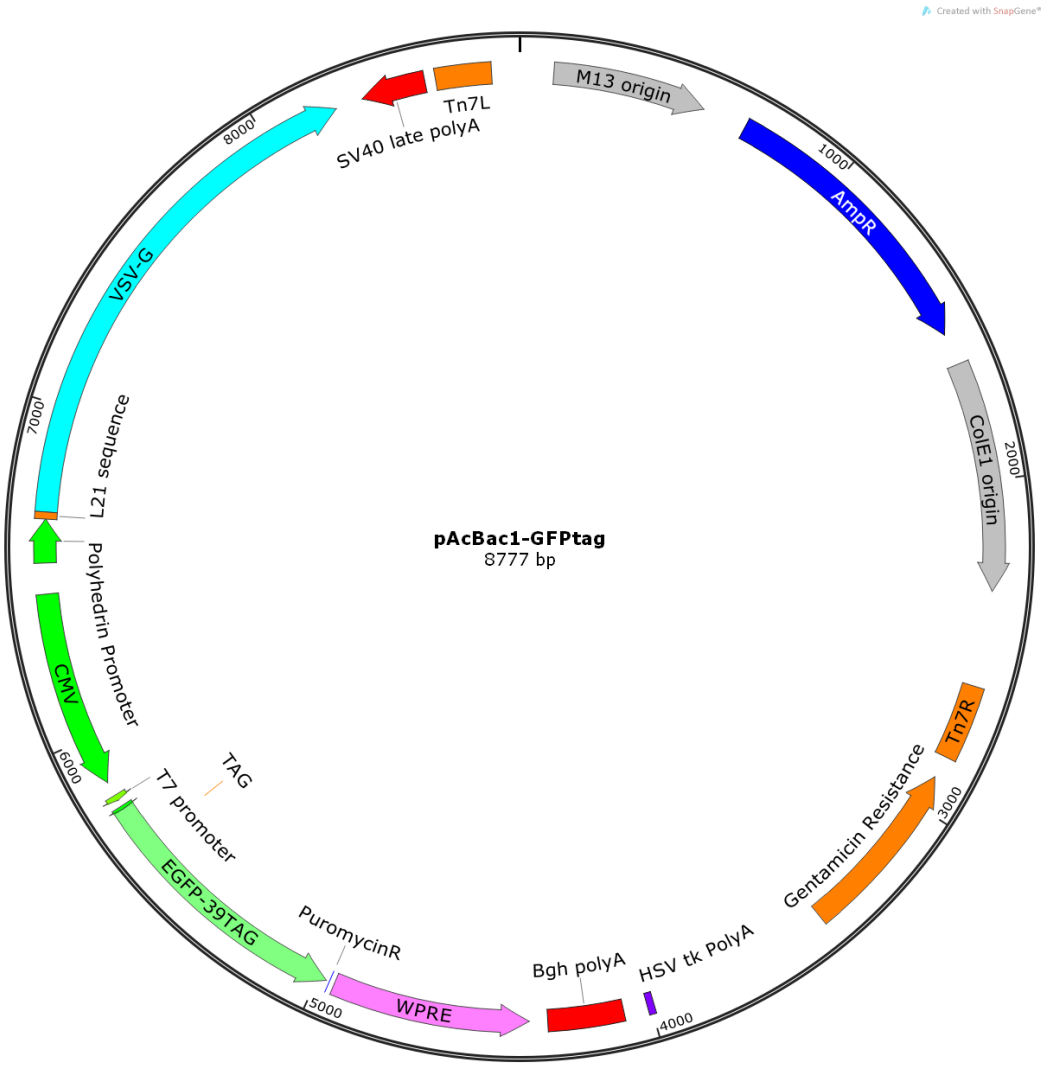
Chapter 5
PheST ZoeF
TAAGTCCAACGAACCAGTGTCACCACTGACACAATGAGGAAAACCGGTGT
TGACAATTAATCATCGGCATAGTATATtGGCATAGTATAA
PheST ZoeR
TCAAACAGATATTCTGACATTTCAGCTTTTGTAAGCGCCATAGGTAGCTTG
CAAATTAAAGCCTTCGAGCGTCCCAAAACCTTCTCAAGC
PheST 5UTR
GCCTGTCTCATCAGGTGTAATTGTCGCAGCCAGACTGGAC
PheST 3UTR
CAGATCGAAGTTACCAAAACCAGAGAGTTTCACCTGTTCG

Chapter 6
ArgH ZeoR F
GGCGCAGCTTTCCGGCATTGAATTTCAAAATAAGGAAACAGAGTTGGTGT
TGACAATTAATCATCGGCATAGTATATtGGCATAGTATAA
ArgH ZeoR R
AAGCCCGGCGATAAGCCAGGCTCAAATTTATACATATAAATGTTCAGCTT
GCAAATTAAAGCCTTCGAGCGTCCCAAAACCTTCTCAAGC
ArgH 5UTR
ATGCGGAGCAGCTTCCGGCACTGTT
ArgH 3UTR
CCTCCATCGCCACGATAGTTCATGG

Appendix III

Chapter 2

pAcBac1-EGFP Y39TAG



TTCTCTGTCACAGAATGAAAATTTTTCTGTCATCTCTTCGTTATTAATGTTT
GTAATTGACTGAATATCAACGCTTATTTGCAGCCTGAATGGCGAATGGGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCT
TCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAA
AAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGA
CGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTT
ATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAG
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCT
AAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG

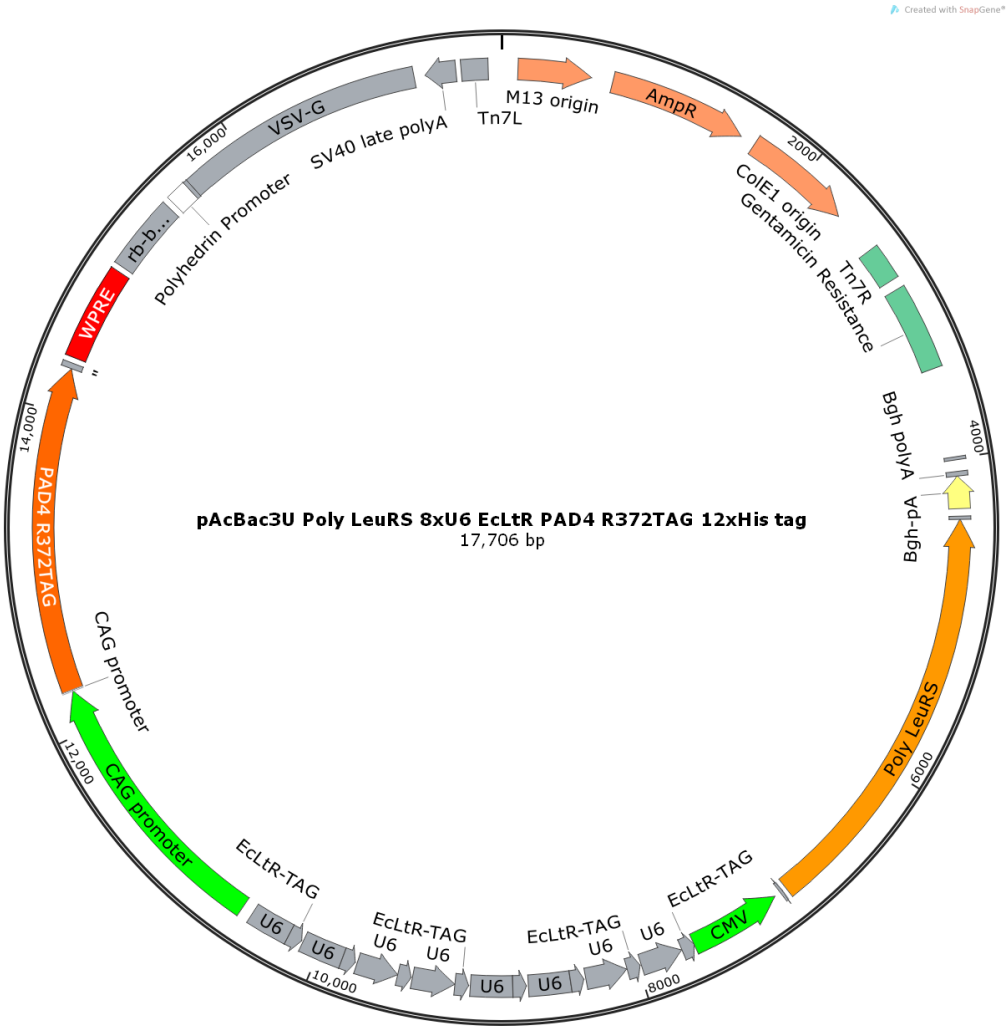
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGT
CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCA
GAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAG
TGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTT
CGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT
GGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCG
CATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAA
AGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGG
ACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTC
GCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG
CGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGA
TGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCT
GGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATC
GCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT
TTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGG
ATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC
TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTC
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGA
AAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGC
GGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACG
CCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTC
GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGC
AACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT
TCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGA
GTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCA
TCTGTGCGGTATTTCACACCGCAGACCAGCCGCGTAACCTGGCAAAATCG
GTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAGCGGGTGTGGGCGGA
CAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT
CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGT

GAAAAAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCT
GAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATG
GTAAAGACTATATTCGCGGCGTTGTGACAATTTACCGAACAACTCCGCGG
CCGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGG
TCGATATCAAAGTGCATCACTTCTTCCCGTATGCCCAACTTTGTATAGAGA
GCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGC
ACCAAGCGCGTTGGCCTCATGCTTGAGGAGATTGATGAGCGCGGTGGCAA
TGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGATATAGATC
TCACTACGCGGCTGCTCAAACCTGGGCAGAACGTAAGCCGCGAGAGCGCC
AACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGATGAATGTCTTACTAC
GGAGCAAGTTCCCGAGGTAATCGGAGTCCGGCTGATGTTGGGAGTAGGTG
GCTACGTCTCCGAACTCACGACCGAAAAGATCAAGAGCAGCCCGCATGGA
TTTGACTTGGTCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTG
AGCCACCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGC
TGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGC
GTAACGCGCTTGCTGCTTGGATGCCCGAGGCATAGACTGTACAAAAAAAC
AGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGT
TCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTAC
AGTTTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTTCATCCGT
TTCCACGGTGTGCGTCACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGC
ATTTCTGTCCTGGCTGGCGAACGAGCGCAAGGTTTCGGTCTCCACGCATCG
TCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGCAAGGTGCTGTGCACGGA
TCTGCCCTGGCTTCAGGAGATCGGTAGACCTCGGCCGTCGCGGCGCTTGC
CGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAA
GGCGAGCATCGTTTGTTCGCCCAGGACTCTAGCTATAGTTCTAGTGGTTGG
CTACGTACCCGTAGTGGCTATGGCAGGGCTTGCGCTTAATGCGCCGCTAC
AGGGCGCGTGGGGATACCCCCTAGAGCCCCAGCTGGTTCTTTCCGCCTCA
GAAGCCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCA
ATCCTCCCCCTTGCTGTCCTGCCCCACCCCACCCCCCAGAATAGAATGACA
CCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTAGGAAAGGACAGT
GGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAACA
ACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGTTTA
AACGGGCCCTCTAGACTCGAGTTAAAGTCGACGCGGGGAGGCGGCCCAA
AGGGAGATCCGACTCGTCTGAGGGCGAAGGCGAAGACGCGGAAGAGGCC
GCAGAGCCGGCAGCAGGCCGCGGGAAGGAAGGTCCGCTGGATTGAGGGC
CGAAGGGACGTAGCAGAAGGACGTCCCGCGCAGAATCCAGGTGGCAACA
CAGGCGAGCAGCCAAGGAAAGGACGATGATTTCCCCGACAACACCACGG
AATTGTCAGTGCCCAACAGCCGAGCCCCTGTCCAGCAGCGGGCAAGGCAG
GCGGCGATGAGTTCCGCCGTGGCAATAGGGAGGGGGAAAGCGAAAGTCC
CGGAAAGGAGCTGACAGGTGGTGGCAATGCCCCAACCAGTGGGGGTTGC
GTCAGCAAACACAGTGCACACCACGCCACGTTGCCTGACAACGGGCCACA

ACTCCTCATAAAGAGACAGCAACCAGGATTTATACAAGGAGGAGAAAAT
GAAAGCCATACGGGAAGCAATAGCATGATACAAAGGCATTAAAGCAGCG
TATCCACATAGCGTAAAAGGAGCAACATAGTTAAGAATACCAGTCAATCT
TTCACAAATTTTGTAATCCAGAGGTTGATTGTCGACTTAACGCGTTGAATT
CTTATCAATGGTGATGGTGATGATGACCGGTATGCATATTCAGATCCTCTT
CTGAGATGAGTTTTTGTTCGAAGGGCCCCTTGTACAGCTCGTCCATGCCGA
GAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGC
TTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTG
TCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTG
GTCGGCGAGCTGCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTT
CACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATATAGACGTTGTGGCT
GTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAA
GTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTT
CACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAGATGGTGC
GCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGC
TTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGGGT
GGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATG
AACTTCAGGGTCAGCTTGCCCTAAGTGGCATCGCCCTCGCCCTCGCCGGA
CACGCTGAACTTGTGGCCGTTTACGTCGCCGTCCAGCTCGACCAGGATGG
GCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCACCATGGTGGCGGCG
CTAGCCAGCTTGGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGCCA
GTAAGCAGTGGGTTCTCTAGTTAGCCAGAGAGCTCTGCTTATATAGACCTC
CCACCGTACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTAC
GACATTTTGGAAAGTCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTGA
CGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCA
CGCCCATTGATGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGAC
TAATACGTAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACT
GGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGC
GTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGT
AAATAGTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTAT
GGGAACATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTC
AGCCAGGCGGGCCATTTACCGTAAGTTATGTAACGCGGAACTCCATATAT
GGGCTATGAACTAATGACCCCGTAATTGATTACTATTAATAACTAGTCAAT
AATCAATGTCAACGCGTATATCTGGCCCGTACATCGCGAAGCAGCGCAAA
ACGGATCCTGCAGGTATTTGCGGCCGCGGTCCGTATACTCCGGAATATTA
ATAGATCATGGAGATAATTAAAATGATAACCATCTCGCAAATAAATAAGT
ATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAATATTCC
GGATTATTCATACCGTCCCACCATCGGGCGCGAACTCCTAAAAAACCGCC
ACCATGAAGTGCCTTTTGTACTTAGCCTTTTTATTCATTGGGGTGAATTGC
AAGTTCACCATAGTTTTTCCACACAACCAAAAAGGAAACTGGAAAAATGT
TCCTTCTAATTACCATTATTGCCCGTCAAGCTCAGATTTAAATTGGCATAA

TGACTTAATAGGCACAGCCTTACAAGTCAAAATGCCCAAGAGTCACAAGG
CTATTCAAGCAGACGGTTGGATGTGTCATGCTTCCAAATGGGTCACTACTT
GTGATTTCCGCTGGTATGGACCGAAGTATATAACACATTCCATCCGATCCT
TCACTCCATCTGTAGAACAATGCAAGGAAAGCATTGAACAAACGAAACA
AGGAACTTGGCTGAATCCAGGCTTCCCTCCTCAAAGTTGTGGATATGCAA
CTGTGACGGATGCCGAAGCAGTGATTGTCCAGGTGACTCCTCACCATGTG
CTGGTTGATGAATACACAGGAGAATGGGTTGATTCACAGTTCATCAACGG
AAAATGCAGCAATTACATATGCCCCACTGTCCATAACTCTACAACCTGGC
ATTCTGACTATAAGGTCAAAGGGCTATGTGATTCTAACCTCATTTCCATGG
ACATCACCTTCTTCTCAGAGGACGGAGAGCTATCATCCCTGGGAAAGGAG
GGCACAGGGTTCAGAAGTAACTACTTTGCTTATGAAACTGGAGGCAAGGC
CTGCAAAATGCAATACTGCAAGCATTGGGGAGTCAGACTCCCATCAGGTG
TCTGGTTCGAGATGGCTGATAAGGATCTCTTTGCTGCAGCCAGATTCCCTG
AATGCCCAGAAGGGTCAAGTATCTCTGCTCCATCTCAGACCTCAGTGGAT
GTAAGTCTAATTCAGGACGTTGAGAGGATCTTGGATTATTCCCTCTGCCAA
GAAACCTGGAGCAAAATCAGAGCGGGTCTTCCAATCTCTCCAGTGGATCT
CAGCTATCTTGCTCCTAAAAACCCAGGAACCGGTCCTGCTTTCACCATAAT
CAATGGTACCCTAAAATACTTTGAGACCAGATACATCAGAGTCGATATTG
CTGCTCCAATCCTCTCAAGAATGGTCGGAATGATCAGTGGAACTACCACA
GAAAGGGAACTGTGGGATGACTGGGCACCATATGAAGACGTGGAAATTG
GACCCAATGGAGTTCTGAGGACCAGTTCAGGATATAAGTTTCCTTTATAC
ATGATTGGACATGGTATGTTGGACTCCGATCTTCATCTTAGCTCAAAGGCT
CAGGTGTTCGAACATCCTCACATTCAAGACGCTGCTTCGCAACTTCCTGAT
GATGAGAGTTTATTTTTTGGTGATACTGGGCTATCCAAAAATCCAATCGAG
CTTGTAGAAGGTTGGTTCAGTAGTTGGAAAAGCTCTATTGCCTCTTTTTTC
TTTATCATAGGGTTAATCATTGGACTATTCTTGGTTCTCCGAGTTGGTATCC
ATCTTTGCATTAAATTAAAGCACACCAAGAAAAGACAGATTTATACAGAC
ATAGAGATGAACCGACTTGGAAAGTGATAAGGCCAGGCCGGCCAAGCTT
GTCGAGAAGTACTAGAGGATCATAATCAGCCATACCACATTTGTAGAGGT
TTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAA
AATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTA
CAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCAC
TGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTG
GATCTGATCACTGCTTGAGCCTAGGAGATCCGAACCAGATAAGTGAAATC
TAGTTCCAAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTA
CACCCAGTTCCCATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAAC
TCCATTTCCACCCCTCCCAGTTCCCAACTATTTTGTCCGCCCACAGCGGGG
CATTTTTCTTCCTGTTATGTTTTTAATCAAACATCCTGCCAACTCCATGTGA
CAAACCGTCATCTTCGGCTACTTT

pAcBac3-UpolyLeuRS 8xU6LeutRNATAG PAD4 R372TAG 12xHis



TTCTCTGTCACAGAATGAAAATTTTTCTGTCATCTCTTCGTTATTAATGTTT
GTAATTGACTGAATATCAACGCTTATTTGCAGCCTGAATGGCGAATGGGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCT
TCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAA
AAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGA
CGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTT
ATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAG
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCT
AAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGT
CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCA

GAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAG
TGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTT
CGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT
GGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCG
CATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAA
AGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGG
ACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTC
GCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG
CGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGA
TGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCT
GGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATC
GCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT
TTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGG
ATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC
TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTC
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGA
AAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGC
GGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACG
CCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTC
GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGC
AACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT
TCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGA
GTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCA
TCTGTGCGGTATTTCACACCGCAGACCAGCCGCGTAACCTGGCAAAATCG
GTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAGCGGGTGTGGGCGGA
CAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT
CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGT
GAAAAAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCT
GAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATG

GTAAAGACTATATTCGCGGCGTTGTGACAATTTACCGAACAACTCCGCGG
CCGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGG
TCGATATCAAAGTGCATCACTTCTTCCCGTATGCCCAACTTTGTATAGAGA
GCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGC
ACCAAGCGCGTTGGCCTCATGCTTGAGGAGATTGATGAGCGCGGTGGCAA
TGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGATATAGATC
TCACTACGCGGCTGCTCAAACCTGGGCAGAACGTAAGCCGCGAGAGCGCC
AACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGATGAATGTCTTACTAC
GGAGCAAGTTCCCGAGGTAATCGGAGTCCGGCTGATGTTGGGAGTAGGTG
GCTACGTCTCCGAACTCACGACCGAAAAGATCAAGAGCAGCCCGCATGGA
TTTGACTTGGTCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTG
AGCCACCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGC
TGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGC
GTAACGCGCTTGCTGCTTGGATGCCCGAGGCATAGACTGTACAAAAAAAC
AGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGT
TCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTAC
AGTTTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTTCATCCGT
TTCCACGGTGTGCGTCACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGC
ATTTCTGTCCTGGCTGGCGAACGAGCGCAAGGTTTCGGTCTCCACGCATCG
TCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGCAAGGTGCTGTGCACGGA
TCTGCCCTGGCTTCAGGAGATCGGTAGACCTCGGCCGTCGCGGCGCTTGC
CGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAA
GGCGAGCATCGTTTGTTCGCCCAGGACTCTAGCTATAGTTCTAGTGGTTGG
CCTACGTACCCGTAGTGGCTATGGCAGGGCTTGCGCTTAATGCGCCGCTA
CAGGGCGCGTGGGGATACCCCCTAGAGCCCCAGCTGGTTCTTTCCGCCTC
AGAAGCCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCC
AATCCTCCCCCTTGCTGTCCTGCCCCACCCCACCCCCCAGAATAGAATGAC
ACCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTAGGAAAGGACA
GTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAA
CAACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGT
TTAAACGGGCCCTCTAGACTCGAGTTAAAGTCGACTTAACGCGTTGAATTT
TAAACGGGCCCGCCAACGACCAGATTGAGGAGTTTACCTGGTACGTAAAT
CACTTTACGTACAGTAACGCCATCAAGATATTTTGCTACCAGATGTTCCTG
GCCAGCACGTTCGCGAACCTGTTCTTCCGTTGCGTCCACCGGAACGGTGAT
TTTGGCACGGACTTTACCGTTAACCTGCACCACGACCAGCGTGGAGTCTTC
CACCATCGCTTTTTCGTCAGCAACCGGCCACGGCGCGTTGTCGATATCGCC
TTCGCCTTTCAGTTCCTGCCACAGCGTGAAGCAGATGTGCGGGGTGAACG
GGTTAAGCATACGGACAACGGCCAGCAGTGCTTCCTGCATCAGAGCGCGA
TCCTGCTCGCCATCGGTTGGTGCTTTCGCCAGTTTGTTCATCAGCTCCATA
ATCGCCGCAATTGCGGTGTTGAAGGTCTGACGACGGCCGATATCATCGGT
CACTTTAGCGATCGTTTTATGCACATCGCGACGCAGCGCTTTCTGATTTTC

AGTCAGCGCATCAACGTTCAGTGCCGCAACATCACCTTTTGCTGTGTGCTC
GTAAACCAGTTTCCAGACACGTTTCAGGAAGCGGTTAGCCCCTTCCACAC
CGGATTCCTGCCATTCGAGAGTCATATCAGCCGGAGAAGCAAACATCATA
AACAGACGAACGGTGTCCGCGCCGTAACGTTCAACCATCACCTGCGGGTC
GATACCGTTGTTCTTCGACTTGGACATTTTGCTCATGCCGGTATAAACCAG
TTCATGGCCTGCCGCATCTTTCGCTTTCACGATACGGCCTTTCTCGTCACGT
TCAACGATAGCATCAACCGGGGAAACCCAGTTACGTTCGCCGTTTTCGCC
AACATAGTAGAAGGCATCTGCCAGCACCATACCCTGACACAGCAACTGTT
TCGCTGGTTCGTCAGAGTTCACCATGCCTGCATCACGCATCAGTTTGTGGA
AGAAGCGGAAGTAGAGCAGACCCATAATGGCGTGTTCAATACCACCAAT
AGCGATATCCACCGGCAGCCAGTAGTTAGCCGCTTCGGAATCCAGCATAC
CTTCTTTGTACTGCGGGCAAGTGTAGCGCGCATAGATCCAGGAGGACTCC
ATAAAGGTGTCGAAAGTGTCGGTTTCACGCAGTGCTGGCATACCGTTAAC
GGTAGTTTTCGCCCACTCCGGATCTGCTTTAATCGGGCTGGTAATGCCGTC
CATTACCACATCTTCCGGCAGGATCACCGGCAGCTGGTCGTCCGGGGTCG
GCATTACGGTACCGTCTTCTAGAGTCACCATCGGAATCGGCGCGCCCCAG
TAACGCTGACGGGAAACACCCCAGTCGCGCAGGCGGTAGTTCACTTTACG
CTCGCCAACGCCCATCGCAGTCAGTTTATCGGCGATGGCGTTGAAGGCCG
CTTCATGGTCAAGACCGTTGAACTCGCCAGAGTTGAACAGCACGCCTTTTT
CAGTCAGGGCTTGCTGAGAAAGATCTGGCTCAGAGCCGTCAGCTGCCAGG
ATAACCGGTTTGATGTTCAGGCCGTATTTAGAGGCAAACTCGTAGTCGCG
CTGGTCGTGCCCCGGTACCGCCATAACTGCGCCCGTGCCGTACTCCATCAA
TACGAAGTTTGCTGCCCAAACGGGAATTTCTTCGCCCGTTAATGGGTGAA
CCGCTTTAAAGCCAGTATCGACGCCTTTTTTCTCCATCGTCGCCATTTCAG
CTTCGGCAACTTTGGTGTTACGGCATTCGTCAATAAAGGCCGCCAGTTCAG
GATTATTTTCCGCCGCTTTCTGCGCCAGCGGATGACCCGCAGCTACCGCCA
GGTAGGTACAACCCATAAACGCGTCCGGGCGGGTAGTGTAAACGGTCAGC
GTGTTGTCATAGTCGTTAACGTTGAAGGTGATCTCCACGCCTTCGGAACGA
CCGATCCAGTTACGCTGCATGGTTTTAACGGTGTCTGGCCAGTGATCCAGT
TTATCCAGATCGTTGAGCAGCTCGTCAGCGTAAGCAGTGATTTTGATAAA
CCACTGCGGGATCTCTTTACGTTCAACTTTGGTATCGCAGCGCCAGCAGCA
GCCGTCGATAACTTGTTCGTTCGCCAGTACGGTCTGGTCGTTCGGACACCA
GTTGACCGCAGAAGTCTTCTTATATACCAGGCCTTTTTTATACAGCTCGGT
GAAGAATTTCTGTTCCCAACGGTAGTATTCCGGCGTACAGGTTGCCAGCTC
GCGGCTCCAGTCATAACCAAAGCCCAGCATTTTGAGCTGGTTTTTCATATA
CGCGATGTTGTCGTACGTCCACGGTGCCGGAGCGGTGTTGTTTTTCACCGC
CGCGCCTTCCGCAGGCAGACCAAACGCGTCCCAGCCGATCGGCTGCAGGA
CGTTTTTGCCCAGCATACGCTGGTAGCGGGCGATCACGTCACCGATGGTG
TAGTTACGTACGTGGCCCATGTGTAGTCGACCAGAAGGATAGGGGAGGAT
AGACAGGCAGTAATACTTCTCTTTGCTCTCGTCTTCGGTTACTTCAAATGT
GCGCTTCTCATCCCAATGAAGCTGTACTTTGGATTCTATCTCTTCCGGGCG

GTATTGCTCTTCCATGGTGGCGCTAGCCAGCTTGGGTCTCCCTATAGTGAG
TCGTATTAATTTCGATAAGCCAGTAAGCAGTGGGTTCTCTAGTTAGCCAGA
GAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCG
TCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTTGGT
GCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCC
CGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCAT
CACCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAA
AGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCG
TCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACT
GCCAAGTGGGCAGTTTACCGTAAATAGTCCACCCATTGACGTCAATGGAA
AGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGG
GCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATG
TAACGCGGAACTCCATATATGGGCTATGAACTAATGACCCCGTAATTGAT
TACTATTAATAACTAGCAAAAAATACCCGGAGCGGGACTTGAACCCGCAC
AGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATTCCACCA
TCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCG
AAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTT
TGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAA
CAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCC
TCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTGAACCCGC
ACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATTCCAC
CATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAAT
CGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATA
ATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTAT
TTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCT
AACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGC
CCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTGAACCC
GCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATTCC
ACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAA
ATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACA
TAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGT
ATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCT
CTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAG
GCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTGAAC
CCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATT
CCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAG
AAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAA
ACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCA
CGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATC
TCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAA

TAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTG
AACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCG
ATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCA
AGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTA
AAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGT
CACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTA
TCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGA
AATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACT
TGAACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTAC
CGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC
CAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTT
AAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTAC
GTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAAT
TATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGG
AAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGAC
TTGAACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTA
CCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAG
CCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATT
TTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTA
CGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAA
TTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGG
GAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGA
CTTGAACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCT
ACCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAA
GCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCAT
TTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCT
ACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTA
ATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATG
GGAAATAGGCCCTCTTCCTGCCCGACCTAGTCAATAATCAATGTCAACGC
GTATATCTGGCCCGTACATCGCGAAGCAGCGCAAAACGGATCCTGCAGGC
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATAT
GGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC
CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTA
ACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTA
AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC
CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC
ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
GCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATC
TCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTG
TGCAGCGATGGGGGCGGGGGGGGGGGGGCGCGCCAGGCGGGGCGG
GGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAG

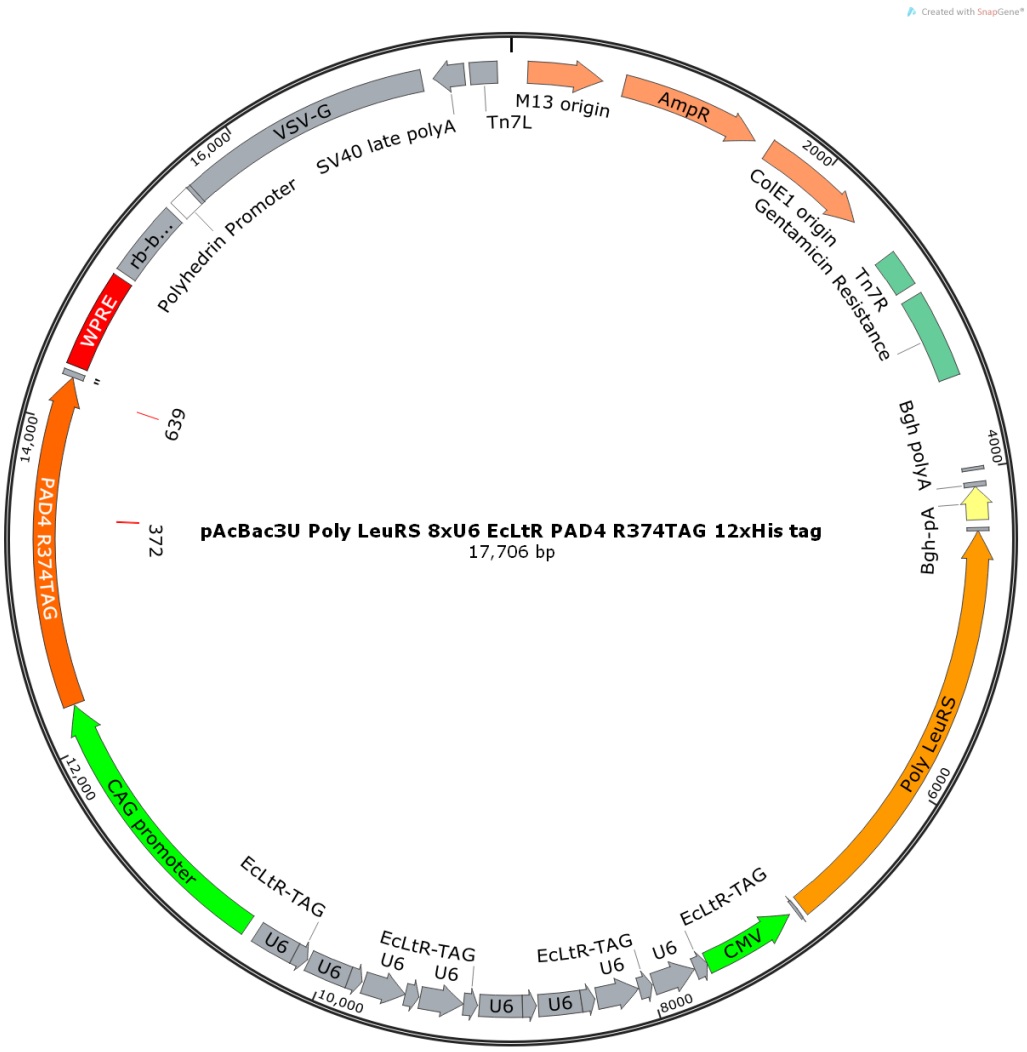
CCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGG
CGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTG
CGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCC
GGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCT
TCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTC
TGTGGCTGCGTGAAAGCCTTAAAGGGCTCCGGGAGGGCCCTTTGTGCGGG
GGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCC
GCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCG
GGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGT
GCCCCGCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGG
TGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTG
TAACCCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTC
GGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGC
GGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGG
CCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGC
TGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGA
GAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGGCGGAGCCGAAATCTGG
GAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCGAAGCGGTGCGGCG
CCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCG
CCGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTG
CCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCG
GCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACA
GCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGA
ATTGGCCAAGGAGGCCACCATGGACTACAAGGACGACGACGACAAGGCC
CAGGGGACATTGATCCGTGTGACCCCAGAGCAGCCCACCCATGCCGTGTG
TGTGCTGGGCACCTTGACTCAGCTTGACATATGCAGCTCTGCCCCTGAGGA
CTGCACGTCCTTCAGCATCAACGCCTCCCCAGGGGTGGTCGTGGATATTGC
CCACAGCCCTCCAGCCAAGAAGAAATCCACAGGTTCCTCCACATGGCCCC
TGGACCCTGGGGTAGAGGTGACCCTGACGATGAAAGCGGCCAGTGGTAG
CACAGGCGACCAGAAGGTTCAGATTTCATACTACGGACCCAAGACTCCAC
CAGTCAAAGCTCTACTCTACCTCACCGGGGTGGAAATCTCCTTGTGCGCA
GACATCACCCGCACCGGCAAAGTGAAGCCAACCAGAGCTGTGAAAGATC
AGAGGACCTGGACCTGGGGCCCTTGTGGACAGGGTGCCATCCTGCTGGTG
AACTGTGACAGAGACAATCTCGAATCTTCTGCCATGGACTGCGAGGATGA
TGAAGTGCTTGACAGCGAAGACCTGCAGGACATGTCGCTGATGACCCTGA
GCACGAAGACCCCCAAGGACTTCTTCACAAACCATACACTGGTGCTACAC
GTGGCCAGGTCTGAGATGGACAAAGTGAGGGTGTTTCAGGCCACACGGG
GCAAACTGTCCTCCAAGTGCAGCGTAGTCTTGGGTCCCAAGTGGCCCTCTC
ACTACCTGATGGTCCCCGGTGGAAAGCACAACATGGACTTCTACGTGGAG
GCCCTCGCTTTCCCGGACACCGACTTCCCGGGGCTCATTACCCTCACCATC
TCCCTGCTGGACACGTCCAACCTGGAGCTCCCCGAGGCTGTGGTGTTCCA

AGACAGCGTGGTCTTCCGCGTGGCGCCCTGGATCATGACCCCCAACACCC
AGCCCCCGCAGGAGGTGTACGCGTGCAGTATTTTTGAAAATGAGGACTTC
CTGAAGTCAGTGACTACTCTGGCCATGAAAGCCAAGTGCAAGCTGACCAT
CTGCCCTGAGGAGGAGAACATGGATGACCAGTGGATGCAGGATGAAATG
GAGATCGGCTACATCCAAGCCCCACACAAAACACTGCCCGTGGTCTTCGA
CTCTCCTTAGAACAGAGGCCTGAAGGAGTTTCCCATCAAACGAGTGATGG
GTCCAGATTTTGGCTATGTAACTCGAGGGCCCCAAACAGGGGGTATCAGT
GGACTGGACTCCTTTGGGAACCTGGAAGTGAGCCCCCCAGTCACAGTCAG
GGGCAAGGAATACCCGCTGGGCAGGATTCTCTTCGGGGACAGCTGTTATC
CCAGCAATGACAGCCGGCAGATGCACCAGGCCCTGCAGGACTTCCTCAGT
GCCCAGCAGGTGCAGGCCCCTGTGAAGCTCTATTCTGACTGGCTGTCCGT
GGGCCACGTGGACGAGTTCCTGAGCTTTGTGCCAGCACCCGACAGGAAGG
GCTTCCGGCTGCTCCTGGCCAGCCCCAGGTCCTGCTACAAACTGTTCCAGG
AGCAGCAGAATGAGGGCCACGGGGAGGCCCTGCTGTTCGAAGGGATCAA
GAAAAAAAAACAGCAGAAAATAAAGAACATTCTGTCAAACAAGACATTG
AGAGAACATAATTCATTTGTGGAGAGATGCATCGACTGGAACCGCGAGCT
GCTGAAGCGGGAGCTGGGCCTGGCCGAGAGTGACATCATTGACATCCCGC
AGCTCTTCAAGCTCAAAGAGTTCTCTAAGGCGGAAGCTTTTTTCCCCAACA
TGGTGAACATGCTGGTGCTAGGGAAGCACCTGGGCATCCCCAAGCCCTTC
GGGCCAGTCATCAACGGCCGCTGCTGCCTGGAGGAGAAGGTGTGTTCCCT
GCTGGAGCCACTGGGCCTCCAGTGCACCTTCATCAACGACTTCTTCACCTA
CCACATCAGGCATGGGGAGGTGCACTGCGGCACCAACGTGCGCAGAAAG
CCCTTCTCCTTCAAGTGGTGGAACATGGTGCCCCACCACCACCACCACCAC
CACCACCACCACCACCACTGAGGCCTCTAAGGCCGAATTCAACGCGTTAA
GTCGACAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAA
ATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACG
TGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCAT
TGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATT
GCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGC
TCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTC
CTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTC
CTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG
CCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGAC
GAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCGTCGACTTTAACTCGGCCA
GCACAGTGGTCGATCGACCAATGCCCTGGCTCACAAATACCACTGAGATC
TTTTTCCCTCTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATC
TGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGA
ATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAA
AACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGC

TGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAA
CAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGAGGT
TAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTAAA
ATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTGACTACTC
CCAGTCATAGCTGTCCCTCTTCTTTGCGGCCGCGGTCCGTATACTCCGGAA
TATTAATAGATCATGGAGATAATTAAAATGATAACCATCTCGCAAATAAA
TAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAAT
ATTCCGGATTATTCATACCGTCCCACCATCGGGCGCGAACTCCTAAAAAA
CCGCCACCATGAAGTGCCTTTTGTACTTAGCCTTTTTATTCATTGGGGTGA
ATTGCAAGTTCACCATAGTTTTTCCACACAACCAAAAAGGAAACTGGAAA
AATGTTCCTTCTAATTACCATTATTGCCCGTCAAGCTCAGATTTAAATTGG
CATAATGACTTAATAGGCACAGCCTTACAAGTCAAAATGCCCAAGAGTCA
CAAGGCTATTCAAGCAGACGGTTGGATGTGTCATGCTTCCAAATGGGTCA
CTACTTGTGATTTCCGCTGGTATGGACCGAAGTATATAACACATTCCATCC
GATCCTTCACTCCATCTGTAGAACAATGCAAGGAAAGCATTGAACAAACG
AAACAAGGAACTTGGCTGAATCCAGGCTTCCCTCCTCAAAGTTGTGGATA
TGCAACTGTGACGGATGCCGAAGCAGTGATTGTCCAGGTGACTCCTCACC
ATGTGCTGGTTGATGAATACACAGGAGAATGGGTTGATTCACAGTTCATC
AACGGAAAATGCAGCAATTACATATGCCCCACTGTCCATAACTCTACAAC
CTGGCATTCTGACTATAAGGTCAAAGGGCTATGTGATTCTAACCTCATTTC
CATGGACATCACCTTCTTCTCAGAGGACGGAGAGCTATCATCCCTGGGAA
AGGAGGGCACAGGGTTCAGAAGTAACTACTTTGCTTATGAAACTGGAGGC
AAGGCCTGCAAAATGCAATACTGCAAGCATTGGGGAGTCAGACTCCCATC
AGGTGTCTGGTTCGAGATGGCTGATAAGGATCTCTTTGCTGCAGCCAGATT
CCCTGAATGCCCAGAAGGGTCAAGTATCTCTGCTCCATCTCAGACCTCAGT
GGATGTAAGTCTAATTCAGGACGTTGAGAGGATCTTGGATTATTCCCTCTG
CCAAGAAACCTGGAGCAAAATCAGAGCGGGTCTTCCAATCTCTCCAGTGG
ATCTCAGCTATCTTGCTCCTAAAAACCCAGGAACCGGTCCTGCTTTCACCA
TAATCAATGGTACCCTAAAATACTTTGAGACCAGATACATCAGAGTCGAT
ATTGCTGCTCCAATCCTCTCAAGAATGGTCGGAATGATCAGTGGAACTAC
CACAGAAAGGGAACTGTGGGATGACTGGGCACCATATGAAGACGTGGAA
ATTGGACCCAATGGAGTTCTGAGGACCAGTTCAGGATATAAGTTTCCTTTA
TACATGATTGGACATGGTATGTTGGACTCCGATCTTCATCTTAGCTCAAAG
GCTCAGGTGTTCGAACATCCTCACATTCAAGACGCTGCTTCGCAACTTCCT
GATGATGAGAGTTTATTTTTTGGTGATACTGGGCTATCCAAAAATCCAATC
GAGCTTGTAGAAGGTTGGTTCAGTAGTTGGAAAAGCTCTATTGCCTCTTTT
TTCTTTATCATAGGGTTAATCATTGGACTATTCTTGGTTCTCCGAGTTGGTA
TCCATCTTTGCATTAAATTAAAGCACACCAAGAAAAGACAGATTTATACA
GACATAGAGATGAACCGACTTGGAAAGTGATAAAAGCTTGTCGAGAAGT
ACTAGAGGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTT
TAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCA

ATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGC
AATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT
TGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTGATCA
CTGCTTGAGCCTAGGAGATCCGAACCAGATAAGTGAAATCTAGTTCCAAA
CTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTC
CCATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCAC
CCCTCCCAGTTCCCAACTATTTTGTCCGCCCACAGCGGGGCATTTTTCTTC
CTGTTATGTTTTTAATCAAACATCCTGCCAACTCCATGTGACAAACCGTCA
TCTTCGGCTACTTT

pAcBac3-UpolyLeuRS 8xU6LeutRNATAG PAD4 R374TAG 12xHis



TTCTCTGTCACAGAATGAAAATTTTTCTGTCATCTCTTCGTTATTAATGTTT
GTAATTGACTGAATATCAACGCTTATTTGCAGCCTGAATGGCGAATGGGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCT

TCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAA
AAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGA
CGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTT
ATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAG
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCT
AAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGT
CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCA
GAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAG
TGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTT
CGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT
GGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCG
CATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAA
AGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGG
ACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTC
GCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG
CGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGA
TGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCT
GGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATC
GCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT
TTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGG
ATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC
TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTC
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGA
AAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGC
GGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACG
CCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTC

GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGC
AACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT
TCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGA
GTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCA
TCTGTGCGGTATTTCACACCGCAGACCAGCCGCGTAACCTGGCAAAATCG
GTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAGCGGGTGTGGGCGGA
CAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT
CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGT
GAAAAAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCT
GAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATG
GTAAAGACTATATTCGCGGCGTTGTGACAATTTACCGAACAACTCCGCGG
CCGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGG
TCGATATCAAAGTGCATCACTTCTTCCCGTATGCCCAACTTTGTATAGAGA
GCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGC
ACCAAGCGCGTTGGCCTCATGCTTGAGGAGATTGATGAGCGCGGTGGCAA
TGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGATATAGATC
TCACTACGCGGCTGCTCAAACCTGGGCAGAACGTAAGCCGCGAGAGCGCC
AACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGATGAATGTCTTACTAC
GGAGCAAGTTCCCGAGGTAATCGGAGTCCGGCTGATGTTGGGAGTAGGTG
GCTACGTCTCCGAACTCACGACCGAAAAGATCAAGAGCAGCCCGCATGGA
TTTGACTTGGTCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTG
AGCCACCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGC
TGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGC
GTAACGCGCTTGCTGCTTGGATGCCCGAGGCATAGACTGTACAAAAAAAC
AGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGT
TCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTAC
AGTTTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTTCATCCGT
TTCCACGGTGTGCGTCACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGC
ATTTCTGTCCTGGCTGGCGAACGAGCGCAAGGTTTCGGTCTCCACGCATCG
TCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGCAAGGTGCTGTGCACGGA
TCTGCCCTGGCTTCAGGAGATCGGTAGACCTCGGCCGTCGCGGCGCTTGC
CGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAA
GGCGAGCATCGTTTGTTCGCCCAGGACTCTAGCTATAGTTCTAGTGGTTGG
CCTACGTACCCGTAGTGGCTATGGCAGGGCTTGCGCTTAATGCGCCGCTA
CAGGGCGCGTGGGGATACCCCCTAGAGCCCCAGCTGGTTCTTTCCGCCTC
AGAAGCCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCC
AATCCTCCCCCTTGCTGTCCTGCCCCACCCCACCCCCCAGAATAGAATGAC
ACCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTAGGAAAGGACA
GTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAA
CAACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGT

TTAAACGGGCCCTCTAGACTCGAGTTAAAGTCGACTTAACGCGTTGAATTT
TAAACGGGCCCGCCAACGACCAGATTGAGGAGTTTACCTGGTACGTAAAT
CACTTTACGTACAGTAACGCCATCAAGATATTTTGCTACCAGATGTTCCTG
GCCAGCACGTTCGCGAACCTGTTCTTCCGTTGCGTCCACCGGAACGGTGAT
TTTGGCACGGACTTTACCGTTAACCTGCACCACGACCAGCGTGGAGTCTTC
CACCATCGCTTTTTCGTCAGCAACCGGCCACGGCGCGTTGTCGATATCGCC
TTCGCCTTTCAGTTCCTGCCACAGCGTGAAGCAGATGTGCGGGGTGAACG
GGTTAAGCATACGGACAACGGCCAGCAGTGCTTCCTGCATCAGAGCGCGA
TCCTGCTCGCCATCGGTTGGTGCTTTCGCCAGTTTGTTCATCAGCTCCATA
ATCGCCGCAATTGCGGTGTTGAAGGTCTGACGACGGCCGATATCATCGGT
CACTTTAGCGATCGTTTTATGCACATCGCGACGCAGCGCTTTCTGATTTTC
AGTCAGCGCATCAACGTTCAGTGCCGCAACATCACCTTTTGCTGTGTGCTC
GTAAACCAGTTTCCAGACACGTTTCAGGAAGCGGTTAGCCCCTTCCACAC
CGGATTCCTGCCATTCGAGAGTCATATCAGCCGGAGAAGCAAACATCATA
AACAGACGAACGGTGTCCGCGCCGTAACGTTCAACCATCACCTGCGGGTC
GATACCGTTGTTCTTCGACTTGGACATTTTGCTCATGCCGGTATAAACCAG
TTCATGGCCTGCCGCATCTTTCGCTTTCACGATACGGCCTTTCTCGTCACGT
TCAACGATAGCATCAACCGGGGAAACCCAGTTACGTTCGCCGTTTTCGCC
AACATAGTAGAAGGCATCTGCCAGCACCATACCCTGACACAGCAACTGTT
TCGCTGGTTCGTCAGAGTTCACCATGCCTGCATCACGCATCAGTTTGTGGA
AGAAGCGGAAGTAGAGCAGACCCATAATGGCGTGTTCAATACCACCAAT
AGCGATATCCACCGGCAGCCAGTAGTTAGCCGCTTCGGAATCCAGCATAC
CTTCTTTGTACTGCGGGCAAGTGTAGCGCGCATAGATCCAGGAGGACTCC
ATAAAGGTGTCGAAAGTGTCGGTTTCACGCAGTGCTGGCATACCGTTAAC
GGTAGTTTTCGCCCACTCCGGATCTGCTTTAATCGGGCTGGTAATGCCGTC
CATTACCACATCTTCCGGCAGGATCACCGGCAGCTGGTCGTCCGGGGTCG
GCATTACGGTACCGTCTTCTAGAGTCACCATCGGAATCGGCGCGCCCCAG
TAACGCTGACGGGAAACACCCCAGTCGCGCAGGCGGTAGTTCACTTTACG
CTCGCCAACGCCCATCGCAGTCAGTTTATCGGCGATGGCGTTGAAGGCCG
CTTCATGGTCAAGACCGTTGAACTCGCCAGAGTTGAACAGCACGCCTTTTT
CAGTCAGGGCTTGCTGAGAAAGATCTGGCTCAGAGCCGTCAGCTGCCAGG
ATAACCGGTTTGATGTTCAGGCCGTATTTAGAGGCAAACTCGTAGTCGCG
CTGGTCGTGCCCCGGTACCGCCATAACTGCGCCCGTGCCGTACTCCATCAA
TACGAAGTTTGCTGCCCAAACGGGAATTTCTTCGCCCGTTAATGGGTGAA
CCGCTTTAAAGCCAGTATCGACGCCTTTTTTCTCCATCGTCGCCATTTCAG
CTTCGGCAACTTTGGTGTTACGGCATTCGTCAATAAAGGCCGCCAGTTCAG
GATTATTTTCCGCCGCTTTCTGCGCCAGCGGATGACCCGCAGCTACCGCCA
GGTAGGTACAACCCATAAACGCGTCCGGGCGGGTAGTGTAAACGGTCAGC
GTGTTGTCATAGTCGTTAACGTTGAAGGTGATCTCCACGCCTTCGGAACGA
CCGATCCAGTTACGCTGCATGGTTTTAACGGTGTCTGGCCAGTGATCCAGT
TTATCCAGATCGTTGAGCAGCTCGTCAGCGTAAGCAGTGATTTTGATAAA

CCACTGCGGGATCTCTTTACGTTCAACTTTGGTATCGCAGCGCCAGCAGCA
GCCGTCGATAACTTGTTCGTTCGCCAGTACGGTCTGGTCGTTCGGACACCA
GTTGACCGCAGAAGTCTTCTTATATACCAGGCCTTTTTTATACAGCTCGGT
GAAGAATTTCTGTTCCCAACGGTAGTATTCCGGCGTACAGGTTGCCAGCTC
GCGGCTCCAGTCATAACCAAAGCCCAGCATTTTGAGCTGGTTTTTCATATA
CGCGATGTTGTCGTACGTCCACGGTGCCGGAGCGGTGTTGTTTTTCACCGC
CGCGCCTTCCGCAGGCAGACCAAACGCGTCCCAGCCGATCGGCTGCAGGA
CGTTTTTGCCCAGCATACGCTGGTAGCGGGCGATCACGTCACCGATGGTG
TAGTTACGTACGTGGCCCATGTGTAGTCGACCAGAAGGATAGGGGAGGAT
AGACAGGCAGTAATACTTCTCTTTGCTCTCGTCTTCGGTTACTTCAAATGT
GCGCTTCTCATCCCAATGAAGCTGTACTTTGGATTCTATCTCTTCCGGGCG
GTATTGCTCTTCCATGGTGGCGCTAGCCAGCTTGGGTCTCCCTATAGTGAG
TCGTATTAATTTCGATAAGCCAGTAAGCAGTGGGTTCTCTAGTTAGCCAGA
GAGCTCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCG
TCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTTGGT
GCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCC
CGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCAT
CACCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAA
AGTCCCATAAGGTCATGTACTGGGCATAATGCCAGGCGGGCCATTTACCG
TCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACT
GCCAAGTGGGCAGTTTACCGTAAATAGTCCACCCATTGACGTCAATGGAA
AGTCCCTATTGGCGTTACTATGGGAACATACGTCATTATTGACGTCAATGG
GCGGGGGTCGTTGGGCGGTCAGCCAGGCGGGCCATTTACCGTAAGTTATG
TAACGCGGAACTCCATATATGGGCTATGAACTAATGACCCCGTAATTGAT
TACTATTAATAACTAGCAAAAAATACCCGGAGCGGGACTTGAACCCGCAC
AGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATTCCACCA
TCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCG
AAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAAT
TTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTT
TGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAA
CAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCC
TCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTGAACCCGC
ACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATTCCAC
CATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAAT
CGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATA
ATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTAT
TTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCT
AACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGC
CCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTGAACCC
GCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATTCC
ACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAA

ATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACA
TAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGT
ATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCT
CTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAG
GCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTGAAC
CCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCGATT
CCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAG
AAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAA
ACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCA
CGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATC
TCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAA
TAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACTTG
AACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTACCG
ATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCA
AGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTA
AAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGT
CACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTA
TCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGA
AATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGACT
TGAACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTAC
CGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC
CAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTT
AAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTAC
GTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAAT
TATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGG
AAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGAC
TTGAACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCTA
CCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAG
CCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATT
TTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTA
CGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAA
TTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGG
GAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATACCCGGAGCGGGA
CTTGAACCCGCACAGCGCGAACGCCGAGGGATTTAGAGTCCCTTGTGTCT
ACCGATTCCACCATCCGGGCGGTGTTTCGTCCTTTCCACAAGATATATAAA
GCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCAT
TTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCT
ACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTA
ATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATG
GGAAATAGGCCCTCTTCCTGCCCGACCTAGTCAATAATCAATGTCAACGC
GTATATCTGGCCCGTACATCGCGAAGCAGCGCAAAACGGATCCTGCAGGC

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATAT
GGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGC
CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTA
ACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTA
AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC
CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC
ATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
GCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATC
TCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTG
TGCAGCGATGGGGGCGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGG
GGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAG
CCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGGCGG
CGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTG
CGTTGCCTTCGCCCCGTGCCCCGCTCCGCGCCGCCTCGCGCCGCCCGCCCC
GGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCT
TCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTCGTTTCTTTTC
TGTGGCTGCGTGAAAGCCTTAAAGGGCTCCGGGAGGGCCCTTTGTGCGGG
GGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCC
GCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCG
GGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGT
GCCCCGCGGTGCGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGG
TGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGGCGGTCGGGCTG
TAACCCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTC
GGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGC
GGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGG
CCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGCCGGCGGC
TGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGA
GAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGGCGGAGCCGAAATCTGG
GAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCGAAGCGGTGCGGCG
CCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCG
CCGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTG
CCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCG
GCGGCTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACA
GCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGA
ATTGGCCAAGGAGGCCACCATGGACTACAAGGACGACGACGACAAGGCC
CAGGGGACATTGATCCGTGTGACCCCAGAGCAGCCCACCCATGCCGTGTG
TGTGCTGGGCACCTTGACTCAGCTTGACATATGCAGCTCTGCCCCTGAGGA
CTGCACGTCCTTCAGCATCAACGCCTCCCCAGGGGTGGTCGTGGATATTGC
CCACAGCCCTCCAGCCAAGAAGAAATCCACAGGTTCCTCCACATGGCCCC
TGGACCCTGGGGTAGAGGTGACCCTGACGATGAAAGCGGCCAGTGGTAG
CACAGGCGACCAGAAGGTTCAGATTTCATACTACGGACCCAAGACTCCAC

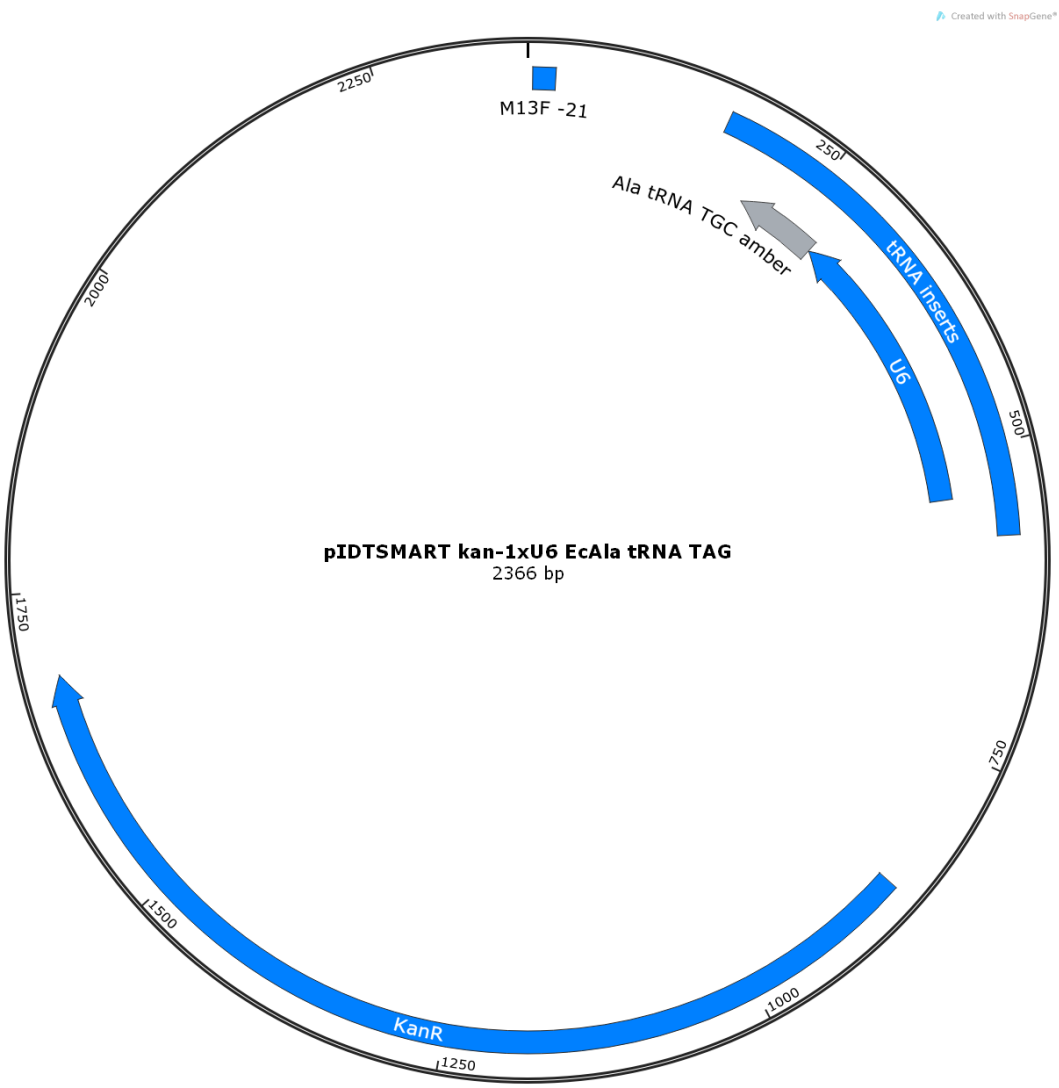
CAGTCAAAGCTCTACTCTACCTCACCGGGGTGGAAATCTCCTTGTGCGCA
GACATCACCCGCACCGGCAAAGTGAAGCCAACCAGAGCTGTGAAAGATC
AGAGGACCTGGACCTGGGGCCCTTGTGGACAGGGTGCCATCCTGCTGGTG
AACTGTGACAGAGACAATCTCGAATCTTCTGCCATGGACTGCGAGGATGA
TGAAGTGCTTGACAGCGAAGACCTGCAGGACATGTCGCTGATGACCCTGA
GCACGAAGACCCCCAAGGACTTCTTCACAAACCATACACTGGTGCTACAC
GTGGCCAGGTCTGAGATGGACAAAGTGAGGGTGTTTCAGGCCACACGGG
GCAAACTGTCCTCCAAGTGCAGCGTAGTCTTGGGTCCCAAGTGGCCCTCTC
ACTACCTGATGGTCCCCGGTGGAAAGCACAACATGGACTTCTACGTGGAG
GCCCTCGCTTTCCCGGACACCGACTTCCCGGGGCTCATTACCCTCACCATC
TCCCTGCTGGACACGTCCAACCTGGAGCTCCCCGAGGCTGTGGTGTTCCA
AGACAGCGTGGTCTTCCGCGTGGCGCCCTGGATCATGACCCCCAACACCC
AGCCCCCGCAGGAGGTGTACGCGTGCAGTATTTTTGAAAATGAGGACTTC
CTGAAGTCAGTGACTACTCTGGCCATGAAAGCCAAGTGCAAGCTGACCAT
CTGCCCTGAGGAGGAGAACATGGATGACCAGTGGATGCAGGATGAAATG
GAGATCGGCTACATCCAAGCCCCACACAAAACACTGCCCGTGGTCTTCGA
CTCTCCTAGGAACTAGGGCCTGAAGGAGTTTCCCATCAAACGAGTGATGG
GTCCAGATTTTGGCTATGTAACTCGAGGGCCCCAAACAGGGGGTATCAGT
GGACTGGACTCCTTTGGGAACCTGGAAGTGAGCCCCCCAGTCACAGTCAG
GGGCAAGGAATACCCGCTGGGCAGGATTCTCTTCGGGGACAGCTGTTATC
CCAGCAATGACAGCCGGCAGATGCACCAGGCCCTGCAGGACTTCCTCAGT
GCCCAGCAGGTGCAGGCCCCTGTGAAGCTCTATTCTGACTGGCTGTCCGT
GGGCCACGTGGACGAGTTCCTGAGCTTTGTGCCAGCACCCGACAGGAAGG
GCTTCCGGCTGCTCCTGGCCAGCCCCAGGTCCTGCTACAAACTGTTCCAGG
AGCAGCAGAATGAGGGCCACGGGGAGGCCCTGCTGTTCGAAGGGATCAA
GAAAAAAAAACAGCAGAAAATAAAGAACATTCTGTCAAACAAGACATTG
AGAGAACATAATTCATTTGTGGAGAGATGCATCGACTGGAACCGCGAGCT
GCTGAAGCGGGAGCTGGGCCTGGCCGAGAGTGACATCATTGACATCCCGC
AGCTCTTCAAGCTCAAAGAGTTCTCTAAGGCGGAAGCTTTTTTCCCCAACA
TGGTGAACATGCTGGTGCTAGGGAAGCACCTGGGCATCCCCAAGCCCTTC
GGGCCAGTCATCAACGGCCGCTGCTGCCTGGAGGAGAAGGTGTGTTCCCT
GCTGGAGCCACTGGGCCTCCAGTGCACCTTCATCAACGACTTCTTCACCTA
CCACATCAGGCATGGGGAGGTGCACTGCGGCACCAACGTGCGCAGAAAG
CCCTTCTCCTTCAAGTGGTGGAACATGGTGCCCCACCACCACCACCACCAC
CACCACCACCACCACCACTGAGGCCTCTAAGGCCGAATTCAACGCGTTAA
GTCGACAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATT
CTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT
TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAA
ATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACG
TGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCAT
TGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATT

GCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGC
TCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTC
CTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTC
CTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG
CCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGAC
GAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCGTCGACTTTAACTCGGCCA
GCACAGTGGTCGATCGACCAATGCCCTGGCTCACAAATACCACTGAGATC
TTTTTCCCTCTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATC
TGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGA
ATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAA
AACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGC
TGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAA
CAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGAGGT
TAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTAAA
ATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTGACTACTC
CCAGTCATAGCTGTCCCTCTTCTTTGCGGCCGCGGTCCGTATACTCCGGAA
TATTAATAGATCATGGAGATAATTAAAATGATAACCATCTCGCAAATAAA
TAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAAT
ATTCCGGATTATTCATACCGTCCCACCATCGGGCGCGAACTCCTAAAAAA
CCGCCACCATGAAGTGCCTTTTGTACTTAGCCTTTTTATTCATTGGGGTGA
ATTGCAAGTTCACCATAGTTTTTCCACACAACCAAAAAGGAAACTGGAAA
AATGTTCCTTCTAATTACCATTATTGCCCGTCAAGCTCAGATTTAAATTGG
CATAATGACTTAATAGGCACAGCCTTACAAGTCAAAATGCCCAAGAGTCA
CAAGGCTATTCAAGCAGACGGTTGGATGTGTCATGCTTCCAAATGGGTCA
CTACTTGTGATTTCCGCTGGTATGGACCGAAGTATATAACACATTCCATCC
GATCCTTCACTCCATCTGTAGAACAATGCAAGGAAAGCATTGAACAAACG
AAACAAGGAACTTGGCTGAATCCAGGCTTCCCTCCTCAAAGTTGTGGATA
TGCAACTGTGACGGATGCCGAAGCAGTGATTGTCCAGGTGACTCCTCACC
ATGTGCTGGTTGATGAATACACAGGAGAATGGGTTGATTCACAGTTCATC
AACGGAAAATGCAGCAATTACATATGCCCCACTGTCCATAACTCTACAAC
CTGGCATTCTGACTATAAGGTCAAAGGGCTATGTGATTCTAACCTCATTTC
CATGGACATCACCTTCTTCTCAGAGGACGGAGAGCTATCATCCCTGGGAA
AGGAGGGCACAGGGTTCAGAAGTAACTACTTTGCTTATGAAACTGGAGGC
AAGGCCTGCAAAATGCAATACTGCAAGCATTGGGGAGTCAGACTCCCATC
AGGTGTCTGGTTCGAGATGGCTGATAAGGATCTCTTTGCTGCAGCCAGATT
CCCTGAATGCCCAGAAGGGTCAAGTATCTCTGCTCCATCTCAGACCTCAGT
GGATGTAAGTCTAATTCAGGACGTTGAGAGGATCTTGGATTATTCCCTCTG
CCAAGAAACCTGGAGCAAAATCAGAGCGGGTCTTCCAATCTCTCCAGTGG
ATCTCAGCTATCTTGCTCCTAAAAACCCAGGAACCGGTCCTGCTTTCACCA
TAATCAATGGTACCCTAAAATACTTTGAGACCAGATACATCAGAGTCGAT
ATTGCTGCTCCAATCCTCTCAAGAATGGTCGGAATGATCAGTGGAACTAC

CACAGAAAGGGAACTGTGGGATGACTGGGCACCATATGAAGACGTGGAA
ATTGGACCCAATGGAGTTCTGAGGACCAGTTCAGGATATAAGTTTCCTTTA
TACATGATTGGACATGGTATGTTGGACTCCGATCTTCATCTTAGCTCAAAG
GCTCAGGTGTTCGAACATCCTCACATTCAAGACGCTGCTTCGCAACTTCCT
GATGATGAGAGTTTATTTTTTGGTGATACTGGGCTATCCAAAAATCCAATC
GAGCTTGTAGAAGGTTGGTTCAGTAGTTGGAAAAGCTCTATTGCCTCTTTT
TTCTTTATCATAGGGTTAATCATTGGACTATTCTTGGTTCTCCGAGTTGGTA
TCCATCTTTGCATTAAATTAAAGCACACCAAGAAAAGACAGATTTATACA
GACATAGAGATGAACCGACTTGGAAAGTGATAAAAGCTTGTCGAGAAGT
ACTAGAGGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTT
TAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCA
ATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGC
AATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGT
TGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTGATCA
CTGCTTGAGCCTAGGAGATCCGAACCAGATAAGTGAAATCTAGTTCCAAA
CTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAGTTC
CCATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTCCAC
CCCTCCCAGTTCCCAACTATTTTGTCCGCCCACAGCGGGGCATTTTTCTTC
CTGTTATGTTTTTAATCAAACATCCTGCCAACTCCATGTGACAAACCGTCA
TCTTCGGCTACTTT

Chapter 3

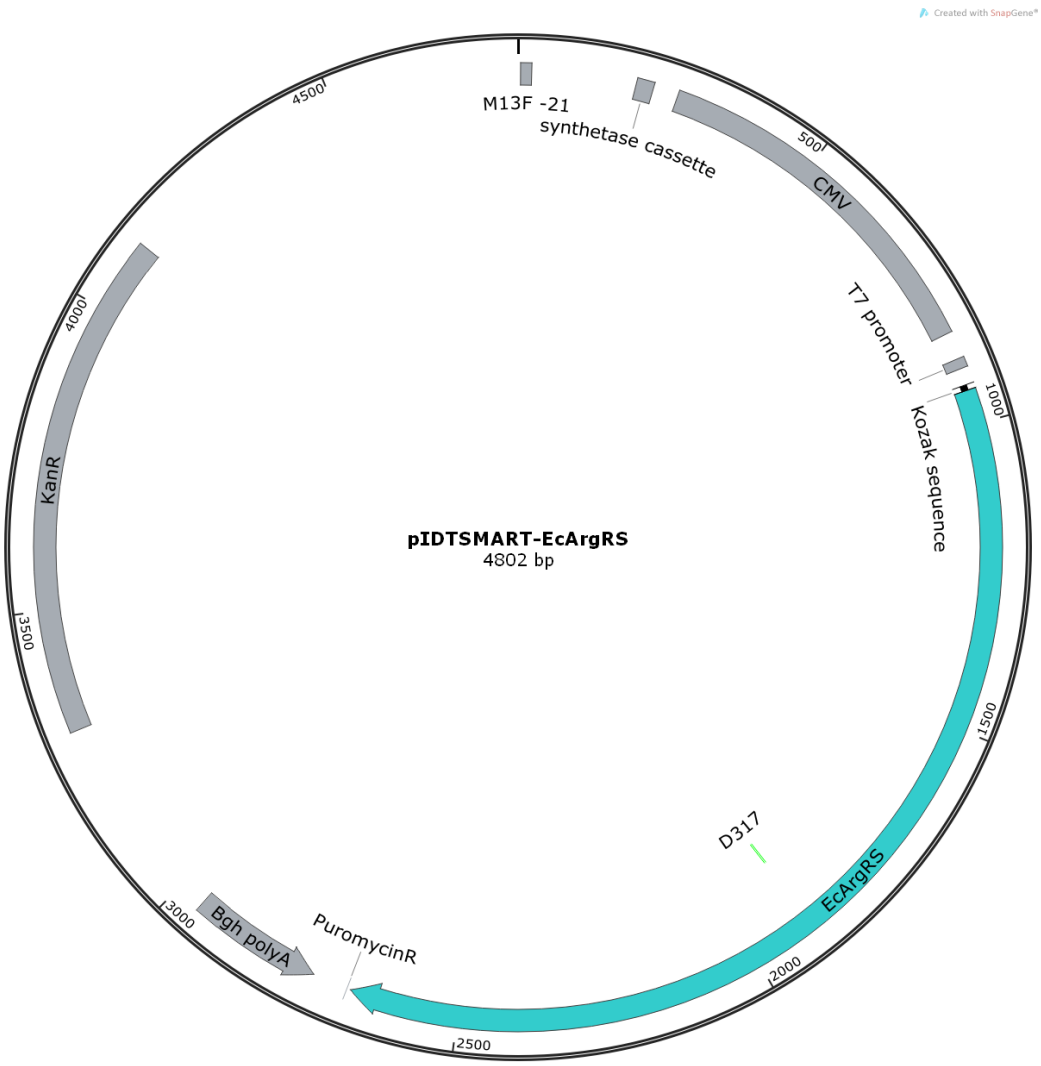
pIDTSMART $tRNA^{Ala}_{CUA}$



CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATC
GTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGA
GTCTGAGGCTCGTCCTGAATGATATGCGaCCGCCGGAGGGTTGCGTTTGAG
ACGGGCGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGCAAAAAA
TGGTGGAGCTATGCGGGATCGAACCGCAGACCTCCTGCTTTAGAGGCAGG
CGCTCTCCCAGCTGAGCTATAGCCCCGGTGTTTCGTCCTTTCCACAAGATA
TATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATA
GTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTA
CTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTA
ATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGA
ATCATGGGAAATAGGCCCTCTTCCTGCCCGAcCTAGGGGTGCGAGCGGAT
CGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGACcCGTGAT
CTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTC
AGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCA
TACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTG

ACTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGAT
CACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAG
CTAAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTA
AATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAA
TGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATG
CGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTT
ACAGATGAGATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCC
GACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCAC
TGCGATCCCAGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATT
CAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATT
CGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCGCGTATTTCGTCTCGC
TCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGATTTTG
ATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCAT
AAACTCTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCA
CTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTT
GGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAA
CTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATA
TGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCACTTGATGCTCGA
TGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCTCACCTTCGGGT
GGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGA
GCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCT
GTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGA
AGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAG
GTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGA
CCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGA
GGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGC
TACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG
TAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAG
GATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pIDTSMART ArgRS

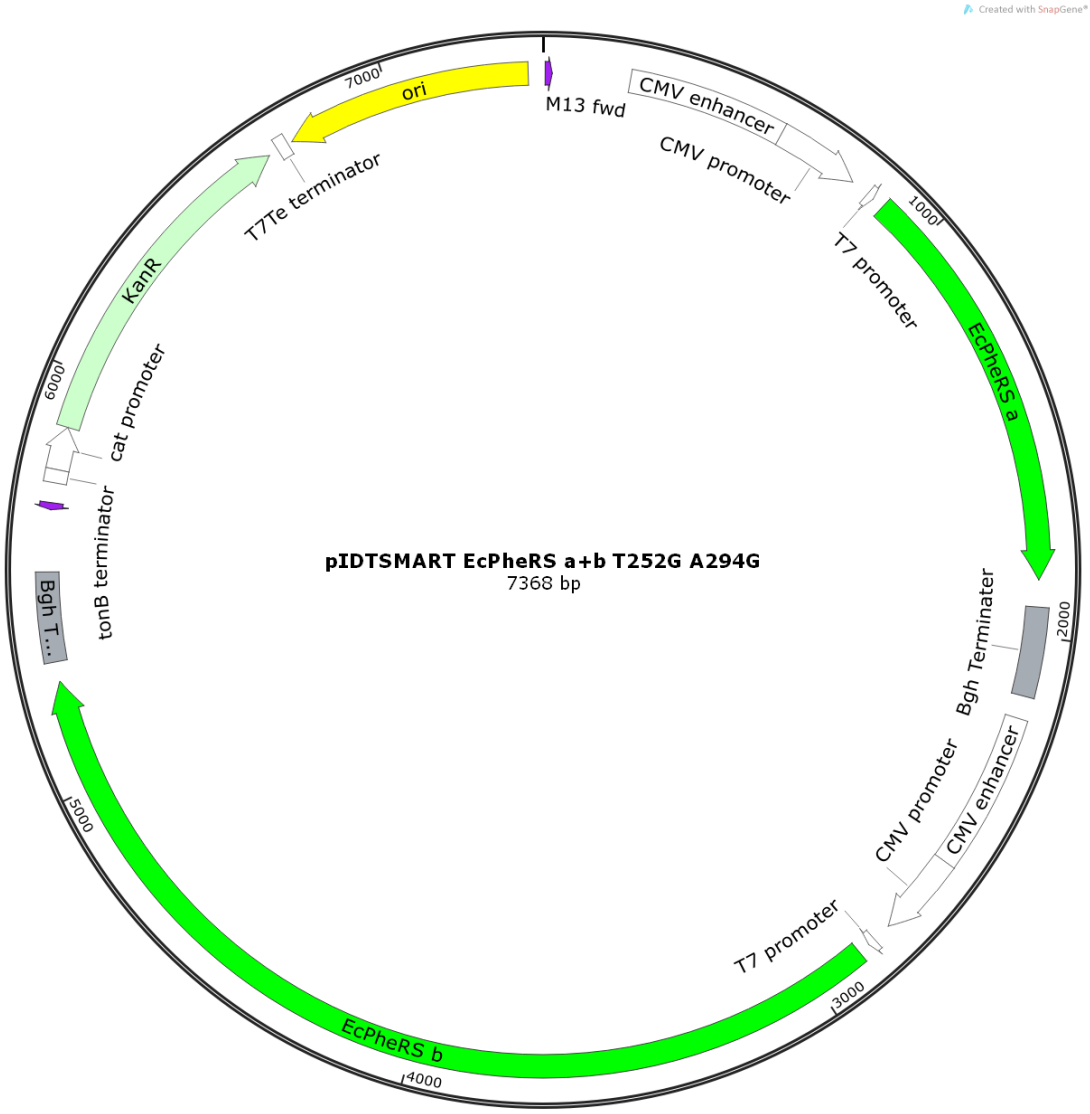


CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATC
GTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGA
GTCTGAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGA
GACGGGCGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGGCGGC
CGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTA
CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTA
CGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACG
TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCA
AGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAAT
GGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTT
GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT
GGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA
AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA
CGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG

CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAA
CTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATA
GGGAGACCCAAGCTGGCTAGCGCCACCGCCGCCACCATGAATATTCAGGC
TCTTCTCTCAGAAAAAGTCCGTCAGGCCATGATTGCGGCAGGCGCGCCTG
CGGATTGCGAACCGCAGGTTCGTCAGTCAGCAAAAGTTCAGTTCGGCGAC
TATCAGGCTAACGGCATGATGGCAGTTGCTAAAAAACTGGGTATGGCACC
GCGACAATTAGCAGAGCAGGTGCTGACTCATCTGGATCTTAACGGTATCG
CCAGCAAAGTTGAGATCGCCGGTCCAGGCTTTATCAACATTTTCCTTGATC
CGGCATTCCTGGCTGAACATGTTCAGCAGGCGCTGGCGTCCGATCGTCTC
GGTGTTGCTACGCCAGAAAAACAGACCATTGTGGTTGACTACTCTGCGCC
AAACGTGGCGAAAGAGATGCATGTCGGTCACCTGCGCTCTACCATTATTG
GTGACGCAGCAGTGCGTACTCTGGAGTTCCTCGGTCACAAAGTGATTCGC
GCAAACCACGTCGGCGACTGGGGCACTCAGTTCGGTATGCTGATTGCATG
GCTGGAAAAGCAGCAGCAGGAAAACGCCGGTGAAATGGAGCTGGCTGAC
CTTGAAGGTTTCTACCGCGATGCGAAAAAGCATTACGATGAAGATGAAGA
GTTCGCCGAGCGCGCACGTAACTACGTGGTAAAACTGCAAAGCGGTGACG
AATATTTCCGCGAGATGTGGCGCAAACTGGTCGACATCACCATGACGCAG
AACCAGATCACCTACGATCGTCTCAACGTGACGCTGACCCGTGATGACGT
GATGGGCGAAAGCCTCTACAACCCGATGCTGCCAGGAATTGTGGCGGATC
TCAAAGCCAAAGGTCTGGCAGTAGAAAGCGAAGGGGCGACCGTCGTATT
CCTTGATGAGTTTAAAAACAAGGAAGGCGAACCGATGGGCGTGATCATTC
AGAAGAAAGATGGCGGCTATCTCTACACCACCACTGATATCGCCTGTGCG
AAATATCGTTATGAAACACTGCATGCCGATCGCGTGCTGTATTACATCGA
CTCCCGTCAGCATCAACACCTGATGCAGGCATGGGCGATCGTCCGTAAAG
CAGGCTATGTACCGGAATCCGTACCGCTGGAACACCACATGTTCGGCATG
ATGCTGGGTAAAGACGGCAAACCGTTCAAAACCCGCGCGGGTGGTACAGT
GAAACTGGCCGATCTGCTGGATGAAGCCCTGGAACGTGCACGCCGTCTGG
TGGCAGAAAAGAACCCGGATATGCCAGCCGACGAGCTGGAAAAACTGGC
TAACGCGGTTGGTATTGGTGCGGTGAAATATGCGGATCTCTCCAAAAACC
GCACCACGGACTACATCTTCGACTGGGACAACATGCTGGCGTTTGAGGGT
AATACCGCGCCATACATGCAGTATGCATACACGCGTGTATTGTCCGTGTTC
CGTAAAGCAGAAATTGACGAAGAGCAACTGGCTGCAGCTCCGGTTATCAT
CCGTGAAGATCGTGAAGCGCAACTGGCAGCTCGCCTGCTGCAGTTTGAAG
AAACCCTCACCGTGGTTGCCCGTGAAGGCACGCCGCATGTAATGTGTGCT
TACCTGTACGATCTGGCCGGTCTGTTCTCTGGCTTCTACGAGCACTGCCCG
ATCCTCAGCGCAGAAAACGAAGAAGTGCGTAACAGCCGTCTAAAACTGG
CACAACTGACGGCGAAGACGCTGAAGCTGGGTCTGGATACGCTGGGTATT
GAGACTGTAGAGCGTATGTAAGAATTCAACGCGTTAAGTCGACTTTAACT
CGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGGTTGCCAGCCATCTG
TTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGT

GTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGA
TTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTT
CTGAGGCGGAAAGAACCCTAGGGGTGCGAGCGGATCGAGCAGTGTCGAT
CACTACTGGACCGCGAGCTGTGCTGCGACCCGTGATCTTACGGCATTATA
CGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTTGATGTCATA
GCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCATACCTGACCTCCAT
AGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTTGATCGGCACG
TAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTA
TTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGAGCCAT
ATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAACATGGA
TGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAG
GTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTT
CTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGT
CAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTT
TATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAA
AACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTG
TTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTA
ATTGTCCTTTTAACGGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCAC
GAATGAATAACGGTTTGGTTGGTGCGAGTGATTTTGATGACGAGCGTAAT
GGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATT
CTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTAT
TTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAA
TCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAG
TTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAAT
CCTGATATGAATAAATTGCAGTTTCACTTGATGCTCGATGAGTTTTTCTAA
TGAGGACCTAAATGTAATCACCTGGCTCACCTTCGGGTGGGCCTTTCTGCG
TTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAAT
CGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACC
AGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGC
CGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTT
CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCA
AGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTA
TCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCC
ACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC
GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
AACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAG
AGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAA
GATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pIDTSMART EcPheRS αβ T252G A294G



CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATC
GTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGA
GTCTGAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGA
GACGGGCGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGGCGGC
CGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGT
CATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAA
ATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATA
ATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT
ATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCC
GCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAG
TACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG
TACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT
CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG

ACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGT
AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAG
AGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGG
AGACCCAAGCTGGCTAGCGCCGCCACCATGTCACATCTCGCAGAACTGGT
TGCCAGTGCGAAGGCGGCCATTAGCCAGGCGTCAGATGTTGCCGCGTTAG
ATAATGTGCGCGTCGAATATTTGGGTAAAAAAGGGCACTTAACCCTTCAG
ATGACGACCCTGCGTGAGCTGCCGCCAGAAGAGCGTCCGGCAGCTGGTGC
GGTTATCAACGAAGCGAAAGAGCAGGTTCAGCAGGCGCTGAATGCGCGT
AAAGCGGAACTGGAAAGCGCTGCACTGAATGCGCGTCTGGCGGCGGAAA
CGATTGATGTCTCTCTGCCAGGTCGTCGCATTGAAAACGGCGGTCTGCATC
CGGTTACCCGTACCATCGACCGTATCGAAAGTTTCTTCGGTGAGCTTGGCT
TTACCGTGGCAACCGGGCCGGAAATCGAAGACGATTATCATAACTTCGAT
GCTCTGAACATTCCTGGTCACCACCCGGCGCGCGCTGACCACGACACTTTC
TGGTTTGACACTACCCGCCTGCTGCGTACCCAGACCTCTGGCGTACAGATC
CGCACCATGAAAGCCCAGCAGCCACCGATTCGTATCATCGCGCCTGGCCG
TGTTTATCGTAACGACTACGACCAGACTCACACGCCGATGTTCCATCAGAT
GGAAGGTCTGATTGTTGATACCAACATCAGCTTTACCAACCTGAAAGGCA
CGCTGCACGACTTCCTGCGTAACTTCTTTGAGGAAGATTTGCAGATTCGCT
TCCGTCCTTCCTACTTCCCGTTTGGCGAACCTTCTGCAGAAGTGGACGTCA
TGGGTAAAAACGGTAAATGGCTGGAAGTGCTGGGCTGCGGGATGGTGCAT
CCGAACGTGTTGCGTAACGTTGGCATCGACCCGGAAGTTTACTCTGGTTTC
GGCTTCGGGATGGGGATGGAGCGTCTGACTATGTTGCGTTACGGCGTCAC
CGACCTGCGTTCATTCTTCGAAAACGATCTGCGTTTCCTCAAACAGTTTAA
ATAAGAATTCAACGCGTTAAGTCGACTTTAACTCGAGTCTAGAGGGCCCG
TTTAAACCCGCTGATCAGGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCG
TGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAA
ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGG
GTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG
GCATGCTGGGGATGCGGTGGGCTCTATGGCCTGCAGGATCCGTTTTGCGC
TGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTA
GTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGG
AGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCC
AACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC
GCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTATTTACGGTAAA
CTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTA
TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATG
ACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCT
ATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA
GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACT
CCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTAT

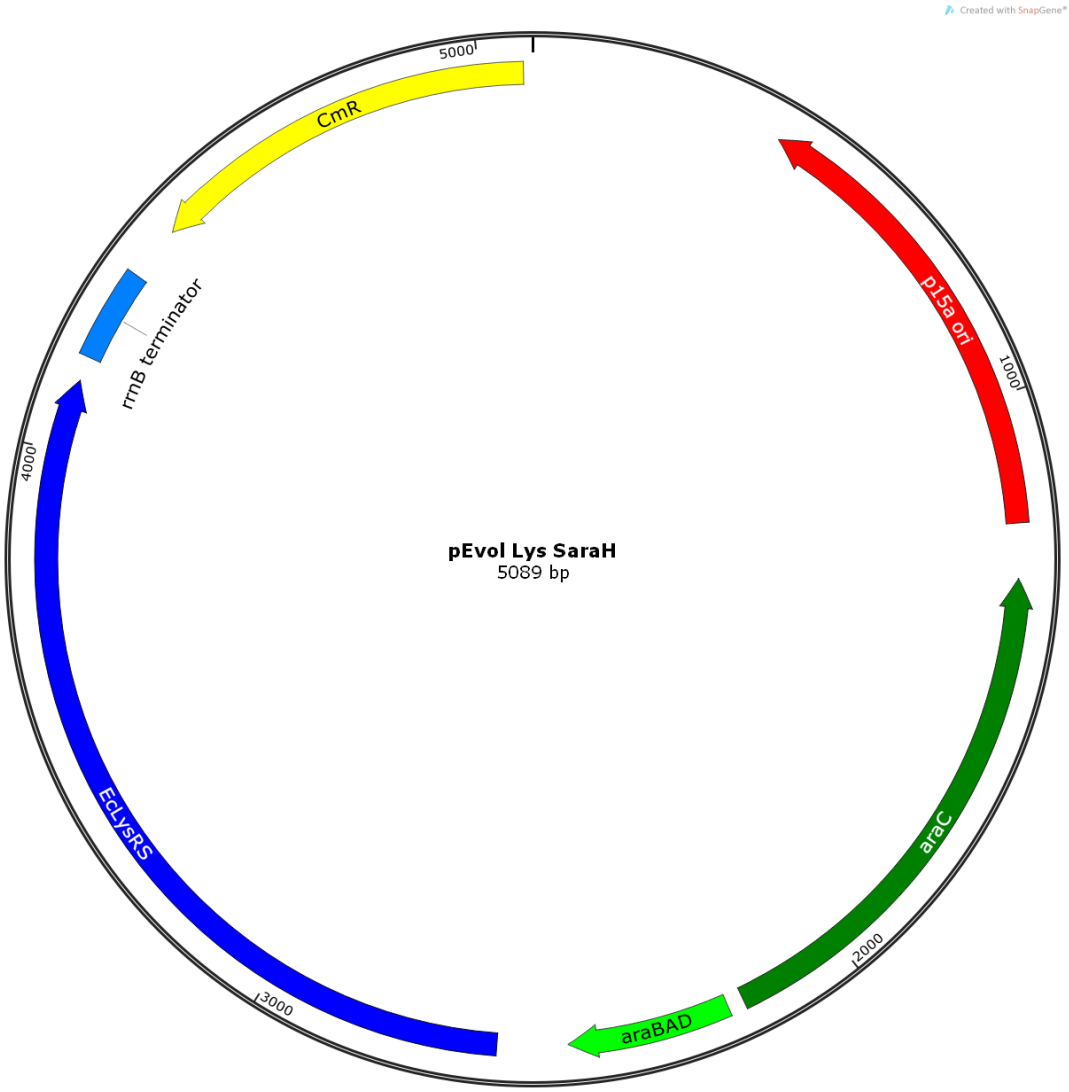
ATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTAT
CGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGCCGCC
ACCATGAAATTCAGTGAACTGTGGTTACGCGAATGGGTGAACCCGGCGAT
TGATAGCGATGCGCTGGCAAATCAAATCACTATGGCGGGCCTGGAAGTTG
ACGGTGTAGAACCGGTTGCCGGCAGCTTCCACGGCGTGGTCGTTGGTGAA
GTGGTTGAGTGTGCGCAGCATCCGAACGCTGACAAACTGCGTGTGACAAA
AGTGAATGTCGGCGGCGATCGCCTGCTGGACATCGTCTGCGGTGCGCCAA
ACTGCCGTCAGGGCCTGCGTGTAGCGGTAGCGACCATTGGTGCTGTTCTG
CCGGGTGATTTCAAAATTAAAGCGGCGAAACTGCGTGGCGAACCGTCTGA
AGGGATGCTGTGCTCCTTCTCTGAACTGGGCATTTCTGACGATCACAGCGG
CATTATCGAACTGCCTGCGGATGCGCCGATTGGCACCGATATCCGTGAAT
ACCTGAAACTTGATGACAACACCATCGAAATCAGCGTGACGCCAAACCGT
GCCGACTGCTTAGGCATCATTGGTGTTGCGCGTGACGTTGCCGTGCTGAAC
CAGCTGCCGCTGGTTCAACCGGAAATCGTTCCGGTTGGTGCGACCATCGA
CGACACGCTGCCGATTACAGTCGAAGCGCCGGAAGCCTGCCCGCGTTATC
TTGGCCGTGTGGTAAAAGGCATTAACGTTAAAGCGCCAACTCCGCTGTGG
ATGAAAGAAAAACTGCGTCGTTGCGGGATCCGTTCTATCGATGCAGTTGT
TGACGTCACCAACTATGTGCTGCTCGAACTGGGCCAGCCGATGCACGCTT
TCGATAAAGATCGCATTGAAGGCGGCATTGTGGTGCGGATGGCGAAAGA
GGGCGAAACGCTGGTGCTGCTCGACGGTACTGAAGCGAAGCTGAATGCTG
ACACTCTGGTCATCGCCGACCACAACAAGGCGCTGGCGATGGGCGGCATC
TTCGGTGGCGAACACTCTGGCGTGAATGACGAAACACAAAACGTGCTGCT
GGAATGCGCGTTCTTTAGCCCGCTGTCTATCACCGGTCGTGCTCGTCGTCA
TGGCCTGCATACCGATGCGTCTCACCGTTATGAGCGTGGCGTTGATCCGGC
ACTGCAGCACAAAGCGATGGAACGTGCGACCCGTCTGCTGATCGACATCT
GCGGTGGTGAGGCTGGCCCGGTAATTGATATCACCAACGAAGCAACGCTG
CCGAAGCGTGCAACCATCACTCTACGTCGTAGCAAACTGGATCGCCTGAT
CGGCCATCATATTGCGGATGAGCAGGTAACTGACATTCTGCGTCGTCTCG
GCTGCGAAGTGACCGAAGGCAAAGACGAATGGCAGGCAGTTGCGCCGAG
CTGGCGTTTCGATATGGAGATTGAAGAAGATCTGGTTGAAGAAGTCGCGC
GTGTTTACGGCTACAACAACATCCCGGATGAGCCGGTACAGGCAAGCCTG
ATTATGGGTACTCACCGTGAAGCTGACCTGTCGCTCAAGCGCGTGAAAAC
GCTGCTCAACGACAAAGGCTATCAGGAAGTGATCACCTACAGCTTCGTTG
ATCCGAAAGTGCAGCAGATGATCCATCCAGGCGTTGAAGCCTTACTGCTG
CCAAGCCCGATCTCTGTTGAAATGTCAGCAATGCGTCTTTCTCTGTGGACT
GGCCTGCTGGCAACCGTGGTGTACAACCAGAACCGTCAGCAGAACCGTGT
GCGCATTTTCGAAAGCGGTCTGCGTTTCGTACCAGATACTCAGGCACCGTT
GGGCATTCGTCAGGATCTGATGTTAGCCGGTGTGATTTGCGGTAACCGTTA
CGAAGAGCACTGGAACCTGGCAAAAGAGACCGTTGATTTCTATGATTTGA
AAGGCGATCTTGAATCCGTTCTCGACCTGACCGGTAAACTGAATGAGGTT
GAGTTCCGTGCAGAAGCGAATCCGGCACTGCATCCGGGGCAATCCGCAGC

GATTTATCTGAAAGGTGAACGTATTGGTTTTGTTGGGGTTGTTCATCCTGA
ACTGGAACGTAAACTGGATCTTAACGGTCGCACTCTGGTGTTCGAACTGG
AGTGGAACAAGCTCGCAGACCGCGTGGTGCCTCAGGCGCGCGAGATTTCT
CGCTTCCCGGCGAACCGTCGTGACATCGCGGTGGTGGTCGCAGAAAACGT
TCCCGCAGCGGATATTTTATCCGAATGTAAGAAAGTTGGCGTAAATCAGG
TAGTTGGCGTAAACTTATTTGACGTGTACCGCGGTAAGGGTGTTGCGGAG
GGGTATAAGAGCCTCGCCATAAGCCTGATCCTGCAAGATACCAGCCGTAC
ACTCGAAGAAGAGGAGATTGCCGCTACCGTCGCCAAATGTGTAGAGGCAT
TAAAAGAGCGATTCCAGGCATCATTGAGGGATTTAGTCGACTTTAACTCG
AGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGGTTGCCAGCCATCTGTT
GTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACT
GTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGT
CATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATT
GGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCT
GAGGCGGAAAGAACCCTAGGGGTGCGAGCGGATCGAGCAGTGTCGATCA
CTACTGGACCGCGAGCTGTGCTGCGACCCGTGATCTTACGGCATTATACGT
ATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTTGATGTCATAGCT
GTTTCCTGAGGCTCAATACTGACCATTTAAATCATACCTGACCTCCATAGC
AGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTTGATCGGCACGTAA
GAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTT
TTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGAGCCATATT
CAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAACATGGATGC
TGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATCAGGTG
CGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTG
AAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAG
GCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTAT
CCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCAGGGAAAA
CAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTT
GATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAAT
TGTCCTTTTAACGGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGA
ATGAATAACGGTTTGGTTGGTGCGAGTGATTTTGATGACGAGCGTAATGG
CTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTCTTGCCATTCT
CACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTTATTT
TTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATC
GCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTT
TTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCC
TGATATGAATAAATTGCAGTTTCACTTGATGCTCGATGAGTTTTTCTAATG
AGGACCTAAATGTAATCACCTGGCTCACCTTCGGGTGGGCCTTTCTGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCA
GGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCC

GCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTC
TCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAA
GCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATC
CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCAC
TGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGG
TGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAA
CAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAAAAAGAG
TTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATTTTCTACCGAAGAAAGGCCCA

Chapter 4

pEvol LysS



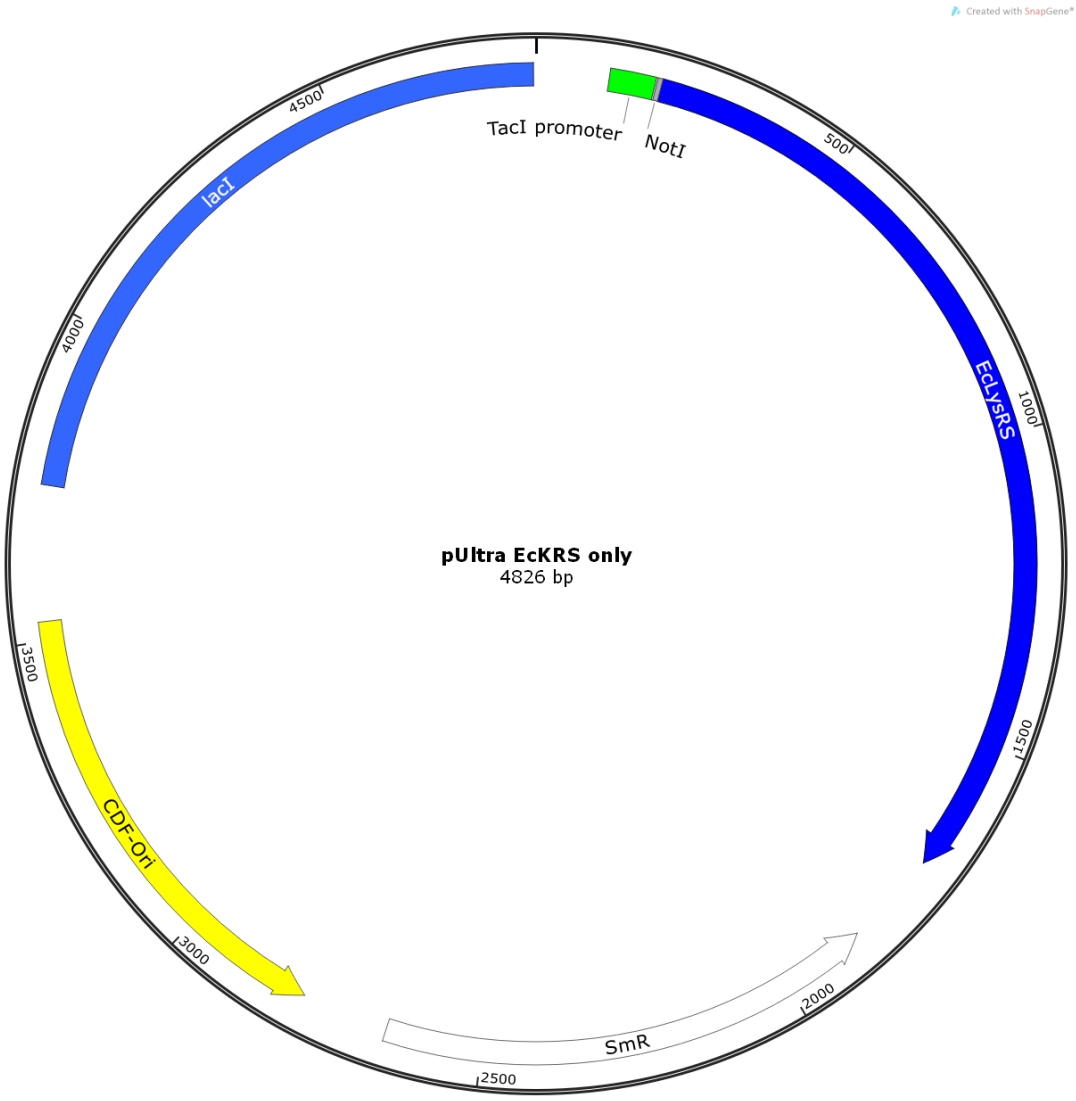
TCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTC
ATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCG

CCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTA
TTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTG
CGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGA
GGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTG
ACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCG
GAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGT
GCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATAT
GTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGT
CGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCC
TGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCA
AAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTG
ACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGT
TTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTG
ACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGA
ACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGT
AATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAA
CTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGG
TTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGC
GGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTC
AAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAAT
TTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGT
AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAATTATGA
CAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTC
GCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTT
GATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTG
GTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTT
AAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACG
GCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCT
GCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCAT
CGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATT
GCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCC
CGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCT
TCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAG
CCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACC
ATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGG
AACAGCAAAATATCACTCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTC
ACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCC
CAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCG
TTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGA

TCATTTTGCGCTTCAGCCATACTTTTCATACTCCCGCCATTCAGAGAAGAA
ACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGG
CTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAA
CAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATA
ATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCT
ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTT
TATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAAGAAA
TAATTTTGTTTAACTTTAAGAAGGAGAATACATCAACTAGTACGCAAGTTC
ACGTAAAAAGGGTATCTAGAGGTTGAGGTGATTTTATGTCTGAACAACAC
GCACAGGGCGCTGACGCGGTAGTCGATCTTAACAATGAACTGAAAACGCG
TCGTGAGAAGCTGGCGAACCTGCGCGAGCAGGGGATTGCCTTCCCGAACG
ATTTCCGTCGCGATCATACCTCTGACCAATTGCACGCAGAATTCGACGGC
AAAGAGAACGAAGAACTGGAAGCGCTGAACATCGAAGTCGCCGTTGCTG
GCCGCATGATGACCCGTCGTATTATGGGTAAAGCGTCTTTCGTTACCCTGC
AGGACGTTGGCGGTCGCATTCAGCTGTACGTTGCCCGTGACGATCTCCCG
GAAGGCGTTTATAACGAGCAGTTCAAAAAATGGGACCTCGGCGACATCCT
CGGCGCGAAAGGTAAGCTGTTCAAAACCAAAACCGGCGAACTGTCTATCC
ACTGCACCGAGTTGCGTCTGCTGACCAAAGCACTGCGTCCGCTGCCGGAT
AAATTCCACGGCTTGCAGGATCAGGAAGCGCGCTATCGTCAGCGTTATCT
CGATCTCATCTCCAACGATGAATCCCGCAACACCTTTAAAGTGCGCTCGC
AGATCCTCTCTGGTATTCGCCAGTTCATGGTGAACCGCGGCTTTATGGAAG
TTGAAACGCCGATGATGCAGGTGATCCCTGGCGGTGCCGCTGCGCGTCCG
TTTATCACCCACCATAACGCGCTGGATCTCGACATGTACCTGCGTATCGCG
CCGGAACTGTACCTCAAGCGTCTGGTGGTTGGTGGCTTCGAGCGTGTATTC
GAAATCAACCGTAACTTCCGTAACGAAGGTATTTCCGTACGTCATAACCC
AGAGTTCACCATGATGGAACTCTACATGGCTTACGCAGATTACAAAGATC
TGATCGAGCTGACCGAATCGCTGTTCCGTACTCTGGCACAGGATATTCTCG
GTAAGACGGAAGTGACCTACGGCGACGTGACGCTGGACTTCGGTAAACCG
TTCGAAAAACTGACCATGCGTGAAGCGATCAAGAAATATCGCCCGGAAAC
CGACATGGCGGATCTGGACAACTTCGACTCTGCGAAAGCAATTGCTGAAT
CTATCGGCATCCACGTTGAGAAGAGCTGGGGTCTGGGCCGTATCGTTACC
GAGATCTTCGAAGAAGTGGCAGAAGCACATCTGATTCAGCCGACCTTCAT
TACTGAATATCCGGCAGAAGTTTCTCCGCTGGCGCGTCGTAACGACGTTA
ACCCGGAAATCACAGACCGCTTTGAGTTCTTCATTGGTGGTCGTGAAATC
GGTAACGGCTTTAGCGAGCTGAATGACGCGGAAGATCAGGCGCAACGCTT
CCTGGATCAGGTTGCCGCGAAAGACGCAGGTGACGACGAAGCGATGTTCT
ACGATGAAGATTACGTCACCGCACTGGAACATGGCTTACCGCCGACAGCA
GGTCTGGGAATTGGTATCGACCGTATGGTAATGCTGTTCACCAACAGCCA
TACCATCCGCGACGTTATTCTGTTCCCGGCGATGCGTCCGGTAAAATAAAC
CGATGCGGCCGCTTGAGAGTCAGCTCCTTCCGGTGGGCGTGCCTGGCGGC
AGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACG

CCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACT
GCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTGTC
GACCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCA
ACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGC
CCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACAT
GGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAG
CACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGA
AGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACC
CAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAA
ATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTA
GAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTT
TCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATAT
CACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCAT
CAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTT
TCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTAT
AGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGC
CATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTA
GCTTCCTTAGC

pUltra LysS

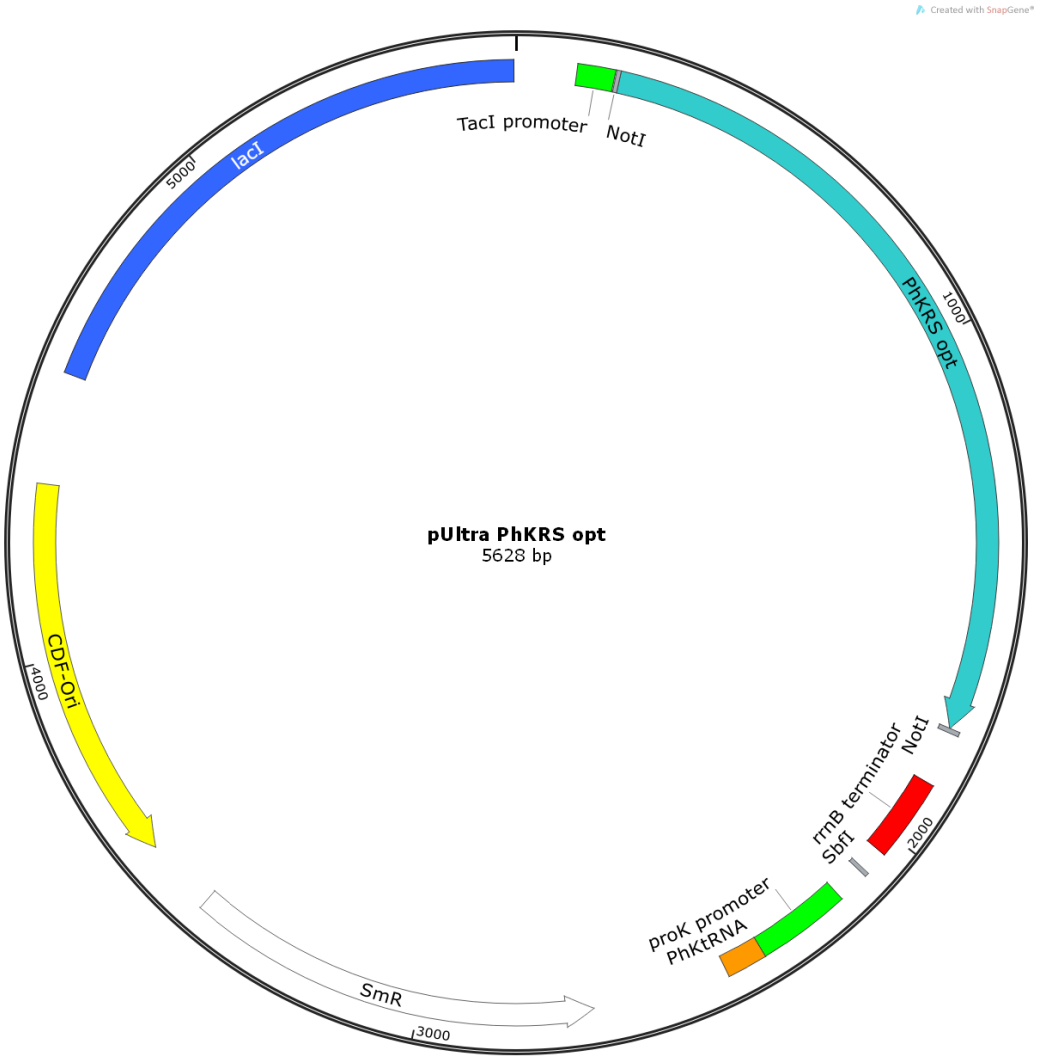


ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCG
ACTCCTGCATTAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTG
TGGAATTGTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGTC
TGAACAACACGCACAGGGCGCTGACGCGGTAGTCGATCTTAACAATGAAC
TGAAAACGCGTCGTGAGAAGCTGGCGAACCTGCGCGAGCAGGGGATTGC
CTTCCCGAACGATTTCCGTCGCGATCATACCTCTGACCAATTGCACGCAGA
ATTCGACGGCAAAGAGAACGAAGAACTGGAAGCGCTGAACATCGAAGTC
GCCGTTGCTGGCCGCATGATGACCCGTCGTATTATGGGTAAAGCGTCTTTC
GTTACCCTGCAGGACGTTGGCGGTCGCATTCAGCTGTACGTTGCCCGTGAC
GATCTCCCGGAAGGCGTTTATAACGAGCAGTTCAAAAAATGGGACCTCGG
CGACATCCTCGGCGCGAAAGGTAAGCTGTTCAAAACCAAAACCGGCGAA
CTGTCTATCCACTGCACCGAGTTGCGTCTGCTGACCAAAGCACTGCGTCCG
CTGCCGGATAAATTCCACGGCTTGCAGGATCAGGAAGCGCGCTATCGTCA
GCGTTATCTCGATCTCATCTCCAACGATGAATCCCGCAACACCTTTAAAGT

GCGCTCGCAGATCCTCTCTGGTATTCGCCAGTTCATGGTGAACCGCGGCTT
TATGGAAGTTGAAACGCCGATGATGCAGGTGATCCCTGGCGGTGCCGCTG
CGCGTCCGTTTATCACCCACCATAACGCGCTGGATCTCGACATGTACCTGC
GTATCGCGCCGGAACTGTACCTCAAGCGTCTGGTGGTTGGTGGCTTCGAG
CGTGTATTCGAAATCAACCGTAACTTCCGTAACGAAGGTATTTCCGTACGT
CATAACCCAGAGTTCACCATGATGGAACTCTACATGGCTTACGCAGATTA
CAAAGATCTGATCGAGCTGACCGAATCGCTGTTCCGTACTCTGGCACAGG
ATATTCTCGGTAAGACGGAAGTGACCTACGGCGACGTGACGCTGGACTTC
GGTAAACCGTTCGAAAAACTGACCATGCGTGAAGCGATCAAGAAATATCG
CCCGGAAACCGACATGGCGGATCTGGACAACTTCGACTCTGCGAAAGCAA
TTGCTGAATCTATCGGCATCCACGTTGAGAAGAGCTGGGGTCTGGGCCGT
ATCGTTACCGAGATCTTCGAAGAAGTGGCAGAAGCACATCTGATTCAGCC
GACCTTCATTACTGAATATCCGGCAGAAGTTTCTCCGCTGGCGCGTCGTAA
CGACGTTAACCCGGAAATCACAGACCGCTTTGAGTTCTTCATTGGTGGTCG
TGAAATCGGTAACGGCTTTAGCGAGCTGAATGACGCGGAAGATCAGGCGC
AACGCTTCCTGGATCAGGTTGCCGCGAAAGACGCAGGTGACGACGAAGC
GATGTTCTACGATGAAGATTACGTCACCGCACTGGAACATGGCTTACCGC
CGACAGCAGGTCTGGGAATTGGTATCGACCGTATGGTAATGCTGTTCACC
AACAGCCATACCATCCGCGACGTTATTCTGTTCCCGGCGATGCGTCCGGTA
AAATAAACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAG
CAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGG
GTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGAC
ATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATT
CTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGAT
AAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGC
TCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACT
GCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCC
AGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCT
CAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGA
CCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGA
TCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCG
CTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAAT
GATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGC
TCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGC
CGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCA
CTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGC
CAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATA
GTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCA
ATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATT
TGAATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACG
CTCCGGGCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCC

ACAGATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAG
GGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCA
CATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACC
ATGAATCTGACAGTACGGGCGAAACCCGACAGGACTTAAAGATCCCCACC
GTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTT
ACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTC
ATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCT
GGGCTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCG
GTAACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTG
GCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAAC
ACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGAC
AGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTC
CCCAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGT
TTACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCA
AATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGT
TAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATT
AATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATT
GCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGG
TTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATA
TAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCA
CCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCAT
CTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCA
TTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTT
CCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCC
AGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGA
TTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCG
TCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATC
AAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCAT
CCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCG
AGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACC
ATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGC
CGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACG
CCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGG
AATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGC
AGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAG
ACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTC

pUltra PhK full length



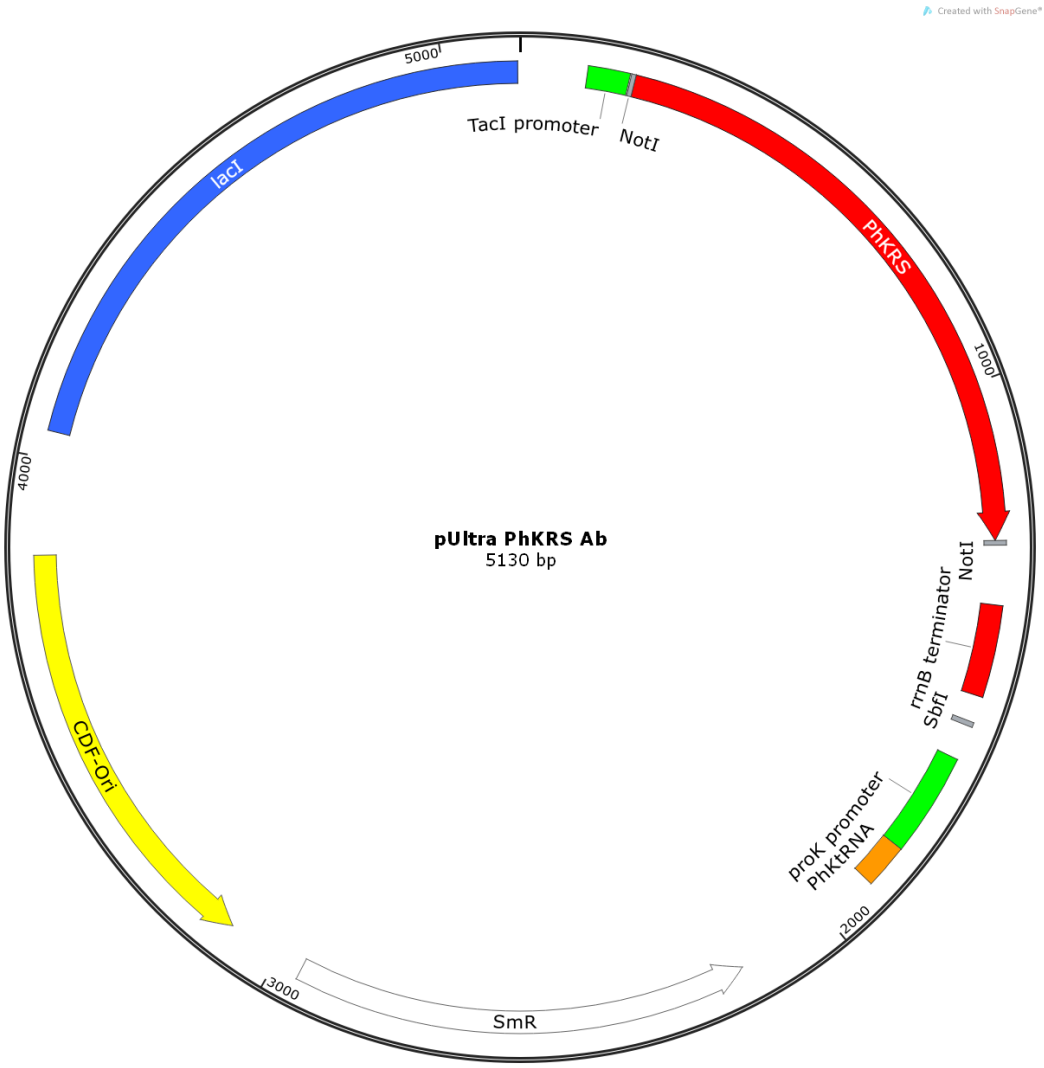
ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCG
ACTCCTGCATTAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTG
TGGAATTGTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGGT
ACATTGGGCTGATTATATTGCGGATAAGATCATTCGTGAACGCGGAGAGA
AAGAGAAATACGTTGTAGAAAGTGGTATCACTCCGTCTGGTTATGTACAT
GTGGGCAATTTCCGTGAGTTATTCACAGCATACATTGTCGGACACGCGTTA
CGCGACAAAGGATATGAAGTTCGCCACATTCACATGTGGGATGACTATGA
TCGCTTTCGCAAAGTACCGCGTAATGTCCCGCAAGAGTGGAAAGACTACT
TAGGCATGCCCATCTCTGAGGTGCCTGACCCTTGGGGGTGCCACGAGTCTT
ATGCAGAGCATTTTATGCGTAAATTCGAGGAAGAGGTAGAAAAACTGGGC
ATTGAGGTCGACTTCTTATATGCTTCCGAATTGTACAAGCGCGGTGAATAC
TCGGAGGAGATCCGCCTTGCATTTGAAAAGCGTGACAAGATTATGGAGAT
CCTGAATAAATATCGTGAAATCGCCAAACAACCGCCATTGCCAGAGAACT
GGTGGCCAGCCATGGTCTACTGTCCGGAACACCGCCGCGAAGCGGAAATT

ATTGAATGGGACGGTGGATGGAAAGTCAAATACAAGTGTCCTGAAGGCC
ATGAAGGCTGGGTTGATATTCGTTCGGGCAACGTAAAACTGCGTTGGCGT
GTTGATTGGCCGATGCGTTGGTCACATTTCGGCGTAGATTTTGAGCCTGCG
GGCAAGGACCACCTTGTAGCTGGATCTTCATATGATACTGGAAAGGAAAT
CATTAAGGAGGTATACGGGAAGGAAGCTCCACTGAGCCTTATGTACGAAT
TTGTAGGAATCAAAGGACAAAAGGGTAAGATGTCAGGGTCTAAAGGTAA
TGTTATTCTTTTGTCGGATCTGTATGAAGTACTGGAGCCAGGGCTTGTCCG
CTTTATCTATGCGCGTCACCGCCCGAACAAGGAGATTAAAATTGATTTGG
GTTTGGGTATCTTGAATTTGTACGACGAGTTCGATAAAGTAGAGCGCATCT
ACTTTGGAGTTGAGGGAGGTAAGGGAGACGACGAGGAACTGCGTCGTAC
ATATGAATTATCCATGCCAAAGAAACCAGAACGCCTTGTAGCACAAGCGC
CTTTTCGTTTCTTAGCAGTACTGGTGCAGCTGCCCCACTTGACTGAAGAGG
ATATCATCAACGTATTGATCAAGCAAGGGCATATTCCACGTGATCTTAGT
AAGGAGGACGTAGAGCGTGTTAAACTTCGCATTAACTTGGCTCGCAACTG
GGTCAAGAAATATGCACCAGAAGATGTTAAGTTCAGCATTTTAGAAAAAC
CTCCCGAAGTTGAGGTCTCCGAAGACGTACGTGAAGCAATGAACGAAGTA
GCCGAGTGGTTGGAAAACCACGAAGAGTTTTCCGTAGAAGAGTTCAACAA
TATTCTGTTTGAGGTCGCGAAGCGTCGCGGAATTTCCAGTCGTGAGTGGTT
TAGCACGCTTTACCGTCTTTTTATCGGGAAAGAACGCGGACCTCGTCTGGC
ATCGTTTCTTGCGTCTCTTGATCGTTCGTTCGTAATTAAACGCCTGCGCCTG
GAAGGCTAAGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGTTTTGGCG
GATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGC
GGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACC
TGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTG
TGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACG
AAAGGCTCAGTCGAAAGACTGGGCCTTGTTTGTGAGCTCCCGGTCATCAA
TCATCCCCATAATCCTTGTTAGCCTGCAGGTAATTCCGCTTCGCAACATGT
GAGCACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAA
ATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGAC
GATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGA
TCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGT
AAGATGCGCCCCGCATTTGGTCCGTAGCTCAGCCTGGTAGAGCGGCGGGC
TTTTAACCCGCAGGTCGCGGGTTCAAATCCCGCCGGACTAGCCAAATTCG
AAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGG
TCGAATTTGCCATGGCGGCCACCAGGTACCACCGGCGCCTCAGGCATTTG
AGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAG
CAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGG
GTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGAC
ATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATT
CTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGAT
AAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGC

TCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACT
GCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCC
AGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCT
CAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGA
CCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGA
TCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCG
CTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAAT
GATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGC
TCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGC
CGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCA
CTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGC
CAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATA
GTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCA
ATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATT
TGAATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACG
CTCCGGGCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCC
ACAGATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAG
GGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCA
CATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACC
ATGAATCTGACAGTACGGGCGAAACCCGACAGGACTTAAAGATCCCCACC
GTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTT
ACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTC
ATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCT
GGGCTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCG
GTAACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTG
GCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAAC
ACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGAC
AGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTC
CCCAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGT
TTACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCA
AATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGT
TAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTC
AAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATT
AATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCA
GCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTG
GGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATT
GCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGG
TTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATA
TAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCA
CCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCAT

CTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCA
TTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTT
CCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCC
AGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGA
TTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCG
TCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATC
AAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCAT
CCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCG
AGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACC
ATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGC
CGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACG
CCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGG
AATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGC
AGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAG
ACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTC

pUltra PhK truncated



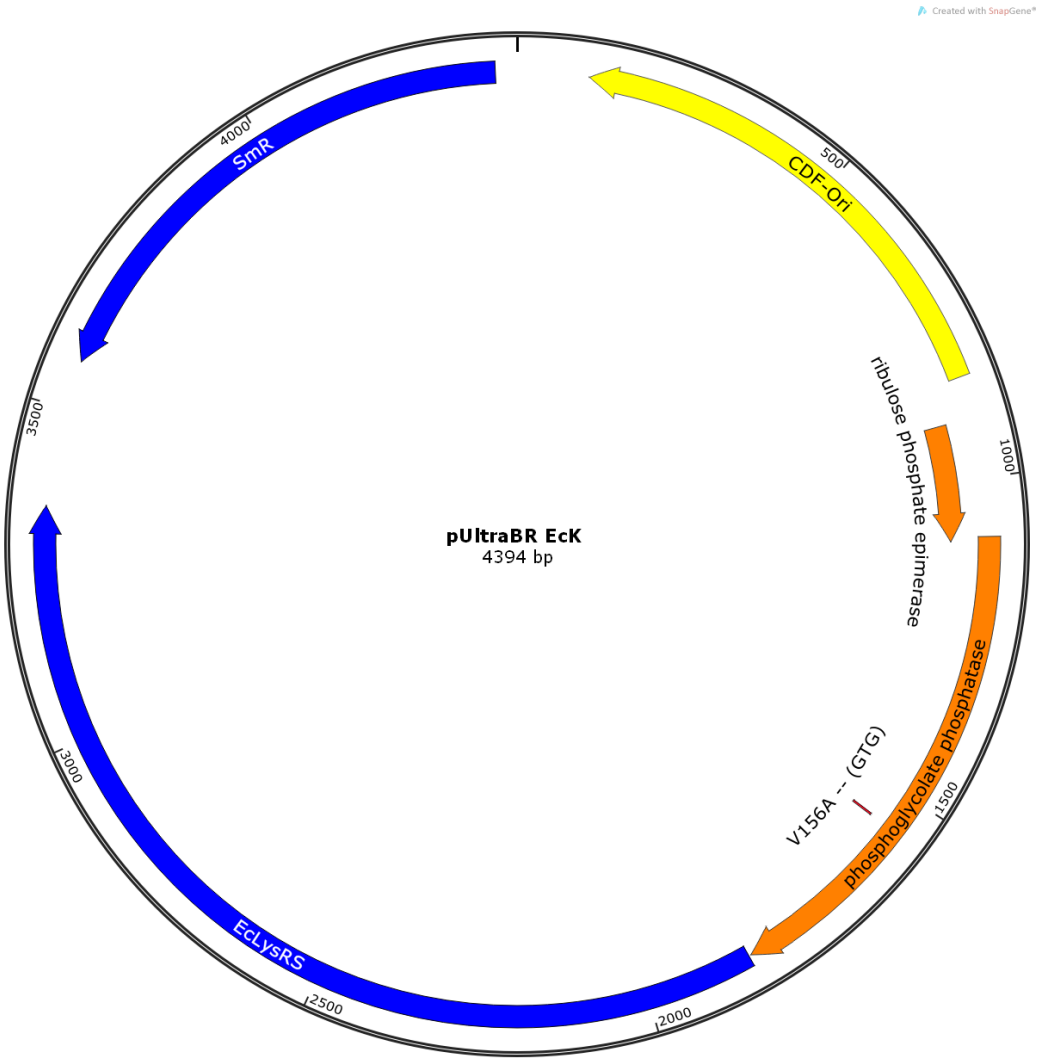
ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCG
ACTCCTGCATTAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTG
TGGAATTGTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGGT
TCATTGGGCCGATTATATTGCTGATAAAATAATTAGAGAGAGGGGGGAGA
AGGAGAAGTACGTTGTTGAGAGTGGAATAACGCCAAGTGGTTACGTTCAC
GTTGGGAACTTTAGGGAGCTTTTTACAGCTTATATTGTGGGCCATGCCCTA
AGGGATAAGGGGTATGAGGTTAGGCACATCCACATGTGGGATGATTATGA
TAGATTTAGGAAGGTTCCAAGGAACGTTCCCCAGGAATGGAAAGATTACC
TGGGAATGCCCATTAGTGAAGTTCCTGATCCCTGGGGATGCCATGAGAGT
TATGCTGAACACTTCATGAGAAAGTTCGAGGAGGAGGTAGAAAAATTAGG
GATCGAAGTTGACTTTCTTTATGCGAGTGAACTCTACAAGAGAGGGGAAT
ATTCTGAGGAGATAAGGTTAGCCTTTGAGAAAAGGGATAAGATAATGGAG
ATACTAAACAAGTATAGGGAAATTGCGAAACAACCTCCCCTTCCAGAGAA
CTGGTGGCCCGCAATGGTTTACTGCCCTGAGCATAGGAGGGAAGCAGAGA

TCATTGAATGGGATGGGGGCTGGAAGGTTAAGTATAAGTGCCCCGAAGGT
CACGAGGGATGGGTTGATATAAGGAGTGGGAACGTGAAACTGAGGTGGC
GTGTTGATTGGCCCATGCGTTGGTCTCACTTTGGCGTTGACTTCGAACCTG
CTGGAAAGGATCATCTTGTGGCTGGTTCAAGCTACGATACGGGAAAGGAG
ATTATAAAGGAAGTTTATGGAAAGGAAGCTCCGTTATCTTTAATGTATGA
GTTTGTTGGAATTAAGGGGCAGAAGGGGAAGATGAGTGGTAGTAAGGGA
AATGTTATTTTACTCAGCGATCTGTATGAGGTTCTTGAGCCAGGTCTCGTT
AGATTTATCTACGCTCGGCATAGGCCAAACAAGGAGATAAAGATAGATCT
AGGTCTTGGCATTCTAAACCTCTACGATGAGTTCGATAAAGTTGAGAGAA
TATACTTCGGGGTTGAGGGTGGTAAAGGTGATGATGAAGAATTAAGGAGG
ACTTACGAGCTTTCATAAGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTG
TTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAAC
GCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTG
GTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGA
TGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAA
ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTGTTTGTGAGCTCCCG
GTCATCAATCATCCCCATAATCCTTGTTAGCCTGCAGGTAATTCCGCTTCG
CAACATGTGAGCACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTGTG
CTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAAT
AATTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTTC
TCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCT
GCGCAGTAAGATGCGCCCCGCATTTGGTCCGTAGCTCAGCCTGGTAGAGC
GGCGGGCTTTTAACCCGCAGGTCGCGGGTTCAAATCCCGCCGGACTAGCC
AAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAG
CTCAGGGTCGAATTTGCCATGGCGGCCACCAGGTACCACCGGCGCCTCAG
GCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGT
AAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGA
CGACCGGGTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATT
GTTAGACATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGG
ACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGT
CCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGG
CAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCC
GGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCT
CATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTT
AGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGC
CGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGAT
AGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGT
CATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCC
ACGGAATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGA
ATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAA
AGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATC

AATATCACTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAAT
GTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACC
TCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTC
CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGA
TACATATTTGAATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGG
TCGCTACGCTCCGGGCGTGAGACTGCGGCGGGCGCTGCGGACACATACAA
AGTTACCCACAGATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAA
ACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCC
AGAGTTCACATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAG
GCTCAACCATGAATCTGACAGTACGGGCGAAACCCGACAGGACTTAAAG
ATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGAC
CCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTG
GCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTC
GCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGC
GCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTAA
AACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGT
TTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACA
CTGGAAGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGG
TTAAGCAGTTCCCCAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGT
GGTTTTTTCGTTTACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTC
AAGAAGATCCTTTGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGCAAT
TTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGT
AATTCTCATGTTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGT
TGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGC
TAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGG
TTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCA
ACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGG
TCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAAC
GGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGA
GATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGC
CCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCC
TCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCG
CCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGC
CAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTA
ACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGT
CGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTC
AGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACA
GCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGAC
GCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGC
TTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGA

GATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGA
GGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCA
CGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCC
GCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTC
TGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTT
CACATTC

pUltraBR EcK

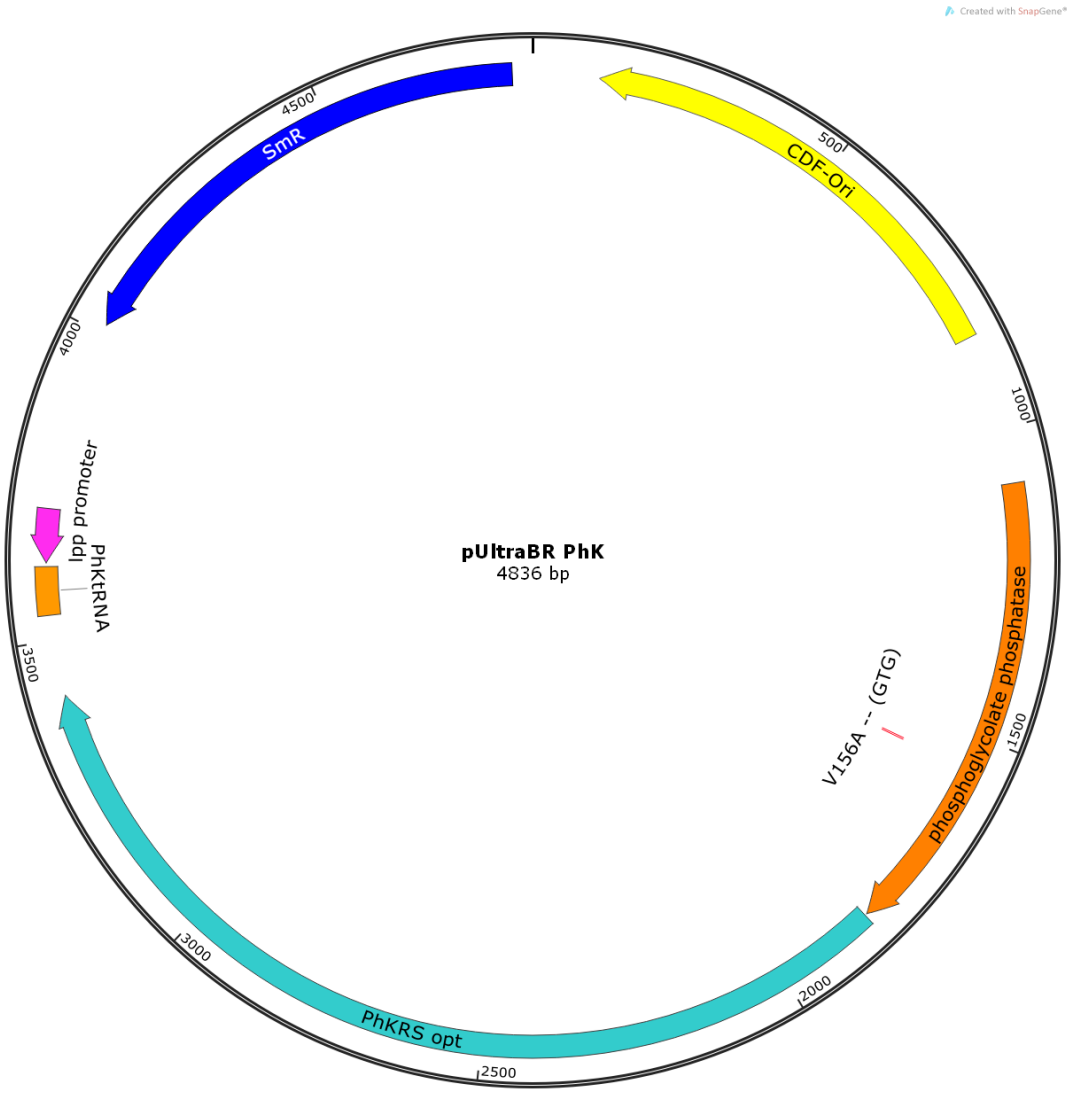


ATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGACTG
CGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAA
GCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGT
TTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTT
CCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACG
GGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCT
CCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCG

CCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGT
ATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAA
CTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAG
TCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTA
ACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGA
AGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTG
CTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTAAC
CTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGA
TTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACT
GAACCGCTCTAGACGTCCATTCCGACAGCATCGTCACTATGGCGTGCTGCT
AGCATCGACGAGTCTGGCTTTGACATTCGACTAGAAGTGGACGGTGGCGT
GAAGGTGAACAACATTGGCGAAATCGCTGCGGCGGGCGCGGATATGTTCG
TCGCCGGTTCGGCAATCTTCGACCAGCCAGACTACAAAAAAGTCATTGAT
GAAATGCGCAGTGAACTGGCAAAGGTAAGTCATGAATAAGTTTGAAGATA
TTCGCGGCGTCGCTTTTGATCTTGATGGTACGCTGGTCGACAGTGCTCCTG
GTCTTGCTGCTGCGGTAGATATGGCGCTGTATGCGCTGGAGTTGCCCGTCG
CAGGTGAAGAACGCGTTATTACCTGGATTGGTAACGGCGCAGATGTTCTG
ATGGAGCGCGCATTGACCTGGGCGCGTCAGGAACGTGCGACTCAGCGTAA
AACAATGGGTAAACCGCCCGTTGATGACGACATTCCGGCAGAAGAACAG
GTACGTATTCTGCGTAAACTGTTCGATCGCTACTATGGCGAGGTTGCCGAA
GAGGGGACGTTTTTGTTCCCGCACGTTGCCGATACGTTGGGCGCGTTGCA
GGCTAAAGGCCTGCCGCTAGGCCTGGTCACCAACAAACCGACGCCGTTCG
TCGCGCCGCTGCTCGAAGCCTTAGATATCGCCAAATACTTCAGCGCGGTG
ATTGGTGGTGATGATGTGCAAAACAAAAAACCGCATCCGGACCCGCTGTT
ACTGGTGGCTGAGCGGATGGGAATTGCCCCACAACAGATGCTGTTTGTCG
GCGACTCACGCAATGATATTCAGGCGGCAAAAGCGGCAGGTTGCCCATCA
GTTGGCTTAACCTACGGATATAACTACGGCGAGGCTATCGATCTCAGCCA
GCCTGATGTAATTTATCAGTCTATAAATGACCTTCTGCCCGCATTAGGGCT
TCCGCATAGCGAAAATCAGGAATCGACATATGTCTGAACAACACGCACAG
GGCGCTGACGCGGTAGTCGATCTTAACAATGAACTGAAAACGCGTCGTGA
GAAGCTGGCGAACCTGCGCGAGCAGGGGATTGCCTTCCCGAACGATTTCC
GTCGCGATCATACCTCTGACCAATTGCACGCAGAATTCGACGGCAAAGAG
AACGAAGAACTGGAAGCGCTGAACATCGAAGTCGCCGTTGCTGGCCGCAT
GATGACCCGTCGTATTATGGGTAAAGCGTCTTTCGTTACCCTGCAGGACGT
TGGCGGTCGCATTCAGCTGTACGTTGCCCGTGACGATCTCCCGGAAGGCG
TTTATAACGAGCAGTTCAAAAAATGGGACCTCGGCGACATCCTCGGCGCG
AAAGGTAAGCTGTTCAAAACCAAAACCGGCGAACTGTCTATCCACTGCAC
CGAGTTGCGTCTGCTGACCAAAGCACTGCGTCCGCTGCCGGATAAATTCC
ACGGCTTGCAGGATCAGGAAGCGCGCTATCGTCAGCGTTATCTCGATCTC
ATCTCCAACGATGAATCCCGCAACACCTTTAAAGTGCGCTCGCAGATCCT
CTCTGGTATTCGCCAGTTCATGGTGAACCGCGGCTTTATGGAAGTTGAAAC

GCCGATGATGCAGGTGATCCCTGGCGGTGCCGCTGCGCGTCCGTTTATCA
CCCACCATAACGCGCTGGATCTCGACATGTACCTGCGTATCGCGCCGGAA
CTGTACCTCAAGCGTCTGGTGGTTGGTGGCTTCGAGCGTGTATTCGAAATC
AACCGTAACTTCCGTAACGAAGGTATTTCCGTACGTCATAACCCAGAGTT
CACCATGATGGAACTCTACATGGCTTACGCAGATTACAAAGATCTGATCG
AGCTGACCGAATCGCTGTTCCGTACTCTGGCACAGGATATTCTCGGTAAG
ACGGAAGTGACCTACGGCGACGTGACGCTGGACTTCGGTAAACCGTTCGA
AAAACTGACCATGCGTGAAGCGATCAAGAAATATCGCCCGGAAACCGAC
ATGGCGGATCTGGACAACTTCGACTCTGCGAAAGCAATTGCTGAATCTAT
CGGCATCCACGTTGAGAAGAGCTGGGGTCTGGGCCGTATCGTTACCGAGA
TCTTCGAAGAAGTGGCAGAAGCACATCTGATTCAGCCGACCTTCATTACT
GAATATCCGGCAGAAGTTTCTCCGCTGGCGCGTCGTAACGACGTTAACCC
GGAAATCACAGACCGCTTTGAGTTCTTCATTGGTGGTCGTGAAATCGGTA
ACGGCTTTAGCGAGCTGAATGACGCGGAAGATCAGGCGCAACGCTTCCTG
GATCAGGTTGCCGCGAAAGACGCAGGTGACGACGAAGCGATGTTCTACG
ATGAAGATTACGTCACCGCACTGGAACATGGCTTACCGCCGACAGCAGGT
CTGGGAATTGGTATCGACCGTATGGTAATGCTGTTCACCAACAGCCATAC
CATCCGCGACGTTATTCTGTTCCCGGCGATGCGTCCGGTAAAATAAGAGC
AGCTCAGGGTCGAATTTGCCATGGCGGCCACCAGGTACCACCGGCGCCTC
AGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCG
GTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACC
GACGACCGGGTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAA
TTGTTAGACATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGT
GGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTT
GTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCC
GGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTT
GCCGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTC
GCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCA
TTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTC
CGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAA
GATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAA
TGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAAC
GCCACGGAATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGG
AGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGAT
CAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCA
AATCAATATCACTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTAC
AAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAAC
TACCTCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACT
CTTCCTTTTTCAATATTATTGAAGCATTT

pUltraBR PhK full length

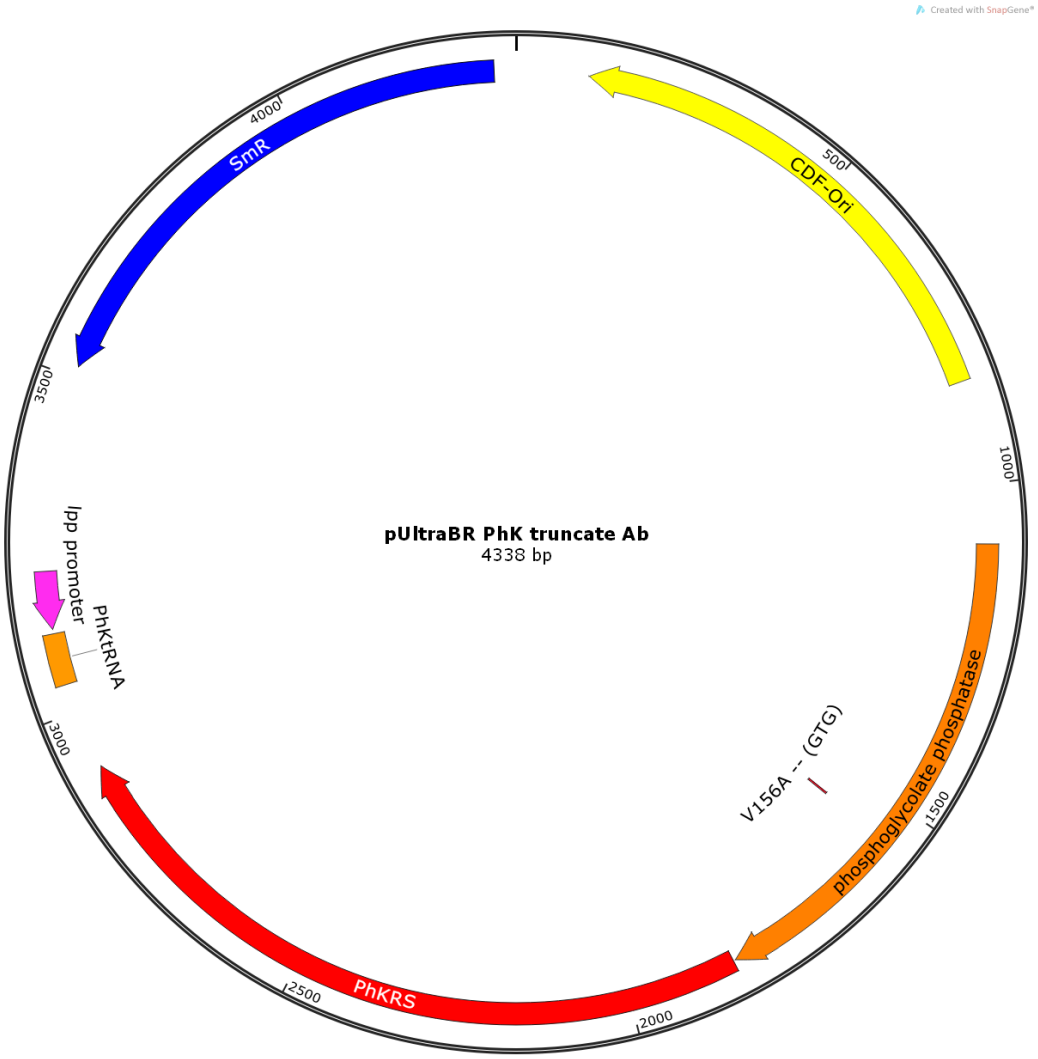


ATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGACTG
CGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAA
GCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGT
TTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTT
CCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACG
GGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCT
CCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCG
CCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGT
ATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAA
CTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAG
TCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTA
ACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGA
AGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTG
CTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTAAC

CTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGA
TTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACT
GAACCGCTCTAGACGTCCATTCCGACAGCATCGTCACTATGGCGTGCTGCT
AGCATCGACGAGTCTGGCTTTGACATTCGACTAGAAGTGGACGGTGGCGT
GAAGGTGAACAACATTGGCGAAATCGCTGCGGCGGGCGCGGATATGTTCG
TCGCCGGTTCGGCAATCTTCGACCAGCCAGACTACAAAAAAGTCATTGAT
GAAATGCGCAGTGAACTGGCAAAGGTAAGTCATGAATAAGTTTGAAGATA
TTCGCGGCGTCGCTTTTGATCTTGATGGTACGCTGGTCGACAGTGCTCCTG
GTCTTGCTGCTGCGGTAGATATGGCGCTGTATGCGCTGGAGTTGCCCGTCG
CAGGTGAAGAACGCGTTATTACCTGGATTGGTAACGGCGCAGATGTTCTG
ATGGAGCGCGCATTGACCTGGGCGCGTCAGGAACGTGCGACTCAGCGTAA
AACAATGGGTAAACCGCCCGTTGATGACGACATTCCGGCAGAAGAACAG
GTACGTATTCTGCGTAAACTGTTCGATCGCTACTATGGCGAGGTTGCCGAA
GAGGGGACGTTTTTGTTCCCGCACGTTGCCGATACGTTGGGCGCGTTGCA
GGCTAAAGGCCTGCCGCTAGGCCTGGTCACCAACAAACCGACGCCGTTCG
TCGCGCCGCTGCTCGAAGCCTTAGATATCGCCAAATACTTCAGCGCGGTG
ATTGGTGGTGATGATGTGCAAAACAAAAAACCGCATCCGGACCCGCTGTT
ACTGGTGGCTGAGCGGATGGGAATTGCCCCACAACAGATGCTGTTTGTCG
GCGACTCACGCAATGATATTCAGGCGGCAAAAGCGGCAGGTTGCCCATCA
GTTGGCTTAACCTACGGATATAACTACGGCGAGGCTATCGATCTCAGCCA
GCCTGATGTAATTTATCAGTCTATAAATGACCTTCTGCCCGCATTAGGGCT
TCCGCATAGCGAAAATCAGGAATCGACATATGGTACATTGGGCTGATTAT
ATTGCGGATAAGATCATTCGTGAACGCGGAGAGAAAGAGAAATACGTTGT
AGAAAGTGGTATCACTCCGTCTGGTTATGTACATGTGGGCAATTTCCGTGA
GTTATTCACAGCATACATTGTCGGACACGCGTTACGCGACAAAGGATATG
AAGTTCGCCACATTCACATGTGGGATGACTATGATCGCTTTCGCAAAGTA
CCGCGTAATGTCCCGCAAGAGTGGAAAGACTACTTAGGCATGCCCATCTC
TGAGGTGCCTGACCCTTGGGGGTGCCACGAGTCTTATGCAGAGCATTTTAT
GCGTAAATTCGAGGAAGAGGTAGAAAAACTGGGCATTGAGGTCGACTTCT
TATATGCTTCCGAATTGTACAAGCGCGGTGAATACTCGGAGGAGATCCGC
CTTGCATTTGAAAAGCGTGACAAGATTATGGAGATCCTGAATAAATATCG
TGAAATCGCCAAACAACCGCCATTGCCAGAGAACTGGTGGCCAGCCATGG
TCTACTGTCCGGAACACCGCCGCGAAGCGGAAATTATTGAATGGGACGGT
GGATGGAAAGTCAAATACAAGTGTCCTGAAGGCCATGAAGGCTGGGTTG
ATATTCGTTCGGGCAACGTAAAACTGCGTTGGCGTGTTGATTGGCCGATG
CGTTGGTCACATTTCGGCGTAGATTTTGAGCCTGCGGGCAAGGACCACCTT
GTAGCTGGATCTTCATATGATACTGGAAAGGAAATCATTAAGGAGGTATA
CGGGAAGGAAGCTCCACTGAGCCTTATGTACGAATTTGTAGGAATCAAAG
GACAAAAGGGTAAGATGTCAGGGTCTAAAGGTAATGTTATTCTTTTGTCG
GATCTGTATGAAGTACTGGAGCCAGGGCTTGTCCGCTTTATCTATGCGCGT
CACCGCCCGAACAAGGAGATTAAAATTGATTTGGGTTTGGGTATCTTGAA

TTTGTACGACGAGTTCGATAAAGTAGAGCGCATCTACTTTGGAGTTGAGG
GAGGTAAGGGAGACGACGAGGAACTGCGTCGTACATATGAATTATCCATG
CCAAAGAAACCAGAACGCCTTGTAGCACAAGCGCCTTTTCGTTTCTTAGC
AGTACTGGTGCAGCTGCCCCACTTGACTGAAGAGGATATCATCAACGTAT
TGATCAAGCAAGGGCATATTCCACGTGATCTTAGTAAGGAGGACGTAGAG
CGTGTTAAACTTCGCATTAACTTGGCTCGCAACTGGGTCAAGAAATATGC
ACCAGAAGATGTTAAGTTCAGCATTTTAGAAAAACCTCCCGAAGTTGAGG
TCTCCGAAGACGTACGTGAAGCAATGAACGAAGTAGCCGAGTGGTTGGA
AAACCACGAAGAGTTTTCCGTAGAAGAGTTCAACAATATTCTGTTTGAGG
TCGCGAAGCGTCGCGGAATTTCCAGTCGTGAGTGGTTTAGCACGCTTTACC
GTCTTTTTATCGGGAAAGAACGCGGACCTCGTCTGGCATCGTTTCTTGCGT
CTCTTGATCGTTCGTTCGTAATTAAACGCCTGCGCCTGGAAGGCTAAAGAT
CTGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATG
AACAGGCAGACATCTGTGAATCGCTTCACGACCACGCATCAAAAAAAATC
CTTAGCTTTCGCTAAGGATTTTATTTTGGCTAGTCCGGCGGGATTTGAACC
CGCGACCTGCGGGTTAAAAGCCCGCCGCTCTACCAGGCTGAGCTACGGAC
CAAGGAAAGTTACAAGTATTACACAAAGTTTTTTATGTTGAGAATATTTTT
TTGATGGGGCGCCACTTATTTTTGATCGTTCGCTCAAAGAAGCCTCGAGTC
ACTTTCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAG
GCTGGATGGCCTTCCCCATTATGGCATGCTCGAGCAGCTCAGGGTCGAAT
TTGCCATGGCGGCCACCAGGTACCACCGGCGCCTCAGGCATTTGAGAAGC
ACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAG
ACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATC
GTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTT
GCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAA
CTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTG
TCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTG
CCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGT
ACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAG
TCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAG
ATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCA
AGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGT
CGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCAT
TCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTC
GTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTC
CAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTT
GTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGT
GGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAA
CGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAG
TCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTA
TTGAAGCATTT

pUltraBR PhK truncated

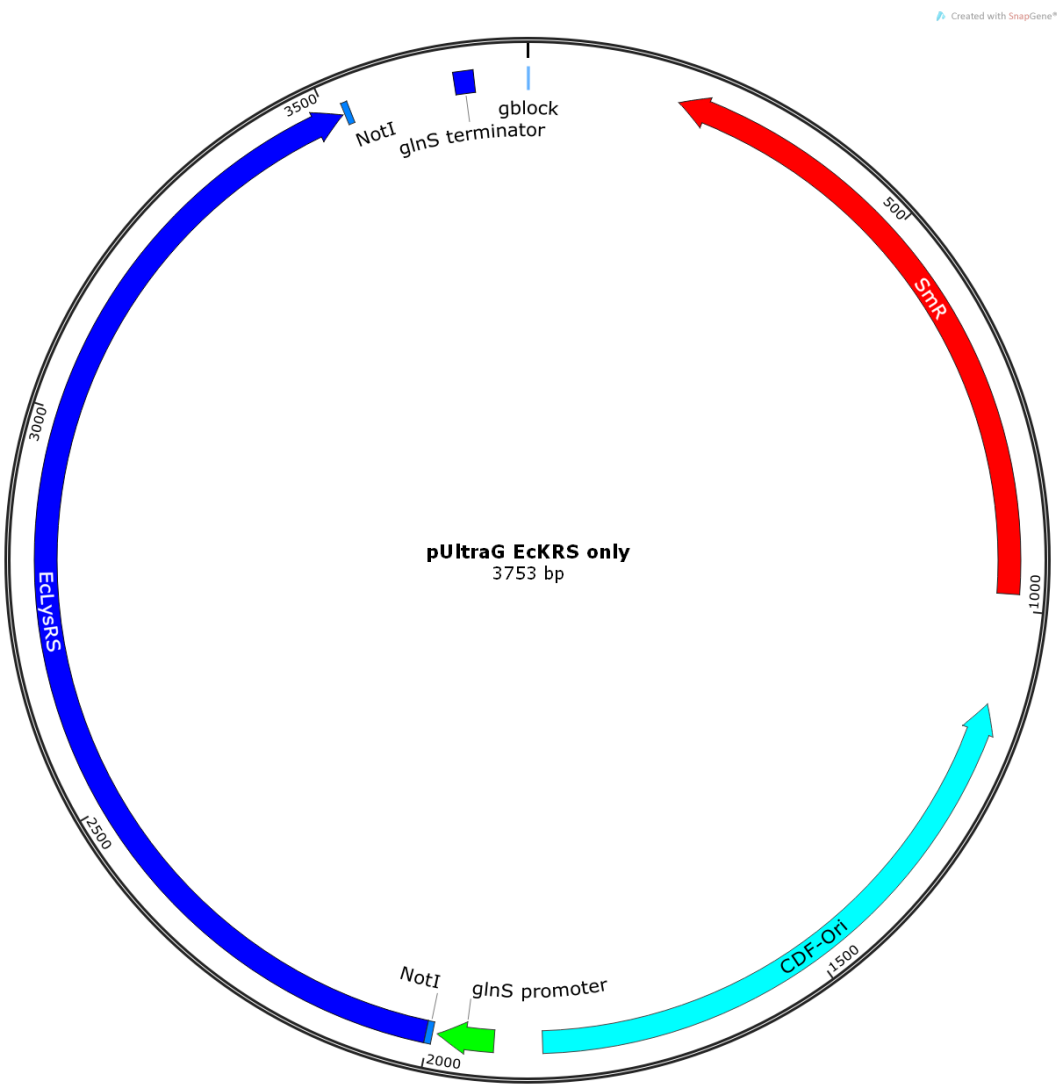


ATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGACTG
CGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAA
GCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGT
TTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTT
CCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACG
GGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCT
CCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCG
CCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGT
ATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAA
CTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAG
TCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTA
ACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGA
AGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTG

CTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTAAC
CTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAAGA
TTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACT
GAACCGCTCTAGACGTCCATTCCGACAGCATCGTCACTATGGCGTGCTGCT
AGCATCGACGAGTCTGGCTTTGACATTCGACTAGAAGTGGACGGTGGCGT
GAAGGTGAACAACATTGGCGAAATCGCTGCGGCGGGCGCGGATATGTTCG
TCGCCGGTTCGGCAATCTTCGACCAGCCAGACTACAAAAAAGTCATTGAT
GAAATGCGCAGTGAACTGGCAAAGGTAAGTCATGAATAAGTTTGAAGATA
TTCGCGGCGTCGCTTTTGATCTTGATGGTACGCTGGTCGACAGTGCTCCTG
GTCTTGCTGCTGCGGTAGATATGGCGCTGTATGCGCTGGAGTTGCCCGTCG
CAGGTGAAGAACGCGTTATTACCTGGATTGGTAACGGCGCAGATGTTCTG
ATGGAGCGCGCATTGACCTGGGCGCGTCAGGAACGTGCGACTCAGCGTAA
AACAATGGGTAAACCGCCCGTTGATGACGACATTCCGGCAGAAGAACAG
GTACGTATTCTGCGTAAACTGTTCGATCGCTACTATGGCGAGGTTGCCGAA
GAGGGGACGTTTTTGTTCCCGCACGTTGCCGATACGTTGGGCGCGTTGCA
GGCTAAAGGCCTGCCGCTAGGCCTGGTCACCAACAAACCGACGCCGTTCG
TCGCGCCGCTGCTCGAAGCCTTAGATATCGCCAAATACTTCAGCGCGGTG
ATTGGTGGTGATGATGTGCAAAACAAAAAACCGCATCCGGACCCGCTGTT
ACTGGTGGCTGAGCGGATGGGAATTGCCCCACAACAGATGCTGTTTGTCG
GCGACTCACGCAATGATATTCAGGCGGCAAAAGCGGCAGGTTGCCCATCA
GTTGGCTTAACCTACGGATATAACTACGGCGAGGCTATCGATCTCAGCCA
GCCTGATGTAATTTATCAGTCTATAAATGACCTTCTGCCCGCATTAGGGCT
TCCGCATAGCGAAAATCAGGAATCGACATATGGTTCATTGGGCCGATTAT
ATTGCTGATAAAATAATTAGAGAGAGGGGGGAGAAGGAGAAGTACGTTG
TTGAGAGTGGAATAACGCCAAGTGGTTACGTTCACGTTGGGAACTTTAGG
GAGCTTTTTACAGCTTATATTGTGGGCCATGCCCTAAGGGATAAGGGGTAT
GAGGTTAGGCACATCCACATGTGGGATGATTATGATAGATTTAGGAAGGT
TCCAAGGAACGTTCCCCAGGAATGGAAAGATTACCTGGGAATGCCCATTA
GTGAAGTTCCTGATCCCTGGGGATGCCATGAGAGTTATGCTGAACACTTC
ATGAGAAAGTTCGAGGAGGAGGTAGAAAAATTAGGGATCGAAGTTGACT
TTCTTTATGCGAGTGAACTCTACAAGAGAGGGGAATATTCTGAGGAGATA
AGGTTAGCCTTTGAGAAAAGGGATAAGATAATGGAGATACTAAACAAGT
ATAGGGAAATTGCGAAACAACCTCCCCTTCCAGAGAACTGGTGGCCCGCA
ATGGTTTACTGCCCTGAGCATAGGAGGGAAGCAGAGATCATTGAATGGGA
TGGGGGCTGGAAGGTTAAGTATAAGTGCCCCGAAGGTCACGAGGGATGG
GTTGATATAAGGAGTGGGAACGTGAAACTGAGGTGGCGTGTTGATTGGCC
CATGCGTTGGTCTCACTTTGGCGTTGACTTCGAACCTGCTGGAAAGGATCA
TCTTGTGGCTGGTTCAAGCTACGATACGGGAAAGGAGATTATAAAGGAAG
TTTATGGAAAGGAAGCTCCGTTATCTTTAATGTATGAGTTTGTTGGAATTA
AGGGGCAGAAGGGGAAGATGAGTGGTAGTAAGGGAAATGTTATTTTACTC
AGCGATCTGTATGAGGTTCTTGAGCCAGGTCTCGTTAGATTTATCTACGCT

CGGCATAGGCCAAACAAGGAGATAAAGATAGATCTAGGTCTTGGCATTCT
AAACCTCTACGATGAGTTCGATAAAGTTGAGAGAATATACTTCGGGGTTG
AGGGTGGTAAAGGTGATGATGAAGAATTAAGGAGGACTTACGAGCTTTCA
TAAAGATCTGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGAC
GCGGATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCATCAAA
AAAAATCCTTAGCTTTCGCTAAGGATTTTATTTTGGCTAGTCCGGCGGGAT
TTGAACCCGCGACCTGCGGGTTAAAAGCCCGCCGCTCTACCAGGCTGAGC
TACGGACCAAGGAAAGTTACAAGTATTACACAAAGTTTTTTATGTTGAGA
ATATTTTTTTGATGGGGCGCCACTTATTTTGATCGTTCGCTCAAAGAAGC
CTCGAGTCACTTTCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCG
ACGCGAGGCTGGATGGCCTTCCCCATTATGGCATGCTCGAGCAGCTCAGG
GTCGAATTTGCCATGGCGGCCACCAGGTACCACCGGCGCCTCAGGCATTT
GAGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCA
GCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCG
GGTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGA
CATTATTTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAAT
TCTTCCAACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGA
TAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCG
CTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTAC
TGCGCTGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGC
CAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCC
TCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGG
ACCTACCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAG
ATCAATGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGC
GCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGA
ATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTC
GCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTC
GCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATAT
CACTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACG
GCCAGCAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGA
TAGTTGAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTT
CAATATTATTGAAGCATTT

pUltraG EcKRS

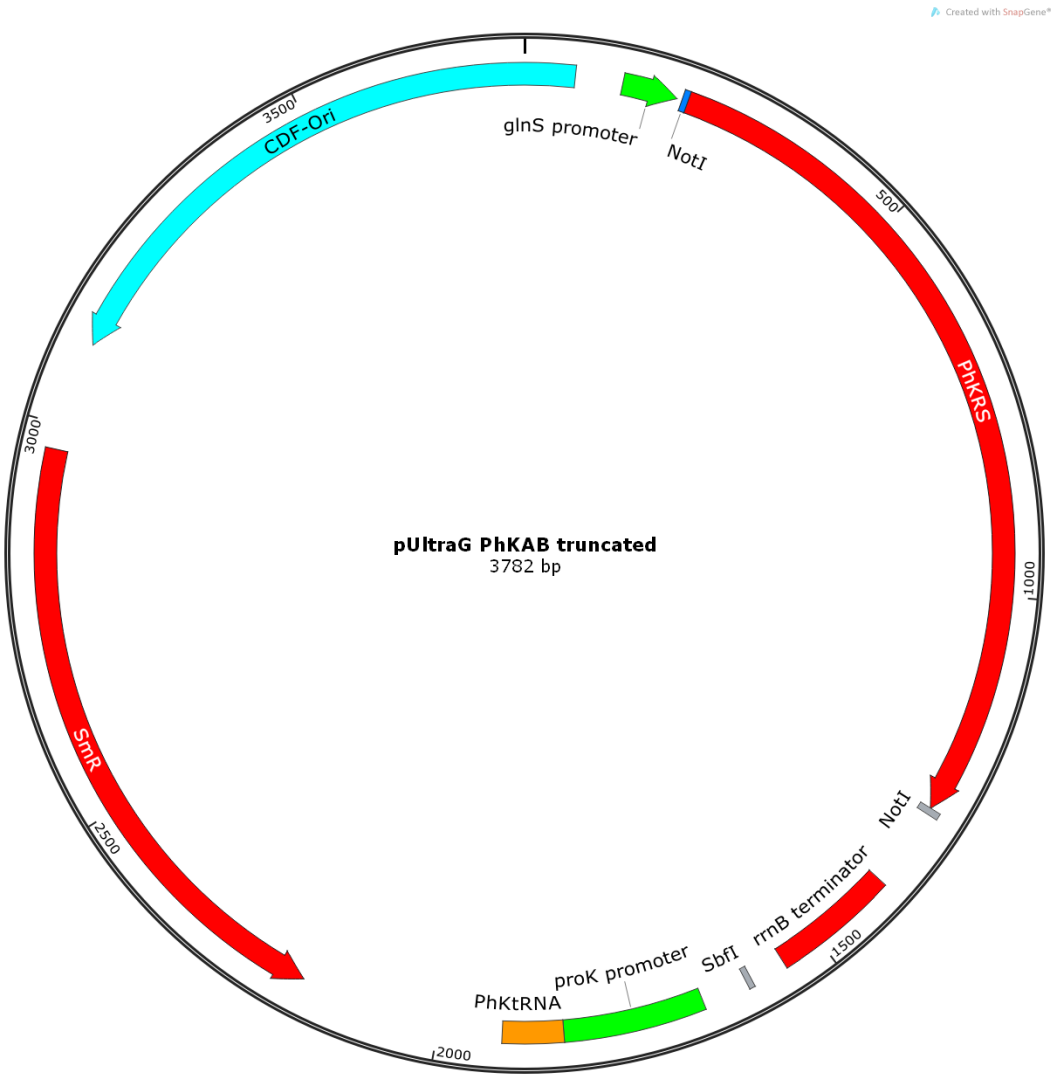


GGCGGCCACCAGGTACCACCGGCGCCTCAGGCATTTGAGAAGCACACGGT
CACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAG
CGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATCGTGGCCG
GATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTTGCCGACT
ACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACTGATCT
GCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCT
TCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCCCAGTC
GGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGTACCAAAT
GCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCGGGCG
GCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTT
CAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACG
CTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCGATCGTG
GCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAA
TTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGTGCA
CAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAA

GCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCA
AGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAG
GCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTT
CGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACT
TCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAGCA
TTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGA
AAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGA
CTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGA
TAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGG
CGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCT
TTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGT
ACGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTC
GCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTT
CCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACT
GGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAA
GAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTT
GAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTG
GTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTT
GAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTG
TGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTA
ACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAA
GATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
CTGAACCGCTCTAGAGTCATCAATCATCCCCATAATCCTTGTTAGATTATC
AATTTTAAAAAACTAACAGTTGTCAGCCTGTCCCGCTTTAATATCATACGC
CGTTATACGTTGTTTACGCTTTAAGGAGGCGGCCGCATGTCTGAACAACA
CGCACAGGGCGCTGACGCGGTAGTCGATCTTAACAATGAACTGAAAACGC
GTCGTGAGAAGCTGGCGAACCTGCGCGAGCAGGGGATTGCCTTCCCGAAC
GATTTCCGTCGCGATCATACCTCTGACCAATTGCACGCAGAATTCGACGG
CAAAGAGAACGAAGAACTGGAAGCGCTGAACATCGAAGTCGCCGTTGCT
GGCCGCATGATGACCCGTCGTATTATGGGTAAAGCGTCTTTCGTTACCCTG
CAGGACGTTGGCGGTCGCATTCAGCTGTACGTTGCCCGTGACGATCTCCC
GGAAGGCGTTTATAACGAGCAGTTCAAAAAATGGGACCTCGGCGACATCC
TCGGCGCGAAAGGTAAGCTGTTCAAAACCAAAACCGGCGAACTGTCTATC
CACTGCACCGAGTTGCGTCTGCTGACCAAAGCACTGCGTCCGCTGCCGGA
TAAATTCCACGGCTTGCAGGATCAGGAAGCGCGCTATCGTCAGCGTTATC
TCGATCTCATCTCCAACGATGAATCCCGCAACACCTTTAAAGTGCGCTCGC
AGATCCTCTCTGGTATTCGCCAGTTCATGGTGAACCGCGGCTTTATGGAAG
TTGAAACGCCGATGATGCAGGTGATCCCTGGCGGTGCCGCTGCGCGTCCG
TTTATCACCCACCATAACGCGCTGGATCTCGACATGTACCTGCGTATCGCG
CCGGAACTGTACCTCAAGCGTCTGGTGGTTGGTGGCTTCGAGCGTGTATTC
GAAATCAACCGTAACTTCCGTAACGAAGGTATTTCCGTACGTCATAACCC

AGAGTTCACCATGATGGAACTCTACATGGCTTACGCAGATTACAAAGATC
TGATCGAGCTGACCGAATCGCTGTTCCGTACTCTGGCACAGGATATTCTCG
GTAAGACGGAAGTGACCTACGGCGACGTGACGCTGGACTTCGGTAAACCG
TTCGAAAAACTGACCATGCGTGAAGCGATCAAGAAATATCGCCCGGAAAC
CGACATGGCGGATCTGGACAACTTCGACTCTGCGAAAGCAATTGCTGAAT
CTATCGGCATCCACGTTGAGAAGAGCTGGGGTCTGGGCCGTATCGTTACC
GAGATCTTCGAAGAAGTGGCAGAAGCACATCTGATTCAGCCGACCTTCAT
TACTGAATATCCGGCAGAAGTTTCTCCGCTGGCGCGTCGTAACGACGTTA
ACCCGGAAATCACAGACCGCTTTGAGTTCTTCATTGGTGGTCGTGAAATC
GGTAACGGCTTTAGCGAGCTGAATGACGCGGAAGATCAGGCGCAACGCTT
CCTGGATCAGGTTGCCGCGAAAGACGCAGGTGACGACGAAGCGATGTTCT
ACGATGAAGATTACGTCACCGCACTGGAACATGGCTTACCGCCGACAGCA
GGTCTGGGAATTGGTATCGACCGTATGGTAATGCTGTTCACCAACAGCCA
TACCATCCGCGACGTTATTCTGTTCCCGGCGATGCGTCCGGTAAAATAAGC
GGCCGCTTTCAAACGCTAAATTGCCTGATGCGCTACGCTTATCAGGCCTAC
ATGATCTCTGCAATATATTGAGTTTGCGTGCTTTTGTAGGCCGGATAAGGC
GTTCACGCCGCATCCGGCAAGAAACAGCAAACAATCCAAAACGCCGCGTT
CAGCGGCGTTTTTTCTGCTTTTCTTCGCGAATTAATTCCGCTTCGCAACATG
TGAGCACCGGTTTATTGACTACCCAT

pUltraG PhKRS truncated

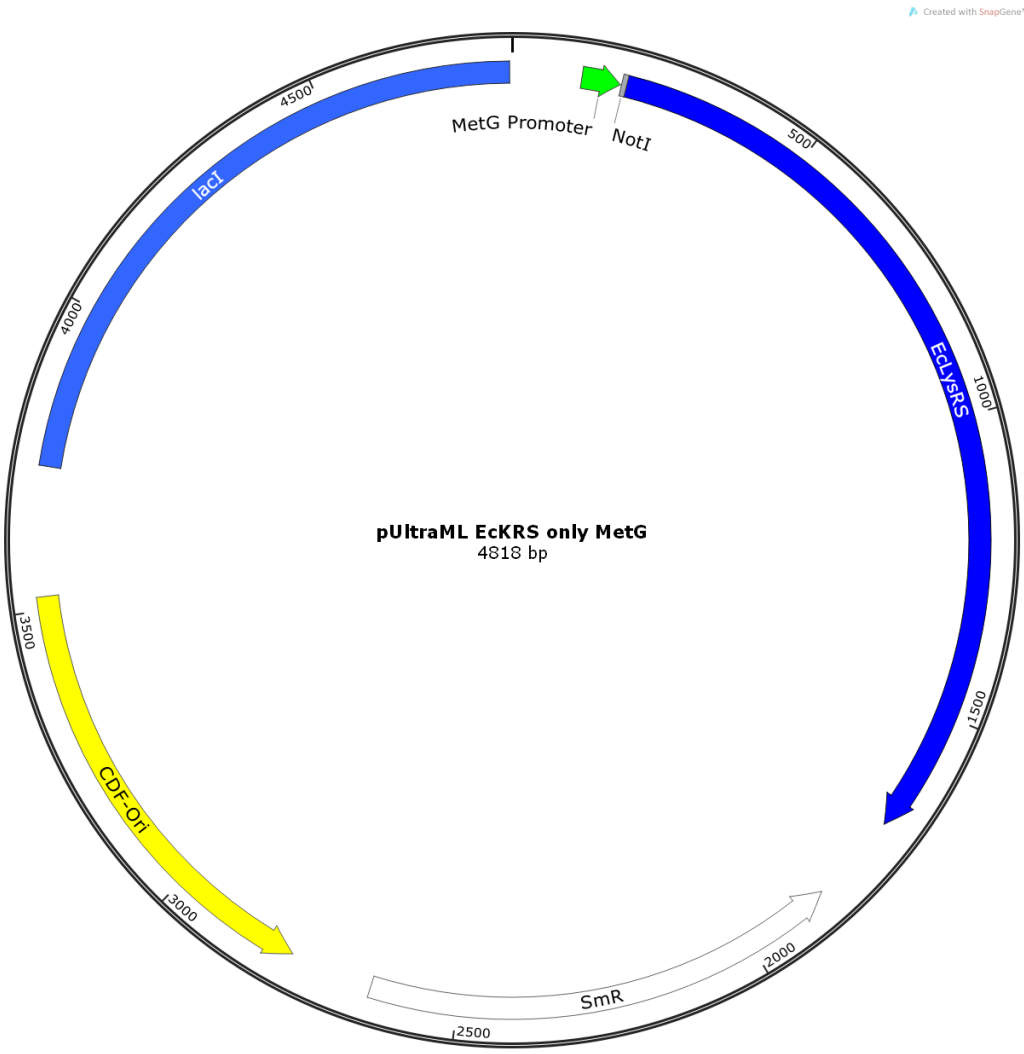


TTTTTCGTTTACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACTGAACCGCTCTAGAGTCATCAATCATCCC
CATAATCCTTGTTAGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCTG
TCCCGCTTTAATATCATACGCCGTTATACGTTGTTTACGCTTTAAGGAGGC
GGCCGCATGGTTCATTGGGCCGATTATATTGCTGATAAAATAATTAGAGA
GAGGGGGGAGAAGGAGAAGTACGTTGTTGAGAGTGGAATAACGCCAAGT
GGTTACGTTCACGTTGGGAACTTTAGGGAGCTTTTTACAGCTTATATTGTG
GGCCATGCCCTAAGGGATAAGGGGTATGAGGTTAGGCACATCCACATGTG
GGATGATTATGATAGATTTAGGAAGGTTCCAAGGAACGTTCCCCAGGAAT
GGAAAGATTACCTGGGAATGCCCATTAGTGAAGTTCCTGATCCCTGGGGA
TGCCATGAGAGTTATGCTGAACACTTCATGAGAAAGTTCGAGGAGGAGGT
AGAAAAATTAGGGATCGAAGTTGACTTTCTTTATGCGAGTGAACTCTACA
AGAGAGGGGAATATTCTGAGGAGATAAGGTTAGCCTTTGAGAAAAGGGA
TAAGATAATGGAGATACTAAACAAGTATAGGGAAATTGCGAAACAACCT
CCCCTTCCAGAGAACTGGTGGCCCGCAATGGTTTACTGCCCTGAGCATAG

GAGGGAAGCAGAGATCATTGAATGGGATGGGGGCTGGAAGGTTAAGTAT
AAGTGCCCCGAAGGTCACGAGGGATGGGTTGATATAAGGAGTGGGAACG
TGAAACTGAGGTGGCGTGTTGATTGGCCCATGCGTTGGTCTCACTTTGGCG
TTGACTTCGAACCTGCTGGAAAGGATCATCTTGTGGCTGGTTCAAGCTACG
ATACGGGAAAGGAGATTATAAAGGAAGTTTATGGAAAGGAAGCTCCGTT
ATCTTTAATGTATGAGTTTGTTGGAATTAAGGGGCAGAAGGGGAAGATGA
GTGGTAGTAAGGGAAATGTTATTTTACTCAGCGATCTGTATGAGGTTCTTG
AGCCAGGTCTCGTTAGATTTATCTACGCTCGGCATAGGCCAAACAAGGAG
ATAAAGATAGATCTAGGTCTTGGCATTCTAAACCTCTACGATGAGTTCGAT
AAAGTTGAGAGAATATACTTCGGGGTTGAGGGTGGTAAAGGTGATGATGA
AGAATTAAGGAGGACTTACGAGCTTTCATAAGCGGCCGCGTTTAAACGGT
CTCCAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACA
GATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCG
GCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAA
CGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAA
CTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTG
TTTGTGAGCTCCCGGTCATCAATCATCCCCATAATCCTTGTTAGCCTGCAG
GTAATTCCGCTTCGCAACATGTGAGCACCGGTTTATTGACTACCGGAAGC
AGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTC
CCCGTGGAGGTAATAATTGACGATATGATCAGTGCACGGCTAACTAAGCG
GCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGA
CGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTTGGTCCGTAGCTC
AGCCTGGTAGAGCGGCGGGCTTTTAACCCGCAGGTCGCGGGTTCAAATCC
CGCCGGACTAGCCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTG
CATGCTCGAGCAGCTCAGGGTCGAATTTGCCATGGCGGCCACCAGGTACC
ACCGGCGCCTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAG
TCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCC
TGCCCTGAACCGACGACCGGGTCATCGTGGCCGGATCTTGCGGCCCCTCG
GCTTGAACGAATTGTTAGACATTATTTGCCGACTACCTTGGTGATCTCGCC
TTTCACGTAGTGGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAAGC
GATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCT
GATACTGGGCCGGCAGGCGCTCCATTGCCCAGTCGGCAGCGACATCCTTC
GGCGCGATTTTGCCGGTTACTGCGCTGTACCAAATGCGGGACAACGTAAG
CACTACATTTCGCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCG
TTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAA
AGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACGCTATGTTCTCTTGCTT
TTGTCAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGAAGATA
CCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTA
GCTGGATAACGCCACGGAATGATGTCGTCGTGCACAACAATGGTGACTTC
TACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAAGCCGAAGTTTCCAAAA
GGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCG

TAACCAGCAAATCAATATCACTGTGTGGCTTCAGGCCGCCATCCACTGCG
GAGCCGTACAAATGTACGGCCAGCAACGTCGGTTCGAGATGGCGCTCGAT
GACGCCAACTACCTCTGATAGTTGAGTCGATACTTCGGCGATCACCGCTTC
CCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGC
TAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGACTGCGGCGGGCGCTGC
GGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGGACTAAC
ATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTC
CGCCCTCCTGCCAGAGTTCACATAAACAGACGCTTTTCCGGTGCATCTGTG
GGAGCCGTGAGGCTCAACCATGAATCTGACAGTACGGGCGAAACCCGAC
AGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCT
CCTGTTCCGACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTAC
GGGAAGTGTGGCGCTTTCTCATAGCTCACACACTGGTATCTCGGCTCGGTG
TAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTCAGCC
CGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAA
AGCACGGTAAAACGCCACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCA
GAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGAAGTGTGCGCCAAAG
TCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCA
GTTACCACGGTTAAGCAGTTCCCCAACTGACTTAACCTTCGATCAAACCAC
CTCCCCAGGTGGT

pUltraML EcKRS

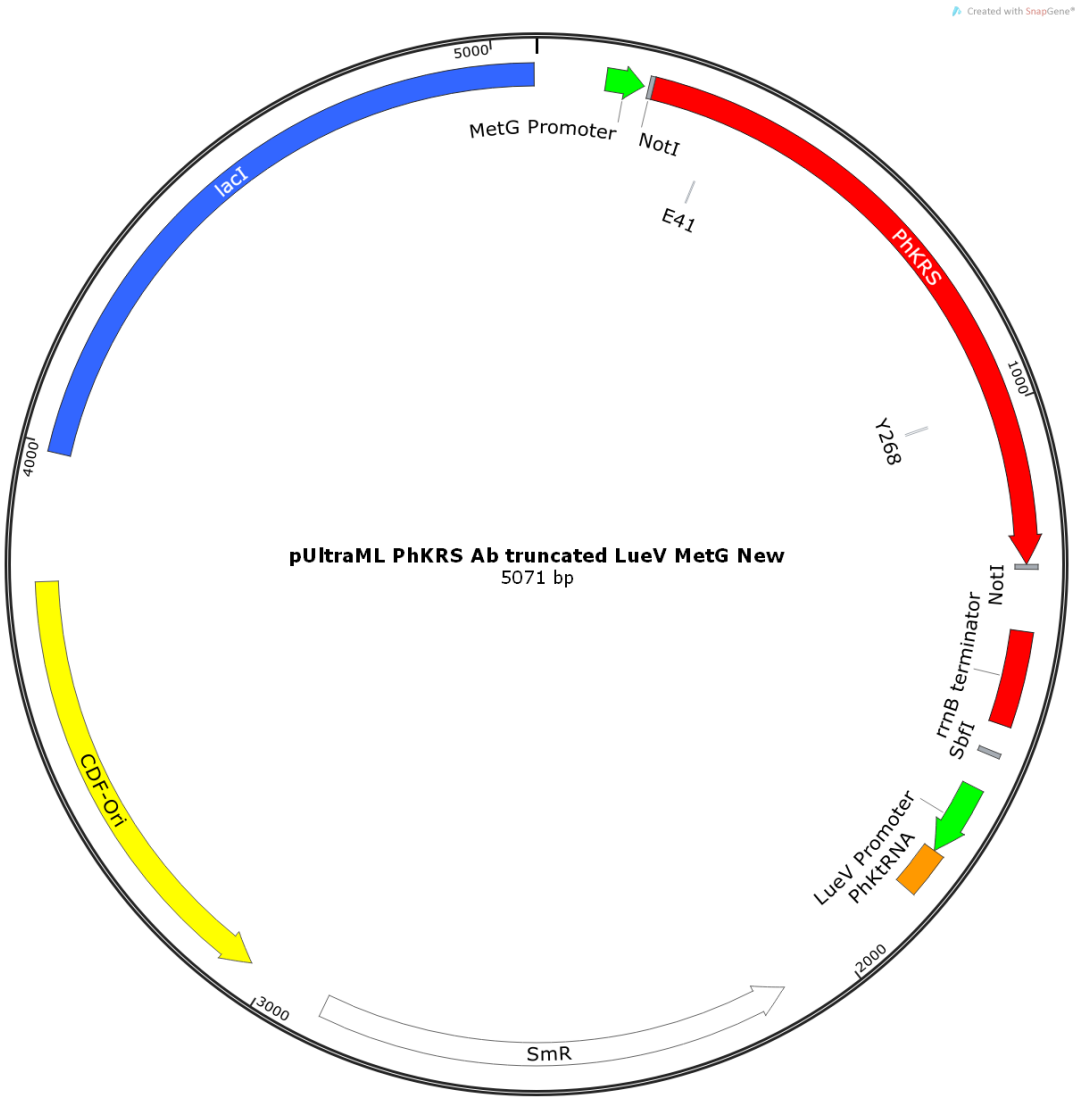


ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCG
ACTCCTGCATTAGGTACTTAACATTTTCCCATTTGGTACTATCTAACCCCTT
TTCACTATTAAGAAGTAATGCCTACTGTGCGGCCGCATGTCTGAACAACA
CGCACAGGGCGCTGACGCGGTAGTCGATCTTAACAATGAACTGAAAACGC
GTCGTGAGAAGCTGGCGAACCTGCGCGAGCAGGGGATTGCCTTCCCGAAC
GATTTCCGTCGCGATCATACCTCTGACCAATTGCACGCAGAATTCGACGG
CAAAGAGAACGAAGAACTGGAAGCGCTGAACATCGAAGTCGCCGTTGCT
GGCCGCATGATGACCCGTCGTATTATGGGTAAAGCGTCTTTCGTTACCCTG
CAGGACGTTGGCGGTCGCATTCAGCTGTACGTTGCCCGTGACGATCTCCC
GGAAGGCGTTTATAACGAGCAGTTCAAAAAATGGGACCTCGGCGACATCC
TCGGCGCGAAAGGTAAGCTGTTCAAAACCAAAACCGGCGAACTGTCTATC
CACTGCACCGAGTTGCGTCTGCTGACCAAAGCACTGCGTCCGCTGCCGGA
TAAATTCCACGGCTTGCAGGATCAGGAAGCGCGCTATCGTCAGCGTTATC
TCGATCTCATCTCCAACGATGAATCCCGCAACACCTTTAAAGTGCGCTCGC
AGATCCTCTCTGGTATTCGCCAGTTCATGGTGAACCGCGGCTTTATGGAAG

TTGAAACGCCGATGATGCAGGTGATCCCTGGCGGTGCCGCTGCGCGTCCG
TTTATCACCCACCATAACGCGCTGGATCTCGACATGTACCTGCGTATCGCG
CCGGAACTGTACCTCAAGCGTCTGGTGGTTGGTGGCTTCGAGCGTGTATTC
GAAATCAACCGTAACTTCCGTAACGAAGGTATTTCCGTACGTCATAACCC
AGAGTTCACCATGATGGAACTCTACATGGCTTACGCAGATTACAAAGATC
TGATCGAGCTGACCGAATCGCTGTTCCGTACTCTGGCACAGGATATTCTCG
GTAAGACGGAAGTGACCTACGGCGACGTGACGCTGGACTTCGGTAAACCG
TTCGAAAAACTGACCATGCGTGAAGCGATCAAGAAATATCGCCCGGAAAC
CGACATGGCGGATCTGGACAACTTCGACTCTGCGAAAGCAATTGCTGAAT
CTATCGGCATCCACGTTGAGAAGAGCTGGGGTCTGGGCCGTATCGTTACC
GAGATCTTCGAAGAAGTGGCAGAAGCACATCTGATTCAGCCGACCTTCAT
TACTGAATATCCGGCAGAAGTTTCTCCGCTGGCGCGTCGTAACGACGTTA
ACCCGGAAATCACAGACCGCTTTGAGTTCTTCATTGGTGGTCGTGAAATC
GGTAACGGCTTTAGCGAGCTGAATGACGCGGAAGATCAGGCGCAACGCTT
CCTGGATCAGGTTGCCGCGAAAGACGCAGGTGACGACGAAGCGATGTTCT
ACGATGAAGATTACGTCACCGCACTGGAACATGGCTTACCGCCGACAGCA
GGTCTGGGAATTGGTATCGACCGTATGGTAATGCTGTTCACCAACAGCCA
TACCATCCGCGACGTTATTCTGTTCCCGGCGATGCGTCCGGTAAAATAAAC
ACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGAC
ATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATCGT
GGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTTGC
CGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACT
GATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTC
TAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCC
CAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGTAC
CAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTC
GGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGAT
CCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAG
GCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCG
ATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCT
CCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTC
GTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAG
GGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTT
TCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGG
CTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACG
TCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTC
GATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATT
GAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTA
TTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGC
GTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTC
CGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGC

CGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACA
GACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCT
GACAGTACGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGG
CGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGAT
ACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTC
ACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGT
AAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTG
TTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAG
CCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGT
TGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTG
GTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACT
GACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGG
GCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCT
TTTCTACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAG
CACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCAT
GCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCA
TCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGT
TGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATT
AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCA
GGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCA
CCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCC
AGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGA
GCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCG
CAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGT
TGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATG
GTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATC
GGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGT
GACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGG
GAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAA
CGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCAT
CCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTG
TGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACC
ACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAAT
TTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGC
AACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTC
AGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGG
CTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATA
CTCTGCGACATCGTATAACGTTACTGGTTTCACATTC

pUltraML PhKRS truncated



ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCG
ACTCCTGCATTAGGTACTTAACATTTTCCCATTTGGTACTATCTAACCCCTT
TTCACTATTAAGAAGTAATGCCTACTAGGAGGTGCGGCCGCATGGTTCAT
TGGGCCGATTATATTGCTGATAAAATAATTAGAGAGAGGGGGGAGAAGG
AGAAGTACGTTGTTGAGAGTGGAATAACGCCAAGTGGTTACGTTCACGTT
GGGAACTTTAGGGAGCTTTTTACAGCTTATATTGTGGGCCATGCCCTAAGG
GATAAGGGGTATGAGGTTAGGCACATCCACATGTGGGATGATTATGATAG
ATTTAGGAAGGTTCCAAGGAACGTTCCCCAGGAATGGAAAGATTACCTGG
GAATGCCCATTAGTGAAGTTCCTGATCCCTGGGGATGCCATGAGAGTTAT
GCTGAACACTTCATGAGAAAGTTCGAGGAGGAGGTAGAAAAATTAGGGA
TCGAAGTTGACTTTCTTTATGCGAGTGAACTCTACAAGAGAGGGGAATAT
TCTGAGGAGATAAGGTTAGCCTTTGAGAAAAGGGATAAGATAATGGAGAT
ACTAAACAAGTATAGGGAAATTGCGAAACAACCTCCCCTTCCAGAGAACT
GGTGGCCCGCAATGGTTTACTGCCCTGAGCATAGGAGGGAAGCAGAGATC

ATTGAATGGGATGGGGGCTGGAAGGTTAAGTATAAGTGCCCCGAAGGTCA
CGAGGGATGGGTTGATATAAGGAGTGGGAACGTGAAACTGAGGTGGCGT
GTTGATTGGCCCATGCGTTGGTCTCACTTTGGCGTTGACTTCGAACCTGCT
GGAAAGGATCATCTTGTGGCTGGTTCAAGCTACGATACGGGAAAGGAGAT
TATAAAGGAAGTTTATGGAAAGGAAGCTCCGTTATCTTTAATGTATGAGTT
TGTTGGAATTAAGGGGCAGAAGGGGAAGATGAGTGGTAGTAAGGGAAAT
GTTATTTTACTCAGCGATCTGTATGAGGTTCTTGAGCCAGGTCTCGTTAGA
TTTATCTACGCTCGGCATAGGCCAAACAAGGAGATAAAGATAGATCTAGG
TCTTGGCATTCTAAACCTCTACGATGAGTTCGATAAAGTTGAGAGAATATA
CTTCGGGGTTGAGGGTGGTAAAGGTGATGATGAAGAATTAAGGAGGACTT
ACGAGCTTTCATAAGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGTTTT
GGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAG
AAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCC
CACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGT
AGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAA
AACGAAAGGCTCAGTCGAAAGACTGGGCCTTGTTTGTGAGCTCCCGGTCA
TCAATCATCCCCATAATCCTTGTTAGCCTGCAGGTAATTCCGCTTCGCAAC
ATGTGAGCACCGGTTTATTGACTACCGGAAGCAGTGTGACCGTAGTTTCCT
GATGGACATTTTTCCAGCAATTACACCTCTGTCGATAATTAACTATTGACG
AAAAGCTGAAAACCACTAGAATGCGCCTCCGTGGTAGCAATTCTTTTTAA
GAATTGATGGTATTGGTCCGTAGCTCAGCCTGGTAGAGCGGCGGGCTTTT
AACCCGCAGGTCGCGGGTTCAAATCCCGCCGGACTAGCCAAATTCGAAAA
GCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGA
ATTTGCCATGGCGGCCACCAGGTACCACCGGCGCCTCAGGCATTTGAGAA
GCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAAT
AGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCA
TCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTAT
TTGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCC
AACTGATCTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGC
CTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCA
TTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGC
TGTACCAAATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCC
CAGTCGGGCGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAA
TAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTA
CCAAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAA
TGTCGATCGTGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGC
CATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGAT
GTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCT
CTCCAGGGGAAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGC
GTTGTTTCATCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTG
TGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAG

CAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTT
GAGTCGATACTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCT
CCGGGCGTGAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCAC
AGATTCCGTGGATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGG
CCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACA
TAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCAT
GAATCTGACAGTACGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGT
TTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTAC
CGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCAT
AGCTCACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGG
GCTGTAAGCAAGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGT
AACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGC
AGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACAC
GCGGTTGCTCTTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAG
ATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCC
CAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTT
ACAGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTT
TGATCTTTTCTACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAA
TGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTA
GTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAA
GGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAA
TTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGC
TGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
CGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCC
CTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTT
GCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAA
CATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCA
ACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTG
ATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTT
GCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCG
CTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGA
CGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTG
CTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTT
CATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGA
AATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTG
GTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAA
GATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCG
ACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCG
ACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAA

TCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGT
AATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAA
CGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCG
GCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTC

Chapter 5

pEvol PheRS

TCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTC
ATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCG
CCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTA
TTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTG
CGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGA
GGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTG
ACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCG
GAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGT
GCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATAT

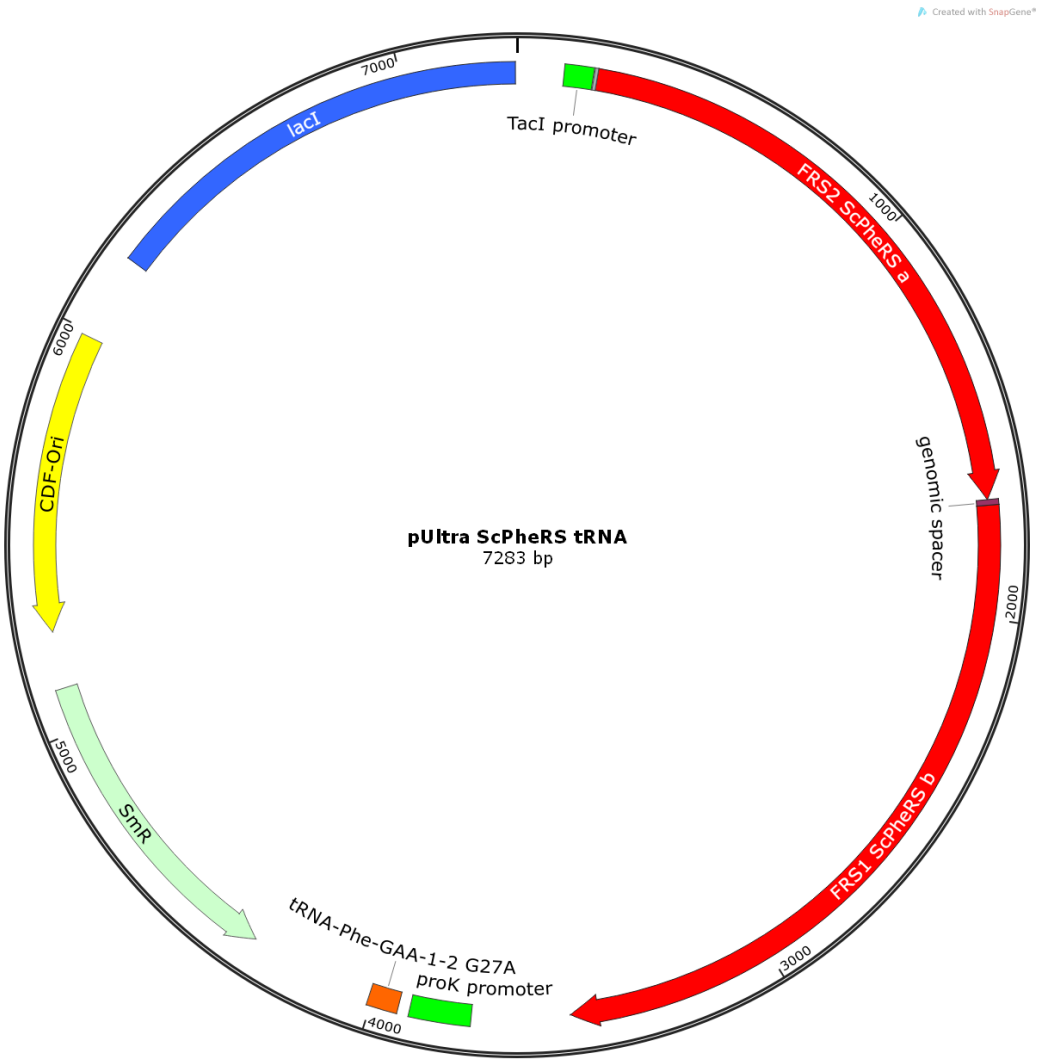
GTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGT
CGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCC
TGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCA
AAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTG
ACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGT
TTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTG
ACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGA
ACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGT
AATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAA
CTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGG
TTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGC
GGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTC
AAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAAT
TTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGT
AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAATTATGA
CAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTC
GCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTT
GATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTG
GTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTT
AAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACG
GCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCT
GCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCAT
CGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATT
GCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCC
CGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCT
TCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAG
CCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACC
ATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGG
AACAGCAAAATATCACTCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTC
ACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCC
CAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCG
TTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGA
TCATTTTGCGCTTCAGCCATACTTTTCATACTCCCGCCATTCAGAGAAGAA
ACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGG
CTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAA
CAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATA
ATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCT
ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTT
TATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAAGAAA

TAATTTTGTTTAACTTTAAGAAGGAGAATACATCAACTAGTACGCAAGTTC
ACGTAAAAAGGGTATCTAGAGGTTGAGGTGATTTTATGTCACATCTCGCA
GAACTGGTTGCCAGTGCGAAGGCGGCCATTAGCCAGGCGTCAGATGTTGC
CGCGTTAGATAATGTGCGCGTCGAATATTTGGGTAAAAAAGGGCACTTAA
CCCTTCAGATGACGACCCTGCGTGAGCTGCCGCCAGAAGAGCGTCCGGCA
GCTGGTGCGGTTATCAACGAAGCGAAAGAGCAGGTTCAGCAGGCGCTGA
ATGCGCGTAAAGCGGAACTGGAAAGCGCTGCACTGAATGCGCGTCTGGCG
GCGGAAACGATTGATGTCTCTCTGCCAGGTCGTCGCATTGAAAACGGCGG
TCTGCATCCGGTTACCCGTACCATCGACCGTATCGAAAGTTTCTTCGGTGA
GCTTGGCTTTACCGTGGCAACCGGGCCGGAAATCGAAGACGATTATCATA
ACTTCGATGCTCTGAACATTCCTGGTCACCACCCGGCGCGCGCTGACCAC
GACACTTTCTGGTTTGACACTACCCGCCTGCTGCGTACCCAGACCTCTGGC
GTACAGATCCGCACCATGAAAGCCCAGCAGCCACCGATTCGTATCATCGC
GCCTGGCCGTGTTTATCGTAACGACTACGACCAGACTCACACGCCGATGT
TCCATCAGATGGAAGGTCTGATTGTTGATACCAACATCAGCTTTACCAACC
TGAAAGGCACGCTGCACGACTTCCTGCGTAACTTCTTTGAGGAAGATTTG
CAGATTCGCTTCCGTCCTTCCTACTTCCCGTTTACCGAACCTTCTGCAGAA
GTGGACGTCATGGGTAAAAACGGTAAATGGCTGGAAGTGCTGGGCTGCG
GGATGGTGCATCCGAACGTGTTGCGTAACGTTGGCATCGACCCGGAAGTT
TACTCTGGTTTCGCCTTCGGGATGGGGATGGAGCGTCTGACTATGTTGCGT
TACGGCGTCACCGACCTGCGTTCATTCTTCGAAAACGATCTGCGTTTCCTC
AAACAGTTTAAATAAGGCAGGAATAGATTATGAAATTCAGTGAACTGTGG
TTACGCGAATGGGTGAACCCGGCGATTGATAGCGATGCGCTGGCAAATCA
AATCACTATGGCGGGCCTGGAAGTTGACGGTGTAGAACCGGTTGCCGGCA
GCTTCCACGGCGTGGTCGTTGGTGAAGTGGTTGAGTGTGCGCAGCATCCG
AACGCTGACAAACTGCGTGTGACAAAAGTGAATGTCGGCGGCGATCGCCT
GCTGGACATCGTCTGCGGTGCGCCAAACTGCCGTCAGGGCCTGCGTGTAG
CGGTAGCGACCATTGGTGCTGTTCTGCCGGGTGATTTCAAAATTAAAGCG
GCGAAACTGCGTGGCGAACCGTCTGAAGGGATGCTGTGCTCCTTCTCTGA
ACTGGGCATTTCTGACGATCACAGCGGCATTATCGAACTGCCTGCGGATG
CGCCGATTGGCACCGATATCCGTGAATACCTGAAACTTGATGACAACACC
ATCGAAATCAGCGTGACGCCAAACCGTGCCGACTGCTTAGGCATCATTGG
TGTTGCGCGTGACGTTGCCGTGCTGAACCAGCTGCCGCTGGTTCAACCGG
AAATCGTTCCGGTTGGTGCGACCATCGACGACACGCTGCCGATTACAGTC
GAAGCGCCGGAAGCCTGCCCGCGTTATCTTGGCCGTGTGGTAAAAGGCAT
TAACGTTAAAGCGCCAACTCCGCTGTGGATGAAAGAAAAACTGCGTCGTT
GCGGGATCCGTTCTATCGATGCAGTTGTTGACGTCACCAACTATGTGCTGC
TCGAACTGGGCCAGCCGATGCACGCTTTCGATAAAGATCGCATTGAAGGC
GGCATTGTGGTGCGGATGGCGAAAGAGGGCGAAACGCTGGTGCTGCTCG
ACGGTACTGAAGCGAAGCTGAATGCTGACACTCTGGTCATCGCCGACCAC
AACAAGGCGCTGGCGATGGGCGGCATCTTCGGTGGCGAACACTCTGGCGT

GAATGACGAAACACAAAACGTGCTGCTGGAATGCGCGTTCTTTAGCCCGC
TGTCTATCACCGGTCGTGCTCGTCGTCATGGCCTGCATACCGATGCGTCTC
ACCGTTATGAGCGTGGCGTTGATCCGGCACTGCAGCACAAAGCGATGGAA
CGTGCGACCCGTCTGCTGATCGACATCTGCGGTGGTGAGGCTGGCCCGGT
AATTGATATCACCAACGAAGCAACGCTGCCGAAGCGTGCAACCATCACTC
TACGTCGTAGCAAACTGGATCGCCTGATCGGCCATCATATTGCGGATGAG
CAGGTAACTGACATTCTGCGTCGTCTCGGCTGCGAAGTGACCGAAGGCAA
AGACGAATGGCAGGCAGTTGCGCCGAGCTGGCGTTTCGATATGGAGATTG
AAGAAGATCTGGTTGAAGAAGTCGCGCGTGTTTACGGCTACAACAACATC
CCGGATGAGCCGGTACAGGCAAGCCTGATTATGGGTACTCACCGTGAAGC
TGACCTGTCGCTCAAGCGCGTGAAAACGCTGCTCAACGACAAAGGCTATC
AGGAAGTGATCACCTACAGCTTCGTTGATCCGAAAGTGCAGCAGATGATC
CATCCAGGCGTTGAAGCCTTACTGCTGCCAAGCCCGATCTCTGTTGAAATG
TCAGCAATGCGTCTTTCTCTGTGGACTGGCCTGCTGGCAACCGTGGTGTAC
AACCAGAACCGTCAGCAGAACCGTGTGCGCATTTTCGAAAGCGGTCTGCG
TTTCGTACCAGATACTCAGGCACCGTTGGGCATTCGTCAGGATCTGATGTT
AGCCGGTGTGATTTGCGGTAACCGTTACGAAGAGCACTGGAACCTGGCAA
AAGAGACCGTTGATTTCTATGATTTGAAAGGCGATCTTGAATCCGTTCTCG
ACCTGACCGGTAAACTGAATGAGGTTGAGTTCCGTGCAGAAGCGAATCCG
GCACTGCATCCGGGGCAATCCGCAGCGATTTATCTGAAAGGTGAACGTAT
TGGTTTTGTTGGGGTTGTTCATCCTGAACTGGAACGTAAACTGGATCTTAA
CGGTCGCACTCTGGTGTTCGAACTGGAGTGGAACAAGCTCGCAGACCGCG
TGGTGCCTCAGGCGCGCGAGATTTCTCGCTTCCCGGCGAACCGTCGTGAC
ATCGCGGTGGTGGTCGCAGAAAACGTTCCCGCAGCGGATATTTTATCCGA
ATGTAAGAAAGTTGGCGTAAATCAGGTAGTTGGCGTAAACTTATTTGACG
TGTACCGCGGTAAGGGTGTTGCGGAGGGGTATAAGAGCCTCGCCATAAGC
CTGATCCTGCAAGATACCAGCCGTACACTCGAAGAAGAGGAGATTGCCGC
TACCGTCGCCAAATGTGTAGAGGCATTAAAAGAGCGATTCCAGGCATCAT
TGAGGGATTAAACCGATGCGGCCGCTTGAGAGTCAGCTCCTTCCGGTGGG
CGTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACT
CAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCG
AGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAA
GACTGGGCCTTGTCGACCGAATTTCTGCCATTCATCCGCTTATTATCACTT
ATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAA
AAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAG
CATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCG
CCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGA
AAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTG
GTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAA
CCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCG
AATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGC

GATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAAC
ACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGG
ATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAAC
TTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGA
ACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATG
TTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTT
TTTCTCCATTTTAGCTTCCTTAGC

pUltra ScPheRS tRNA



ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCG
ACTCCTGCATTAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTG
TGGAATTGTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGTC
TGACTTCCAATTAGAAATTCTAAAAAAACTAGATGAATTGGATGAGATCA
AGTCCACACTGGCAACTTTCCCTCAGCACGGCTCTCAAGATGTTCTTTCCG
CTTTGAACTCTTTGAAAGCCCACAACAAGTTAGAGTTTTCCAAGGTCGAC

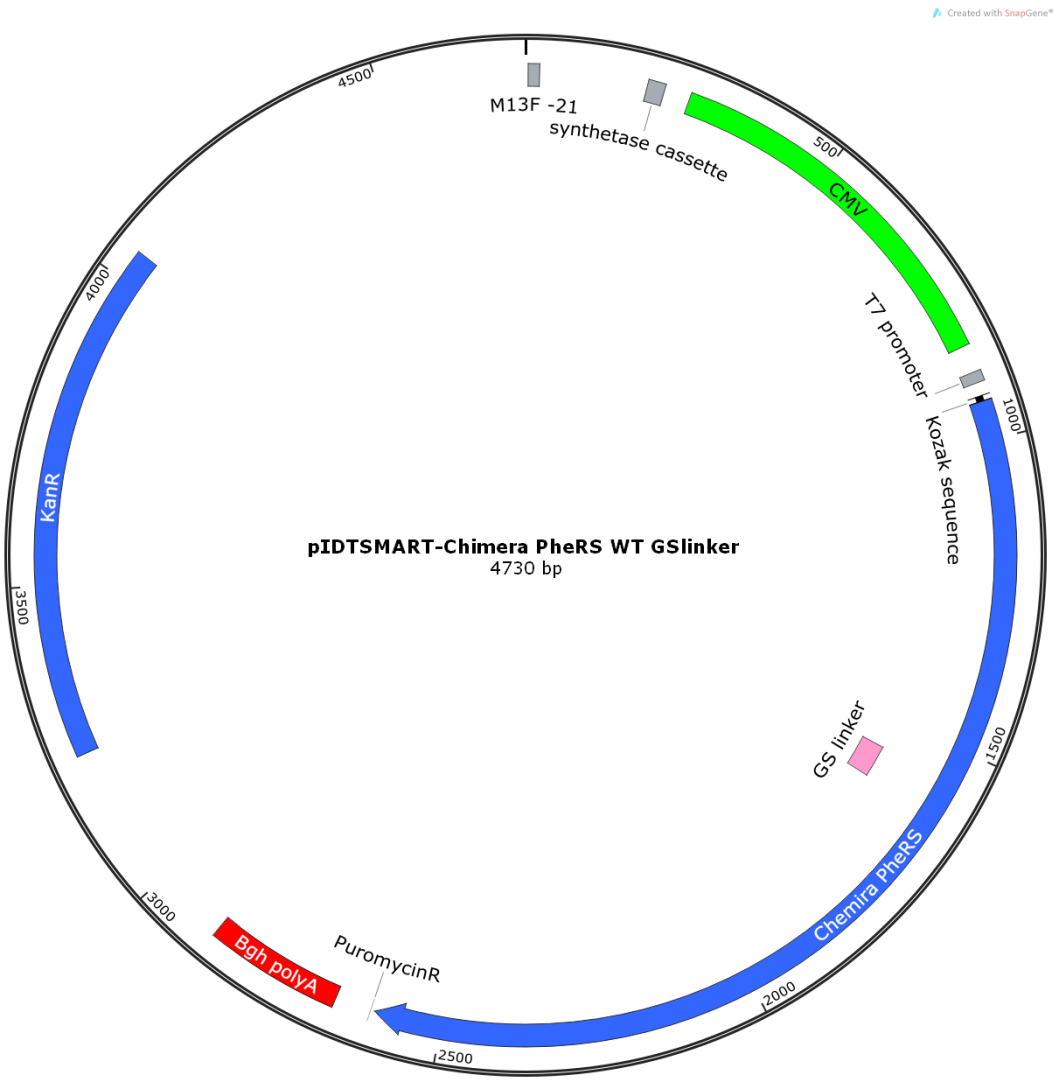
ACGGTTACGTATGACTTGACCAAAGAAGGTGCTCAAATTTTGAATGAAGG
TTCGTACGAAATTAAACTAGTCAAGCTCATCCAAGAGTTGGGTCAACTTC
AAATCAAAGATGTGATGTCCAAACTGGGTCCTCAAGTTGGTAAGGTCGGT
CAGGCTAGAGCTTTCAAGAACGGCTGGATCGCCAAAAACGCCTCAAACGA
GCTTGAACTCTCCGCAAAATTGCAAAATACCGATTTAAATGAGCTTACTG
ATGAAACGCAATCTATTCTAGCGCAAATCAAGAACAACTCGCATCTGGAT
AGCATTGACGCCAAGATTTTGAACGACTTGAAGAAAAGAAAGTTAATTGC
TCAAGGTAAAATCACAGATTTCAACGTTACCAAAGGGCCAGAGTTCTCGA
CCGACCTCACCAAATTGGAAACCGATCTTACCTCCGACATGGTCTCCACC
AATGCATACAAGGACTTGAAGTTCAAGCCTTACAATTTCAATTCTCAAGG
TGTGCAAATATCTTCAGGTGCTCTTCACCCCTTAAACAAAGTCAGAGAGG
AATTTAGACAAATTTTCTTTTCCATGGGATTCACAGAGATGCCCTCGAACC
AATACGTCGAGACAGGTTTCTGGAACTTCGATGCCCTTTACGTCCCACAAC
AGCATCCTGCTCGTGACCTGCAAGACACTTTCTACATCAAGGACCCACTA
ACCGCTGACTTGCCCGATGACAAGACATACATGGACAATATCAAAGCCGT
TCACGAACAGGGGAGATTCGGGTCCATCGGTTATCGTTACAACTGGAAGC
CAGAAGAATGTCAAAAATTGGTCTTGAGAACTCACTCCACAGCCATCTCT
GCCAGAATGCTGCACGATTTGGCCAAAGATCCAAAGCCCACCAGATTGTT
TTCTATCGACCGTGTTTTCCGTAACGAAGCAGTTGACGCCACCCATTTGGC
CGAATTCCACCAGGTGGAAGGTGTTCTTGCCGACTACAACATTACTCTGG
GTGACCTGATCAAGTTCATGGAAGAGTTTTTCGAAAGAATGGGTGTCACC
GGTTTGAGATTCAAGCCTACCTACAATCCTTACACCGAGCCATCAATGGA
AATCTTTTCTTGGCACGAAGGTTTGCAAAAATGGGTCGAAATCGGTAACT
CTGGTATGTTCAGACCAGAAATGCTCGAGTCCATGGGTCTACCAAAGGAT
CTAAGAGTCCTTGGTTGGGGGTTATCCTTGGAAAGACCTACCATGATCAA
ATATAAGGTTCAAAACATCAGAGAACTGTTAGGTCATAAAGTCTCTTTGG
ACTTTATCGAAACCAATCCTGCTGCTAGATTGGACGAAGACTTGTACGAA
TAAGGCAGGAATAGATTATGCCTACCGTCTCCGTGAACAAGCAGCAATTA
TTTGATCTTCTAGGCAAAAACTACACTTCCCAAGAGTTCGATGAATTATGT
TTTGAATTCGGTATGGAAATGGACGAAGACACCACAGAGGAGGCCTTGAA
AACCGGGGAGGAGCCGGAATTGAAGCTTGATATCAGTGCCAATCGTTACG
ATTTGCTTTGTATCGAAGGTATTTCACAGTCGCTGAACGAATACTTGGAAC
GTAAAGAAAGACCTGACTATAAATTAAGCAAGCCAACCACTAAGTTGATC
ATCGACAAATCAACGGAGCAAATTAGACCTTTTGCTACCGCTGCTGTATT
GAGAAATATCAAGCTTAACGAAAAATCTTACGCTTCTTTTATTGCCTTGCA
AGATAAATTACATGCCAATCTATGTAGAAACAGAAGCTTGGTTGCCATGG
GTACTCACGATTTAGATTCAATTGAAGGTCCATTCCATTACAGAGCTCTAC
CACCAAAGGACATCAAGTTCGTACCATTGAATCAAACCCAAGAGTTTACT
GGTGACAAATTGATCGAGTTTTATAAATCTCCAGAACAGAAAAACAACAT
AGGGAGATACGTTCACATTATTGAGGATTCTCCAGTCTTCCCAGTTATTAT
GGACAGCAAAGATCGTGTTTGCTCCCTGCCACCATTAATCAATAGTGAAC

ATTCGAAGATCTCTGTGAACACCCGTAACATTTTGATTGATATAACCGCCA
CCGATAAGACCAAAGCCGAGATCGTTTTGAACATATTAACTACAATGTTC
TCACGTTATTGTGACGAACCATTCACGGTTGAGCCTGTAGAAATTGTCTCT
GAACACAATGGCCAATCCCGTTTGGCGCCAAACTTCAACGATAGAATTAT
GGATGTCTCCATTAAGTACATCAACTCCTGTCTTGGCCTAGATCAATCCGC
TGATGAAATTGCTCATTGTCTAAAGAAGATGTCGTTGCATGCCGTTCAATC
AAAGGAAGACAAGGACATCTTGCACGTTGACATTCCGGTAACTAGACCTG
ATATTTTGCACGCTTGTGATATAATGGAAGATGCCGCTGTCGGTTATGGTT
TCAATAATTTGCCAAAGGGTGAGAAATTATCCAATGCCAACTTCATTGCC
AAACCATTACCAATCAACAAGGTTTCTGATATTTTCAGAGTTGCATCCTCT
CAAGCCACGTGGGTTGAGGTTTTACCATTGACCTTATGTTCGCACGATGAA
AACTTTAAATTTCTAAGACAATCCGACAATGGTGATTTAGCTGTCAAATTG
GCCAACCCAAAGACTTTGGAATACCAAGTTGTTAGAACCACTTTATTGCCT
GGTATCTTAAAGACAGTCAAGGAAAACAGAAAACATTCCTTGCCAATCAA
AGTCTTTGAAACCGGTGACGTTGTATTTAAAGACGACAAACTAGAAAGGA
AGGCGTACAATGAACGTCACTGGGCTGCCATCTACGTGGGTAAGAATTCT
GGGTTTGAAATCATTCAAGGGTTATTGGGTAAAATCATGCAAACTTTTAG
AACAGAGTGGATTGCAGACTACGGTGCTGCTGCTTCTGGCAGAGGTTACT
GGATTGAAGAAGACGATTCTGTGAAAACCTACTTCCCAGGTAGAGGTGCC
AAGGTCATGTTCAGATCCAAAGAAGGCGCTGAGCCAAAGCAAATCGGCC
ACTTGGGTGTCTTGCATCCTGAAGTCATGATGAATTTCGACGTTCCATTCG
CTGCATCCTTTGTAGAGGTTAATGCCGAAGTCTTCCTATAACTGCAGTTTC
AAACGCTAAATTGCCTGATGCGCTACGCTTATCAGGCCTACATGATCTCTG
CAATATATTGAGTTTGCGTGCTTTTGTAGGCCGGATAAGGCGTTCACGCCG
CATCCGGCAAGAAACAGCAAACAATCCAAAACGCCGCGTTCAGCGGCGTT
TTTTCTGCTTTTCTTCGCGAATTAATTCCGCTTCGCAACATGTGAGCACCG
GTTTATTGACTACCGGAAGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAG
GCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAATTGACGATATGATC
AGTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGC
ATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGC
CCCGCATTGCGGACTTAGCTCAGTTGGGAGAGCGCCAGACTGAAAATCTG
GAGGTCCTGTGTTCGATCCACAGAGTTCGCACCAAATTCGAAAAGCCTGC
TCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCT
TTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAA
CCAGGCGTTTAAGGGTACCACCGGCGCCTCAGGCATTTGAGAAGCACACG
GTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATA
AGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATCGTGGC
CGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTTGCCGA
CTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACTGAT
CTGCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAG
CTTCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCCCAG

TCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGTACCAA
ATGCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCGGG
CGGCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTG
TTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAA
CGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCGATCG
TGGCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCA
AATTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGTG
CACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGG
AAGCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCA
TCAAGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTC
AGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGG
TTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATA
CTTCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAG
CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTA
GAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGA
GACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTG
GATAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGT
GGCGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACG
CTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACA
GTACGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGG
TCGCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTG
TTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACA
CTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCA
AGAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACT
TGAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATT
GGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTC
TTGAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGC
TGTGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTT
AACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAA
AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCT
ACTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCT
GAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGCCCC
GCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGT
CGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCG
CTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATG
AATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGT
GGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGC
CTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCA
GGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTG
TCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCAGC
CCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGC

AACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTT
GTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCT
GAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACC
CAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCC
GGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAG
CGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCA
CCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCA
CGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGC
GACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACG
ACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCT
CCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGG
CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCT
GCGACATCGTATAACGTTACTGGTTTCACATTC

pIDTSMART ChiPheRS WT Gslinker

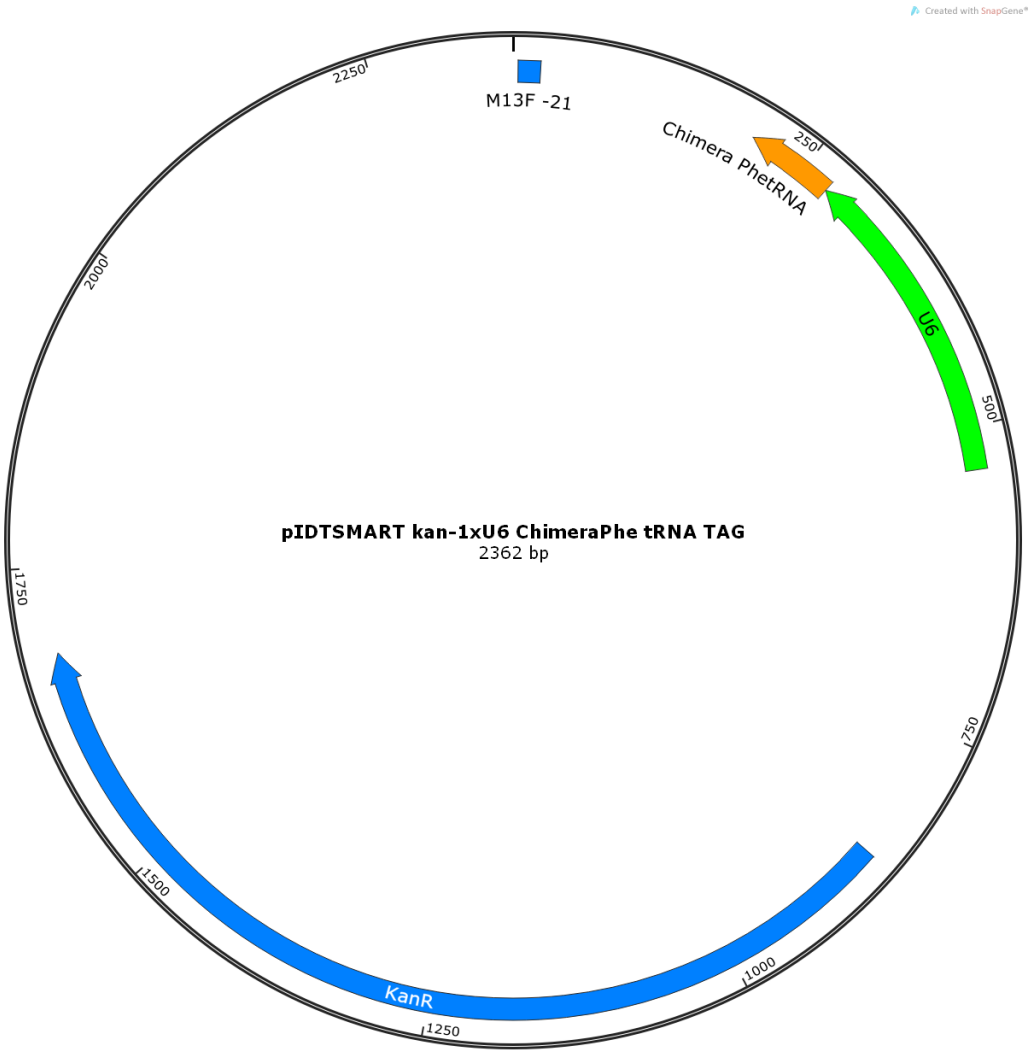


CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATC
GTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGA
GTCTGAGGCTCGTCCTGAATGATATGCGACCGCCGGAGGGTTGCGTTTGA
GACGGGCGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGCTAGGGCGGC
CGCAAATACCTGCAGGATCCGTTTTGCGCTGCTTCGCGATGTACGGGCCA
GATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTA
CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTA
CGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACG
TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCA
AGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAAT
GGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTT
GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT
GGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA
AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA

CGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAA
CTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATA
GGGAGACCCAAGCTGGCTAGCGCCGCCACCATGGATAAGAAGCCGCTGG
ATGTTCTGATCTCTGCGACCGGTCTGTGGATGTCCCGTACCGGCACGCTGC
ACAAGATCAAGCACTATGAGATTTCTCGTTCTAAAATCTACATCGAAATG
GCGTGTGGTGACCATCTGGTTGTGAACAACTCTCGTTCTTGTCGTCCCGCA
CGTGCATTCCGTTATCATAAATACCGTAAAACCTGCAAACGTTGTCGTGTT
TCTGACGAAGATATCAACAACTTCCTGACCCGTTCTACCGAAGGCAAAAC
CTCTGTTAAAGTTAAAGTTGTTTCTGAGCCGAAAGTGAAAAAAGCGATGC
CGAAATCTGTTTCTCGTGCGCCGAAACCGCTGGAAAATCCGGTTTCTGCG
AAAGCGTCTACCGACACCTCTCGTTCTGTTCCGTCTCCGGCGAAATCTACC
CCGAACTCTCCGGTTCCGACCTCTGCAAGCGCCCCAGCTCTGACTAAATCC
CAGACGGACCGTCTGGAGGTGCTGCTGAACCCAAAGGATGAAATCTCTCT
GAACAGCGGCAAGCCTTTCCGTGAGCTGGAAAGCGAGCTGCTGTCTCGTC
GTAAAAAGGATCTGCAACAGATCTACGCTGAGGAACGCGAGGGAGGCGG
TGGGAGTGGGGGCGGTGGAAGCGGAGGTGGCGGGAGTGGAGGAGGAGG
AAGTCAGGCCTGGGGATCGAGGCCTCCTGCAGCAGAGTGTGCCACCCAAA
GAGCTCCAGGCAGTGTGGTGGAGCTGCTGGGCAAATCCTACCCTCAGGAC
GACCACAGCAACCTCACCCGGAAGGTCCTCACCAGAGTTGGCAGGAACCT
GCACAACCAGCAGCATCACCCTCTGTGGCTGATCAAGGAGAGGGTGAAG
GAGCACTTCTACAAGCAGTATGTGGGCCGCTTTGGGACCCCGTTGTTCTCG
GTCTACGACAACCTTTCTCCAGTGGTCACGACCTGGCAGAACTTTGACAG
CCTGCTCATCCCAGCTGATCACCCCAGCAGGAAGAAGGGGGACAACTATT
ACCTGAATCGGACTCACATGCTGAGAGCGCACACGTCTGCACACCAGTGG
GACTTGCTGCACGCGGGACTGGATGCCTTCCTGGTGGTGGGTGATGTCTA
CAGGCGTGACCAGATCGACTCCCAGCACTACCCTATTTTCCACCAGCTGG
AGGCCGTGCGGCTCTTCTCCAAGCATGAGTTATTTGCTGGTATAAAGGAT
GGAGAAAGCCTGCAGCTCTTTGAACAAAGTTCTCGCTCTGCGCATAAACA
AGAGACACACACCATGGAGGCCGTGAAGCTTGTAGAGTTTGATCTTAAGC
AAACGCTTACCAGGCTCATGGCACATCTTTTTGGAGATGAGCTGGAGATA
AGATGGGTAGACTGCTACTTCCCTTTTACACATCCTTCCTTTGAGATGGAG
ATCAACTTTCATGGAGAATGGCTGGAAGTTCTTGGCTGCGGGGTGATGGA
ACAACAACTGGTCAATTCAGCTGGTGCTCAAGACCGAATCGGCTGGGCTT
TTGGCCTAGGATTAGAAAGGCTAGCCATGATCCTCTACGACATCCCTGAT
ATCCGTCTCTTCTGGTGTGAGGACGAGCGCTTCCTGAAGCAGTTCTGTGTA
TCCAACATTAATCAGAAGGTGAAGTTTCAGCCTCTTAGCAAATAAGAATT
CAACGCGTTAAGTCGACTTTAACTCGAGTCTAGAGGGCCCGTTTAAACCC
GCTGATCAGGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAA
TTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGG

GGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGG
GGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCCTAGGGGTGC
GAGCGGATCGAGCAGTGTCGATCACTACTGGACCGCGAGCTGTGCTGCGA
CCCGTGATCTTACGGCATTATACGTATGATCGGTCCACGATCAGCTAGATT
ATCTAGTCAGCTTGATGTCATAGCTGTTTCCTGAGGCTCAATACTGACCAT
TTAAATCATACCTGACCTCCATAGCAGAAAGTCAAAAGCCTCCGACCGGA
GGCTTTTGACTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAA
ATAAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCT
AAGGAAGCTAAAATGAGCCATATTCAACGGGAAACGTCTTGCTTGAAGCC
GCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTC
GCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAG
CCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAA
TGATGTTACAGATGAGATGGTCAGGCTAAACTGGCTGACGGAATTTATGC
CTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTAC
TCACCACTGCGATCCCAGGGAAAACAGCATTCCAGGTATTAGAAGAATAT
CCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGG
TTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACGGCGATCGCGTATTTC
GTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGGTGCGAGT
GATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGA
AATGCATAAACTCTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGA
TTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTAT
TGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCC
TATGGAACTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTC
AAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCACTTGA
TGCTCGATGAGTTTTTCTAATGAGGACCTAAATGTAATCACCTGGCTCACC
TTCGGGTGGGCCTTTCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCC
CTGACGAGCATCACAAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCC
GACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCC
TTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTC
GGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTC
AGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAA
CTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGC
CAGTTACCTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA
CCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGA
AAAAAAGGATCTCAAGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

pIDTSAMRT ChiPhetRNATAG



CCCGTGTAAAACGACGGCCAGTTTATCTAGTCAGCTTGATTCTAGCTGATC
GTGGACCGGAAGGTGAGCCAGTGAGTTGATTGCAGTCCAGTTACGCTGGA
GTCTGAGGCTCGTCCTGAATGATATGCGaCCGCCGGAGGGTTGCGTTTGAG
ACGGGCGACAGATCCAGTCGCGCTGCTCTCGTCGATCCGctagcaaaaaaTGGT
GCCGAAACCGGGAATCTAACCCGGCTGAACGGATTTAGAGTCCGTTCGAT
CTACATGATCATCTCGGCGGTGTTTCGTCCTTTCCACAAGATATATAAAGC
CAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTT
AAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTAC
GTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAAT
TATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGG
AAATAGGCCCTCTTCCTGCCCGAcCTAGGGGTGCGAGCGGATCGAGCAGT
GTCGATCACTACTGGACCGCGAGCTGTGCTGCGACcCGTGATCTTACGGCA
TTATACGTATGATCGGTCCACGATCAGCTAGATTATCTAGTCAGCTTGATG
TCATAGCTGTTTCCTGAGGCTCAATACTGACCATTTAAATCATACCTGACC
TCCATAGCAGAAAGTCAAAAGCCTCCGACCGGAGGCTTTTGACTTGATCG
GCACGTAAGAGGTTCCAACTTTCACCATAATGAAATAAGATCACTACCGG

GCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATG
AGCCATATTCAACGGGAAACGTCTTGCTTGAAGCCGCGATTAAATTCCAA
CATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGC
AATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAG
TTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGA
GATGGTCAGGCTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCA
AGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCC
CAGGGAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAA
AATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCT
GTTTGTAATTGTCCTTTTAACGGCGATCGCGTATTTCGTCTCGCTCAGGCG
CAATCACGAATGAATAACGGTTTGGTTGGTGCGAGTGATTTTGATGACGA
GCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTCT
TGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATA
ACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAG
TCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTC
GGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATT
GATAATCCTGATATGAATAAATTGCAGTTTCACTTGATGCTCGATGAGTTT
TTCTAATGAGGACCTAAATGTAATCACCTGGCTCACCTTCGGGTGGGCCTT
TCTGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCAC
AAAAATCGATGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGA
CCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGG
CGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTC
GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC
GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTA
TCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGT
AGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTCGGAA
AAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCA
AGAAGATCCTTTGATTTTCTACCGAAGAAAGGCCCA

Chapter 6
pEvol ArgRS

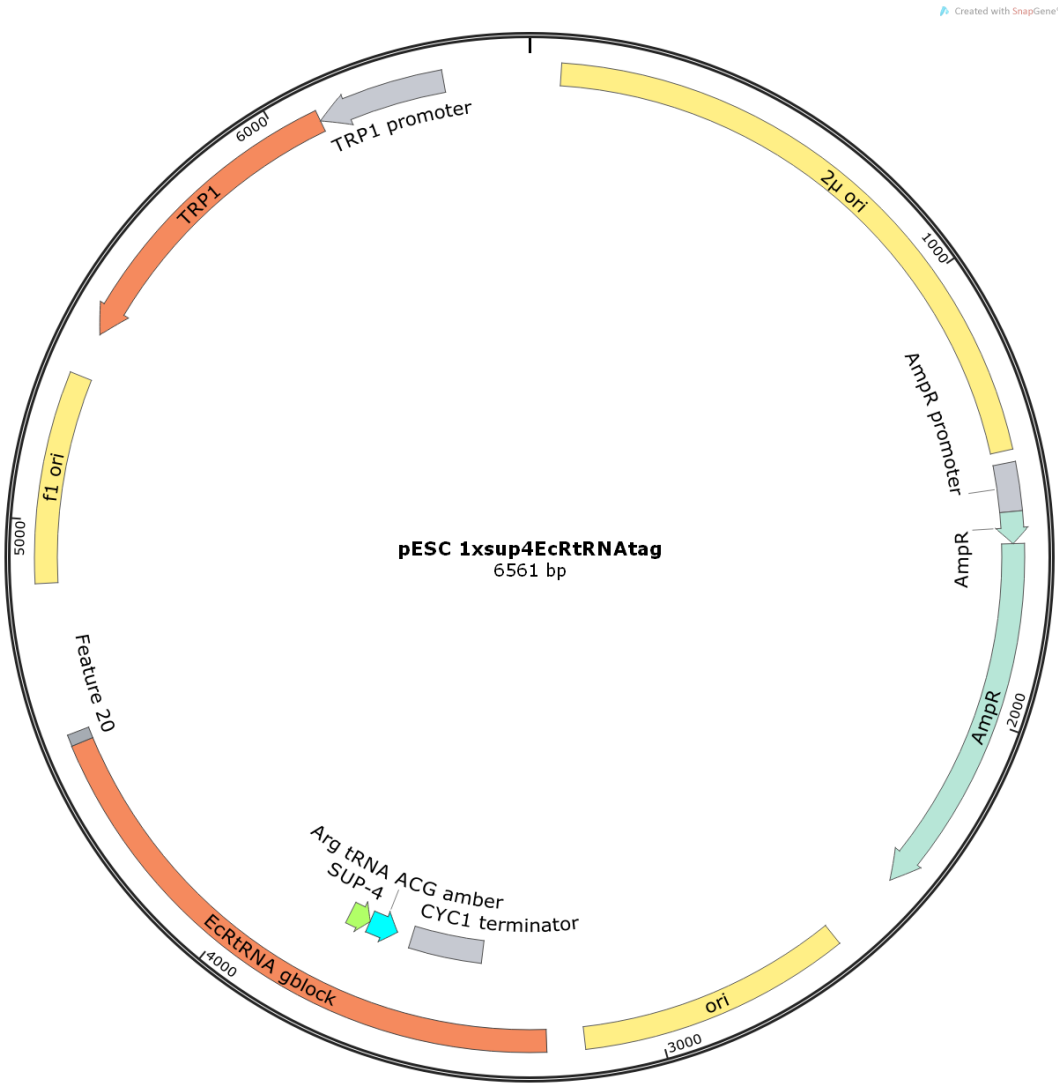
TCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTC
ATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCG
CCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTA
TTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTG
CGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGA
GGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTG
ACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCG
GAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGT
GCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATAT
GTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGT
CGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCC
TGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCA
AAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTG
ACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGT

TTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTG
ACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGA
ACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGT
AATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAA
CTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGG
TTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGC
GGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTC
AAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAAT
TTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGT
AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAATTATGA
CAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTC
GCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTT
GATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTG
GTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTT
AAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACG
GCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCT
GCCAGGTGATCGCTGATGTACTGACAAGCCTCGGAAACCAATTGTCCATA
TTGCATCAGACATTGCCGTCACTGGCAGTAACAATTGCTCAAGCAGATTT
ATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATT
TGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAA
GAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGT
AGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCC
GGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATC
ACTCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACC
GCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGA
TAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCGTTAAACCCGCCACC
AGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTC
AGCCATACTTTTCATACTCCCGCCATTCAGAGAAGAAACCAATTGTCCATA
TTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAAC
CAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCA
AAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAA
AGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATT
TTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTC
TACTGTTTCTCCATACCCGTTTTTTTGGGCTAAAGAAATAATTTTGTTTAAC
TTTAAGAAGGAGAATACATCAACTAGTACGCAAGTTCACGTAAAAAGGGT
ATCTAGAGGTTGAGGTGATTTTATGAATATTCAGGCTCTTCTCTCAGAAAA
AGTCCGTCAGGCCATGATTGCGGCAGGCGCGCCTGCGGATTGCGAACCGC
AGGTTCGTCAGTCAGCAAAAGTTCAGTTCGGCGACTATCAGGCTAACGGC
ATGATGGCAGTTGCTAAAAAACTGGGTATGGCACCGCGACAATTAGCAGA
GCAGGTGCTGACTCATCTGGATCTTAACGGTATCGCCAGCAAAGTTGAGA

TCGCCGGTCCAGGCTTTATCAACATTTTCCTTGATCCGGCATTCCTGGCTG
AACATGTTCAGCAGGCGCTGGCGTCCGATCGTCTCGGTGTTGCTACGCCA
GAAAAACAGACCATTGTGGTTGACTACTCTGCGCCAAACGTGGCGAAAGA
GATGCATGTCGGTCACCTGCGCTCTACCATTATTGGTGACGCAGCAGTGC
GTACTCTGGAGTTCCTCGGTCACAAAGTGATTCGCGCAAACCACGTCGGC
GACTGGGGCACTCAGTTCGGTATGCTGATTGCATGGCTGGAAAAGCAGCA
GCAGGAAAACGCCGGTGAAATGGAGCTGGCTGACCTTGAAGGTTTCTACC
GCGATGCGAAAAAGCATTACGATGAAGATGAAGAGTTCGCCGAGCGCGC
ACGTAACTACGTGGTAAAACTGCAAAGCGGTGACGAATATTTCCGCGAGA
TGTGGCGCAAACTGGTCGACATCACCATGACGCAGAACCAGATCACCTAC
GATCGTCTCAACGTGACGCTGACCCGTGATGACGTGATGGGCGAAAGCCT
CTACAACCCGATGCTGCCAGGAATTGTGGCGGATCTCAAAGCCAAAGGTC
TGGCAGTAGAAAGCGAAGGGGCGACCGTCGTATTCCTTGATGAGTTTAAA
AACAAGGAAGGCGAACCGATGGGCGTGATCATTCAGAAGAAAGATGGCG
GCTATCTCTACACCACCACTGATATCGCCTGTGCGAAATATCGTTATGAAA
CACTGCATGCCGATCGCGTGCTGTATTACATCGACTCCCGTCAGCATCAAC
ACCTGATGCAGGCATGGGCGATCGTCCGTAAAGCAGGCTATGTACCGGAA
TCCGTACCGCTGGAACACCACATGTTCGGCATGATGCTGGGTAAAGACGG
CAAACCGTTCAAAACCCGCGCGGGTGGTACAGTGAAACTGGCCGATCTGC
TGGATGAAGCCCTGGAACGTGCACGCCGTCTGGTGGCAGAAAAGAACCC
GGATATGCCAGCCGACGAGCTGGAAAAACTGGCTAACGCGGTTGGTATTG
GTGCGGTGAAATATGCGGATCTCTCCAAAAACCGCACCACGGACTACATC
TTCGACTGGGACAACATGCTGGCGTTTGAGGGTAATACCGCGCCATACAT
GCAGTATGCATACACGCGTGTATTGTCCGTGTTCCGTAAAGCAGAAATTG
ACGAAGAGCAACTGGCTGCAGCTCCGGTTATCATCCGTGAAGATCGTGAA
GCGCAACTGGCAGCTCGCCTGCTGCAGTTTGAAGAAACCCTCACCGTGGT
TGCCCGTGAAGGCACGCCGCATGTAATGTGTGCTTACCTGTACGATCTGG
CCGGTCTGTTCTCTGGCTTCTACGAGCACTGCCCGATCCTCAGCGCAGAAA
ACGAAGAAGTGCGTAACAGCCGTCTAAAACTGGCACAACTGACGGCGAA
GACGCTGAAGCTGGGTCTGGATACGCTGGGTATTGAGACTGTAGAGCGTA
TGTAACATCACCATCACCACCATTAAACCGATGCGGCCGCTTGAGAGTCA
GCTCCTTCCGGTGGGCGTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTG
ACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTG
GGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGA
AAGGCTCAGTCGAAAGACTGGGCCTTGTCGACCGAATTTCTGCCATTCAT
CCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACC
AATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTAC
TGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGC
ATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTAT
AATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCC
ACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAA

AAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGT
AACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCG
TGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAAC
GGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCAT
TGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAA
TAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGG
CCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGAC
TGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTG
GTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGC

pESC 1xSUP4 EcRtRNA TAG



GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAA
TAATGGTTTCTTAATATGATCCAATATCAAAGGAAATGATAGCATTGAAG
GATGAGACTAATCCAATTGAGGAGTGGCAGCATATAGAACAGCTAAAGG
GTAGTGCTGAAGGAAGCATACGATACCCCGCATGGAATGGGATAATATCA

CAGGAGGTACTAGACTACCTTTCATCCTACATAAATAGACGCATATAAGT
ACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATATAT
ATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGT
GCGCAGCTCGCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAAACGCTTTGA
AGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCAGAGC
GCTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAA
ACGCACCGGACTGTAACGAGCTACTAAAATATTGCGAATACCGCTTCCAC
AAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCT
ATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGA
CCTCTACATTTTTTATGTTTATCTCTAGTATTACTCTTTAGACAAAAAAATT
GTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACGAAAATGTAAA
CATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAA
TTTTCTGACCAATGAAGAATCATCAACGCTATCACTTTCTGTTCACAAAGT
ATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTTTATCTTG
AAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGGAAGTGG
AGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGCCTTCTT
CTAACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCATTAT
AGAGCGCACAAAGGAGAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAA
ATAGCGCTCTCGGGATGCATTTTTGTAGAACAAAAAAGAAGTATAGATTC
TTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAATGC
AGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTT
TACAAAAATGAAGCACAGATTCTTCGTTGGTAAAATAGCGCTTTCGCGTT
GCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAG
CGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTT
CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTT
TCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAA
ATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCG
TGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACC
CAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACG
AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTAT
GTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGC
CGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA
AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCA
TAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAAC
TCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACG
AGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTA
TTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGG
ATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGC
TGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGATCTCGCG

GTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTT
ATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGA
TCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAA
GTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAAC
GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGA
TCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA
AAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG
TCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCAC
CGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT
AAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGA
GAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAA
GCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA
CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCG
TCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTT
GAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT
CAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGCTGGATCTTCGAGCGTCCCAAAA
CCTTCTCAAGCAAGGTTTTCAGTATAATGTTACATGCGTACACGCGTCTGT
ACAGAAAAAAAAGAAAAATTTGAAATATAAATAACGTTCTTAATACTAAC
ATAACTATAAAAAAATAAATAGGGACCTAGACTTCAGGTTGTCTAACTCC
TTCCTTTTCGGTTAGAGCGGATCTTAGCTAGGAAGTGAATGGAGACATAA
AAAACAAAAAAATGGTGCATCCGGGAGGATTCGAACCTCCGACCGCTCG
GTTTAGAGCCGAGTACTCTATCCAGCTGAGCTACGGATGCTATTTAATTGT
TGAAGAAAGAGTATACTACATAACACATATACAATTGAAAAAGAGGCTA
GCACCACCACCGGTCGGGATCGAAGAAATGATGGTAAATGAAATAGGAA
ATCAAGGAGCATGAAGGCAAAAGACAAATATAAGGGTCGAACGAAAAAT
AAAGTGAAAAGTGTTGATATGATGTATTTGGCTTTGCGGCGCCGAAAAAA
CGAGTTTACGCAATTGCACAATCATGCTGACTCTGTGGCGGACCCGCGCT
CTTGCCGGCCCGGCGATAACGCTGGGCGTGAGGCTGTGCCCGGCGGAGTT
TTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATTGG
AGAAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGACA
ACTGGTGTCATTATTTAAGTTGCCGAAAGAACCTGAGTGCATTTGCAACAT
GAGTATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGT
GGTGCGAACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATAAC
CGCTAGAGTACTTTGAAGAGGAAACAGCAATAGGGTTGCTACCAGTATAA

ATAGACAGGTACATACAACACTGGAAATGGTTGTCTGTTTGAGTACGCTTT
CAATTCATTTGGGTGTGCACTTTATTATGTTACAATATGGAAGGGAACTTT
ACACTTCTCCTATGCACATATATTAATTAAAGTCCAATGCTAGTAGAGAA
GGGGGGTAACACCCCTCCGCGCTCTTTTCCGATTTTTTTCTAAGTCTCCAA
TCAAGGTTGTCGGCTTGTCTACCTTGCCAGAAATTTACGAAAAGATGGAA
AAGGGTCAAATCGTTGGTAGATACGTTGTTGACACTTCTAAATAAGCGAA
TTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAG
TGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATT
CTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGA
GGTCGCTCCAATTCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGC
CCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTG
TAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCG
CTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTT
TCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGA
TTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCG
CCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC
TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGAT
TTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATT
TAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTAT
TTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAGGCAAGTGCACA
AACAATACTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTT
TGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACC
TCCGCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCACCAA
CATTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGC
TCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCAC
CTGTCCCACCTGCTTCTGAATCAAACAAGGGAATAAACGAATGAGGTTTC
TGTGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTT
TTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCT
CCATGCAGTTGGACGATATCAATGCCGTAATCATTGACCAGAGCCAAAAC
ATCCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGC
CTGAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAAC
CGGGTCAATTGTTCTCTTTCTATTGGGCACACATATAATACCCAGCAAGTC
AGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGC
AAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATA
CTCCAAGCTGCCTTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGT
CACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAA
TTCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACGTAAGG
TGACAAGCTATTTTTCAATAAAGAATATCTTCCACTACTGCCATCTGGCGT
CATAACTGCAAAGTACACATATATTACGATGCTGTCTATTAAATGCTTCCT
ATATTATATATATAGTAATGTCGTTTATGGTGCACTCTCAGTACAATCTGC

TCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGA
CGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGT
GACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGA
AACGCGCGA

pESC 1xSUP4 EcRtRNA TGA

GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAA
TAATGGTTTCTTAATATGATCCAATATCAAAGGAAATGATAGCATTGAAG
GATGAGACTAATCCAATTGAGGAGTGGCAGCATATAGAACAGCTAAAGG
GTAGTGCTGAAGGAAGCATACGATACCCCGCATGGAATGGGATAATATCA
CAGGAGGTACTAGACTACCTTTCATCCTACATAAATAGACGCATATAAGT
ACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATATAT
ATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGT
GCGCAGCTCGCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAAACGCTTTGA

AGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCAGAGC
GCTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAA
ACGCACCGGACTGTAACGAGCTACTAAAATATTGCGAATACCGCTTCCAC
AAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCT
ATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGA
CCTCTACATTTTTTATGTTTATCTCTAGTATTACTCTTTAGACAAAAAAATT
GTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACGAAAATGTAAA
CATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAA
TTTTCTGACCAATGAAGAATCATCAACGCTATCACTTTCTGTTCACAAAGT
ATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTTTATCTTG
AAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGGAAGTGG
AGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGCCTTCTT
CTAACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCATTAT
AGAGCGCACAAAGGAGAAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAA
ATAGCGCTCTCGGGATGCATTTTTGTAGAACAAAAAAGAAGTATAGATTC
TTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAATGC
AGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTT
TACAAAAATGAAGCACAGATTCTTCGTTGGTAAAATAGCGCTTTCGCGTT
GCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAG
CGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTT
CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTT
TCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAA
ATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCG
TGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACC
CAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACG
AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTAT
GTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGC
CGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA
AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCA
TAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAAC
TCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACG
AGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTA
TTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGG
ATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGC
TGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGATCTCGCG
GTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTT
ATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGA
TCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAA
GTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA

GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAAC
GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGA
TCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA
AAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG
TCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCAC
CGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT
AAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGA
GAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAA
GCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA
CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCG
TCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTT
GAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT
CAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGCTGGATCTTCGAGCGTCCCAAAA
CCTTCTCAAGCAAGGTTTTCAGTATAATGTTACATGCGTACACGCGTCTGT
ACAGAAAAAAAAGAAAAATTTGAAATATAAATAACGTTCTTAATACTAAC
ATAACTATAAAAAAATAAATAGGGACCTAGACTTCAGGTTGTCTAACTCC
TTCCTTTTCGGTTAGAGCGGATCTTAGCTAGGAAGTGAATGGAGACATAA
AAAACAAAAAAATGGTGCATCCGGGAGGATTCGAACCTCCGACCGCTCG
GTTTGAAGCCGAGTACTCTATCCAGCTGAGCTACGGATGCTATTTAATTGT
TGAAGAAAGAGTATACTACATAACACATATACAATTGAAAAAGAGGCTA
GCACCACCACCGGTCGGGATCGAAGAAATGATGGTAAATGAAATAGGAA
ATCAAGGAGCATGAAGGCAAAAGACAAATATAAGGGTCGAACGAAAAAT
AAAGTGAAAAGTGTTGATATGATGTATTTGGCTTTGCGGCGCCGAAAAAA
CGAGTTTACGCAATTGCACAATCATGCTGACTCTGTGGCGGACCCGCGCT
CTTGCCGGCCCGGCGATAACGCTGGGCGTGAGGCTGTGCCCGGCGGAGTT
TTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATTGG
AGAAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGACA
ACTGGTGTCATTATTTAAGTTGCCGAAAGAACCTGAGTGCATTTGCAACAT
GAGTATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGT
GGTGCGAACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATAAC
CGCTAGAGTACTTTGAAGAGGAAACAGCAATAGGGTTGCTACCAGTATAA
ATAGACAGGTACATACAACACTGGAAATGGTTGTCTGTTTGAGTACGCTTT
CAATTCATTTGGGTGTGCACTTTATTATGTTACAATATGGAAGGGAACTTT
ACACTTCTCCTATGCACATATATTAATTAAAGTCCAATGCTAGTAGAGAA
GGGGGGTAACACCCCTCCGCGCTCTTTTCCGATTTTTTTCTAAGTCTCCAA

TCAAGGTTGTCGGCTTGTCTACCTTGCCAGAAATTTACGAAAAGATGGAA
AAGGGTCAAATCGTTGGTAGATACGTTGTTGACACTTCTAAATAAGCGAA
TTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAG
TGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATT
CTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGA
GGTCGCTCCAATTCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGC
CCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTG
TAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCG
CTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTT
TCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCC
TTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGA
TTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCG
CCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAAC
TGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGAT
TTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATT
TAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTAT
TTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAGGCAAGTGCACA
AACAATACTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTT
TGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACC
TCCGCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCACCAA
CATTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGC
TCTCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCAC
CTGTCCCACCTGCTTCTGAATCAAACAAGGGAATAAACGAATGAGGTTTC
TGTGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTT
TTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCT
CCATGCAGTTGGACGATATCAATGCCGTAATCATTGACCAGAGCCAAAAC
ATCCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGC
CTGAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAAC
CGGGTCAATTGTTCTCTTTCTATTGGGCACACATATAATACCCAGCAAGTC
AGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGC
AAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATA
CTCCAAGCTGCCTTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGT
CACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAA
TTCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACGTAAGG
TGACAAGCTATTTTTCAATAAAGAATATCTTCCACTACTGCCATCTGGCGT
CATAACTGCAAAGTACACATATATTACGATGCTGTCTATTAAATGCTTCCT
ATATTATATATATAGTAATGTCGTTTATGGTGCACTCTCAGTACAATCTGC
TCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGA
CGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGT
GACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGA
AACGCGCGA

pESC EcRRSWT 1xSUP4 EcRtRNA$_{UCCU}$

GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAA
TAATGGTTTCTTAATATGATCCAATATCAAAGGAAATGATAGCATTGAAG
GATGAGACTAATCCAATTGAGGAGTGGCAGCATATAGAACAGCTAAAGG
GTAGTGCTGAAGGAAGCATACGATACCCCGCATGGAATGGGATAATATCA
CAGGAGGTACTAGACTACCTTTCATCCTACATAAATAGACGCATATAAGT
ACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATATAT
ATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGT
GCGCAGCTCGCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAAACGCTTTGA
AGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCAGAGC
GCTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAA
ACGCACCGGACTGTAACGAGCTACTAAAATATTGCGAATACCGCTTCCAC
AAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCT
ATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGA

CCTCTACATTTTTTATGTTTATCTCTAGTATTACTCTTTAGACAAAAAAATT
GTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACGAAAATGTAAA
CATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAA
TTTTCTGACCAATGAAGAATCATCAACGCTATCACTTTCTGTTCACAAAGT
ATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTTTATCTTG
AAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGGAAGTGG
AGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGCCTTCTT
CTAACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCATTAT
AGAGCGCACAAAGGAGAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAA
ATAGCGCTCTCGGGATGCATTTTTGTAGAACAAAAAAGAAGTATAGATTC
TTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAATGC
AGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTT
TACAAAAATGAAGCACAGATTCTTCGTTGGTAAAATAGCGCTTTCGCGTT
GCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAG
CGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTT
CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTT
TCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAA
ATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCG
TGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACC
CAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACG
AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTAT
GTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGC
CGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA
AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCA
TAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGA
GGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAAC
TCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACG
AGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTA
TTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGG
ATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGC
TGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGATCTCGCG
GTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTT
ATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGA
TCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAA
GTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAA
GGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAAC
GTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGA
TCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAA
AAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTG

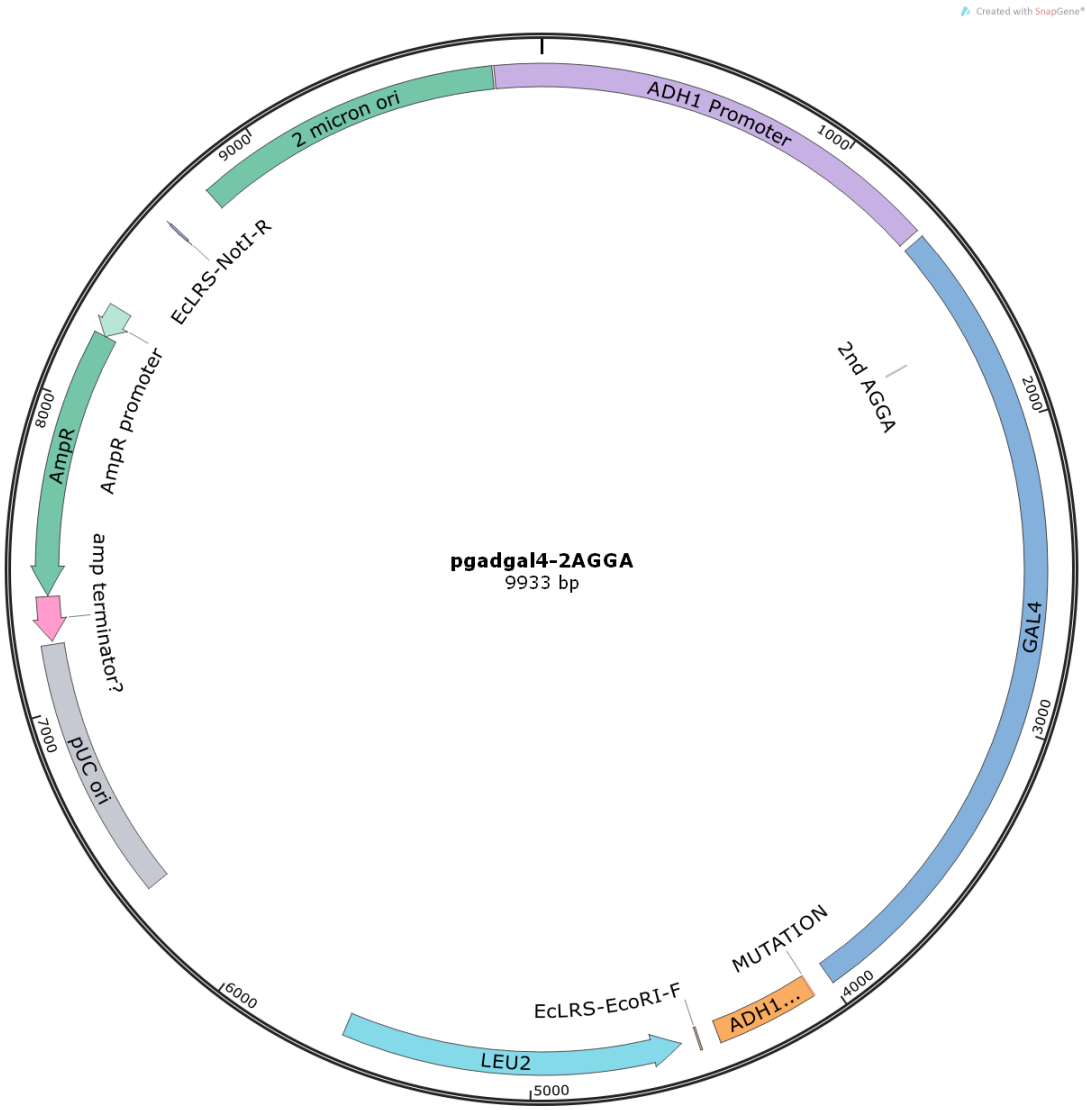
TCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCAC
CGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT
AAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTT
GGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGA
GAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAA
GCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA
CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCG
TCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCA
GCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACAT
GTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTT
GAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGT
CAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGCTGGATCTTCGAGCGTCCCAAAA
CCTTCTCAAGCAAGGTTTTCAGTATAATGTTACATGCGTACACGCGTCTGT
ACAGAAAAAAAAGAAAAATTTGAAATATAAATAACGTTCTTAATACTAAC
ATAACTATAAAAAAATAAATAGGGACCTAGACTTCAGGTTGTCTAACTCC
TTCCTTTTCGGTTAGAGCGGATCTTAGCTAGGAAGTGAATGGAGACATAA
AAAACAAAAAAATGGTGCATCCGGGAGGATTCGAACCTCCGACCGCTCG
GTTAGGAAGCCGAGTACTCTATCCAGCTGAGCTACGGATGCTATTTAATT
GTTGAAGAAAGAGTATACTACATAACACATATACAATTGAAAAAGAGGCT
AGCACCACCACCGGTCGGGATCGAAGAAATGATGGTAAATGAAATAGGA
AATCAAGGAGCATGAAGGCAAAAGACAAATATAAGGGTCGAACGAAAAA
TAAAGTGAAAAGTGTTGATATGATGTATTTGGCTTTGCGGCGCCGAAAAA
ACGAGTTTACGCAATTGCACAATCATGCTGACTCTGTGGCGGACCCGCGC
TCTTGCCGGCCCGGCGATAACGCTGGGCGTGAGGCTGTGCCCGGCGGAGT
TTTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATTG
GAGAAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGAC
AACTGGTGTCATTATTTAAGTTGCCGAAAGAACCTGAGTGCATTTGCAAC
ATGAGTATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCG
GTGGTGCGAACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATA
ACCGCTAGAGTACTTTGAAGAGGAAACAGCAATAGGGTTGCTACCAGTAT
AAATAGACAGGTACATACAACACTGGAAATGGTTGTCTGTTTGAGTACGC
TTTCAATTCATTTGGGTGTGCACTTTATTATGTTACAATATGGAAGGGAAC
TTTACACTTCTCCTATGCACATATATTAATTAAAGTCCAATGCTAGTAGAG
AAGGGGGGTAACACCCCTCCGCGCTCTTTTCCGATTTTTTTCTAAACCGTG
GAATATTTCGGATATCCTTTTGTTGTTTCCGGGTGTACAATATGGACTTCCT
CTTTTCTGGCAACCAAACCCATACATCGGGATTCCTATAATACCTTCGTTA
GTCTCCCTAACATGTAGGTGGCGGAGGGGAGATATACAATAGAACAGATA
CCAGACAAGACATAATGGGCTAAACAAGACTACACCAATTACACTGCCTC
ATTGATGGTGGTACATAACGAACTAATACTGTAGCCCTAGACTTGATAGC

CATCATCATATCGAAGTTTCACTACCCTTTTTCCATTTGCCATCTATTGAAG
TAATAATAGGCGCATGCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCC
GTTGTTGTCTCACCATATCCGCAATGACAAAAAAAATGATGGAAGACACTA
AAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCA
GAGCTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTCCTTCCTTCATT
CACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTTCCAGTTA
CTTGAATTTGAAATAAAAAAAAGTTTGCTGTCTTGCTATCAAGTATAAATA
GACCTGCAATTATTAATCTTTTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCT
TCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATC
AACTGAATTCGCCGCCACCATGAATATTCAGGCTCTTCTCTCAGAAAAAG
TCCGTCAGGCCATGATTGCGGCAGGCGCGCCTGCGGATTGCGAACCGCAG
GTTCGTCAGTCAGCAAAAGTTCAGTTCGGCGACTATCAGGCTAACGGCAT
GATGGCAGTTGCTAAAAAACTGGGTATGGCACCGCGACAATTAGCAGAGC
AGGTGCTGACTCATCTGGATCTTAACGGTATCGCCAGCAAAGTTGAGATC
GCCGGTCCAGGCTTTATCAACATTTTCCTTGATCCGGCATTCCTGGCTGAA
CATGTTCAGCAGGCGCTGGCGTCCGATCGTCTCGGTGTTGCTACGCCAGA
AAAACAGACCATTGTGGTTGACTACTCTGCGCCAAACGTGGCGAAAGAGA
TGCATGTCGGTCACCTGCGCTCTACCATTATTGGTGACGCAGCAGTGCGTA
CTCTGGAGTTCCTCGGTCACAAAGTGATTCGCGCAAACCACGTCGGCGAC
TGGGGCACTCAGTTCGGTATGCTGATTGCATGGCTGGAAAAGCAGCAGCA
GGAAAACGCCGGTGAAATGGAGCTGGCTGACCTTGAAGGTTTCTACCGCG
ATGCGAAAAAGCATTACGATGAAGATGAAGAGTTCGCCGAGCGCGCACG
TAACTACGTGGTAAAACTGCAAAGCGGTGACGAATATTTCCGCGAGATGT
GGCGCAAACTGGTCGACATCACCATGACGCAGAACCAGATCACCTACGAT
CGTCTCAACGTGACGCTGACCCGTGATGACGTGATGGGCGAAAGCCTCTA
CAACCCGATGCTGCCAGGAATTGTGGCGGATCTCAAAGCCAAAGGTCTGG
CAGTAGAAAGCGAAGGGGCGACCGTCGTATTCCTTGATGAGTTTAAAAAC
AAGGAAGGCGAACCGATGGGCGTGATCATTCAGAAGAAAGATGGCGGCT
ATCTCTACACCACCACTGATATCGCCTGTGCGAAATATCGTTATGAAACAC
TGCATGCCGATCGCGTGCTGTATTACATCGACTCCCGTCAGCATCAACACC
TGATGCAGGCATGGGCGATCGTCCGTAAAGCAGGCTATGTACCGGAATCC
GTACCGCTGGAACACCACATGTTCGGCATGATGCTGGGTAAAGACGGCAA
ACCGTTCAAAACCCGCGCGGGTGGTACAGTGAAACTGGCCGATCTGCTGG
ATGAAGCCCTGGAACGTGCACGCCGTCTGGTGGCAGAAAAGAACCCGGA
TATGCCAGCCGACGAGCTGGAAAAACTGGCTAACGCGGTTGGTATTGGTG
CGGTGAAATATGCGGATCTCTCCAAAAACCGCACCACGGACTACATCTTC
GACTGGGACAACATGCTGGCGTTTGAGGGTAATACCGCGCCATACATGCA
GTATGCATACACGCGTGTATTGTCCGTGTTCCGTAAAGCAGAAATTGACG
AAGAGCAACTGGCTGCAGCTCCGGTTATCATCCGTGAAGATCGTGAAGCG
CAACTGGCAGCTCGCCTGCTGCAGTTTGAAGAAACCCTCACCGTGGTTGC
CCGTGAAGGCACGCCGCATGTAATGTGTGCTTACCTGTACGATCTGGCCG

GTCTGTTCTCTGGCTTCTACGAGCACTGCCCGATCCTCAGCGCAGAAAACG
AAGAAGTGCGTAACAGCCGTCTAAAACTGGCACAACTGACGGCGAAGAC
GCTGAAGCTGGGTCTGGATACGCTGGGTATTGAGACTGTAGAGCGTATGT
AAGCGGCCGCACTAGTATCGATGGATTACAAGGATGACGACGATAAGATC
TGAGCTCTTAATTAACAATTCTTCGCCAGAGGTTTGGTCAAGTCTCCAATC
AAGGTTGTCGGCTTGTCTACCTTGCCAGAAATTTACGAAAAGATGGAAAA
GGGTCAAATCGTTGGTAGATACGTTGTTGACACTTCTAAATAAGCGAATTT
CTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGT
ATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATTCTT
GAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGT
CGCTCCAATTCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCT
TCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAG
CGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTA
CACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCT
CGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTT
AGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTA
GGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCC
TTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGG
AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTT
GCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTA
ACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCGGTATTT
TCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAGGCAAGTGCACAAA
CAATACTTAAATAAATACTACTCAGTAATAACCTATTTCTTAGCATTTTTG
ACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTC
CGCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCACCAACA
TTTTCTGGCGTCAGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGCTC
TCTTGCCTTCCAACCCAGTCAGAAATCGAGTTCCAATCCAAAAGTTCACCT
GTCCCACCTGCTTCTGAATCAAACAAGGGAATAAACGAATGAGGTTTCTG
TGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATACGAGTCTTTT
AATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCTCC
ATGCAGTTGGACGATATCAATGCCGTAATCATTGACCAGAGCCAAAACAT
CCTCCTTAGGTTGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCT
GAACTATTTTTATATGCTTTTACAAGACTTGAAATTTTCCTTGCAATAACC
GGGTCAATTGTTCTCTTTCTATTGGGCACACATATAATACCCAGCAAGTCA
GCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGCAAGCCGCA
AACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATAC
TCCAAGCTGCCTTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGTC
ACCAATGCCCTCCCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAAT
TCTTAATCGGCAAAAAAAGAAAAGCTCCGGATCAAGATTGTACGTAAGGT
GACAAGCTATTTTTCAATAAAGAATATCTTCCACTACTGCCATCTGGCGTC
ATAACTGCAAAGTACACATATATTACGATGCTGTCTATTAAATGCTTCCTA

TATTATATATATAGTAATGTCGTTTATGGTGCACTCTCAGTACAATCTGCT
CTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGAC
GCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGT
GACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGA
AACGCGCGA

pGADGAL4 2xAGGA



ACGAGTTTACGCAATTGCACAATCATGCTGACTCTGTGGCGGACCCGCGC
TCTTGCCGGCCCGGCGATAACGCTGGGCGTGAGGCTGTGCCCGGCGGAGT
TTTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATTG
GAGAAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGAC
AACTGGTGTCATTATTTAAGTTGCCGAAAGAACCTGAGTGCATTTGCAAC
ATGAGTATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCG
GTGGTGCGAACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATA
ACCGCTAGAGTACTTTGAAGAGGAAACAGCAATAGGGTTGCTACCAGTAT

AAATAGACAGGTACATACAACACTGGAAATGGTTGTCTGTTTGAGTACGC
TTTCAATTCATTTGGGTGTGCACTTTATTATGTTACAATATGGAAGGGAAC
TTTACACTTCTCCTATGCACATATATTAATTAAAGTCCAATGCTAGTAGAG
AAGGGGGGTAACACCCCTCCGCGCTCTTTTCCGATTTTTTTCTAAACCGTG
GAATATTTCGGATATCCTTTTGTTGTTTCCGGGTGTACAATATGGACTTCCT
CTTTTCTGGCAACCAAACCCATACATCGGGATTCCTATAATACCTTCGTTG
GTCTCCCTAACATGTAGGTGGCGGAGGGGAGATATACAATAGAACAGATA
CCAGACAAGACATAATGGGCTAAACAAGACTACACCAATTACACTGCCTC
ATTGATGGTGGTACATAACGAACTAATACTGTAGCCCTAGACTTGATAGC
CATCATCATATCGAAGTTTCACTACCCTTTTTCCATTTGCCATCTATTGAAG
TAATAATAGGCGCATGCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCC
GTTGTTGTCTCACCATATCCGCAATGACAAAAAAATGATGGAAGACACTA
AAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTGTTCCA
GAGCTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTCCTTCCTTCATT
CACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTTCCAGTTA
CTTGAATTTGAAATAAAAAAAAGTTTGCTGTCTTGCTATCAAGTATAAATA
GACCTGCAATTATTAATCTTTTGTTTCCTCGTCATTGTTCTCGTTCCCTTTCT
TCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATC
AACTCCAAGCTTGAAGCAAGCCTCCTGAAAGATGAAGCTACTGTCTTCTA
TCGAACAAGCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGCTCCAAA
GAAAAACCGAAGTGCGCCAAGTGTCTGAAGAACAACTGGGAGTGTCGCT
ACTCTCCCAAAAGGAAAAAGGTCTCCGCTGACTAGGGCACATCTGACAGA
AGTGGAATCAAGGCTAGAAAGACTGGAACAGCTATTTCTACTGATTTTTC
CTCGAGAAGACCTTGACATGATTTTGAAAATGGATTCTTTACAGGATATA
AAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGATGC
CGTCACAGATAGGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTG
AGACAGCATAGAATAAGTGCGACATCATCATCGGAAGAGAGTAGTAACA
AAGGTCAAAGACAGTTGACTGTATCGATTGACTCGGCAGCTCATCATGAT
AACTCCACAATTCCGTTGGATTTTATGCCCAGGGATGCTCTTCATGGATTT
GATTGGTCTGAAGAGGATGACATGTCGGATGGCTTGCCCTTCCTGAAAAC
GGACCCCAACAATAATGGGTTCTTTGGCGACGGTTCTCTCTTATGTATTCT
TCGATCTATTGGCTTTAAACCGGAAAATTACACGAACTCTAACGTTAACA
GGCTCCCGACCATGATTACGGATAGATACACGTTGGCTTCTAGATCCACA
ACATCCCGTTTACTTCAAAGTTATCTCAATAATTTTCACCCCTACTGCCCT
ATCGTGCACTCACCGACGCTAATGATGTTGTATAATAACCAGATTGAAAT
CGCGTCGAAGGATCAATGGCAAATCCTTTTTAACTGCATATTAGCCATTGG
AGCCTGGTGTATAGAGGGGGAATCTACTGATATAGATGTTTTTTACTATCA
AAATGCTAAATCTCATTTGACGAGCAAGGTCTTCGAGTCAGGTTCCATAA
TTTTGGTGACAGCCCTACATCTTCTGTCGCGATATACACAGTGGAGGCAG
AAAACAAATACTAGCTATAATTTTCACAGCTTTTCCATAAGAATGGCCAT
ATCATTGGGCTTGAATAGGGACCTCCCCTCGTCCTTCAGTGATAGCAGCAT

TCTGGAACAAAGACGCCGAATTTGGTGGTCTGTCTACTCTTGGGAGATCC
AATTGTCCCTGCTTTATGGTCGATCCATCCAGCTTTCTCAGAATACAATCT
CCTTCCCTTCTTCTGTCGACGATGTGCAGCGTACCACAACAGGTCCCACCA
TATATCATGGCATCATTGAAACAGCAAGGCTCTTACAAGTTTTCACAAAA
ATCTATGAACTAGACAAAACAGTAACTGCAGAAAAAAGTCCTATATGTGC
AAAAAAATGCTTGATGATTTGTAATGAGATTGAGGAGGTTTCGAGACAGG
CACCAAAGTTTTTACAAATGGATATTTCCACCACCGCTCTAACCAATTTGT
TGAAGGAACACCCTTGGCTATCCTTTACAAGATTCGAACTGAAGTGGAAA
CAGTTGTCTCTTATCATTTATGTATTAAGAGATTTTTTCACTAATTTTACCC
AGAAAAAGTCACAACTAGAACAGGATCAAAATGATCATCAAAGTTATGA
AGTTAAACGATGCTCCATCATGTTAAGCGATGCAGCACAAAGAACTGTTA
TGTCTGTAAGTAGCTATATGGACAATCATAATGTCACCCCATATTTTGCCT
GGAATTGTTCTTATTACTTGTTCAATGCAGTCCTAGTACCCATAAAGACTC
TACTCTCAAACTCAAAATCGAATGCTGAGAATAACGAGACCGCACAATTA
TTACAACAAATTAACACTGTTCTGATGCTATTAAAAAAACTGGCCACTTTT
AAAATCCAGACTTGTGAAAAATACATTCAAGTACTGGAAGAGGTATGTGC
GCCGTTTCTGTTATCACAGTGTGCAATCCCATTACCGCATATCAGTTATAA
CAATAGTAATGGTAGCGCCATTAAAAATATTGTCGGTTCTGCAACTATCG
CCCAATACCCTACTCTTCCGGAGGAAAATGTCAACAATATCAGTGTTAAA
TATGTTTCTCCTGGCTCAGTAGGGCCTTCACCTGTGCCATTGAAATCAGGA
GCAAGTTTCAGTGATCTAGTCAAGCTGCTATCTAACCGTCCACCCTCTCGT
AACTCTCCAGTGACAATACCAAGAAGCACACCTTCGCATCGCTCAGTCAC
GCCTTTTCTAGGGCAACAGCAACAGCTGCAATCATTAGTGCCACTGACCC
CGTCTGCTTTGTTTGGTGGCGCCAATTTTAATCAAAGTGGGAATATTGCTG
ATAGCTCATTGTCCTTCACTTTCACTAACAGTAGCAACGGTCCGAACCTCA
TAACAACTCAAACAAATTCTCAAGCGCTTTCACAACCAATTGCCTCCTCTA
ACGTTCATGATAACTTCATGAATAATGAAATCACGGCTAGTAAAATTGAT
GATGGTAATAATTCAAAACCACTGTCACCTGGTTGGACGGACCAAACTGC
GTATAACGCGTTTGGAATCACTACAGGGATGTTTAATACCACTACAATGG
ATGATGTATATAACTATCTATTCGATGATGAAGATaCCCCACCAAACCCAA
AAAAAGAGTAAGCTAGCAAGGCCTTGTAAGGGCGAATTCCAGCACACTG
GCGGCCGTTACTAGTGGATCCGAGCTCGGTACCCCAAGCTTTGGACTTCTT
CGCCAGAGGTTTGGTCAAGTCTCCAATCAAGGTTGTCGGCTTGTCTACCTT
GCCAGAAATTTACGAAAAGATGGAAAAGGGTCAAATCGTTGGTAGATAC
GTTGTTGACACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTTATTA
TTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCT
TAGGTTTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTC
AGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGACCACACCTCT
ACCGGCCGGTCGAAATTCCCCTACCCTATGAACATATTCCATTTTGTAATT
TCGTGTCGTTTCTATTATGAATTTCATTTATAAAGTTTATGTACAAATATCA
TAAAAAAAGAGAATCTTTTTAAGCAAGGATTTTCTTAACTTCTTCGGCGAC

AGCATCACCGACTTCGGTGGTACTGTTGGAACCACCTAAATCACCAGTTCT
GATACCTGCATCCAAAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTC
TTCAGGCAAGTTCAATGACAATTTCAACATCATTGCAGCAGACAAGATAG
TGGCGATAGGGTTGACCTTATTCTTTGGCAAATCTGGAGCAGAACCGTGG
CATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGCCAA
GGACGCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCA
TCGGAGATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGG
TGGGTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATG
TTGAACCTTCAATGTAGGAAATTCGTTCTTGATGGTTTCCTCCACAGTTTTT
CTCCATAATCTTGAAGAGGCCAAAACATTAGCTTTATCCAAGGACCAAAT
AGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGGCCATTCTTGTGA
TTCTTTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCAT
CACCATCGTCTTCCTTTCTCTTACCAAAGTAAATACCTCCCACTAATTCTCT
GACAACAACGAAGTCAGTACCTTTAGCAAATTGTGGCTTGATTGGAGATA
AGTCTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAGTTGGCGTAC
AATTGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAGGTCTAACACTA
CCTGTACCCCATTTAGGACCACCCACAGCACCTAACAAAACGGCATCAGC
CTTCTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCTTC
GATAGCAGCACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGG
AACGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATT
TCTTGACCAACGTGGTCACCTGGCAAAACGACGATCTTCTTAGGGGCAGA
CATTAGAATGGTATATCCTTGAAATATATATATATATTGCTGAAATGTAAA
AGGTAAGAAAAGTTAGAAAGTAAGACGATTGCTAACCACCTATTGGAAA
AAACAATAGGTCCTTAAATAATATTGTCAACTTCAAGTATTGTGATGCAA
GCATTTAGTCATGAACGCTTCTCTATTCTATATGAAAAGCCGGTTCCGGCG
CTCTCACCTTTCCTTTTTCTCCCAATTTTTCAGTTGAAAAAGGTATATGCGT
CAGGCGACCTCTGAAATTAACAAAAAATTTCCAGTCATCGAATTTGATTCT
GTGCGATAGCGCCCCTGTGTGTTCTCGTTATGTTGAGGAAAAAAATAATG
GTTGCTAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTCCCAC
AGTTGGGGGATCTCGACTCTAGCTAGAGGATCAATTCGTAATCATGTCAT
AGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC
GAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTA
ACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCT
GTCGTGCCAGCTGATAACTTCGTATAATGTATGCTATACGAAGTTATTAGG
TCTGAAGAGGAGTTTACGTCCAGCCAAGCTAGCTTGGCTGCAGGTCGAGC
GGCCGCGATCCGGAACCCTTAATATAACTTCGTATAATGTATGCTATACG
AAGTTATCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTT
TGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGG
TCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGG
TTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAG
GCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTC

CATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCA
GAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG
GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCT
GTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGC
ACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCA
GCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTT
CTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTG
GTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAA
TGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGT
TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGT
TCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGA
GGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCT
CACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAG
CGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGT
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGT
TGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGC
TTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCAT
GTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAA
GTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATT
CTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACT
CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC
CCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT
GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTT
CAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGG
CAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC
TCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCT
CATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGG
TTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATT
ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGT
CTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCC
GGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCC
CGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTA
TGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATAACGCATTTAAGC
ATAAACACGCACTATGCCGTTCTTCTCATGTATATATATATACAGGCAACA

CGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGTGCGCAGCTCGCG
TTGCATTTTCGGAAGCGCTCGTTTTCGGAAACGCTTTGAAGTTCCTATTCC
GAAGTTCCTATTCTCTAGCTAGAAAGTATAGGAACTTCAGAGCGCTTTTGA
AAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAAACGCACC
GGACTGTAACGAGCTACTAAAATATTGCGAATACCGCTTCCACAAACATT
GCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCTATATAAC
CTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGACCTCTAC
ATTTTTTATGTTTATCTCTAGTATTACTCTTTAGACAAAAAAATTGTAGTAA
GAACTATTCATAGAGTGAATCGAAAACAATACGAAAATGTAAACATTTCC
TATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAATTTTCTG
ACCAATGAAGAATCATCAACGCTATCACTTTCTGTTCACAAAGTATGCGC
AATCCACATCGGTATAGAATATAATCGGGGATGCCTTTATCTTGAAAAAA
TGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGGAAGTGGAGTCAGG
CTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGCCTTCTTCTAACCTT
AACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCATTATAGAGCGCA
CAAAGGAGAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAAATAGCGCTC
TCGGGATGCATTTTTGTAGAACAAAAAAGAAGTATAGATTCTTTGTTGGTA
AAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATT
CTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTTTACAAAAAT
GAAGCACAGATTCTTCGTTGGTAAAATAGCGCTTTCGCGTTGCATTTCTGT
TCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCG
TTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTTGCTTGCATGC
CTGCAGGTCGAGATCCGGGATCGAAGAAATGATGGTAAATGAAATAGGA
AATCAAGGAGCATGAAGGCAAAAGACAAATATAAGGGTCGAACGAAAAA
TAAAGTGAAAAGTGTTGATATGATGTATTTGGCTTTGCGGCGCCGAAAAA

PGADyEGFP AGGA

TATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGC
CACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAA
AGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGG
TTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAG
AAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACT
CACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAG
ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAG
TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCA
GCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAG
ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGAT
ACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGC
CAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCC
ATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT

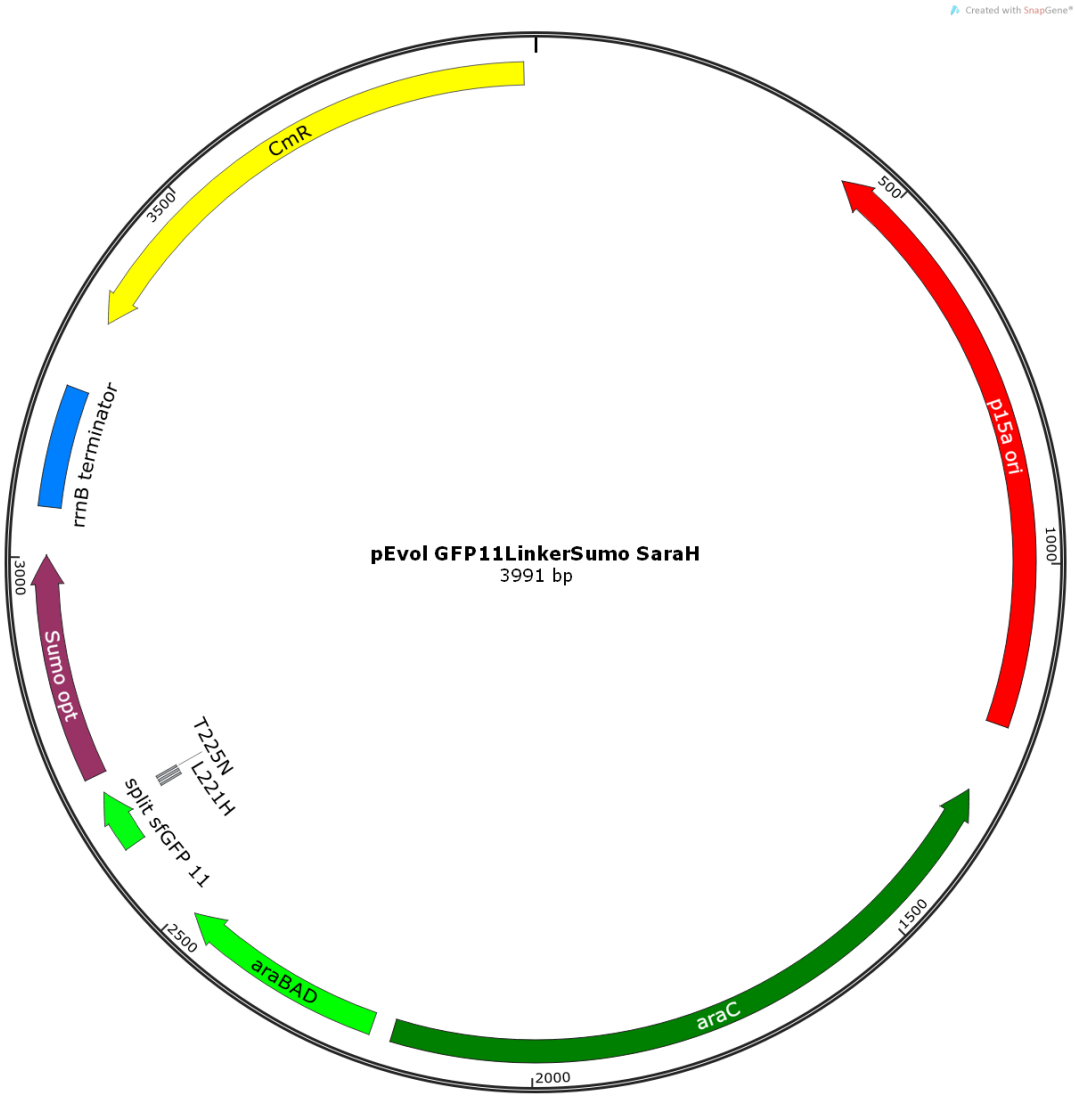
AATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGC
TCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGA
GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCT
CCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATG
GCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCT
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCG
ACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATA
GCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAA
CTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGT
GCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGA
GCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACA
CGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTT
ATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAA
ATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
GTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTAT
CACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTG
ACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCG
GGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCG
GGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACC
ATAACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATA
TATATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTA
TGTGCGCAGCTCGCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAAACGCTT
TGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGCTAGAAAGTATAGGAACT
TCAGAGCGCTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAA
ACCAAAAAACGCACCGGACTGTAACGAGCTACTAAAATATTGCGAATACCG
CTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTA
TATCCCTATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTA
AACTCGACCTCTACATTTTTTATGTTTATCTCTAGTATTACTCTTTAGACAA
AAAAATTGTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACGAAA
ATGTAAACATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCG
TTCATAATTTTCTGACCAATGAAGAATCATCAACGCTATCACTTTCTGTTC
ACAAAGTATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTT
TATCTTGAAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGG
AAGTGGAGTCAGGCTTTTTTTATGGAAGAGAAAATAGACACCAAAGTAGC
CTTCTTCTAACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTG
CATTATAGAGCGCACAAAGGAGAAAAAAAAGTAATCTAAGATGCTTTGTTA
GAAAAATAGCGCTCTCGGGATGCATTTTTGTAGAACAAAAAAGAAGTATA
GATTCTTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAA
AATGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTT
TTGTTTTACAAAAATGAAGCACAGATTCTTCGTTGGTAAAATAGCGCTTTC
GCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAA

ATTAGCGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCT
TCGTTGCTTGCATGCAGTTTATCATTATCAATACTCGCCATTTCAAAGAAT
ACGTAAATAATTAATAGTAGTGATTTTCCTAACTTTATTTAGTCAAAAAAT
TAGCCTTTTAATTCTGCTGTAACCCGTACATGCCCAAAATAGGGGGCGGG
TTACACAGAATATATAACATCGTAGGTGTCTGGGTGAACAGTTTATTCCTG
GCATCCACTAAATATAATGGAGCCCGCTTTTTAAGCTGGCATCCAGAAAA
AAAAAGAATCCCAGCACCAAAATATTGTTTTCTTCACCAACCATCAGTTC
ATAGGTCCATTCTCTTAGCGCAACTACAGAGAACAGGGGCACAAACAGGC
AAAAAACGGGCACAACCTCAATGGAGTGATGCAACCTGCCTGGAGTAAA
TGATGACACAAGGCAATTGACCCACGCATGTATCTATCTCATTTTCTTACA
CCTTCTATTACCTTCTGCTCTCTCTGATTTGGAAAAAGCTGAAAAAAAAGG
TTGAAACCAGTTCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAA
AGACGGTAGGTATTGATTGTAATTCTGTAAATCTATTTCTTAAACTTCTTA
AATTCTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACCAGAACTTA
GTTTCGACGGATTCTAGAGAATTCCAGCACACTGGCGGCCGTTACTAGTA
TGTCTAAAAGGAGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGTTGAA
TTAGATGGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAA
GGTGATGCTACTTACGGTAAATTGACCTTAAAATTTATTTGTACTACTGGT
AAATTGCCAGTTCCATGGCCAACCTTAGTCACTACTTTCGGTTATGGTGTT
CAATGTTTTGCTAGATACCCAGATCATATGAAACAACATGACTTTTTCAAG
TCTGCCATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTCAAAGAT
GACGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACCTT
AGTTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACA
TTTTAGGTCACAAATTGGAATACAACTATAACTCTCACAATGTTTACATCA
TGGCTGACAAACAAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACAC
AACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACT
CCAATTGGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACT
CAATCTGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCTT
GTTAGAATTTGTTACTGCTGCTGGTATTACCCATGGTATGGATGAATTGTA
CAAACACCATCACCATCACCATTAAGGTACCAAGCTTTGGACTTCTTCGCC
AGAGGTTTGGTCAAGTCTCCAATCAAGGTTGTCGGCTTGTCTACCTTGCCA
GAAATTTACGAAAAGATGGAAAAGGGTCAAATCGTTGGTAGATACGTTGT
TGACACTTCTAAATAAGCGAATTTCTTATGATTTATGATTTTTATTATTAAA
TAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGT
TTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTT
GCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGACCACACCTCTACCGG
CCGGTCGAAATTCCCCTACCCTATGAACATATTCCATTTTGTAATTTCGTG
TCGTTTCTATTATGAATTTCATTTATAAAGTTTATGTACAAATATCATAAA
AAAAGAGAATCTTTTTAAGCAAGGATTTTCTTAACTTCTTCGGCGACAGCA
TCACCGACTTCGGTGGTACTGTTGGAACCACCTAAATCACCAGTTCTGATA
CCTGCATCCAAAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCA

GGCAAGTTCAATGACAATTTCAACATCATTGCAGCAGACAAGATAGTGGC
GATAGGGTTGACCTTATTCTTTGGCAAATCTGGAGCAGAACCGTGGCATG
GTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGCCAAGGAC
GCAGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATCGG
AGATGATATCACCAAACATGTTGCTGGTGATTATAATACCATTTAGGTGG
GTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTG
AACCTTCAATGTAGGAAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTC
CATAATCTTGAAGAGGCCAAAACATTAGCTTTATCCAAGGACCAAATAGG
CAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGGCCATTCTTGTGATTCT
TTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACC
ATCGTCTTCCTTTCTCTTACCAAAGTAAATACCTCCCACTAATTCTCTGAC
AACAACGAAGTCAGTACCTTTAGCAAATTGTGGCTTGATTGGAGATAAGT
CTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAGTTGGCGTACAAT
TGAAGTTCTTTACGGATTTTTAGTAAACCTTGTTCAGGTCTAACACTACCT
GTACCCCATTTAGGACCACCCACAGCACCTAACAAAACGGCATCAGCCTT
CTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCTTCGA
TAGCAGCACCACCAATTAAATGATTTTCGAAATCGAACTTGACATTGGAA
CGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCT
TGACCAACGTGGTCACCTGGCAAAACGACGATCTTCTTAGGGGCAGACAT
TAGAATGGTATATCCTTGAAATATATATATATATTGCTGAAATGTAAAAG
GTAAGAAAAGTTAGAAAGTAAGACGATTGCTAACCACCTATTGGAAAAA
ACAATAGGTCCTTAAATAATATTGTCAACTTCAAGTATTGTGATGCAAGCA
TTTAGTCATGAACGCTTCTCTATTCTATATGAAAAGCCGGTTCCGGCGCTC
TCACCTTTCCTTTTTCTCCCAATTTTTCAGTTGAAAAAGGTATATGCGTCAG
GCGACCTCTGAAATTAACAAAAAATTTCCAGTCATCGAATTTGATTCTGTG
CGATAGCGCCCCTGTGTGTTCTCGTTATGTTGAGGAAAAAAATAATGGTTG
CTAAGAGATTCGAACTCTTGCATCTTACGATACCTGAGTATTCCCACAGTT
GGGGGATCTCGACTCTAGCTAGAGGATCAATTCGTAATCATGTCATAGCT
GTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGC
CGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCA
CATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
GCCAGCTGATAACTTCGTATAATGTATGCTATACGAAGTTATTAGGTCTGA
AGAGGAGTTTACGTCCAGCCAAGCTAGCTTGGCTGCAGGTCGAGCGGCCG
CGATCCGGAACCCTTAATATAACTTCGTATAATGTATGCTATACGAAGTTA
TCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTA
TTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTC
GGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCC
ACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGC
AAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGG
CTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTG
GCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCT

CCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG
CCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGT
ATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAAC
CCCCCGTTCAGCCCGACCGCTGCGCCT

pEvol SUMOLinkerGFP11

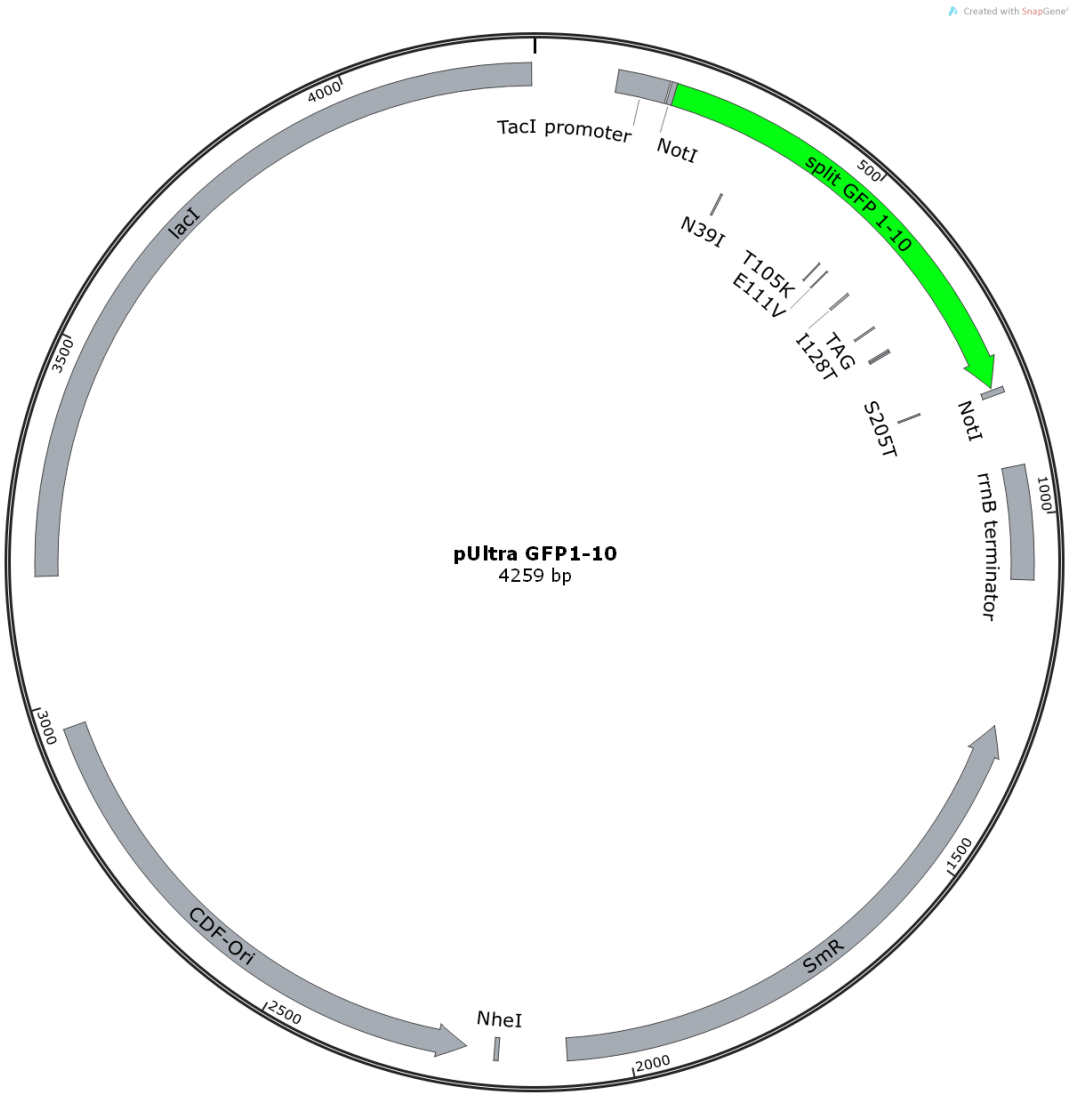


TCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTC
ATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCG
CCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTA
TTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTG
CGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGA
GGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTG
ACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCG
GAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGT
GCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATAT

GTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGT
CGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCC
TGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCA
AAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTG
ACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGT
TTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTG
ACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGA
ACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGT
AATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAA
CTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGG
TTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGC
GGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTC
AAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAAT
TTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGT
AATTCTCATGTTTGACAGCTTATCATCGATAAGCTTGGTACCCAATTATGA
CAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTC
GCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTT
GATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTG
GTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTT
AAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACG
GCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCT
GCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCAT
CGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATT
GCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCC
CGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCT
TCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAG
CCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACC
ATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGG
AACAGCAAAATATCACTCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTC
ACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCC
CAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCG
TTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGA
TCATTTTGCGCTTCAGCCATACTTTTCATACTCCCGCCATTCAGAGAAGAA
ACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGG
CTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAA
CAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATA
ATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCT
ATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTT
TATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAAAGAAA

TAATTTTGTTTAACTTTAAGAAGGAGAATACATCAACTAGTACGCAAGTTC
ACGTAAAAAGGGTATCTAGAGGTTGAGGTGATTTTATGGTCCTTTCGAAA
GATCCCAACGAAAAGCGTGACCACATGGTCCTTCATGAGTACGTAACTGC
TGCTGGGATTACAGGTGGAGGCTCTGGCGGAGGTTCGTCATTACAGGATT
CAGAAGTCAATCAAGAAGCGAAGCCCGAGGTGAAACCTGAAGTAAAACC
GGAAACACACATTAACCTGAAGGTTTCGGATGGATCGAGCGAAATTTTCT
TCAAGATTAAAAAGACTACTCCGCTTCGCCGTCTGATGGAAGCGTTCGCA
AAGCGCCAAGGTAAGGAAATGGACTCATTGACATTTTTGTACGACGGCAT
TGAAATCCAGGCGGACCAAACCCCTGAAGACTTAGATATGGAGGATAATG
ACATTATTGAGGCCCATCGCGAACAGATTGGGGGGCACCACCACCACCAC
CACTAAACCGATGCGGCCGCTTGAGAGTCAGCTCCTTCCGGTGGGCGTGC
CTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAA
GTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGT
AGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTG
GGCCTTGTCGACCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCA
GGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATT
ACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTC
TGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGC
GGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACG
GGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAA
ACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTT
TAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATAT
ATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGA
AAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTAT
CCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAG
CATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGC
TTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTC
TGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTA
CGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCC
ATTTTAGCTTCCTTAGC

pUltra sGFP1-10

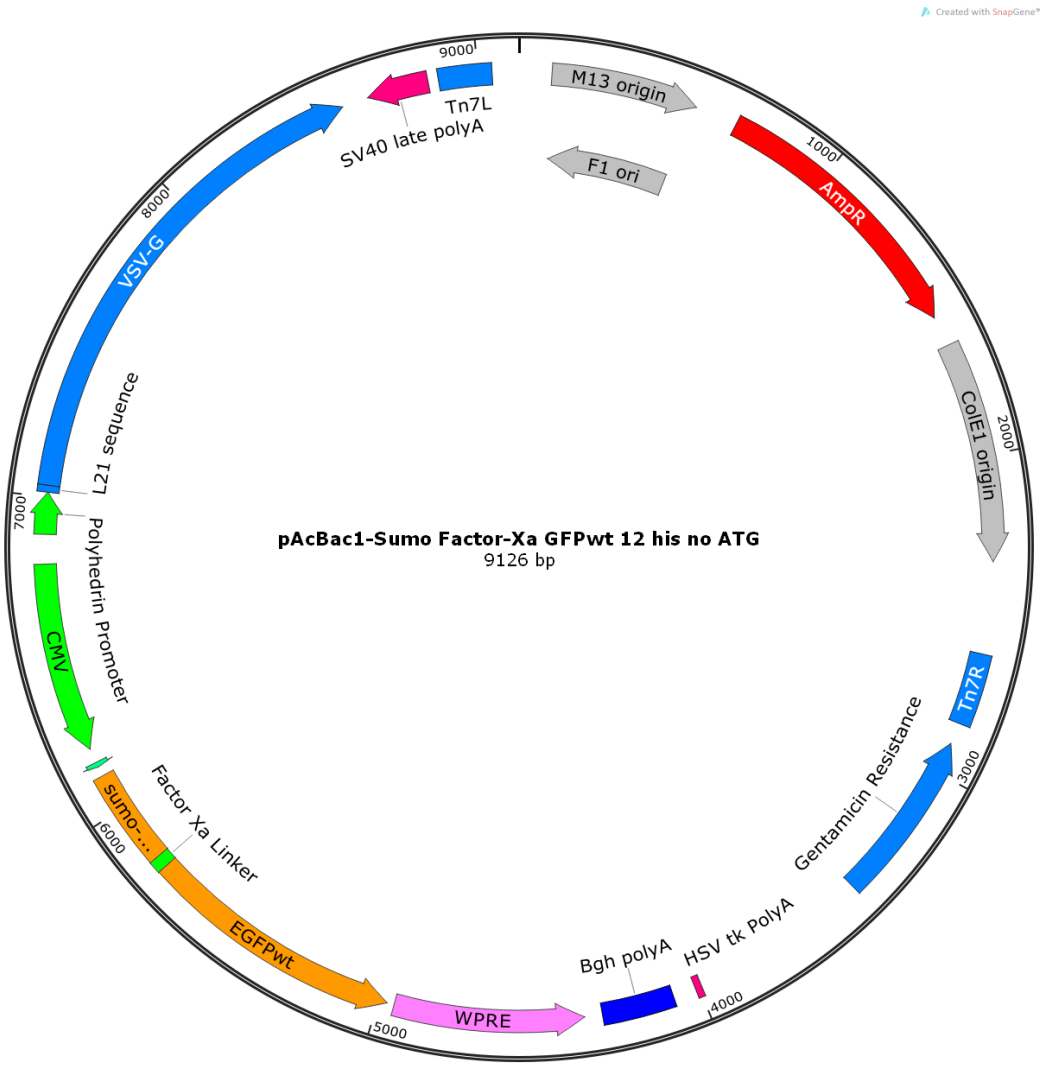


ACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCG
ACTCCTGCATTAGGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTG
TGGAATTGTGAGCGGATAACAATTTCACAAAGGAGGTGCGGCCGCATGGC
TTATAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGA
ATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTG
AAGGTGATGCTACAATAGGAAAACTCACCCTTAAATTTATTTGCACTACT
GGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGT
GTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTC
AAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAA
AGATGACGGGAAGTACAAGACGCGTGCTGTTGTCAAGTTTGAAGGTGATA
CCCTTGTTAATCGTATCGAGTTAAAGGGTACCGATTTTAAAGAAGATGGA
AACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATA
CATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCACAGTC
CGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACA

AAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCT
GTCGACACAAACGTAAGCGGCCGCGTTTAAACGGTCTCCAGCTTGGCTGT
TTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACG
CAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGG
TCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGAT
GGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAA
TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTGTTTGTGAGCTCCCAT
GGCGGCCACCAGGTACCACCGGCGCCTCAGGCATTTGAGAAGCACACGGT
CACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAG
CGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCATCGTGGCCG
GATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTTGCCGACT
ACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACTGATCT
GCGCGCGAGGCCAAGCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCT
TCAAGTATGACGGGCTGATACTGGGCCGGCAGGCGCTCCATTGCCCAGTC
GGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGTACCAAAT
GCGGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCGGGCG
GCGAGTTCCATAGCGTTAAGGTTTCATTTAGCGCCTCAAATAGATCCTGTT
CAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACCAAGGCAACG
CTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCGATCGTG
GCTGGCTCGAAGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAA
TTGCAGTTCGCGCTTAGCTGGATAACGCCACGGAATGATGTCGTCGTGCA
CAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAA
GCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCA
AGCCTTACGGTCACCGTAACCAGCAAATCAATATCACTGTGTGGCTTCAG
GCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAGCAACGTCGGTT
CGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACT
TCGGCGATCACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAGCA
TTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGA
AAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGTGAGA
CTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGA
TAAGCAGGGGACTAACATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGG
CGTTTTTCCATAGGCTCCGCCCTCCTGCCAGAGTTCACATAAACAGACGCT
TTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGT
ACGGGCGAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTC
GCTCCCTCTTGCGCTCTCCTGTTCCGACCCTGCCGTTTACCGGATACCTGTT
CCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTCACACACT
GGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAA
GAACTCCCCGTTCAGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTT
GAGTCCAACCCGGAAAAGCACGGTAAAACGCCACTGGCAGCAGCCATTG
GTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTT
GAAGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTG

TGCTCTGCGAAAGCCAGTTACCACGGTTAAGCAGTTCCCCAACTGACTTA
ACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTACAGGGCAAAA
GATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
CTGAACCGCTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTG
AAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGCCCCG
CGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTC
GAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCT
CACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA
ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTG
GTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCC
TGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAG
GCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGT
CTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCAGC
CCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGC
AACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTT
GTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCT
GAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACC
CAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGA
AAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCC
GGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAG
CGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCA
CCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCA
CGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGC
GACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACG
ACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCT
CCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGG
CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCT
GCGACATCGTATAACGTTACTGGTTTCACATTC

pAcBac1-SUMO-FxaGFPWT 12xHis



TTCTCTGTCACAGAATGAAAATTTTTCTGTCATCTCTTCGTTATTAATGTTT
GTAATTGACTGAATATCAACGCTTATTTGCAGCCTGAATGGCGAATGGGA
CGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA
GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCT
TCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCG
GGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAA
AAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGA
CGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTT
GTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTT
ATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAG
GTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCT
AAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATG
CTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGT
CGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCA
GAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAG

TGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTT
CGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGT
GGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCG
CATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAA
AGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGG
ACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTC
GCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAG
CGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGA
TGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCT
GGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGG
TATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTAT
CTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATC
GCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGT
TTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGG
ATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGT
GAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATC
TTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTC
TTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTC
CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCG
CCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA
GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGG
AGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCATTGAGA
AAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGC
GGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACG
CCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTC
GATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGC
AACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGT
TCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGA
GTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCA
GTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCA
TCTGTGCGGTATTTCACACCGCAGACCAGCCGCGTAACCTGGCAAAATCG
GTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAGCGGGTGTGGGCGGA
CAATAAAGTCTTAAACTGAACAAAATAGATCTAAACTATGACAATAAAGT
CTTAAACTAGACAGAATAGTTGTAAACTGAAATCAGTCCAGTTATGCTGT
GAAAAAGCATACTGGACTTTTGTTATGGCTAAAGCAAACTCTTCATTTTCT
GAAGTGCAAATTGCCCGTCGTATTAAAGAGGGGCGTGGCCAAGGGCATG
GTAAAGACTATATTCGCGGCGTTGTGACAATTTACCGAACAACTCCGCGG

CCGGGAAGCCGATCTCGGCTTGAACGAATTGTTAGGTGGCGGTACTTGGG
TCGATATCAAAGTGCATCACTTCTTCCCGTATGCCCAACTTTGTATAGAGA
GCCACTGCGGGATCGTCACCGTAATCTGCTTGCACGTAGATCACATAAGC
ACCAAGCGCGTTGGCCTCATGCTTGAGGAGATTGATGAGCGCGGTGGCAA
TGCCCTGCCTCCGGTGCTCGCCGGAGACTGCGAGATCATAGATATAGATC
TCACTACGCGGCTGCTCAAACCTGGGCAGAACGTAAGCCGCGAGAGCGCC
AACAACCGCTTCTTGGTCGAAGGCAGCAAGCGCGATGAATGTCTTACTAC
GGAGCAAGTTCCCGAGGTAATCGGAGTCCGGCTGATGTTGGGAGTAGGTG
GCTACGTCTCCGAACTCACGACCGAAAAGATCAAGAGCAGCCCGCATGGA
TTTGACTTGGTCAGGGCCGAGCCTACATGTGCGAATGATGCCCATACTTG
AGCCACCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCGTTGC
TGCTGCGTAACATCGTTGCTGCTCCATAACATCAAACATCGACCCACGGC
GTAACGCGCTTGCTGCTTGGATGCCCGAGGCATAGACTGTACAAAAAAAC
AGTCATAACAAGCCATGAAAACCGCCACTGCGCCGTTACCACCGCTGCGT
TCGGTCAAGGTTCTGGACCAGTTGCGTGAGCGCATACGCTACTTGCATTAC
AGTTTACGAACCGAACAGGCTTATGTCAACTGGGTTCGTGCCTTCATCCGT
TTCCACGGTGTGCGTCACCCGGCAACCTTGGGCAGCAGCGAAGTCGAGGC
ATTTCTGTCCTGGCTGGCGAACGAGCGCAAGGTTTCGGTCTCCACGCATCG
TCAGGCATTGGCGGCCTTGCTGTTCTTCTACGGCAAGGTGCTGTGCACGGA
TCTGCCCTGGCTTCAGGAGATCGGTAGACCTCGGCCGTCGCGGCGCTTGC
CGGTGGTGCTGACCCCGGATGAAGTGGTTCGCATCCTCGGTTTTCTGGAA
GGCGAGCATCGTTTGTTCGCCCAGGACTCTAGCTATAGTTCTAGTGGTTGG
CCTACGTACCCGTAGTGGCTATGGCAGGGCTTGCGCTTAATGCGCCGCTA
CAGGGCGCGTGGGGATACCCCCTAGAGCCCCAGCTGGTTCTTTCCGCCTC
AGAAGCCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCC
AATCCTCCCCCTTGCTGTCCTGCCCCACCCCACCCCCCAGAATAGAATGAC
ACCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTAGGAAAGGACA
GTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAA
CAACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGT
TTAAACGGGCCCTCTAGACTCGAGTTAAAGTCGACGCGGGGAGGCGGCCC
AAAGGGAGATCCGACTCGTCTGAGGGCGAAGGCGAAGACGCGGAAGAGG
CCGCAGAGCCGGCAGCAGGCCGCGGGAAGGAAGGTCCGCTGGATTGAGG
GCCGAAGGGACGTAGCAGAAGGACGTCCCGCGCAGAATCCAGGTGGCAA
CACAGGCGAGCAGCCAAGGAAAGGACGATGATTTCCCCGACAACACCAC
GGAATTGTCAGTGCCCAACAGCCGAGCCCCTGTCCAGCAGCGGGCAAGGC
AGGCGGCGATGAGTTCCGCCGTGGCAATAGGGAGGGGGAAAGCGAAAGT
CCCGGAAAGGAGCTGACAGGTGGTGGCAATGCCCCAACCAGTGGGGGTT
GCGTCAGCAAACACAGTGCACACCACGCCACGTTGCCTGACAACGGGCCA
CAACTCCTCATAAAGAGACAGCAACCAGGATTTATACAAGGAGGAGAAA
ATGAAAGCCATACGGGAAGCAATAGCATGATACAAAGGCATTAAAGCAG
CGTATCCACATAGCGTAAAAGGAGCAACATAGTTAAGAATACCAGTCAAT

CTTTCACAAATTTTGTAATCCAGAGGTTGATTGTCGACTTAACGCGTTGAA
TTCTTAATGGTGATGGTGATGATGATGGTGATGGTGATGATGACCGGTAT
GCATATTCAGATCCTCTTCTGAGATGAGTTTTTGTTCGAAGGGCCCCTTGT
ACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGC
AGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTG
GGTGCTCAGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGG
GGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTCGATGT
TGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCA
TGATATAGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGA
TGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTTCACCA
GGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGT
CCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGAC
TTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTG
CACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGCACGGGCAGC
TTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCGTAGGTGGCATC
GCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCCGTC
CAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGC
TCACTCCGCCGCCTCCTCAGCCATCAATACCTCCACCTCCACCTCCAATCT
GTTCTCTGTGAGCCTCAATAATATCGTTATCCTCCATGTCCAAATCTTCAG
GGGTCTGATCAGCTTGAATTTCAATACCGTCGTACAAGAACGTTAAGGAG
TCCATTTCCTTACCCTGTCTTTTAGCGAACGCTTCCATCAGCCTTCTTAAAG
GAGTGGTCTTTTTGATCTTGAAGAAGATCTCTGAAGATCCATCGGACACCT
TTAAATTGATGTGAGTCTCAGGCTTGACTTCTGGCTTGACCTCTGGCTTAG
CTTCTTGATTGACTTCTGAGTCCTGCAGGGACATGGTGGCGGCGCTAGCCA
GCTTGGGTCTCCCTATAGTGAGTCGTATTAATTTCGATAAGCCAGTAAGCA
GTGGGTTCTCTAGTTAGCCAGAGAGCTCTGCTTATATAGACCTCCCACCGT
ACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTT
GGAAAGTCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTGACGTCAAT
GGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCAT
TGATGTACTGCCAAAACCGCATCACCATGGTAATAGCGATGACTAATACG
TAGATGTACTGCCAAGTAGGAAAGTCCCATAAGGTCATGTACTGGGCATA
ATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTG
GCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATAG
TCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAAC
ATACGTCATTATTGACGTCAATGGGCGGGGGTCGTTGGGCGGTCAGCCAG
GCGGGCCATTTACCGTAAGTTATGTAACGCGGAACTCCATATATGGGCTA
TGAACTAATGACCCCGTAATTGATTACTATTAATAACTAGTCAATAATCAA
TGTCAACGCGTATATCTGGCCCGTACATCGCGAAGCAGCGCAAAACGGAT
CCTGCAGGTATTTGCGGCCGCGGTCCGTATACTCCGGAATATTAATAGATC
ATGGAGATAATTAAAATGATAACCATCTCGCAAATAAATAAGTATTTTAC
TGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAATATTCCGGATTAT

TCATACCGTCCCACCATCGGGCGCGAACTCCTAAAAAACCGCCACCATGA
AGTGCCTTTTGTACTTAGCCTTTTTATTCATTGGGGTGAATTGCAAGTTCAC
CATAGTTTTTCCACACAACCAAAAAGGAAACTGGAAAAATGTTCCTTCTA
ATTACCATTATTGCCCGTCAAGCTCAGATTTAAATTGGCATAATGACTTAA
TAGGCACAGCCTTACAAGTCAAAATGCCCAAGAGTCACAAGGCTATTCAA
GCAGACGGTTGGATGTGTCATGCTTCCAAATGGGTCACTACTTGTGATTTC
CGCTGGTATGGACCGAAGTATATAACACATTCCATCCGATCCTTCACTCCA
TCTGTAGAACAATGCAAGGAAAGCATTGAACAAACGAAACAAGGAACTT
GGCTGAATCCAGGCTTCCCTCCTCAAAGTTGTGGATATGCAACTGTGACG
GATGCCGAAGCAGTGATTGTCCAGGTGACTCCTCACCATGTGCTGGTTGA
TGAATACACAGGAGAATGGGTTGATTCACAGTTCATCAACGGAAAATGCA
GCAATTACATATGCCCCACTGTCCATAACTCTACAACCTGGCATTCTGACT
ATAAGGTCAAAGGGCTATGTGATTCTAACCTCATTTCCATGGACATCACCT
TCTTCTCAGAGGACGGAGAGCTATCATCCCTGGGAAAGGAGGGCACAGG
GTTCAGAAGTAACTACTTTGCTTATGAAACTGGAGGCAAGGCCTGCAAAA
TGCAATACTGCAAGCATTGGGGAGTCAGACTCCCATCAGGTGTCTGGTTC
GAGATGGCTGATAAGGATCTCTTTGCTGCAGCCAGATTCCCTGAATGCCC
AGAAGGGTCAAGTATCTCTGCTCCATCTCAGACCTCAGTGGATGTAAGTC
TAATTCAGGACGTTGAGAGGATCTTGGATTATTCCCTCTGCCAAGAAACCT
GGAGCAAAATCAGAGCGGGTCTTCCAATCTCTCCAGTGGATCTCAGCTAT
CTTGCTCCTAAAAACCCAGGAACCGGTCCTGCTTTCACCATAATCAATGGT
ACCCTAAAATACTTTGAGACCAGATACATCAGAGTCGATATTGCTGCTCC
AATCCTCTCAAGAATGGTCGGAATGATCAGTGGAACTACCACAGAAAGGG
AACTGTGGGATGACTGGGCACCATATGAAGACGTGGAAATTGGACCCAAT
GGAGTTCTGAGGACCAGTTCAGGATATAAGTTTCCTTTATACATGATTGGA
CATGGTATGTTGGACTCCGATCTTCATCTTAGCTCAAAGGCTCAGGTGTTC
GAACATCCTCACATTCAAGACGCTGCTTCGCAACTTCCTGATGATGAGAG
TTTATTTTTTGGTGATACTGGGCTATCCAAAAATCCAATCGAGCTTGTAGA
AGGTTGGTTCAGTAGTTGGAAAAGCTCTATTGCCTCTTTTTTCTTTATCATA
GGGTTAATCATTGGACTATTCTTGGTTCTCCGAGTTGGTATCCATCTTTGC
ATTAAATTAAAGCACACCAAGAAAAGACAGATTTATACAGACATAGAGA
TGAACCGACTTGGAAAGTGATAAGGCCAGGCCGGCCAAGCTTGTCGAGA
AGTACTAGAGGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTG
CTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAAT
GCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAA
AGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCT
AGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTGA
TCACTGCTTGAGCCTAGGAGATCCGAACCAGATAAGTGAAATCTAGTTCC
AAACTATTTTGTCATTTTTAATTTTCGTATTAGCTTACGACGCTACACCCAG
TTCCCATCTATTTTGTCACTCTTCCCTAAATAATCCTTAAAAACTCCATTTC
CACCCCTCCCAGTTCCCAACTATTTTGTCCGCCCACAGCGGGGCATTTTTC

TTCCTGTTATGTTTTTAATCAAACATCCTGCCAACTCCATGTGACAAACCG
TCATCTTCGGCTACTTT