DEVELOPMENT AND APPLICATIONS OF UNIVERSAL GENETIC CODE EXPANSION PLATFORMS

by

JAMES S. ITALIA

A dissertation

submitted to the Faculty of

the department of Chemistry

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Morrissey School of Arts and Sciences
Graduate School

May 2019

© copyright by JAMES S. ITALIA
2019

DEVELOPMENT AND APPLICATIONS OF UNIVERSAL GENETIC CODE EXPANSION PLATFORMS

James S. Italia

Thesis Advisor: Abhishek Chatterjee, Ph.D.

Abstract

The emergence of genetic code expansion (GCE) technology, which enables site-specific incorporation of unnatural amino acids (UAAs) into proteins, has facilitated powerful new ways to probe and engineer protein structure and function. Using engineered orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pairs that suppress repurposed nonsense codons, a variety of structurally diverse UAAs have been incorporated into proteins in living cells. This technology offers tremendous potential for deciphering the complex biology of eukaryotes, but its scope in eukaryotic systems remains restricted due to several technical limitations. For example, development of the engineered tRNA/aaRS pairs for eukaryotic GCE traditionally relied on a eukaryotic cell-based directed evolution system, which are significantly less efficient relative to bacteria-based engineering platforms. The work described in this thesis establishes a new paradigm in GCE through the development of a novel class of universal tRNA/aaRS pairs, which can be used for ncAA incorporation in both *E. coli* and eukaryotes. We achieve this by developing engineered strains of *E. coli*, where one of its endogenous tRNA/aaRS pair is functionally replaced with an evolutionarily distant counterpart. The liberated pair can then be used for GCE in the resulting altered translational machinery (ATM) strain, as well as any eukaryote. Using this strategy, we have been able to genetically encode new bioconjugation chemistries, post-translational modifications, and facilitate the incorporation of multiple, distinct ncAAs into a single protein. The ATM technology holds enormous promise for significantly expanding the scope of the GCE technology in both bacteria and eukaryotes.

Dedication

I dedicate this thesis to my parents,

Chris and Cheryl Italia,

as well as my fiancé Haley DiBiase.

Christopher A. Italia
Dad

Cheryl Italia
Mom

Haley DiBiase

Acknowledgements

One of the biggest lessons I learned throughout my PhD is that it takes a team to accomplish anything great. I could not have finished this PhD without the help of my colleagues, family, and friends. I truly believe this accomplishment is as much of a testament to my support network as it is my own. Thank you all from the bottom of my heart.

I'd like to start by thanking my mentor and PhD advisor, Abhishek Chatterjee. You are not only a leader in the lab but a mentor in my life. 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 character. It is not often you find an advisor or boss with whom you mesh so well and I was truly blessed to have spent the last few years cranking out science with you. I look forward to our next ventures together and expect to see continual success from you and your lab.

Regarding the lab, I am so thankful for the highly collaborative and team-oriented environment we have developed. I was always confident that I could find someone to ask for help if I needed assistance with a new protocol, equipment, or just needed to chat through an idea. While everyone deserves acknowledgement, I'd like to thank Yunan Zheng and Rachel Kelemen, in particular, for helping to start the lab and continually providing feedback. Additionally, I'd like to thank Partha Sarathi Addy for his help with all synthesis projects. I additionally need to thank my team of undergraduates Chester Wrobel (C1), Chris Latour (C2), and Christen Hillenbrand (C3). I could not have been nearly as productive without you and expect great things in years to come from all three. Lastly, thank you to Katie Grasso for being a great friend in the lab and good luck keeping team ATM going!

Thank you to Eranthie Weerapana and Jianmin Gao for their guidance in both undergraduate and graduate school. I've been with you for a while! I'd like to thank all

members of the Weerapana and Gao labs their help over the years. The accessibility and collegial atmosphere in the department is inspiring. Particularly, thank you to Jenny Peeler for her help with the mass spec, as well as Tyler Bechtel for "getting that CRISPY." Thank you to Kyle Cole, as well, for his guidance and friendship. From the Gao lab, I'd like to thank the "dude bro" Mike Kelly, who has emerged as a lifetime friend who was instrumental in keeping me sane. 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, allowing us to focus on the science!

I'd like to broadly thank my family and friends who have supported me throughout these challenging years. I love my brother Nick and sister Lauren, who constantly push me to do my best. Thank you to my relatives and friends who were always supportive: Auntie Lynne, Uncle Tim, Uncle Gary, the BC Woofers, and many more. I am very thankful to still have the support of my Grammie, Nana, and Pop who are all entering the ripe young age of 80+ years old. I can say for certainty that I feel supported by my loved ones and am very grateful.

I'd like to briefly thank my future family, the Devitt and DiBiase clans. Janice, Michael, Joslin, John, and the rest of the family have afforded me nothing but the kindest regards since first meeting me. I look forward to becoming an official part of the family in June.

Lastly, I'd like to thank the people to whom this thesis is dedicated. Unfortunately, my father is no longer with us after his battle with a gliobastoma. He was truly determined, selfless, and is a large inspiration for me as a man. Thank you mom, for teaching me how to remain steady in challenging times. You're the toughest and most loving woman I know. Finally, thank you Haley for sticking with me through some of the most intense years of our lives. I would not be half the person I am today without your support and love. I love you all so much and thank you again.

Table of Contents

List of Figures

List of Abbreviations

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

AA	amino acid
4-CDZ	4-carboxydiazonium
5AmW	5-aminotryptophan
5AzW	5-azidotryptophan
5BrW	5-bromotryptophan
5HTP	5-hydroxytryptophan
5MTP	5-methyoxytryptophan
5PrW	5-propargyltryptophan
aaRS	aminoacyl-tRNA synthetase (aa can be replaced with 3- letter code or UAA shorthand)
Alexa488-PCA	alexa488-picolyl azide
Amp	ampicillin
AND	N-terminal domain
arab	arabinose
ATM	Altered Translational Machinery
ATMW	Altered Translational Machinery Tryptophanyl Strain
ATMY	Altered Translational Machinery Tyrosyl Strain
ATP	adenosine tri-phosphate
AzK	azidolysine
BocK	Boc-lysine
BoF	p-boronophenylalanine
CAT	chloramphenicol acetyl transferase
Chlor	chloramphenicol
CpK	cyclopropene lysine

CRACR	chemoselective rapid azo-coupling reaction
DBCO-Cy5	dibenzocyclooctyne-amine-Cy5
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	dimethylsulfoxide
EGFP	enhanced green fluorescent protein
ESI-MS	electrospray ionization mass spectrometry
Fab	fragment antigen binding
FBS	Fetal bovine serum
FRET	Förster resonance electron transfer
GAG	glycosaminoglycans
galK	galactokinase
GBD	glycosoaminoglycan binding domain
GCE	genetic code expansion
Gent	gentamycin
$Gent^R$	gentamycin resistance
GMO	genetically modified organism
gRNA	guide RNA
HCII	Heparin cofactor II
HEK293T	Human embryonic kidney cell line
HER2	human epidermal growth factor receptor 2
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
IPTG	isopropyl β-D-1-thiogalactopyranoside
Kan	kanamycin
LB	luria broth
MG	malachite green

Mj	*M. jannaschii*
MS	mass spectrometry
MWCO	molecular weight cutoff
ncAA	non-canonical amino acid
NTP	nitrotryptophan
OmeY	O-methyltyrosine
OPrY	O-propargyltyrosine
pAcF	p-acetylphenylalanine
PAPS	3'-phosphoadenosine-5'-phosphosulfate
pAzF	4-azido-L-phenylalanine
PEI	Polyethylamine
Pen	Penicillin
Pi	inorganic phosphate
pIF	p-iodophenylalanine
Ppiase	inorganic pyrophosphatase
PTM	post-translational modification
pTyr	phosphotyrosine
PVDF	polyvinylidene fluoride
Pyl	pyrrolysyl
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sfGFP	super-folder green fluorescent protein
SPAAC	strain-promoted azide-alkyne cycloadditions
Spec	spectinomycin
SPIEDAC	strain-promoted inverse electron-demand Diels-Alder cycloaddition
Strep	Streptinomycin
sTyr	sulfotyrosine

Tet	tetracycline
TEV	tobacco etch virus
TIC	total ion count
tolC	outer membrane protein tolC precursor
TPST	tyrosylprotein sulfotransferase
tRNA	transfer RNA
trpS	tryptophanyl-tRNA ligase
trpT	tryptophanyl tRNA
tyrS	tyrosine-tRNA ligase
tyrTV	tyrosyl tRNA T/V
tyrU	tyrosyl tRNA U
UAA	unnatural amino acid
Zeo	zeocin
Zeo^R	zeocin resistance

Chapter 1

Introduction

A portion of this chapter has been published in:

Italia, J.S., Zheng, Y. Kelemen, R.E., Erickson, S.B., Addy, P.S., Chatterjee, A. Expanding the genetic code of mammalian cells. *Biochem. Soc. Trans.,* **2017,** 45, 555-562.

1.1 Genetic code expansion: Established suppression pairs and selection of new pairs in bacteria and eukaryotes

Natural protein function is diverse, varying between protein-protein/DNA/RNA interactions, cell signaling, and enzymatic processes. While protein structures are highly variable, their primary sequence is restricted to just 20 amino acid building blocks. These 20 canonical amino acids are limited in their chemical functionalities, but can be post-translationally modified (PTM) *in vivo* or *in vitro* to diversify their biological function for natural cellular processes or biotherapeutic applications.[1–3] Over the last 20 years, techniques such as mass spectrometry based proteomics have increased the available knowledge of existing protein post-translational modifications,[1–3] but many questions regarding their function remain unanswered due to limitations in homogeneous protein expression both *in vivo* and *in vitro,* which is required to study the protein's natural function. Techniques which allow for the expression of proteins in bacterial or eukaryotic expression hosts are improving but there is a constant push to diversify the toolkit of recombinantly produced proteins. In particular, there is a critical need for technology to introduce non-natural functionalities precisely into proteins to both study or engineer their function. Eukaryotic cells have become increasingly popular platforms for protein production, as the resultant proteins contain desired modifications and processing absent in bacterial expression platforms. The use of eukaryotic expression hosts for protein production enables the manufacturing of new protein based-therapeutics[4] and there is consequentially a drive to improve strategies which can homogenously customize the recombinant protein composition in a predetermined fashion.

The site-specific incorporation of non-canonical amino acids (ncAAs) is an attractive approach to study a protein's effect on the cell biology as well as produce therapeutically relevant biologics.[5–8] This technology has made great strides over the last 20 years by enabling

the incorporation of ncAAs which function as bioconjugation handles,[6,9–15] small fluorescent/redox probes,[6,16–23] photocrosslinkers,[6,16,24–27] PTMs,[16,28–36] and photocaged amino acids,[5,16,37–39] many of which will be discussed in this thesis (Figure 1-1). Two essential elements are central to this technology: an engineered tRNA/aminoacyl-tRNA synthetase (aaRS) pair and a unique codon directing the incorporation of the ncAA.[5–8] The strategy begins with the import of an orthogonal (non-cross reactive) tRNA/aaRS pair into the desired expression host from a different domain of life. This is because the pairs used for ncAA incorporation must be orthogonal to the host to avoid cross-reactivity with host translational machinery, which would result in misaminoacylation with an undesired amino acid on host tRNAs, or visa versa. Imported tRNA/aaRS pairs must use a unique codon for the incorporation of the ncAA into a site-specific spot in a protein, which requires the mutagenesis of the anticodon loop of the tRNA to a unique noncoding, liberated, or frameshift anticodon (Figure 1-2). A repurposed TAG stop codon has been the most commonly used codon for direction of the ncAA of interest. Efforts to use other stop codons such as TGA/TAA[40–43] or four base frameshift codons[44] have also been demonstrated but are generally less effective. Once imported tRNA/aaRS pairs are identified as orthogonal nonsense codon suppressors, they can be evolved to modify their substrate specificity from a natural amino acid to a ncAA of interest, usually with a structural similarity to the original amino acid (e.g. TrpRS incorporates Trp analogs, etc. with exception). Directed evolution of the synthetases occurs through a series of positive and negative selection stages, most commonly performed in bacteria or in yeast depending on the tRNA/aaRS pair and with varied levels of success.[7]

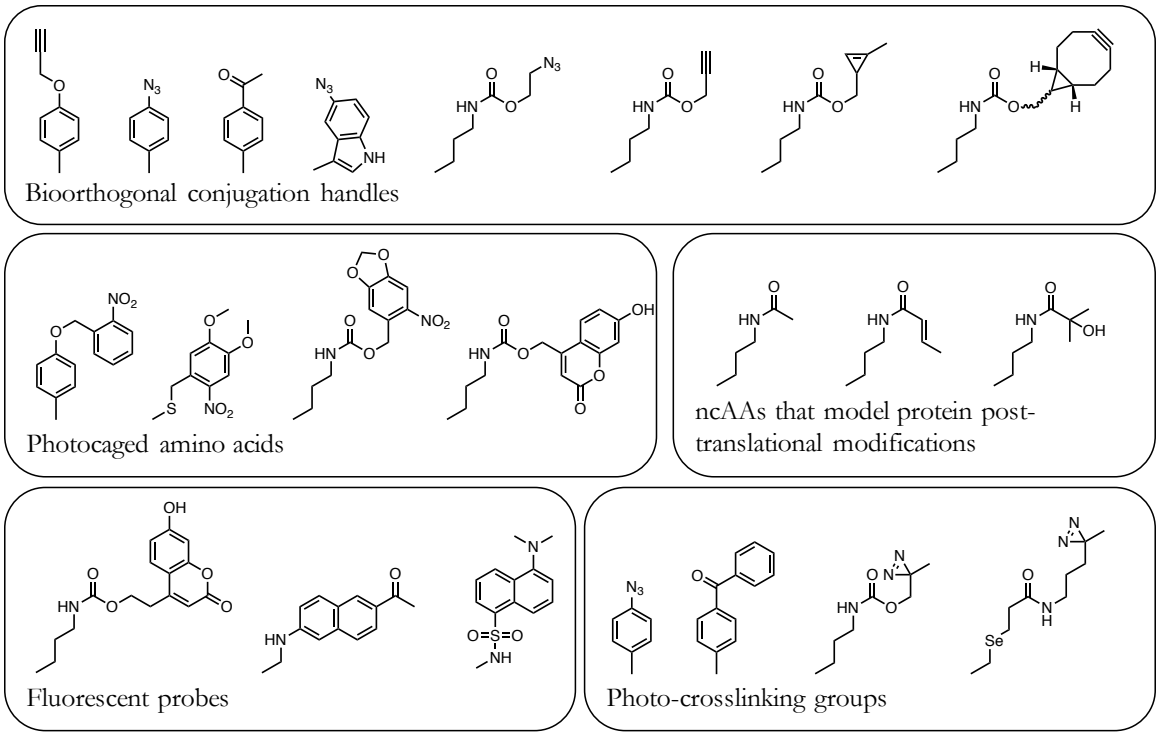


Figure 1-1. A representative list of useful ncAAs.

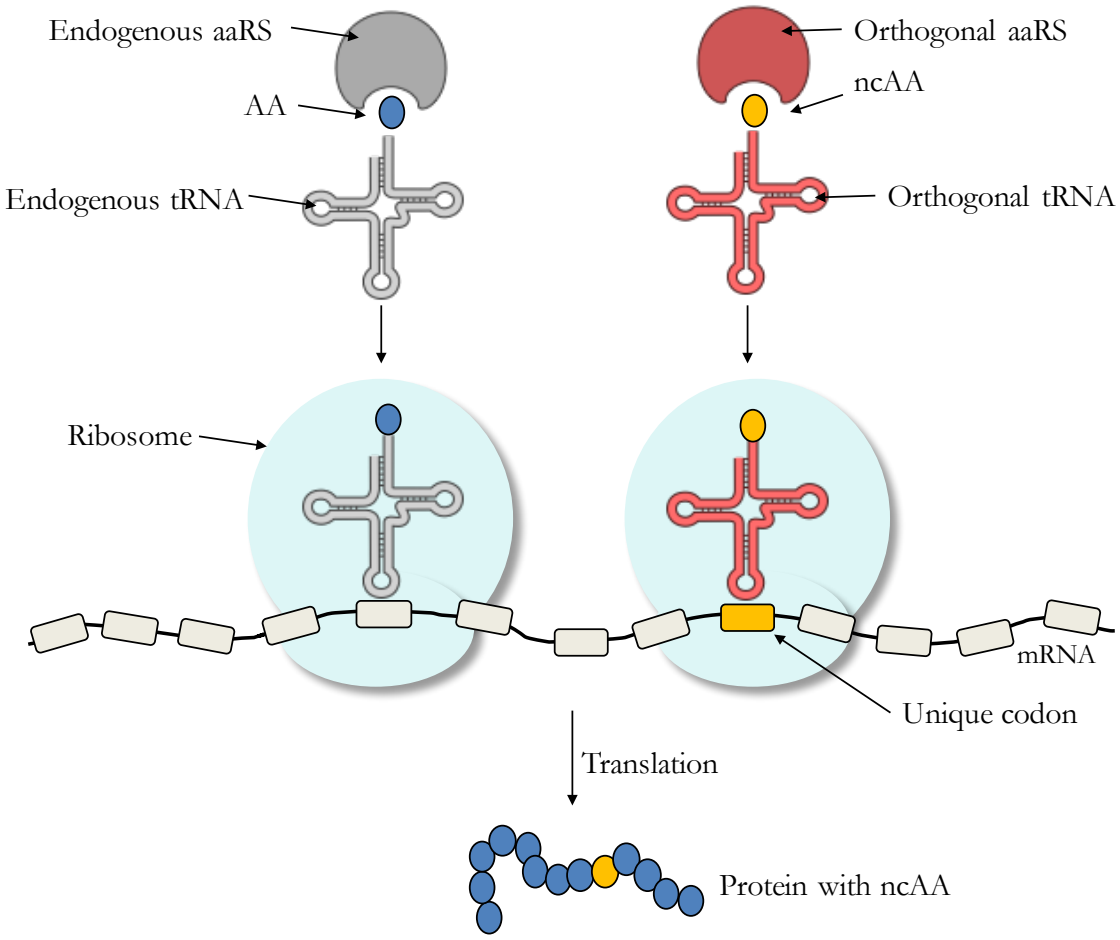


Figure 1-2. Overview of ncAA incorporation. An orthogonal (i.e. non-crossreactive) tRNA/aaRS pair (red) from a different domain of life direct the incorporation of a non-canonical/unnatural amino acid to a site-specific spot in a protein of interest. The ncAA is directed to the site-specific location by a unique codon, usually the TAG nonsense codon. The imported orthogonal tRNA/aaRS must not cross react with the host machinery (grey).

The evolution of new tRNA/aaRS pairs for genetic code expansion varies in different domains of life due to the orthogonality requirements of the introduced pair (Figure 1-3).[5–8] Generally, bacterial tRNA/aaRS pairs are orthogonal in eukaryotes, while eukaryotic/archaeal pairs are orthogonal in bacteria (Figure 1-3). There are therefore two main platforms for developing engineered tRNA/aaRS pairs for GCE: directed evolution of bacterial pairs in yeast and eukaryotic/archaeal pairs in *E. coli*. Evolution of the tRNA to create a more orthogonal tRNA may be necessary if the introduced tRNA is not completely orthogonal or inefficient as a nonsense codon suppressor in the host organism.[45] In order to engineer the aaRS and shift the specificity from an endogenous amino acid to a ncAA, further directed evolution must be performed in the orthogonal host (Figure 1-3). For GCE in bacteria, orthogonal eukaryotic or archaeal pairs can be developed using the facile *E. coli* platform (Figure 1-3).[7] The *E. coli* selection system leverages an well-established double-sieve selections scheme. In the positive selection, a library of eukaryotic/archaeal aaRS variants are subjected to a gradient of the antibiotic chloramphenicol in the presence of its cognate tRNA and the desired ncAA. A chloramphenicol acetyl transferase gene containing a stop codon (at a permissive site) ensures selection of any active mutant that charges the ncAA or an endogenous amino acid. The subsequent negative selection eliminates aaRS variants that charge canonical amino acids from this pool. It uses the toxic barnase gene containing in-frame stop codons in the absence of the ncAA.[7] Generally, multiple rounds of positive and negative selection are used to isolate selective aaRS mutants.

The selection process of bacterial pairs (which are typically orthogonal in eukaryotes) in eukaryotic hosts for eukaryotic GCE is more complicated. The engineering of bacterial pairs has historically been performed in *Saccharomyces cerevisiae*,[7] but technical challenges associated with this platform has limited the structural diversity of ncAAs incorporated in eukaryotic cells

due to poor selection efficiency relative to *E. coli*.[6] Positive and negative selections are performed similarly to the *E. coli* based selection, but metabolic markers are used instead of antibiotics.[7]

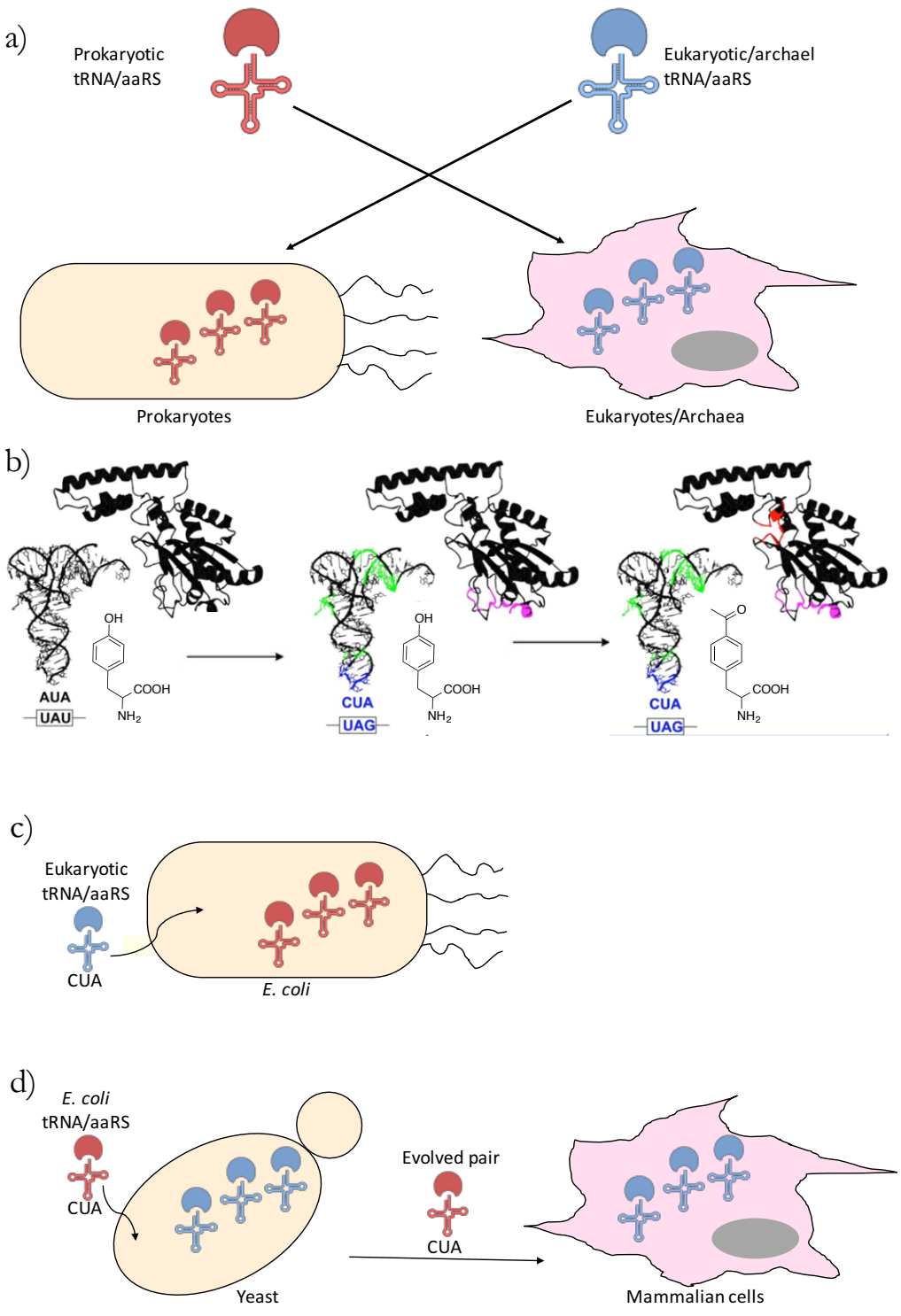


Figure 1-3. Evolution of new tRNA/aaRS pairs in different domains of life. a) Overview of orthogonality trends between prokaryotes and eukaryotes for genetic code expansion. b) Step 1: An orthogonal tRNA/aaRS pair is imported from a different domain of life. Step 2: The

tRNA/aaRS pair is optimized for nonsense codon suppression in the new host. Step 3: The aaRS substrate specificity is modified to shift from a natural amino acid to a ncAA of interest via positive[7] (chloramphenicol resistance) and negative selection (barnase).[7] c) Eukaryotic pairs are used for GCE in bacteria and can be evolved in the facile *E. coli* host/selection platform. d) In order to perform GCE in mammalian cells, one must import an *E. coli* derived pair into the challenging yeast selection platform. Hits from the yeast selection can subsequently be used in mammalian cells. The yeast selection system has been shown to be less efficient than the bacterial pair, evident by many fewer mammalian encoded ncAAs.[6]

The traditional paradigm of GCE results in two issues: 1) Two separate iterations of directed evolution must be performed for the same ncAA for application in bacteria and eukaryotes. 2) The inefficiencies of the yeast based selection system severely limits the collection of ncAAs that can be genetically encoded in eukaryotes using bacterial pairs. Conveniently, nature has provided a solution to these challenges: the pyrrolyl tRNA/aaRS pair.[46] This naturally occurring TAG suppressor maintains orthogonality in both bacteria and eukaryotes due to its unique structural features.[47,48] Therefore, the pyrrolysyl tRNA/aaRS pair can be evolved in the facile *E. coli* selection system and discovered hits can be universally used in both bacteria and eukaryotes. The advantage of this univerasal orthogonality is evident in the heavy reliance on the pyrrolysyl platform for eukaryotic GCE after its discovery (Figure 1-4).[49] While the advancement of the pyrrolysine system has been beneficial for the field, a near-exclusive dependence on one tRNA/aaRS pair has resulted in a bias toward ncAAs which are structurally similar to pyrrolysine analogs. While some non-lysine based amino acids have been incorporated, their efficiency is generally poor.[29,50] Additionally, dependence on a single tRNA/aaRS pair limits the scope of site-specific incorporation of multiple distinct ncAAs

(Discussed in Chapter 5 and Section 1.5), which requires multiple orthogonal tRNA/aaRS pairs.

It is useful to briefly describe the tRNA/aaRS pairs used for GCE in bacteria and eukaryotes as a primer to our work which would be covered in the next chapters. In bacteria, the predominant pairs are the *M. jannaschii* derived tyrosyl and *Methanosarcina* derived pyrrolysyl pairs. These pairs have been evolved in *E. coli* and can be used to encode a variety of ncAA, including bioconjugation handles,[6,9–15] photocrosslinkers,[6,16,24–27] etc. Less frequently used pairs include the phosphoserine/threonine encoding SepRS,[29,51,52] *Saccharomyces cerevisiae* tryptophanyl tRNA pair,[53] *P. horikoshii* aaRS/*A. fulgidus* tRNA encoding prolyl analogs,[45] and the histidyl *C. crescentus* pair.[54] Existing pairs for eukaryotic GCE are limited to the pyrrolysyl,[46] *E. coli* tyrosyl,[6,7,55] and *E. coli* leucyl pairs,[7,42] with the engineering of the *E. coli* pairs performed in yeast. As mentioned, the overwhelming majority of ncAAs encoded in eukaryotes use the pyrrolysyl platform (Figure 1-4). The introduction of additional universal tRNA/aaRS pairs would improve the genetically encodable structural space, providing access to ncAAs that are challenging to incorporate using by the pyrrolysyl pair.

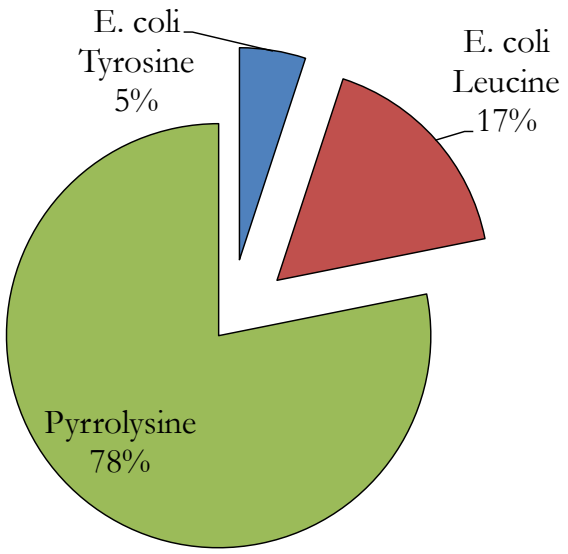


Figure 1-4. Proportion of ncAAs genetically encoded in eukaryotes using the three most established tRNA/aaRS pairs, compiled via Liu *et al.*[7] and Dumas *et al.*[6] See figure 2-2 for comparison by year.

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

We envisioned overcoming these limitations through the development of additional universally orthogonal tRNA/aaRS pairs. These pairs, like the pyrrolysyl pair, can be evolved using the facile *E. coli* selection system, and subsequently imported into eukaryotic cells for GCE. The facile *E. coli* selection system can enable incorporation of highly engineered aaRS mutants selective for ncAAs that are structurally distant from its original substrate. We developed a novel and general strategy for creating such universal pairs by developing a strain of *E. coli*, where one of its endogenous tRNA/aaRS pairs is functionally replaced with a pair from an alternate domain of life (Figure 1-5).[49,56] By doing so, we can then reimport this

liberated *E. coli* tRNA/aaRS pair into the resulting strain of *E. coli* (Altered Translational Machinery, or ATM strain) as a nonsense suppressor, which can subsequently be engineered to encode new ncAAs. The resultant bacterial tRNA/aaRS pairs can also be used in eukaryotes for GCE, as these are typically orthogonal in eukaryotes. We were able to complement the removal of the endogenous tRNA/aaRS pairs for the tryptophan and tyrosyl platforms with the corresponding archaeal/eukaryotic pair (encoded in a plasmid). We were then able to reintroduce the liberated *E. coli* tRNA/aaRS pairs for directed evolution of tryptophanyl and tyrosyl based pairs, which is outlined extensively in Chapter 2[49] and Chapter 3,[56] respectively (Figure 1-5). In future efforts, the pursuit of additional ATM strains for other endogenous amino acids would drastically improve the chemical diversity of the eukaryotic toolkit due to the generation of similarly universal tRNA/aaRS pairs.

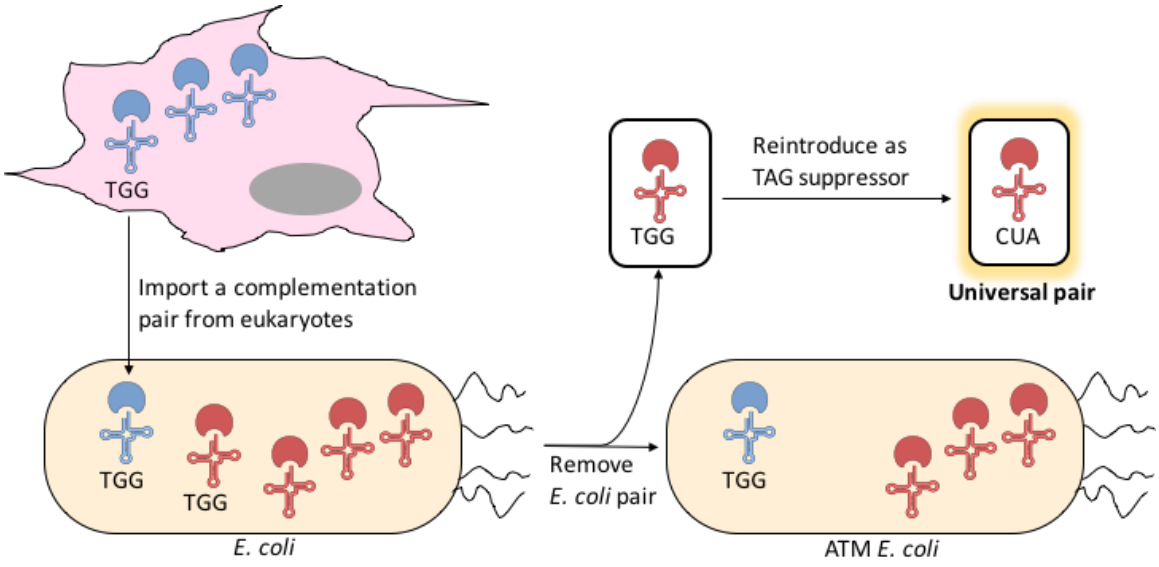


Figure 1-5. Overview of the strategy to create an aaRS-tRNA pair that can drive genetic code expansion in both eukaryotes and *E. coli*.

1.3 Post-translational modifications and their exploration using genetic code expansion

While developing these ATM strains, we had a long-term goal of the incorporation of underrepresented ncAAs in eukaryotic GCE, such as post-translational modifications (PTMs). The catalog and knowledge of the prevalence of eukaryotic post-translational modifications has grown dramatically in recent years, partially due to the advances in mass spectrometry of the proteome.[16,28–36] However, discovery of these modifications has only led to further curiosity into their function which can have strong implications in human health and disease (Figure 1-6). Assigning their function remains challenging due to the lack of methods to produce homogeneously modified proteins containing the specific PTM at the site of interest. The biochemical origin of many PTMs remain unclear or challenging to reconstitute *in vitro* or *in vivo*. Co-expression with a PTM inducing system (kinase, TPST, etc),[57] is sometimes feasible but provide no positional control for proteins with multiple modification sites[58–60] (eg. two tyrosine residues that can be phosphorylated). Therefore, strategies to improve the generation of recombinant post-translationally modified proteins in a homogenous state would significantly enhance our knowledge about the direct effect of the individual PTM function.

The ability to genetically encode new PTMs can circumvent challenges in the installation mechanisms of PTMs into recombinant proteins. While there is a broad variety of PTMs in human biology, a significant portion of the GCE field has revolved around phosphorylation[51,31,52] or lysine modifications,[61,32,33] with the latter being the more frequently incorporated subclass in eukaryotic GCE. Particularly, there has been a significant amount of work in histones when using the pyrrolysine pair to incorporate lysine based PTMs, highlighting the effects of acetylation,[32] crotonylation,[33] and hydroxyisobutyrylation[61] on chromatin function (Figure 1-6). This research has effectively shown the potential that GCE

provides when studying these epigenetic PTMs but is limited to analogs from the pyrrolysine platform.

The lysine based PTMs have been the most abundantly genetically encoded and studied (using the versatile pyrrolysyl pair) but other PTMs have also been encoded, including phosphoserine, [29,51,52] phosphothreonine,[36] sulfotyrosine,[35,58] phosphotyrosine,[31] and nitrotyrosine[62,63] (Figure 1-6). Additionally, nonhydrolyzable mimics of some of these PTMs have also been incorporated.[64] In some cases, for example phosphotyrosine, such a mimetic is necessary due to the host's natural tendency to remove the modification (e.g., due to phosphatase activity).[31] While the list of PTMs that have been genetically encoded is quite extensive, the majority of these cannot be extended to eukaryotes, because the necessary tRNA/aaRS pairs are incompatible with the eukaryotic expression platform. Indeed, the only class of PTMs that can be efficiently incorporated in eukaryotic cells are acylated lysine derivatives, charged using the universal pyrrolysyl platform. The ability to extend genetically encoded PTMs to eukaryotes is essential to: 1) facilitate the expression of eukaryotic proteins that cannot be expressed in bacteria due to the lack of necessary processing, and 2) to evaluate the in vivo consequences of a PTM. The tRNA/aaRS pairs derived from the ATM technology should enhance the PTM toolkit available in eukaryotic expressions.

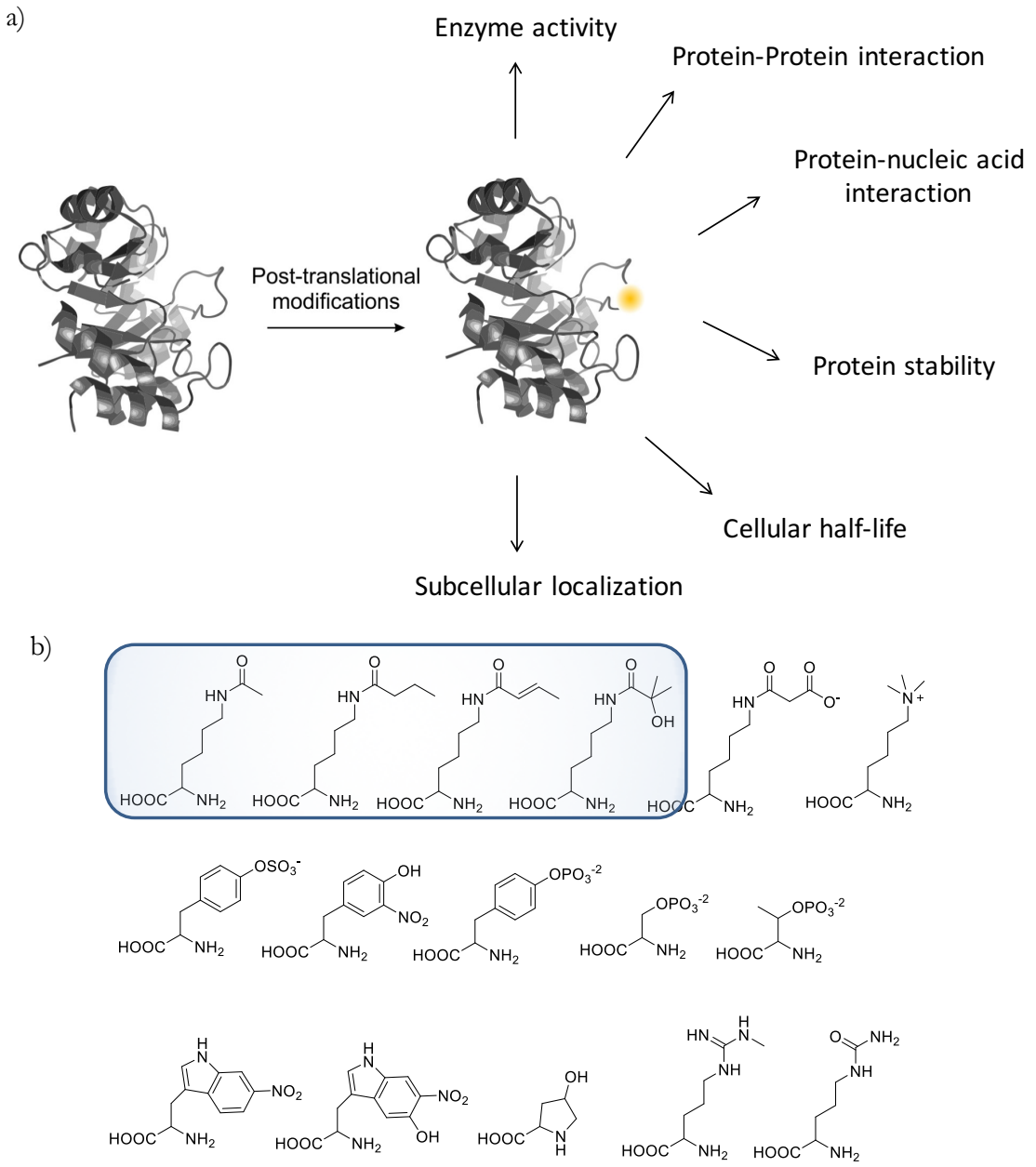


Figure 1-6. Overview of post-translational modifications. a) Post translational modifications can have a variety of functions. b) A representative sample of PTMs. Lysine based PTMs that have been encoded in mammalian cells have been highlighted. Most of the other PTMs have been encoded in bacteria, while others like 5-hydroxy-6-nitrotryptophan has not yet been encoded.

1.4 Chemoselective bioconjugation handles

Another exciting avenue of GCE is the incorporation of bioconjugation handles, which provide the ability to attach a variety of ligands to a protein of interest. Introducing these orthogonal conjugation functionalities absent in nature have enabled researchers to conjugate biophysical probes[19] (eg. fluorophores) *in vivo* and *in vitro,* introduce glycosylation/pegylation, develop homogenous antibody drug conjugates,[65] and modify the capsids of virus.[66] A diverse subset of ncAA have been incorporated and are compatible with a variety of bioconjugation techniques, with some of the main methods being: Staudinger ligation,[7,67] strain-promoted azide-alkyne cycloadditions (SPAAC),[10] strain-promoted inverse electron-demand Diels-Alder cycloaddition (SPIEDAC),[68] as well as conjugation via ketones[69] and boronic acid[70] reactive groups, which can be used to introduce reactive handles for modification (Figure 1-7). However, not all of the strategies can be used in combination due to cross-reactivity among the reagents. Development of new mutually compatible chemistries would be beneficial for applications where multiple orthogonal bioconjugation handles were desired, as one could direct multiple appendages to specific residues in a single protein (Chapter 5).

While this work does not directly focus on the development of new bio-orthogonal reactions such as CRACR[13] and oxidative coupling,[71,72] an important pre-requisite for such work is the ability to incorporate an increasing diversity of bio-orthogonally reactive ncAAs. As an example, 5-hydroxytrypothan (5HTP), which was incorporated for the first time using the ATM-Trp strain (Chapter 3), has since been used in our lab to develop new classes of bio-orthogonal reactions.

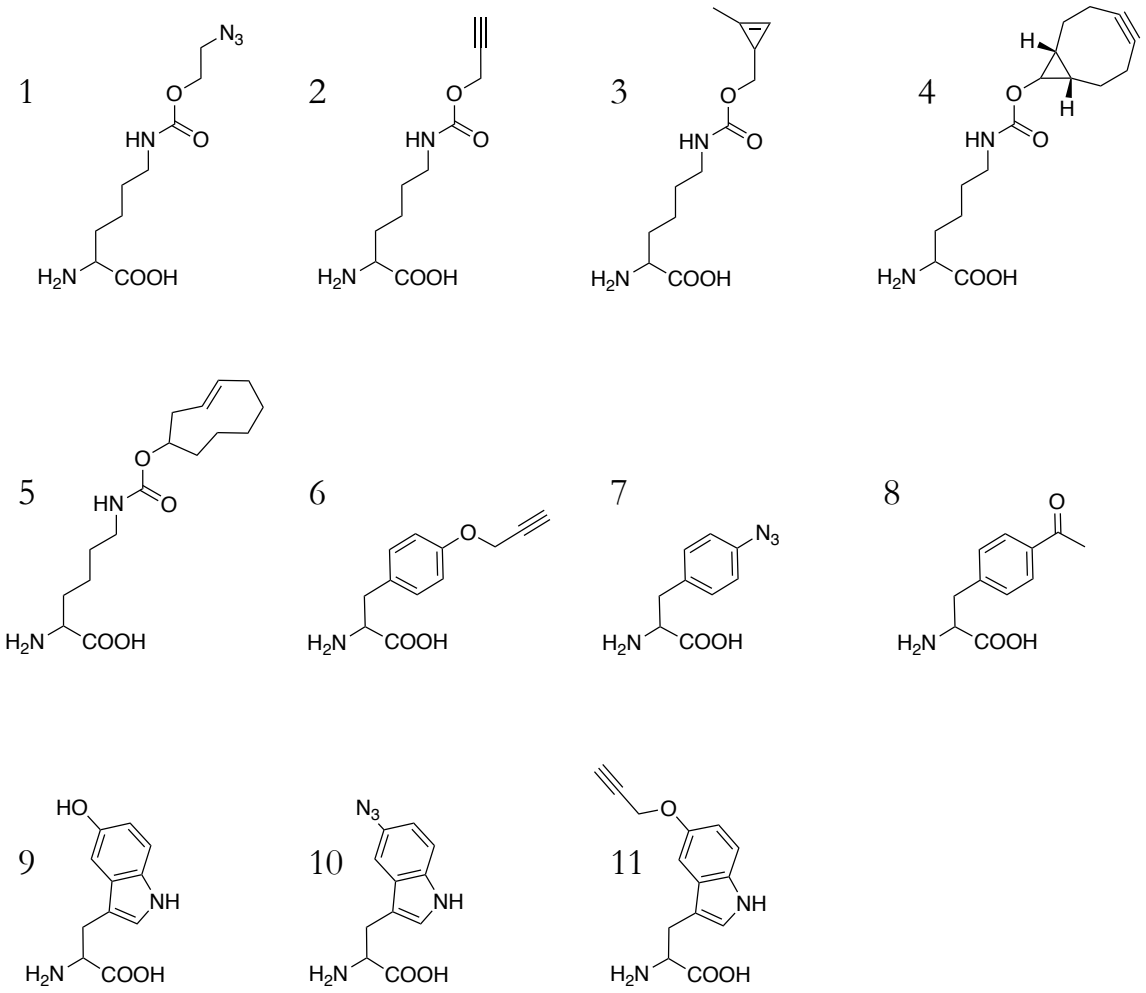


Figure 1-7. Subset of genetically encoded bioconjugation handles. 1-5 are encoded with the pyrrolysl platform. 6-8 are encoded with the tyrosyl platforms in both eukaryotes and bacteria. 9-11 are encoded with the tryptophanyl platform[73] in eukaryotes and bacteria.

1.5 Multiple site incorporation of ncAAs in bacteria and eukaryotes

While the majority of the GCE field has revolved around the incorporation of a single amino acid into one site of a protein in bacteria or eukaryotes, the incorporation of multiple distinct ncAAs at specific locations in a single protein is of great interest for many biological applications, both therapeutic and investigative. For example, incorporation of multiple distinct ncAAs opens up some exciting applications such as FRET,[74] conjugation of multiple ligands to an antibody,[13] probing a PTM's function,[75] photocrosslinking,[26,27] and other creative uses. In order to incorporate multiple distinct ncAAs, a few requirements must be met. Each ncAA must be charged by its own tRNA/aaRS pair and a distinct codon directing its incorporation. The different engineered tRNA/aaRS pairs used in such a system must not cross-react with each other (Figure 1-8) or with their counterparts from the host. It is important to remember that two pairs can cross-react at various levels (Figure 1-8). The limited number of pairs currently available for GCE and the intrinsic inefficiency of the engineered pairs restricts the scope of multiple distinct ncAA incorporation. Additionally, all tRNA/aaRS pairs are typically developed as TAG suppressors. Consequently, a subset of these pairs must be reengineered to charge other nonsense/frameshift codons to enable multisite ncAA incorporation.

So far, it has been possible to site-specifically incorporate two distinct ncAAs into proteins in *E. coli*.[43,74,76–79] The highly evolvable *M. jannaschii* and pyrrolysyl tRNA/aaRS pair have been demonstrated to be mutually orthogonal, and each has been used to incorporate numerous enabling ncAAs. Additionally, the pyrrolysyl pair has been shown to allow efficient suppression of TGA and TAA codons, providing a facile route for incorporating two distinct ncAAs.[43,74,76–79] Additional effort has also demonstrated the ability to use quadruplet codons to direct the incorporation of these residues in parallel with the TAG codon.[76,80] Generally, MjTyr

is used to suppress the TAG codons while the pyrrolysyl pair is used for the TAA or quadruplet codons. This strategy has also been used to site-specifically incorporate two distinct bioconjugation handles that are mutually compatible, paving the way for precise labeling of the resulting protein at two different sites.[43,74,77,78] While their use as a proof of concept has been effective, their efficiencies can be limited by the poor suppression of the pyrrolysine pair. Additionally, incorporation of more than two distinct ncAAs is currently not feasible due the lack of additional mutually orthogonal pairs. Therefore, the ability to develop new, more efficient tRNA/aaRS pairs can dramatically broaden the scope of multisite incorporation. In Chapter 5, the use of the tryptophanyl tRNA/aaRS pair for triple suppression and triple labeling of a single protein is extensively discussed and highlights the improvement afforded by this novel TGA suppressor. Our tryptophan TGA suppressor facilitated increased double suppression efficiency due to its superior TGA suppression efficiency relative to other pairs.

While this thesis does not thoroughly discuss multisite ncAA incorporation in mammalian cells, it is important to briefly discuss the advances and challenges faced by the field and comment on the benefit of the ATM system. Generally, researchers have only been able to show the incorporation of two distinct ncAA in eukaryotic systems using the pyrrolysine system in combination with the *E. coli* tyrosyl or *E. coli* leucyl pairs.[81] Our tryptophanyl tRNA/aaRS pair has been introduced into mammalian cells and was found to be an exceptional TGA suppressor. Additionally, it is orthogonal to other pairs currently used for eukaryotic genetic code expansion. Thus, it provides exciting new opportunities for multisite ncAA incorporation into eukaryotic proteins.

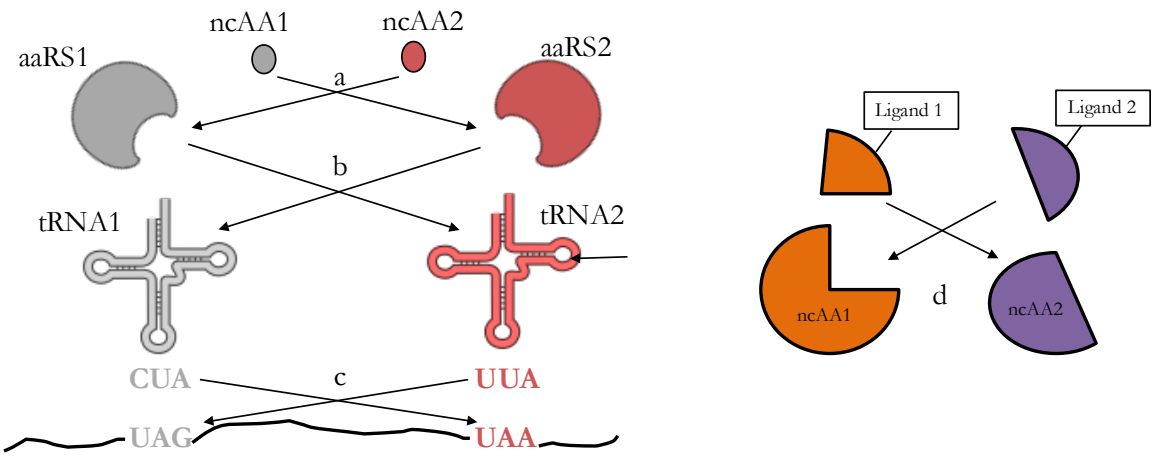


Figure 1-8. Potential cross-reactivity pain points for multiple suppression. a) ncAA substrate specificity of the aaRS1 and aaRS2. b) Between tRNA/aaRS for ncAA1 and tRNA/aaRS for ncAA2. c) Between codons/anticodons. d) Between bioconjugation handles (if used) or ncAA specificity of aaRS1 and aaRS2.

1.6 Conclusions

Overall, this thesis will discuss the development and applications of the ATM technology. The two ATM strains (tryptophan and tyrosyl) have opened up the doors for a variety of exciting applications and should be able to accelerate the GCE field in both bacteria and eukaryotes. The roadmap for this thesis should bring the reader through the generation of the tryptophan and tyrosyl strains, demonstrating their benefit through minor applications in Chapters 2 and 3. These chapters highlight the challenges of strain building and show the ability of the ATM platform to select for novel and efficient tRNA/aaRS variants. Chapters 4 and 5 focus more on the applications and improvements provided to the PTM (sulfotyrosine) and multisite incorporation space. Chapter 6 is dedicated to unpublished work completed during this PhD and would hopefully help future graduate students and researchers build upon

the current work developed from the ATM platform. I hope this thesis efficiently depicts the origins of the ATM platform and extends to future applications of the technology in the GCE field.

1.7 Future Directions and Considerations

While platforms are individually discussed in Chapter 2 and 3 respectively, our results suggest some general considerations for the development of future strains: complementation plasmid components, the presence of multiple codons for a particular amino acid, the genomic recombination site, and the parent strain. The complementation plasmid must provide the appropriate expression levels of tRNA/aaRS to complement the removal of the endogenous pair. This pair must also be highly efficient in the context of the *E. coli* translation system to support its protein synthesis. Although efficient suppressor pairs from archaea/eukaryote were already available for tryptophan[49] and tyrosine[56], we found that the use of the corresponding optimized nonsense suppressor tRNAs were not always optimal for decoding the sense codon. In future efforts, it may be best to test many different pairs and expression levels when determining the optimal complementation system (for future student: pCond plasmid). The presence of multiple codons for a particular amino acid was not an issue for the tryptophanyl or tyrosyl system (some challenges with tyrosyl which has three tRNAs with the same anticodon that can wobble base pair with the two tyrosine codons, see Chapter 2 and 3) but would be challenging for a system such as the leucyl pair. The leucyl pair contains six codons, multiple anticodons, and multiple tRNAs in >3 genomic locations, prompting the concern over the appropriate expression levels of the complementation tRNA/aaRS as well as which anticodons are optimal in the complementation tRNAs.

The last two considerations are less concerning but important nonetheless. Recombination is performed via the λ-RED mechanism[82] (Chapter 2 and Chapter 3) and one must take consideration into the genomic context of the KO site. As seen with the tyrosyl system (Chapter 3), growth defects can occur if a cumbersome cassette (long, contains a terminator, etc) is introduced into a polycistronic tRNA cassette or in front of another essential gene for cell viability. Lastly, the downstream applications will dictate which parent strain is optimal for development of the ATM strain. Efforts in this area are in progress (Chapter 6), but "specialized" ATM strains may be beneficial, such as the creation of an ATM strain derived from BL21[83] or C321,[84] whose parent strains have benefits when used for protein expression or multisite incorporation of ncAAs, for example.

1.8 References

1. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *Comput. Biol.* 1–23 (2015). doi:10.1371/journal.pcbi.1004049
2. Editorial, S. I. Orchestrating the proteome with post-translational modifications. *J. Exp. Bot.* **69,** 4499–4503 (2018).
3. Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-translational modification: Nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4,** 565–583 (2012).
4. Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat. Rev. Mol. Cell Biol.* **15,** 95–107 (2014).
5. Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* 1–30 (2014). doi:10.1146/annurev-biochem-060713-035737
6. Dumas, a., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical reassignment - expanding the chemistry in biology. *Chem. Sci.* **6,** 50–69 (2014).
7. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).
8. Xiao, H. & Schultz, P. G. At the Interface of Chemical and Biological Synthesis : An Expanded Genetic Code. (2019).
9. Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4,** 239–244 (2007).
10. Kim, C. H., Axup, J. Y. & Schultz, P. G. Protein conjugation with genetically encoded

unnatural amino acids. *Curr. Opin. Chem. Biol.* **17,** 412–419 (2013).
11. Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **7,** 876–884 (2011).
12. Lang, K. & Chin, J. W. Bioorthogonal reactions for labeling proteins. *ACS Chem. Biol.* **9,** 16–20 (2014).
13. Addy, P. S., Erickson, S. B., Italia, J. S. & Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **139,** 11670–11673 (2017).
14. Francis, M. B. & Carrico, I. S. New frontiers in protein bioconjugation. *Curr. Opin. Chem. Biol.* **14,** 771–773 (2010).
15. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chemie - Int. Ed.* **48,** 6974–6998 (2009).
16. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550,** 53–60 (2017).
17. Wang, J., Xie, J. & Schultz, P. G. A genetically encoded fluorescent amino acid SI. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9785–9789 (2006).
18. Summerer, D. *et al.* A genetically encoded fluorescent amino acid. *Pnas* **103,** 9785–9789 (2006).
19. Lee, H. S., Guo, J., Lemke, E. A., Dimla, R. D. & Schultz, P. G. Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in Saccharomyces cerevisiae. *J. Am. Chem. Soc.* **131,** 12921–3 (2009).
20. Luo, J. *et al.* Genetically Encoded Optochemical Probes for Simultaneous Fluorescence Reporting and Light Activation of Protein Function with Two-Photon Excitation. *J. Am. Chem. Soc.* **136,** 15551–15558 (2014).
21. Chatterjee, A., Guo, J., Lee, H. S. & Schultz, P. G. A genetically encoded fluorescent probe in mammalian cells. *J. Am. Chem. Soc.* **135,** 12540–12543 (2013).
22. Ren, W. & Ai, H. W. Genetically encoded fluorescent redox probes. *Sensors (Switzerland)* **13,** 15422–15433 (2013).
23. Mueller, V. *et al.* A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* **5,** 132–139 (2013).
24. Yang, Y., Song, H. & Chen, P. R. Genetically encoded photocrosslinkers for identifying and mapping protein-protein interactions in living cells. *IUBMB Life* **68,** 879–886 (2016).
25. Chou, C., Uprety, R., Davis, L., Chin, J. W. & Deiters, A. Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem. Sci.* **2,** 480–483 (2011).
26. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci.* **99,** 11020–11024 (2002).
27. Koole, C. *et al.* Genetically-Encoded Photocrosslinkers Determine the Biological Binding Site of Exendin-4 in the N-Terminal Domain of the Intact Human Glucagon-Like Peptide-1 Receptor (GLP-1R). *J. Biol. Chem.* jbc.M117.779496 (2017). doi:10.1074/jbc.M117.779496
28. Chen, H., Venkat, S., McGuire, P., Gan, Q. & Fan, C. Recent development of genetic code expansion for posttranslational modification studies. *Molecules* **23,** 1–19 (2018).
29. Kym, G. *et al.* Genetically Encoded Protein Phosphorylation in Mammalian Cells. *Cell Chem. Biol.* **25,** 1067–1074.e5 (2018).
30. Fan, C., Ip, K. & Söll, D. Expanding the genetic code of Escherichia coli with phosphotyrosine. *FEBS Lett.* **292,** 498–500 (2016).

31. Luo, X. *et al.* Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* 1–7 (2017). doi:10.1038/nchembio.2405
32. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding Nε-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4,** 232–234 (2008).
33. Kim, C. H., Kang, M., Kim, H. J., Chatterjee, A. & Schultz, P. G. Site-specific incorporation of ??-N-crotonyllysine into histones. *Angew. Chemie - Int. Ed.* **51,** 7246–7249 (2012).
34. Rogerson, D. T. *et al.* Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* **11,** 496–503 (2015).
35. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24,** 1436–1440 (2006).
36. Zhang, M. S. *et al.* Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* (2017). doi:10.1038/nmeth.4302
37. Baker, A. S. & Deiters, A. Optical control of protein function through unnatural amino acid mutagenesis and other optogenetic approaches. *ACS Chem. Biol.* **9,** 1398–1407 (2014).
38. Erickson, S. B. *et al.* Precise Photoremovable Perturbation of a Virus???Host Interaction. *Angew. Chemie - Int. Ed.* **56,** 4234–4237 (2017).
39. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A genetically encoded photocaged tyrosine. *Angew. Chemie - Int. Ed.* **45,** 2728–2731 (2006).
40. Kopelowitz, J., Hampe, C., Goldman, R., Reches, M. & Engelberg-Kulka, H. Influence of codon context on UGA suppression and readthrough. *J. Mol. Biol.* **225,** 261–269 (1992).
41. O'Donoghue, P. *et al.* Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. *FEBS Lett.* **586,** 3931–3937 (2012).
42. 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* **57,** 441–445 (2018).
43. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry* **52,** 1828–1837 (2013).
44. Wang, N., Shang, X., Cerny, R., Niu, W. & Guo, J. Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria. *Sci. Rep.* **6,** 21898 (2016).
45. Chatterjee, A., Xiao, H. & Schultz, P. G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. *Proc. Natl. Acad. Sci.* **109,** 14841–14846 (2012).
46. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 1059–1070 (2014).
47. Crnković, A., Suzuki, T., Söll, D. & Reynolds, N. M. Pyrrolysyl-tRNA Synthetase, an Aminoacyl-tRNA Synthetase for Genetic Code Expansion. *Croat. Chem. Acta* **89,** (2016).
48. Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci.* **104,** 11268–11273 (2007).
49. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both

bacteria and eukaryotes. *Nat. Chem. Biol.* 1–8 (2017). doi:10.1038/nchembio.2312
50. Tuley, A., Lee, Y.-J., Wu, B., Wang, Z. U. & Liu, W. R. A genetically encoded aldehyde for rapid protein labelling. *Chem. Commun.* **50,** 7424–7426 (2014).
51. Park, H.-S. *et al.* Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science (80-.).* **333,** 1151–1154 (2011).
52. Zhang, M. S. *et al.* Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat. Methods* (2017). doi:10.1038/nmeth.4302
53. Chatterjee, A., Xiao, H., Yang, P.-Y., Soundararajan, G. & Schultz, P. G. A Tryptophanyl-tRNA Synthetase/tRNA Pair for Unnatural Amino Acid Mutagenesis in *E. coli*. *Angew. Chemie Int. Ed.* **52,** 5106–5109 (2013).
54. Englert, M. *et al.* A genomically modified Escherichia coli strain carrying an orthogonal E. coli histidyl-tRNA synthetase•tRNAHis pair. *Biochim. Biophys. Acta - Gen. Subj.* **1861,** 3009–3015 (2017).
55. Chin, J. W. An Expanded Eukaryotic Genetic Code. *Science (80-.).* **301,** 964–967 (2003).
56. Italia, J. S., Latour, C., Wrobel, C. J. J. & Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. *Cell Chem. Biol.* **25,** 1304–1312.e5 (2018).
57. Khow, O. & Suntrarachun, S. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pac. J. Trop. Biomed.* **2,** 159–162 (2012).
58. Liu, C. C., Choe, H., Farzan, M., Smider, V. V. & Schultz, P. G. Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. *Biochemistry* **48,** 8891–8898 (2009).
59. Li, X. *et al.* Characterization of a sulfated anti-HIV antibody using an expanded genetic code Characterization of a sulfated anti-HIV antibody using an expanded genetic code. (2018). doi:10.1021/acs.biochem.8b00374
60. Hortin, G., Tollefsen, D. M. & Strauss, A. W. Identification of two sites of sulfation of human heparin cofactor II. *J. Biol. Chem.* **261,** 15827–15830 (1986).
61. Xiao, H., Xuan, W., Shao, S., Liu, T. & Schultz, P. G. Genetic Incorporation of ε- N -2-Hydroxyisobutyryl-lysine into Recombinant Histones. *ACS Chem. Biol.* **10,** 1599–1603 (2015).
62. Neumann, H., Hazen, J. L., Weinstein, J., Mehl, R. A. & Chin, J. W. Genetically encoding protein oxidative damage. *J. Am. Chem. Soc.* **130,** 4028–4033 (2008).
63. Cooley, R. B. *et al.* Structural basis of improved second-generation 3-nitro-tyrosine tRNA synthetases. *Biochemistry* **53,** 1916–1924 (2014).
64. Xie, J., Supekova, L. & Schultz, P. G. A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli. *ACS Chem. Biol.* **2,** 474–478 (2007).
65. Xiao, H. *et al.* Genetic incorporation of multiple unnatural amino acids into proteins in Mammalian cells. *Angew. Chem. Int. Ed. Engl.* **52,** 14080–14083 (2013).
66. Kelemen, R. E. *et al.* A Precise Chemical Strategy To Alter the Receptor Specificity of the Adeno-Associated Virus. *Angew. Chemie Int. Ed.* 10–15 (2016). doi:10.1002/anie.201604067
67. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126,** 15046–15047 (2004).
68. Seitchik, J. L. *et al.* Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. *J. Am. Chem. Soc.* **134,** 2898–2901 (2012).

69. Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 56–61 (2003).
70. Brustad, E. *et al.* A genetically encoded boronate-containing amino acid. *Angew. Chemie - Int. Ed.* **47,** 8220–8223 (2008).
71. Addy, P. S., Italia, J. S. & Chatterjee, A. An oxidative bioconjugation strategy targeted to a genetically encoded 5-hydroxytryptophan. *ChemBioChem* **19,** 1375–1378 (2018).
72. Addy, P. S., Zheng, Y., Italia, J. S. & Chatterjee, A. A 'quenchergenic' chemoselective protein labeling strategy. *ChemBioChem* (2019). doi:10.1002/cbic.201800817
73. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat Chem Biolfile///Users/jamesitalia/Documents/Chatterjee Lab/Papers/c4cc02000f.pdf* **13,** 446–450 (2017).
74. Sachdeva, A., Wang, K., Elliott, T. & Chin, J. W. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *J. Am. Chem. Soc.* **136,** 7785–7788 (2014).
75. Baumann, T. *et al.* Prospects of in vivo incorporation of non-canonical amino acids for the chemical diversification of antimicrobial peptides. *Front. Microbiol.* **8,** 1–9 (2017).
76. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464,** 441–444 (2010).
77. Wan, W. *et al.* A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chemie - Int. Ed.* **49,** 3211–3214 (2010).
78. Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* **6,** 393–403 (2014).
79. Zheng, Y., Gilgenast, M. J., Hauc, S. & Chatterjee, A. Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis. *ACS Chem. Biol.* **13,** 1137–1141 (2018).
80. Niu, W., Schultz, P. G. & Guo, J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **8,** 1640–1645 (2013).
81. Zheng, Y., Addy, P. S., Mukherjee, R. & Chatterjee, A. Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells. *Chem. Sci.* **8,** 7211–7217 (2017).
82. Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda Red Recombineering in Escherichia coli Occurs Through a Fully Single-Stranded Intermediate. *Genetics* **186,** 791–799 (2010).
83. Fathi-Roudsari, M., Akhavian-Tehrani, A. & Maghsoudi, N. Comparison of three Escherichia coli strains in recombinant production of reteplase. *Avicenna J. Med. Biotechnol.* **8,** 16–22 (2016).
84. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science (80-.).* **342,** 357–360 (2013).

Chapter 2

An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes

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

Italia, J.S., Addy, P.S., Wrobel, C.J., Crawford, L.A., Lajoie, M.J., Zheng, Y., Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat. Chem. Biol.,* **2017,** 13, 446-450.

2.1 Introduction

2.1.1 Overview

In this study, we demonstrate the feasibility of expanding the genetic code of *Escherichia coli* using its own tryptophanyl–tRNA synthetase and tRNA (TrpRS–$tRNA^{Trp}$) pair. This was made possible by first functionally replacing this endogenous pair with an *E. coli*–optimized counterpart from *Saccharomyces cerevisiae*, and then reintroducing the liberated E. coli TrpRS–$tRNA^{Trp}$ pair into the resulting strain as a nonsense suppressor, which was then followed by its directed evolution to genetically encode several new unnatural amino acids (UAAs). These engineered TrpRS– $tRNA^{Trp}$ variants were also able to drive efficient UAA mutagenesis in mammalian cells. Since bacteria-derived aminoacyl–tRNA synthetase (aaRS)–tRNA pairs are typically orthogonal in eukaryotes, our work provides a general strategy to develop additional aaRS–tRNA pairs that can be used for UAA mutagenesis of proteins expressed in both *E. coli* and eukaryotes.

2.1.2 Genetic code expansion and previous selection system paradigms

The ability to site-specifically incorporate unnatural amino acids (UAAs) into a protein in living cells has emerged as a powerful method to probe and manipulate its structure and function.[1–3] Central to this technology is an engineered aminoacyl–tRNA synthetase (aaRS)–tRNA pair that delivers a desired UAA in response to a nonsense or frameshift codon. Such a UAA-specific aaRS–tRNA pair must not cross-react with its host counterparts (i.e., must be orthogonal) to maintain the fidelity of translation. To ensure the lack of such cross-reactivity, candidates for the development of UAA-specific orthogonal aaRS–tRNA pairs are imported into a host cell from a different domain of life harboring evolutionarily divergent translational components.[2,3] Thus, genetic code expansion of bacteria relies upon aaRS–tRNA pairs derived

from eukaryotes or archaea, while in eukaryotes such expansion generally utilizes bacterial pairs (archaeal aaRS–tRNA pairs are typically cross-reactive with their eukaryotic counterparts). The use of two distinct sets of aaRS–tRNA pairs for genetic code expansion in eukaryotes and bacteria leads to a substantial disadvantage: each desirable UAA must be separately genetically encoded using two distinct platforms for applications in bacterial and eukaryotic cells.

Two *in vivo*–selection systems have been developed so far to enable the generation of UAA-specific aaRS variants; these use *E. coli*[4] or *S. cerevisiae*[5] (yeast) as selection hosts and enable the directed evolution of eukaryotic–archaeal or bacterial aaRS–tRNA pairs, respectively. Because of its facile nature, the *E. coli*–based selection platform has been substantially more successful for genetically encoding new UAAs relative to its yeast counterpart (Figure 2-1). Nearly half of the UAAs genetically encoded with the yeast selection system use a single *E. coli* leucyl-tRNA synthetase variant that exhibits high polyspecificity towards structurally homologous substrates.

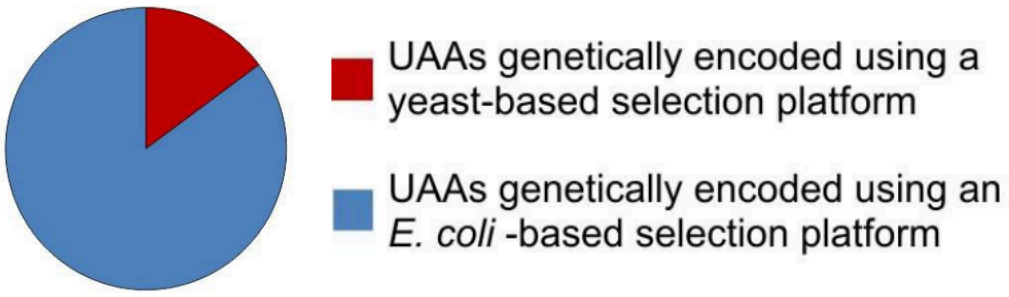


Figure 2-1. Relative comparison of UAA incorporated with yeast or bacterial based selection platforms, derived from Dumas *et. al*[85] in 2016. Comparing the number of UAAs genetically encoded using the *E. coli* based selection system relative to those using a yeast based selection system.

However, the *E. coli*-based selection platform cannot be used to evolve bacteria-derived aaRS–tRNA pairs, which are generally used to drive eukaryotic genetic code expansion.[7,85] Only the archaea-derived pyrrolysyl–tRNA synthetase (PylRS)–tRNAPyl pair, which is orthogonal in both bacteria and eukaryotes owing to its unique structural features,[1–3,6] overcomes this limitation and enables the use of the facile *E. coli*–based selection system to genetically encode new UAAs in eukaryotic cells. [1–3,6] As a result, the introduction of new UAAs into the eukaryotic genetic code has become exclusively dependent on this platform (Figure 2-2).[2] Indeed, all new UAAs genetically encoded in eukaryotes over the last 6 years have taken advantage of the pyrrolysyl system (Figure 2-2).

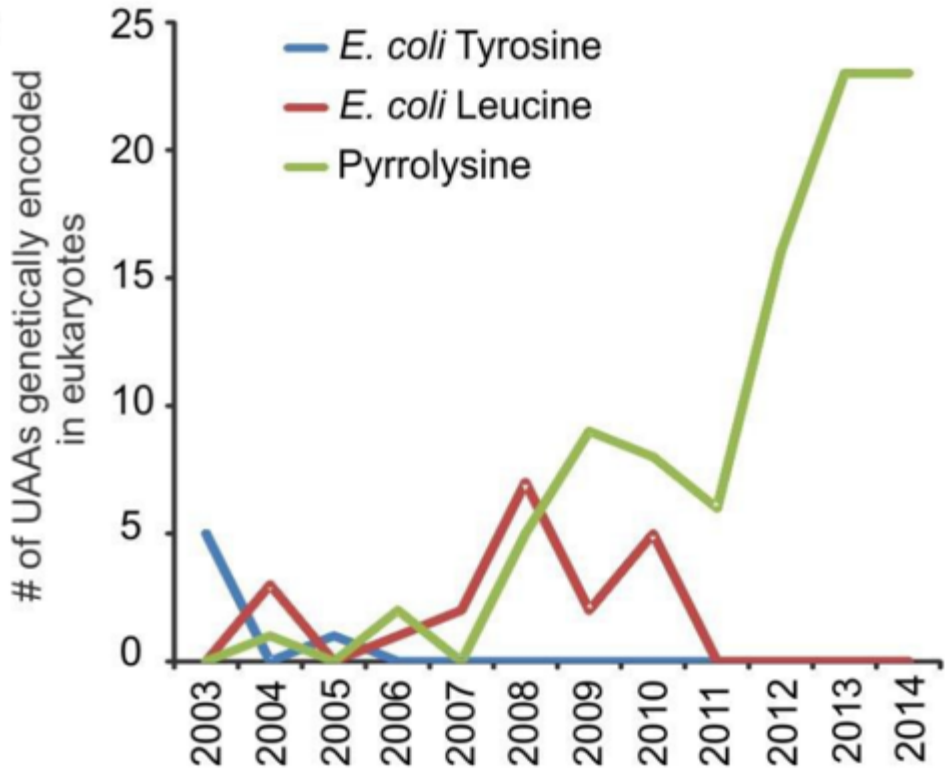


Figure 2-2. Number of UAAs genetically encoded in eukaryotes using the three most established tRNA/aaRS pairs, compiled via Liu *et al.*[3] and Dumas *et al.*[2]

2.1.3 Altered Translational Machinery Overview

Here we report a novel strategy to circumvent these limitations by taking advantage of an *E. coli* strain in which one of the native aaRS–tRNA pairs is functionally replaced with a eukaryotic–archaeal counterpart (Figure 2-3). The 'liberated' aaRS–tRNA pair can then be reintroduced into the resulting 'altered translational machinery (ATM)' *E. coli* as a nonsense suppressor, and can be evolved to charge desirable UAAs. Owing to its bacterial origin, the same pair can also be directly used for eukaryotic genetic code expansion. Our strategy will enable the development of additional 'universal' aaRS–tRNA pairs that share the unique advantages of the pyrrolysyl system but provide access to new active site topologies for genetically encoding UAAs inaccessible to PylRS. The development of multiple such aaRS–tRNA pairs—each capable of incorporating a rich set of UAAs—will also facilitate site-specific incorporation of multiple distinct UAAs into proteins in both eukaryotes and bacteria.[7–10]

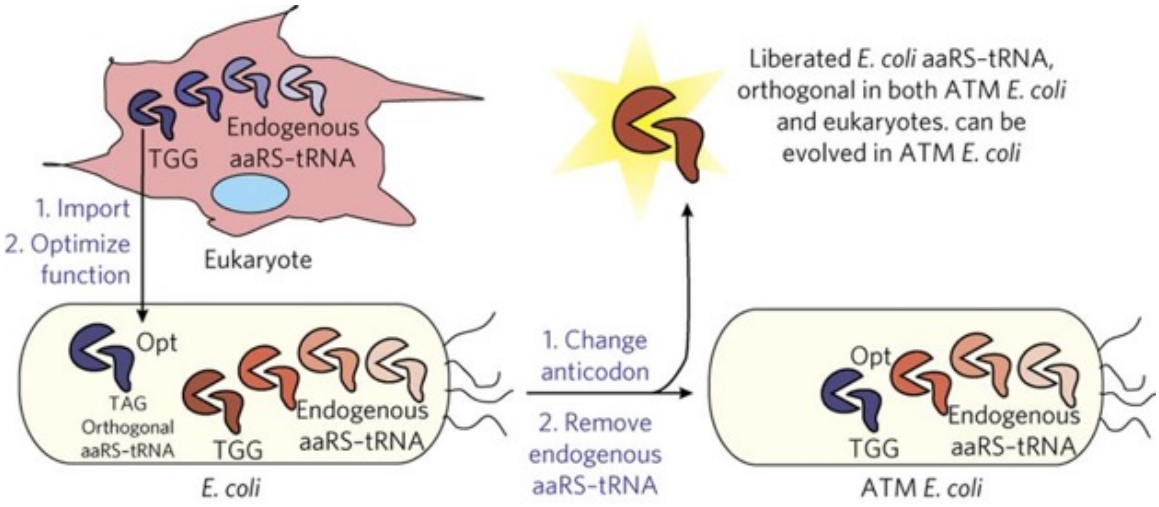


Figure 2-3. Overview of the strategy to create an aaRS-tRNA pair that can drive genetic code expansion in both eukaryotes and *E. coli.*

2.2. Results and Discussion

2.2.1 Functionally replacing the TrpRS-tRNA^{Trp} pair of E. coli

The feasibility of functionally replacing an endogenous aaRS–tRNA pair of E. coli with an evolutionarily distant counterpart has been shown before.[11] However, the resulting strains often exhibit growth defects, presumably due to the suboptimal interaction of the heterologous aaRS–tRNA with the translational apparatus of E. coli. Moreover, the suitability of the resulting ATM *E. coli* strain as the selection host for evolving variants of the 'liberated' endogenous aaRS–tRNA pair with distinct substrate specificities has not yet been demonstrated. We envisioned that it should be possible to achieve more efficient functional replacement of the endogenous pair by first optimizing the performance of the substituting aaRS–tRNA pair in *E. coli*, which may alleviate the growth defect associated with the resulting engineered strain (Figure 2-3). A number of heterologous aaRS–tRNA pairs have already been engineered for enhanced performance in *E. coli*, providing a pool of potential candidates.

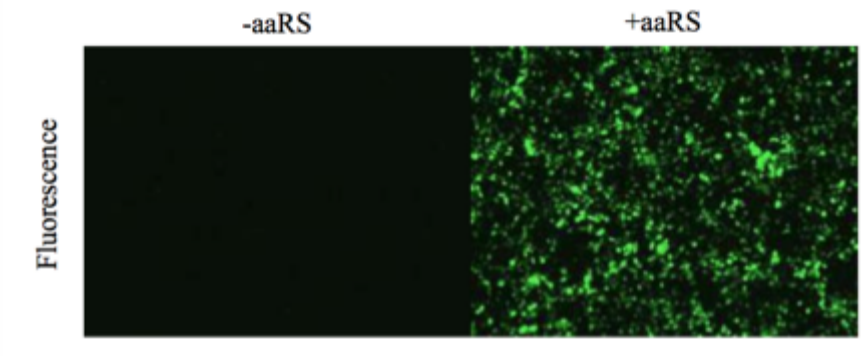


Figure 2-4. Fluorescence microscopy showing the expression of EGFP-39-TAG reporter in HEK293T cells co-expressing $tRNA^{EcTrp}_{CUA}$, in the presence and absence of its cognate EcTrpRS.

We targeted the development of the tryptophanyl–tRNA synthetase (TrpRS)–$tRNA^{Trp}$ pair of *E. coli* for genetic code expansion of both *E. coli* and eukaryotes using this approach. The unique active site of TrpRS should enable the introduction of structurally novel UAAs into the genetic code of eukaryotes as well as *E. coli*. To verify whether the *E. coli* TrpRS–$tRNA^{Trp}$ pair is indeed suitable for eukaryotic genetic code expansion, we co-expressed in HEK293T cells the TAG-suppressing *E. coli* $tRNA^{Trp}$ ($tRNA^{EcTrp}_{CUA}$) and an enhanced green fluorescent protein (EGFP) reporter harboring a stop codon at a permissive site (EGFP-39-TAG), with or without the cognate synthetase. Robust EGFP expression was only observed in the presence of the EcTrpRS (Figure 2-4), suggesting that (1) EcTrpRS–$tRNA^{EcTrp}_{CUA}$ is capable of efficient TAG suppression in eukaryotic cells, and (2) the $tRNA^{EcTrp}_{CUA}$ is non-cross-reactive in eukaryotic cells.

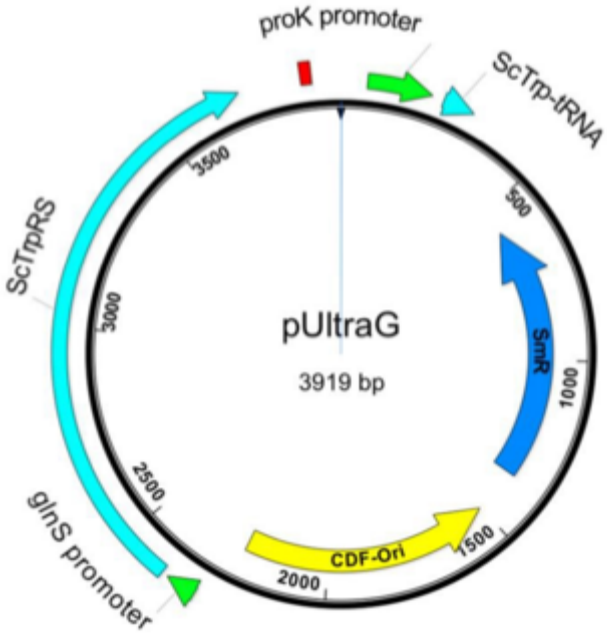


Figure 2-5. Complementation plasmid map. Note: In this first ATM system, the complementation plasmid did not hinder our ability to generate the ATMW strain, but it was later shown that the tyrosine strain contains growth defects which are likely due to poor

complementation efficiency (see Chapter 3). It is currently unclear whether stronger expression of the tRNA/aaRS pair would benefit the complementation, but rather, a sense codon optimized orthogonal pair would likely create the largest benefit for recombination. For those interested in pursuing this avenue of research, please look into the conditional aaRS expression plasmids I cloned. These plasmids can be used to turn on/off *E. coli* aaRS expression and can therefore be used to select and characterize the best complementation plasmids.

An *S. cerevisiae*–derived ScTrpRS–$tRNA^{ScTrp}$ pair has already been optimized in *E. coli* for highly efficient nonsense suppression, providing a good candidate for substituting its endogenous counterpart.[12,13] To enable the deletion of the endogenous tryptophanyl pair from the *E. coli* genome, we first generated a plasmid (pUltraG-ScW40CCA) expressing the aforementioned engineered ScTrpRS–$tRNA^{ScTrp}_{CCA}$ pair derived from yeast (Figure 2-5).[12,13] This plasmid harbors a unique CloDF13 origin of replication that allows stable propagation of most commonly used plasmids, which use the ColE1 or p15a origin of replication, in the same cell. In addition, pUltraG-ScW40$_{CCA}$ contains a *glnS'* promoted wild-type yeast TrpRS, *proK* promoted engineered yeast Trp-$tRNA_{CCA,}$[13] and spectinomycin selectable marker.

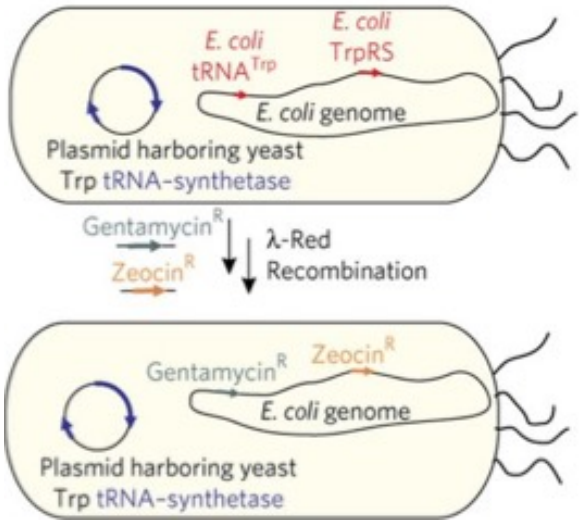


Figure 2-6. Depiction of the strategy to replace the TrpRS-tRNA^{Trp} pair from the *E. coli* genome.

The K12-derived EcNR1 strain of *E. coli*, which encodes a temperature-inducible λ-Red recombination system, was used as the host.[14] We aimed to replace the *trpS* (encoding EcTrpRS) and *trpT* (encoding tRNA$^{EcTrp}_{CCA}$) genes from the EcNR1 genome using zeocin- and gentamycin-selectable markers (Figure 2-6), respectively. We chose to use antibiotic cassettes due to the simplicity of selection for recombinant hits at the time. We first replaced EcTrpRS with zeocin resistance cassettes.

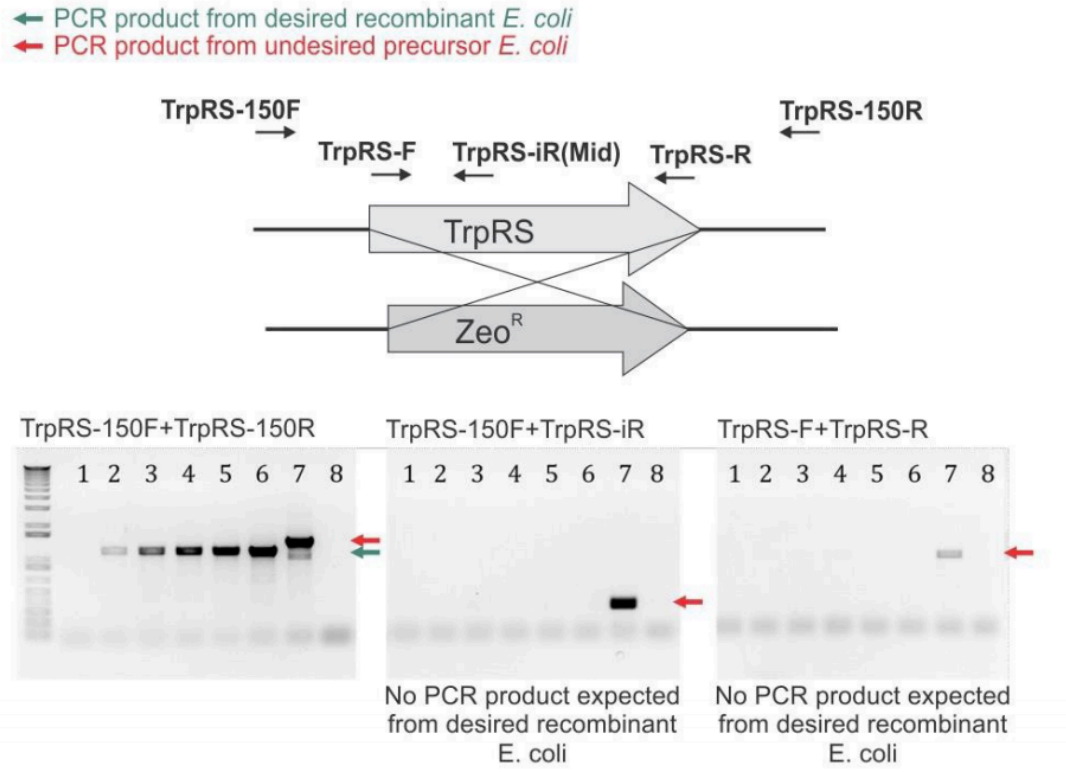


Figure 2-7. Colony PCR based verification of successful genomic alterations of TrpRS. Lanes 1-6 are potential hits, lane 7 is the parent strain EcNR1, and lane 8 is a no *E. coli* control (Figure 2-7). It is important to note that a major troubleshooting solution was discovered here: when performing the colony PCR, make sure to reduce the number of cycles to ~24-26. Too many cycles will result in amplification of minimal amounts of bacterial DNA. Additionally, make sure to streak potential hits prior to analysis, as there can be wild type contaminant present in the colony PCR.

Next, a gentamycin-resistance cassette was used to replace $tRNA^{EcTrp}_{CCA}$ (*trpT*). Successful hits contain larger PCR products due to the introduction of the larger gentamycin resistance cassette.

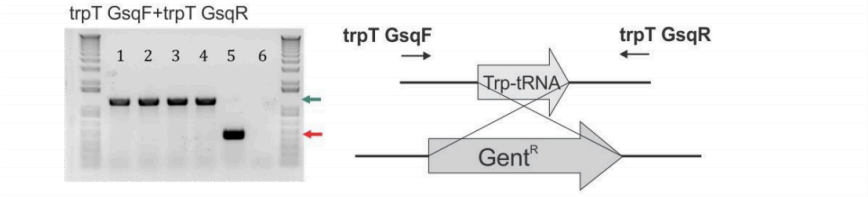


Figure 2-8. Colony PCR based verification of successful genomic alterations of trpT. Lanes 1-4 are potential hits, lane 5 is the precursor strain, and lane 6 is the no *E. coli* control.

At this stage, our strain was spectinomycin, zeocin, and gentamycin resistant but it was necessary to remove the Red recombination system due to its temperature sensitiviy. We began to run out of antibiotic cassettes, as we needed to save kanamycin (pBK), tetracycline (pRep), ampicillin (pNeg), and chloramphenicol (pEvol) for plasmid propogation. Therefore, we turned to a metabolically selectable genetic marker, galK.[15] Using a 90bp oligo, we were able to remove *galK* from its endogenous location in the *E. coli* genome, liberating its use for deletion of the Red recombination system (Figure 2-9).

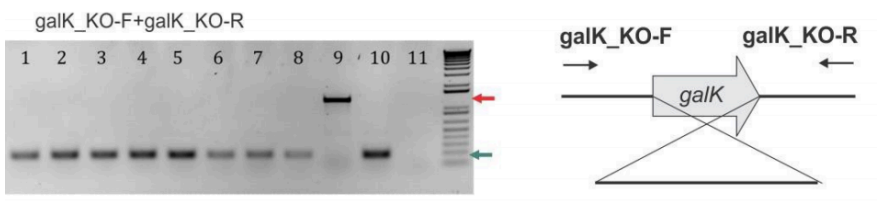


Figure 2-9. Colony PCR based verification of successful genomic alterations of galK. Successful hits (lanes 1-8) contain smaller PCR products. Lane 9 is the wild-type *E. coli*, while lane 10 was a previously deleted *galK* strain. Lane 11 is the no *E. coli* control.

Finally, we removed the λ prophage, which encodes the Red recombination system, from the genome using a galactose-selectable galK marker (Figure 2-10)[15] to create a strain named ATMW1. In Figure 2-10, The final ATMW1 strain has the following genotype: EcNR1 pUltraG-ScW40CCA *trpS*::ZeoR *trpT*::GentR ⊿*galK* λRED::*galK*.

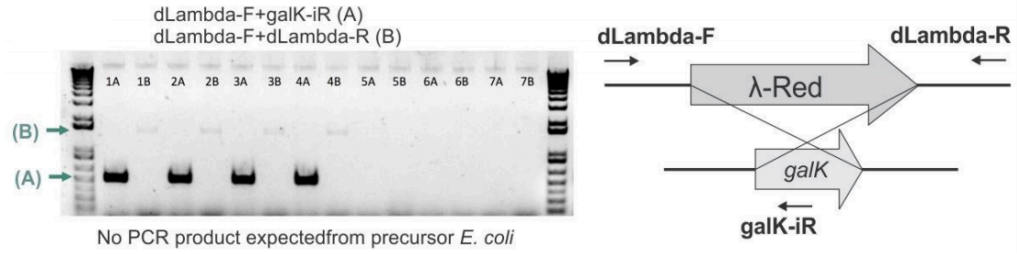


Figure 2-10. Colony PCR based verification of successful genomic alterations of λ-Red. Two different primer combinations were used to test the proper deletion of the prophage. Primer combo A results in a smaller band, as the forward primer anneals upstream of the recombination site while the internal reverse primer sits on an internal site of *galK*. Primer combo B anneals 150bp up and downstream from the recombination site. A band will only be present if galK has replaced λ-RED, as the λ-RED site is too large for PCR amplification. Lanes 1-4 are successful ATMW1 hits, 5 is EcNR1 precurser, 6 is Top10 containing no λ-RED, and 7 is PBS.

The ATMW1 strain exhibited no observable growth defect relative to its progenitor strain, confirming efficient functional substitution by the engineered ScTrpRS–tRNAScTrpCCA pair (Figure 2-11). EcNR1G is identical to EcNR1 but contains the *galK* knockout. This lack of a growth defect may be due to the low abundance of TGG codons in

comparison to other tRNA/aaRS pairs. Care should be taken when optimizing higher complexity ATM strains.

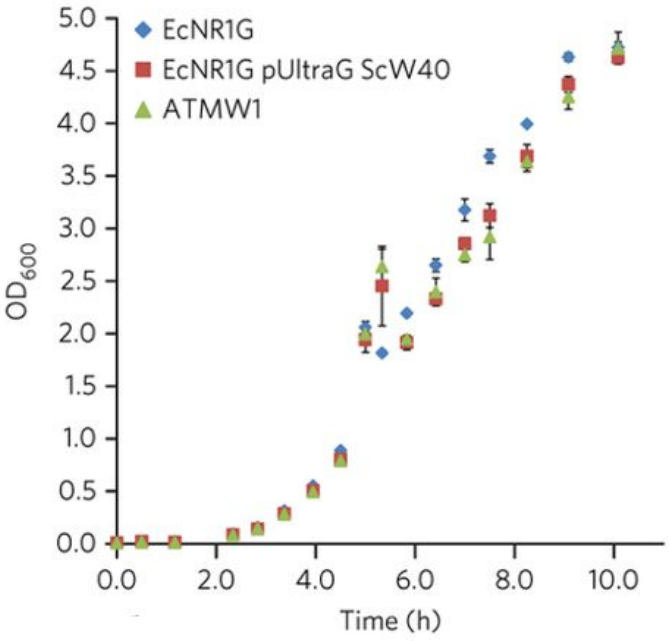


Figure 2-11. Growth rate comparison of ATMW1 to progenitor EcNR1G strain.

2.2.2 Reintroducing the EcTrpRS-tRNA^{EcTrp} pair into ATMW1

With ATMW1 in hand, we investigated the possibility of reintroducing the liberated EcTrpRS–tRNA$^{EcTrp}_{CUA}$ pair into this strain as an orthogonal nonsense suppressor. It has been previously reported that the TAG suppressor tRNA$^{EcTrp}_{CUA}$ is mischarged by *E. coli* glutaminyl–tRNA synthetase (EcGlnRS).[16,17] To evaluate the impact of this cross-reactivity, we expressed tRNA$^{EcTrp}_{CUA}$ in ATMW1 along with a superfolder GFP reporter (sfGFP-151-TAG), and observed robust protein expression in the absence of EcTrpRS (Figure 2-12). An analogous experiment with a chloramphenicol acetyltransferase (CAT) reporter containing a TAG codon further confirmed the substantial nature of this cross-reactivity (Figure 2-12). Next, we isolated the sfGFP-151-TAG reporter by Ni–NTA affinity purification, and analyzed it by mass

spectrometry (whole protein, as well as tryptic digestion–MS) to confirm the incorporation of glutamine into the TAG site (Figure 2-13). It was clear that the CUA anticodon was the cause of the misincorporation in the absence of the EcTrpRS.

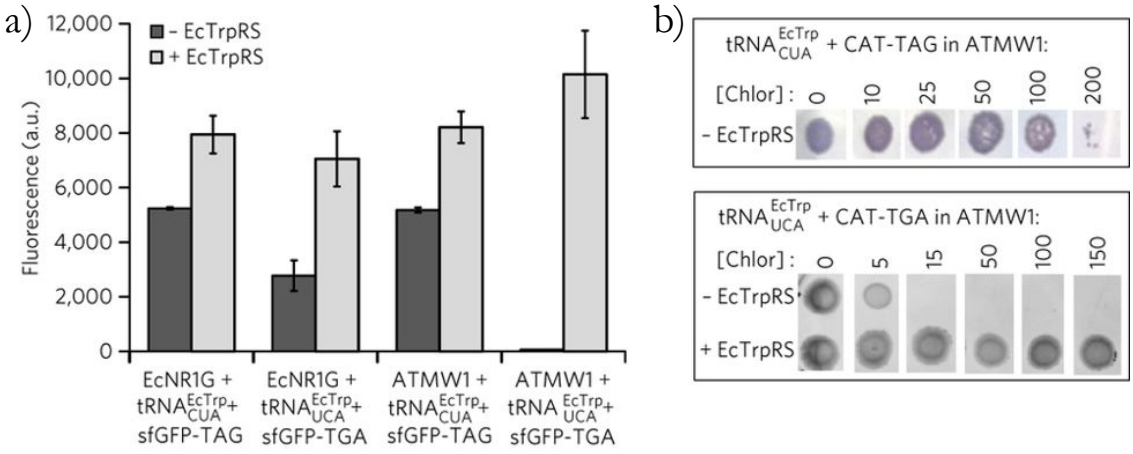


Figure 2-12. Establishing the EcTrpRS-tRNA^{EcTrp} pair as an orthogonal nonsense suppressor in ATMW1. a) Evaluating the nonsense-suppression efficiencies of tRNA$^{EcTrp}_{CUA}$ and tRNA$^{EcTrp}_{UCA}$ using sfGFP-151-TAG or sfGFP-151TGA as reporters, respectively, in ATMW1 and its progenitor strain EcNR1G in the presence or absence of EcTrpRS overexpression. tRNA$^{EcTrp}_{CUA}$ is active in the absence of EcTrpRS in ATMW1, but tRNA$^{EcTrp}_{UCA}$ is not. However, in the presence of EcTrpRS co-expression, tRNA$^{EcTrp}_{UCA}$ can facilitate the expression of sfGFP-151-TGA at high levels in ATMW1. b) This is further confirmed by evaluating the nonsense-suppression efficiencies of tRNA$^{EcTrp}_{CUA}$ and tRNA$^{EcTrp}_{UCA}$ in ATMW1 using chloramphenicol acetyltransferase (CAT) reporters harboring a TAG or a TGA codon, and measuring the survival of the resulting strains on solid medium supplemented with increasing concentration of chloramphenicol.

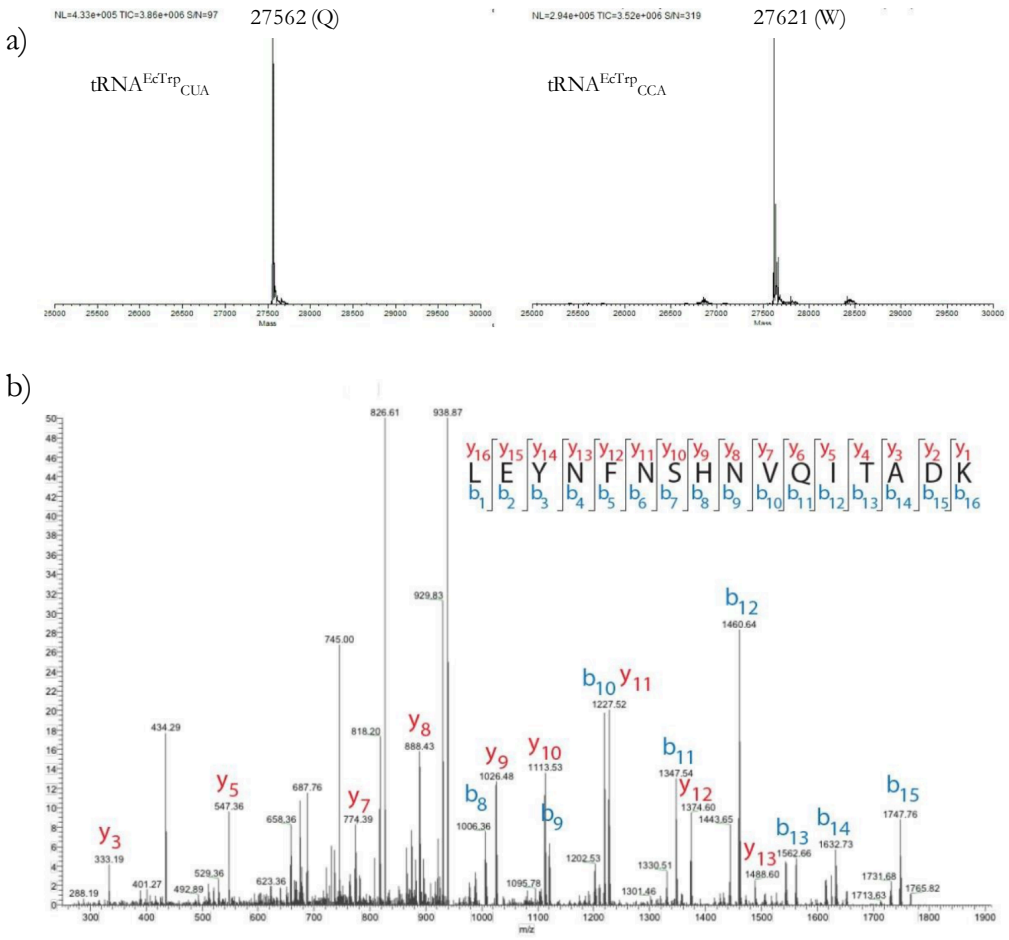


Figure 2-13. Mass spectrometry analysis of glutamine misincorporation. a) ESI-MS analysis of sfGFP-151-TAG purified from ATMW1 co-expressing tRNA^{EcTrp} CUA alone (left) or UCA with EcTrpRS (right). The observed mass (27562) is consistent with the incorporation glutamine, glutamate, or a lysine residue at the nonsense codon. The observed mass (27621) is consistent with the incorporation of tryptophan in response to TGA at the sfGFP-151 site. b) Tryptic digestion of this protein followed by LCMS/MS analysis of the resulting peptides confirms the incorporation of glutamine in response to TAG.

Despite the remarkable homology between the tryptophanyl and glutaminyl tRNAs of *E. coli*, EcGlnRS selectively recognizes its cognate tRNA$^{EcGln}_{CUG}$ using the middle U residue in

the anticodon of $tRNA^{EcGln}_{CUG}$ as one of the major identity elements (Figure 2-14).[17,18] While the endogenous $tRNA^{EcTrp}_{CCA}$ lacks this residue, enabling it to be distinguished from $tRNA^{EcGln}_{CUG}$, its presence in the TAG suppressor $tRNA^{EcTrp}_{CUA}$ leads to cross-reactivity (Figure 2-14). Although TAG suppression nearly exclusively forms the basis of genetic code expansion in *E. coli*,[2,3] the aforementioned cross-reactivity precludes its use with the EcTrpRS-$tRNA^{EcTrp}$ pair. We circumvented this issue by generating a TGA suppressor $tRNA^{EcTrp}_{UCA}$ that avoids introducing the middle U residue in the anticodon. However, termination at the TGA stop codon in E. coli is often 'leaky'—a result of nonspecific suppression by the endogenous tryptophanyl tRNA—making it a suboptimal choice for genetic code expansion.[19,20] We found that in our ATMW1 strain, in which the endogenous tryptophanyl pair was replaced with the yeast counterpart, TGA did not exhibit such leaky behavior (Figure 2-15), suggesting that it could feasibly be used for genetic code expansion with high fidelity.

When the sfGFP-151-TGA reporter and $tRNA^{EcTrp}_{UCA}$ were co-expressed in ATMW1, no reporter expression was observed unless the EcTrpRS was also present, confirming the lack of cross-reactivity of $tRNA^{EcTrp}_{UCA}$ in ATMW1, as well as the efficient opal suppression activity of the EcTrpRS– $tRNA^{EcTrp}_{UCA}$ pair (Figure 2-12a). MS analysis of the isolated protein further confirmed incorporation of tryptophan in response to TGA (Figure 2-13b). The lack of cross-reactivity of $tRNA^{EcTrp}_{UCA}$ in ATMW1 was further confirmed using a CAT reporter harboring a TGA codon (Figure 2-12b).

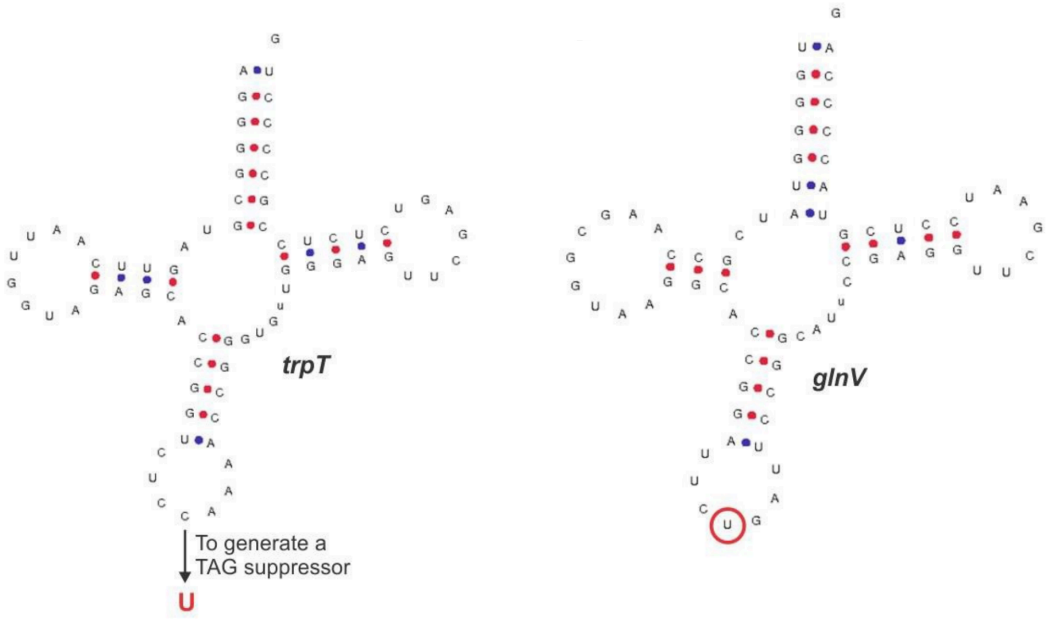


Figure 2-14. Predicted secondary structures of trytophan tRNA *trpT* and glutamine tRNA *glnV*.

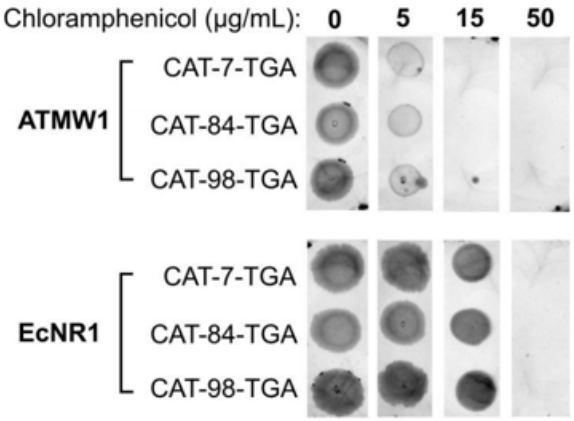
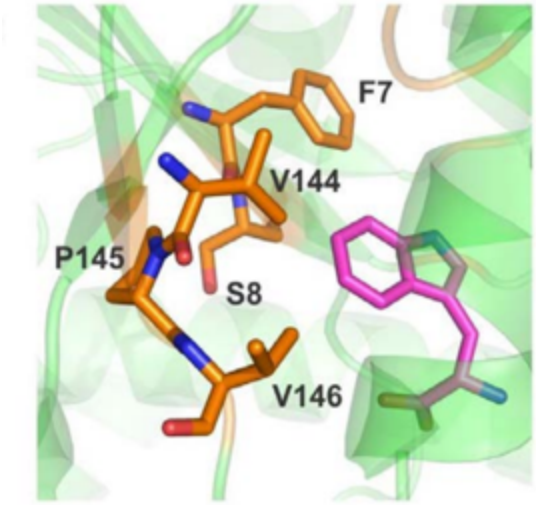


Figure 2-15. Chloramphenicol assay to assess TGA leakiness in *E. coli*.

2.2.3 Evolution of EcTrpRS in ATMW1 to selectively charge 5HTP

Establishment of an efficient, orthogonal opal-suppressing EcTrpRS–$tRNA^{EcTrp}_{UCA}$ pair in the ATMW1 *E. coli* creates an exciting opportunity to alter its substrate specificity using the facile *E. coli*–based selection system. The existing reporter plasmids used for this double-sieved selection scheme,[4,21] which enables either enrichment (positive selection) or depletion (negative selection) of aaRS variants on the basis of their ability to charge its amber (TAG)-suppressing cognate $tRNA_{CUA}$, were mutated to generate variants that would instead allow analogous selection based on opal (TGA) suppression. Guided by the crystal structure (Figure 2-16) of the highly homologous *Geobacillus stearothermophilus* TrpRS (PDB ID 1I6M), we constructed a library of 3.15 × 106 EcTrpRS mutants by simultaneously randomizing Phe7(NBT), Ser8(NST), Val144(NNK), Pro145(NNK), and Val146(NNK) residues using site-saturation mutagenesis, and covered the library using >3 × 107 unique transformants. The randomized residues point at C4–C5–C6 of the indole ring of the substrate tryptophan. We first attempted to identify a mutant from this library that selectively charges 5-hydroxytryptophan (5HTP). This UAA was previously genetically encoded in bacteria using a yeast-derived tryptophanyl pair.[13] A report claiming its incorporation in eukaryotic cells using a bacteria-derived tRNA–aaRS pair[22] was recently refuted.[23]



	F7	S8	V144	P145	V146
h9	F	A	S	P	A
h10	F	A	G	P	I
h13	F	A	A	P	A
h14	F	A	G	P	C

Figure 2-16. *Geobacillus sterothermophilus* TrpRS crystal structure (PDB 1I6M) active site and mutations associated with 5HTP-specific EcTrpRS variants.

We subjected the aforementioned library of EcTrpRS variants to three rounds of selections (positive selection in the presence of 1 mM 5HTP, negative selection in the absence of the UAA, and then another round of positive selection) in the ATMW1 strain. Ninety-six colonies from the surviving pool were individually screened for conditional survival under the positive-selection conditions (40 μg/ml chloramphenicol) in the presence of 5HTP. Sequence analysis of four of the most successful mutants show substantial sequence convergence, in which Phe7 and Pro145 are conserved, Ser8 is mutated to Ala, Val144 changes to a small amino acid (Gly, Ser, or Ala), and Val146 is mutated to different small and hydrophobic amino acids (Figure 2-16). Next, we evaluated the ability of these mutant EcTrpRS variants to drive the expression of a sfGFP-151-TGA reporter along with its cognate $tRNA^{EcTrp}_{UCA}$. All mutants

were able to facilitate efficient reporter expression in the presence of 1 mM 5HTP, but EcTrpRS-h14 exhibited the least background reporter expression in the absence of the UAA (Figure 2-17a). The reporter protein was isolated using a C-terminal $(His)_6$ tag and subjected to SDS–PAGE (Figure 2-17b) and to MS analysis to confirm 5HTP incorporation. Yields of the mutant proteins were up to 65% of the wild-type sfGFP reporter.

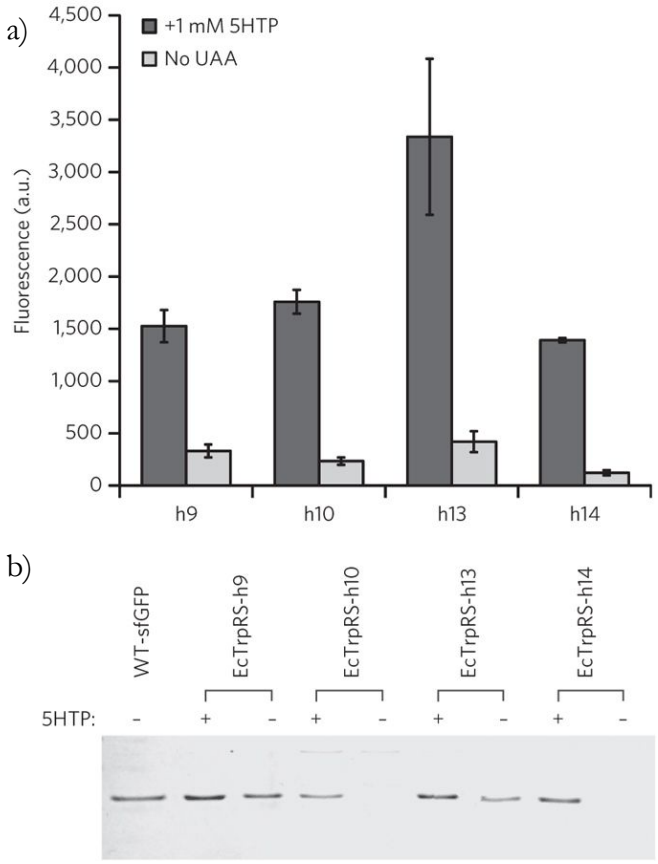


Figure 2-17. Evolution of EcTrpRS to develop mutants charging 5HTP with high fidelity and efficiency. a) 5HTP-dependent expression of sfGFP-151-TGA in ATMW1 *E. coli* using EcTrpRS-h9, h10, h13, and 14, measured as OD600-normalized fluorescence of sfGFP in cells resuspended in PBS. b) SDS-PGE analysis of sfGFP-151-TGA expressed in ATMW1 *E. coli* facilitated by various EcTrpRS variants in the presence or absence of 1mM 5HTP.

2.2.4 Substrate Polyspecificity of the evolved EcTrpRS variants

Since the negative selection step in the aforementioned scheme discriminates only against natural amino acids, but not other UAAs, the isolated mutants are sometimes capable of polyspecificity: the ability to charge a number of structurally similar UAAs, while discriminating against the 20 canonical amino acids.[24,25] We screened our isolated EcTrpRS mutants for polyspecificity toward other 5-substituted tryptophan derivatives (Figure 2-18) using an assay that measures the enhancement of sfGFP-151-TGA expression in the presence of a particular UAA, relative to a no-UAA control. EcTrpRS-h14 exhibited high polyspecificity toward four additional amino acids, whereas EcTrpRS-h9 also enabled the incorporation of 5AmW (Figure 2-19a). In all cases, the reporter protein was isolated using a C-terminal $(His)_6$ tag and subjected to SDS–PAGE (Figure 2-19b) and MS analysis to confirm the incorporation of the corresponding UAAs (Figure 2-19c).

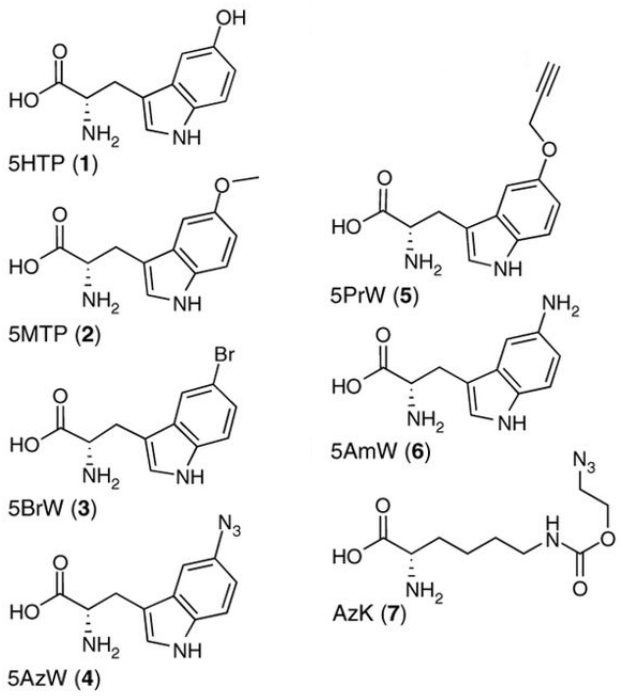
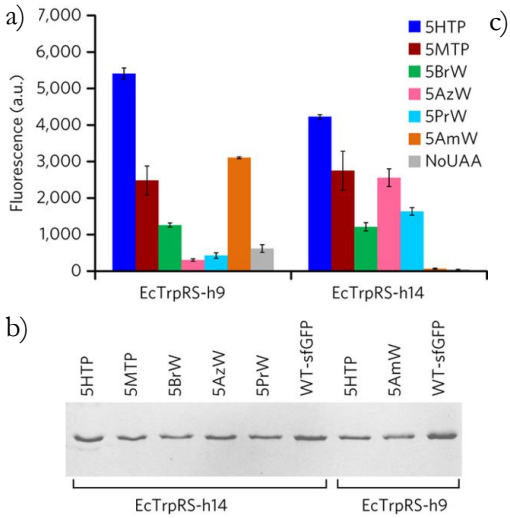


Figure 2-18. Structures of UAA used in this study



c)

Reporter	Amino acid	aaRS	Expected mass	Observed mass	Note
sfGFP-151TAG	Gln	none	27620	27562	pBK/tac system in ATMW1 (Glnx-reactive)
sfGFP-151-TGA	Trp	EcWRSwt	27620	27621	pBK/tac system in ATMW1
sfGFP-151-TGA	5HTP	h14	27636	27636	pBK/tac system in ATMW1
sfGFP-151-TGA	5MTP	h14	27652	27650	pBK/tac system in ATMW1
sfGFP-151-TGA	5BrW	h14	27700	27699	pBK/tac system in ATMW1
sfGFP-151-TGA	5AzW	h14	27663	27660	pBK/tac system in ATMW1
sfGFP-151-TGA	5PrW	h14	27676	27674	pBK/tac system in ATMW1
sfGFP-151-TGA	5HTP	h9	27636	27637	pBK/tac system in ATMW1
sfGFP-151-TGA	5AmW	h9	27636	27635	pBK/tac system in ATMW1
sfGFP-151-TGG	Trp	h14	27620	27620	pBK/tac system in ATMW1
EGFPwt	Tyr	None	29683	29683	pAcBac1 system in HEK293T
EGFP-39-TAG	BocK	MbPyl	29748	29748	pAcBac1 system in HEK293T
EGFP-39-TAG	AzK	MbPyl	29761	29761	pAcBac1 system in HEK293T
EGFP-39-TAG	Trp	EcWRSwt	29707	29708	pAcBac1 system in HEK293T
EGFP-39-TAG	5HTP	h14	29723	29724	pAcBac1 system in HEK293T
EGFP-39-TAG	5MTP	h14	29736	29736	pAcBac1 system in HEK293T
EGFP-39-TAG	5BrW	h14	29786	29785	pAcBac1 system in HEK293T
EGFP-39-TAG	5AzW	h14	29748	29747	pAcBac1 system in HEK293T
EGFP-39-TAG	5PrW	h14	29761	29760	pAcBac1 system in HEK293T
EGFP-39-TAG	5HTP	h9	29723	29724	pAcBac1 system in HEK293T
EGFP-39-TAG	5AmW	h9	29722	29721	pAcBac1 system in HEK293T
sfGFP-151-TAG	5AzW	h14	28368	28366	Labeled with DBCO-biotin
EGFP-39-TAG	5AzW	h14	30453	30453	Labeled with DBCO-biotin

Figure 2-19. Site specific incorporation of tryptophan analogs into proteins expressed in ATMW1 *E. coli*. a) Demonstration of substrate polyspecificity associated with EcTrpRS-h9 and -14 using sfGFP-151-TGA expression assay. b) SDS-PAGE analysis of sfGFP-151-TGA reporters, isolated from ATMW1 strain incorporating various UAA (Figure 2-18). c) ESI-MS analysis of reporter proteins isolated from ATMW1 or HEK293T cells. Yields were as follows (mg/L): 5HTP (92), 5MTP (89), 5BrW (25), 5AzW (80), 5PrW (61), 5AmW (68), EGFPwt (140).

Further evolution of the large EcTrpRS active site holds the potential to genetically encode structurally unique UAAs in both eukaryotes and *E. coli*. One concern associated with the use of TGA suppression for genetic code expansion in bacteria arises from previous reports demonstrating the ability of such tRNAs to also suppress TGG codons.[26] However, careful MS analysis of purified reporter proteins confirmed the absence of detectable UAA misincorporation at TGG codons by our TGA suppression system (Figure 2-20).

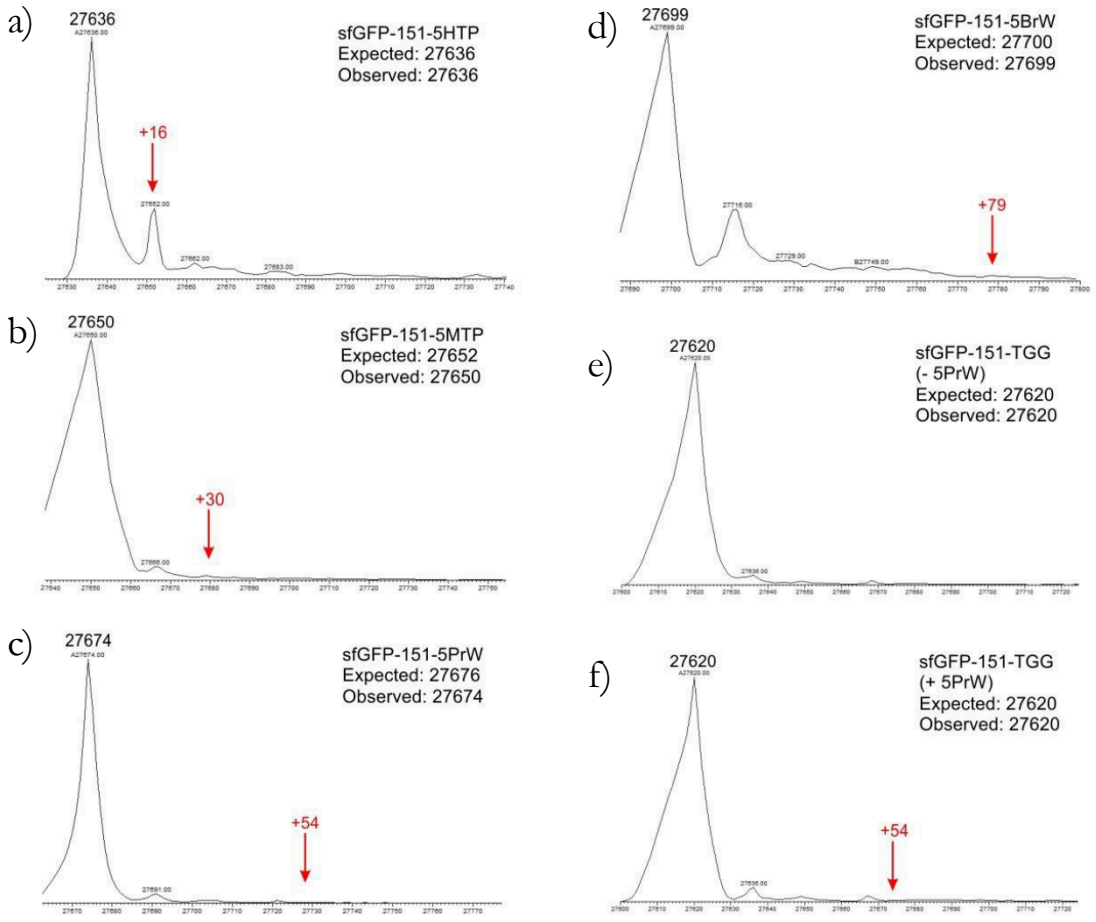


Figure 2-20. Deconvoluted whole-protein mass-spectrometry analysis of sfGFP reporters indicate the absence of TGG mischarging by tRNAEcTrp UCA in ATM E. coli. Panels (a)-(d) represent sfGFP-151-TGA (harbors one TGG codon) co-expressed with EcTrpRS-h14/tRNAEcTrp UCA in the presence of 1 mM 5HTP, 5MTP, 5PrW, or 5BrW, respectively. Panels (e) and (f) represent sfGFP-151-TGG (harbors two TGG codons) co-expressed with EcTrpRS-h14/tRNAEcTrp UCA in the absence, or presence of 1 mM 5PrW. While the presence of low levels of oxidized protein species (+16, +32, +48) complicates the detection of potential TGG-mischarging product (expected mass for the single misincorporation product is indicated with a red arrow in each case) for 5HTP (a) or 5MTP (b), the absence of such misincorporation product can be confirmed for 5PrW (c and f), and 5BrW (d). Note: We

additionally performed western blotting with a GFP standard curve to identify the low level of TGG incorporation (data not shown here). We found that there was actually <1% misincorporation. We rationalized that this was similar to the MjY mischarging of proline and that it was not significant. Most researchers do not even look into this low level of background, but we wanted to be sure it was a negligible level of misincorporation which is likely present with all UAA translational systems used for genetic code expansion.

2.2.5 UAA mutagenesis of proteins expressed in mammalian cells

To demonstrate the feasibility of using the evolved EcTrpRS variants for UAA incorporation into proteins in mammalian cells, EcTrpRS-h14 and EcTrpRS-h9 were cloned into the previously described pAcBac1[27] plasmid system together with its cognate $tRNA^{EcTrp}$ driven by CMV and U6 promoters, respectively. Although the opal suppressor $tRNA^{EcTrp}_{UCA}$ was orthogonal in mammalian cells, and exhibited comparable nonsense suppression efficiency (Figure 2-21) relative to $tRNA^{EcTrp}_{CUA}$ (Figure 2-4), for the proof-of-concept demonstration of UAA incorporation in mammalian cells, we used the more well-established amber (TAG) suppression.

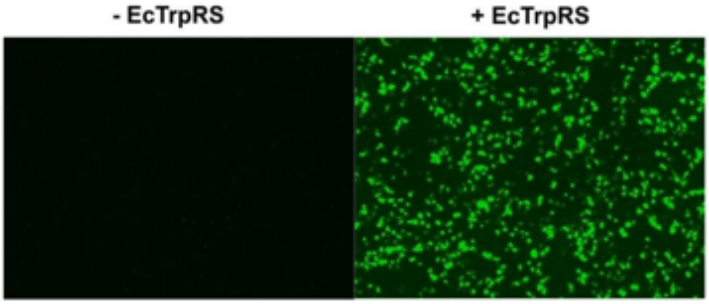
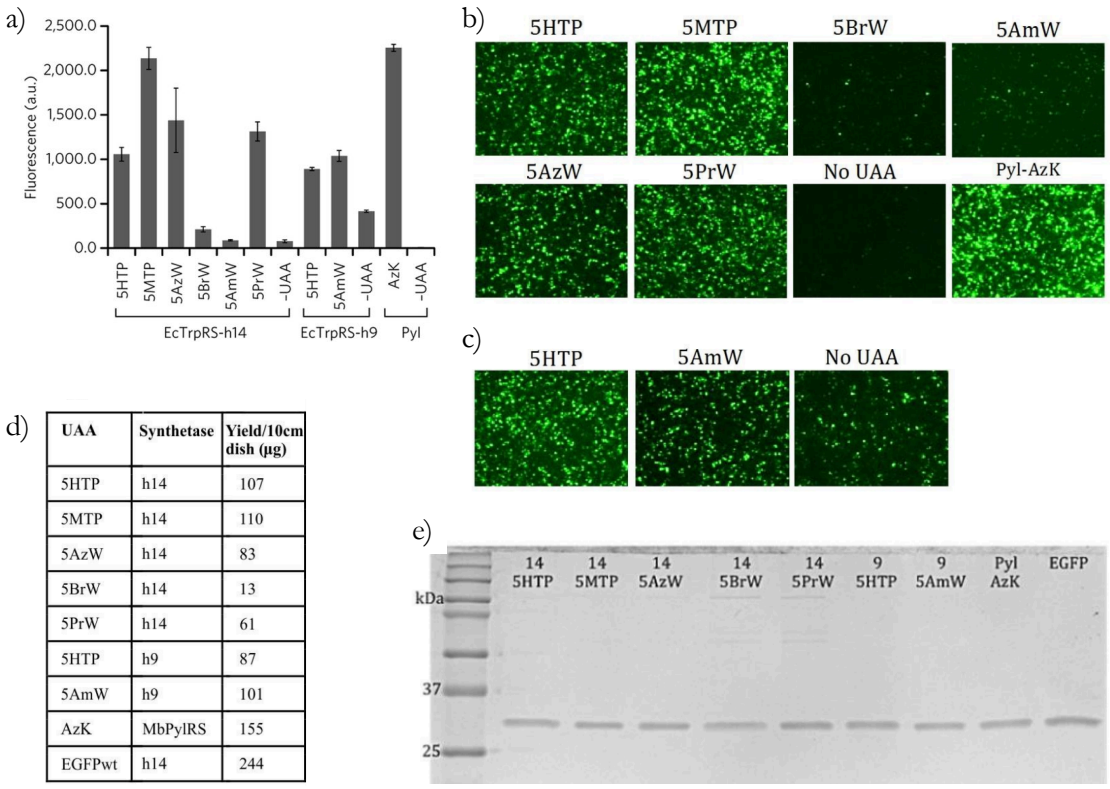


Figure 2-21. Fluorescence microscopy images showing the expression of EGFP-39-TGA reporter in HEK293T cells co-expressing $tRNA^{EcTrp}_{UCA}$, in the presence and the absence of its

cognate EcTrpRS. Robust expression and lack of cross reactivity for TGA suppression in HEK293T.

The resulting plasmid was co-transfected into HEK293T cells along with another plasmid expressing EGFP-39-TAG, and the expression of this reporter was monitored by fluorescence microscopy (Figure 2-22b,c), as well as by measuring EGFP fluorescence in cell-free extract (Figure 2-22a). Apart from 5-bromotryptophan (5BrW), addition of all other UAAs led to robust reporter expression relative to a no-UAA control. The expressed protein was isolated using Ni–NTA affinity chromatography, and subsequent SDS–PAGE (Figure 2-22e) and ESI–MS analysis (Figure 2-19c) revealed pure protein samples with a mass consistent with the incorporation of the desired UAAs. Expression levels were up to 45% those of a wild-type EGFP reporter (no TAG codon) and were comparable with the same achieved using the well-established pyrrolysyl system (Figure 2-22d).



d)

UAA	Synthetase	Yield/10cm dish (μg)
5HTP	h14	107
5MTP	h14	110
5AzW	h14	83
5BrW	h14	13
5PrW	h14	61
5HTP	h9	87
5AmW	h9	101
AzK	MbPylRS	155
EGFPwt	h14	244

Figure 2-22. Mammalian expression and characterization of TrpRS hits. (a) Expression of EGFP-39-TAG reporter incorporating various UAAs in HEK293T cells using EcTrpRS (h14 or h9)–$tRNA^{EcTrp}_{CUA}$ pair (measured as EGFP fluorescence in clarified lysate). Expression of the same reporter with the pyrrolysyl–tRNA–aaRS pair, using AzK as the substrate, is also reported for comparison. (b,c) Fluorescence images of HEK293T cells transfected with EGFP-39-TAG, tRNAEcTrp CUA, and either EcWRS-h14 (b) or -h9 (c) in the presence or absence of indicated UAAs. Expression of the same reporter facilitated by *M. barkeri* pyrrolysyl-tRNA synthetase/M. mazei pyrrolysyl-tRNA pair in the presence of its substrate AzK. (d) Yields of aforementioned purified EGFP-39-TAG from a 10 cm dish (~107 cells). (e) SDS-PAGE analysis of purified EGFP39* incorporating the indicated UAA, expressed in HEK293T cells (shown in panel a and b).

2.2.6 5AzW and 5PrW enable site-specific bioconjugation

Our work reports, for the first time, the genetic incorporation of the novel tryptophan variants 5-azidotryptophan (5AzW) and 5-propargyloxytryptophan (5PrW) into proteins expressed in both *E. coli* as well as mammalian cells. Incorporation of 5PrW and 5AzW into proteins introduces unique azido and alkyne functionalities, respectively, that can be used for bioorthogonal conjugation reactions using Cu-mediated as well as Cu-free click chemistry. This was demonstrated by conjugating DBCO–Cy5 or Alexa Fluor 488 picolyl azide (Alexa 488-PCA; Figure 2-23a) to the 5AzW or 5PrW residues in reporter proteins (EGFP-39-TAG expressed in mammalian cells or sfGFP-151-TGA expressed in ATMW1 *E. coli*), using strain-promoted or Cu-dependent click conjugation, respectively (Figure 2-23b,c). In each case, corresponding wild-type reporter proteins lacking the unnatural residue was used as a control, which did not show labeling.

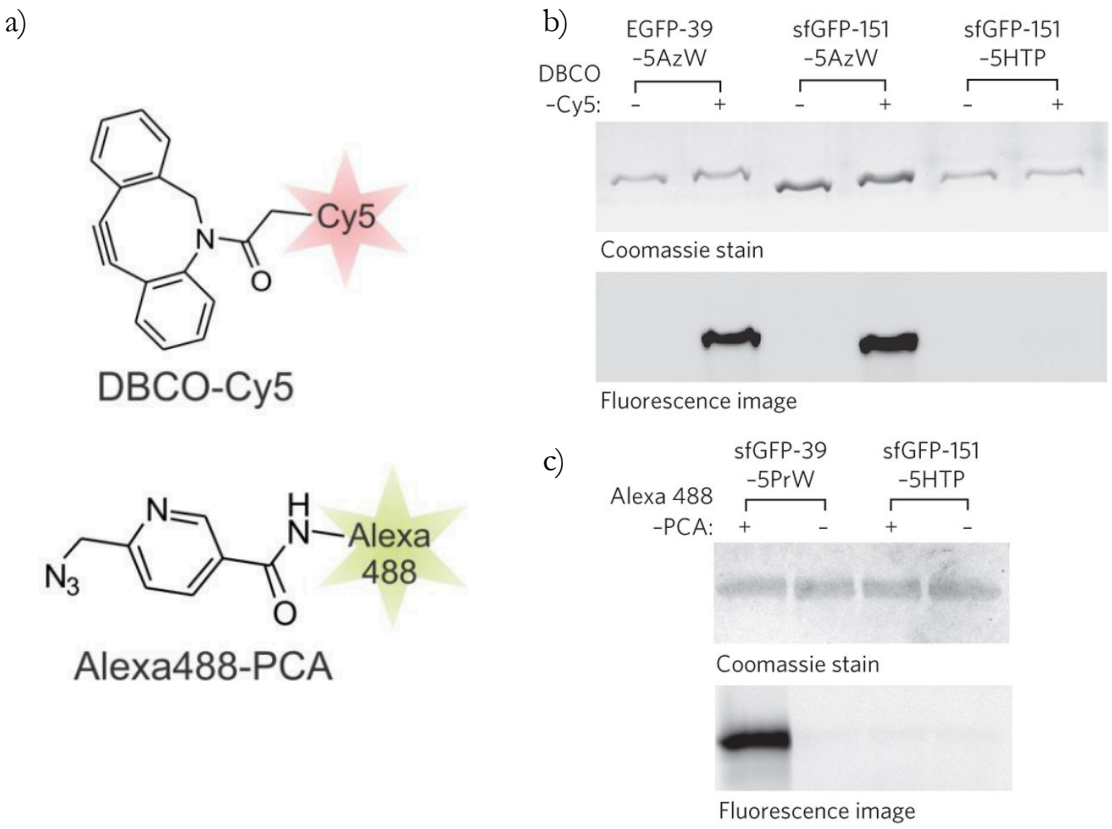


Figure 2-23. Site-specific bioconjugation of Trp analog containing GFP. (a) Structures of DBCO-Cy5 and Alexa488-PCA. (b) EGFP and sfGFP proteins containing site-specifically incorporated 5AzW or 5HTP were labeled with DBCO–Cy5, resolved using SDS–PAGE, and subjected to fluorescence imaging followed by Coomassie staining. (c) Preparations of sfGFP containing site-specifically incorporated 5PrW or 5HTP were labeled with Alexa 488-PCA, resolved using SDS–PAGE, and subjected to fluorescence imaging followed by Coomassie staining.

2.3 Conclusions

We demonstrate here the feasibility of functionally replacing an endogenous aaRS–tRNA in *E. coli* with an *E. coli*–optimized eukaryotic counterpart without causing a growth penalty, and show that the resulting strain can be used as a selection host for evolving variants of the liberated aaRS–tRNA pair for charging UAAs. These variants enable genetic code expansion in eukaryotes as well as in *E. coli*. Since bacterial aaRS–tRNA pairs are generally orthogonal in eukaryotic cells, this approach holds the potential to provide additional 'universal' aaRS–tRNA platforms. We also highlight a potential challenge with this approach: altering the anticodon of the liberated *E. coli* tRNA (to enable nonsense suppression) may lead to cross-reactivity with other endogenous aaRSs. However, understanding the molecular basis of the observed cross-reactivity and rational selection of an alternative nonsense codon allowed us to circumvent this issue with the EcTrpRS– $tRNA^{EcTrp}$ pair. While the same solution may not be suitable for other pairs, a directed evolution approach can be adopted to develop variants of the liberated nonsense suppressor tRNA with attenuated cross-reactivity.[12,13,28,29]

Our work also introduces a new tryptophanyl–tRNA synthetase–tRNA platform for genetic code expansion, whose utility was illustrated by introducing several new UAA additions to the genetic code of *E. coli* as well as mammalian cells, including 5AzW and 5PrW, which enable site-selective bioconjugation reactions. Access to new universal tRNA–aaRS pairs will augment the structural diversity of the genetically encoded UAA toolbox, particularly for application in eukaryotic cells, and facilitate the development of powerful new technology involving simultaneous incorporation of multiple UAAs into a polypeptide in both eukaryotes and *E. coli*.

2.4 Experimental procedures

2.4.1 General methods

For cloning and plasmid propagation, the DH10b (Life Technologies) strain of E. coli was used. Polymerase chain reaction (PCR) was performed using the Phusion Hot Start II DNA Polymerase (Fisher Scientific) using the manufacturer's protocol. For purification of DNA (plasmid as well as PCR products, etc.) spin columns from Epoch Life Science were used. Restriction enzymes and T4 DNA ligase were obtained from New England BioLabs (NEB). DNA oligomers for PCR were purchased from Integrated DNA Technologies (IDT). Sequences of the oligonucleotides used in this work are shown in the oligo appendix. Verification of cloned DNA by Sanger sequencing was performed by Eton Biosciences. Antibiotics, isopropyl β-D-1- thiogalactopyranoside (IPTG), and L-arabinose were purchased from Sigma-Aldrich or Fisher Scientific. Components of media were obtained from Fisher Scientific. Bacteria were grown on LB or M63 agar plates17 and LB liquid medium with the following antibiotic concentrations unless otherwise mentioned: 95 μg/ml spectinomycin, 20 μg/ml chloramphenicol, and 100 μg/ml ampicillin, 30 μg/ml kanamycin, 15 μg/ml zeocin, 12 μg/ml tetracycline, 10 μg/ml gentamycin.

2.4.2 Statistical methods

For all expression analyses (sfGFP in bacteria or EGFP in HEK293T cells), the mean of three independent experiments was reported, and error bars represent s.d. For the growth rate analysis, each data point represents the mean OD600 of three independent cultures of the same strain (error bars represent s.d.). In our experience, a mean of three experiments provides adequate levels of accuracy for these experiments.

2.4.3 Strains, cell lines

The EcNR1 strain was a gift from G.M. Church. The HEK293T cell line was obtained from ATCC, and propagated without further confirmation. Cell lines are regularly monitored for Mycoplasma contamination. Even though HEK293T is listed under misidentified cell lines in the ICLAC database, we used it for our expression analyses as a representative mammalian cell line. Given the widespread use of this cell line as a model mammalian expression host, and since our conclusions do not rely on its specific identity (beyond a representative mammalian cell-line), we believe that the use of this strain is justified.

2.4.4 Lambda Red recombination

All strains were derived from EcNR1.[14] This strain contains temperature-inducible lambda recombinase genes (Exo, Beta, Gam) and a constitutive ampicillin resistance gene disrupting the bioA–bioB genomic locus. Strains were grown in 10 ml LB at 30 °C to 0.5 OD600 and then induced for 15 min in a 42 °C water bath (250 r.p.m.). The cell pellet was then washed twice with 10 ml ddH2O by centrifuging at 4,500 × g. Cells were resuspended in residual ddH2O (~250 μl), and 50 μl aliquots were electroporated with 50 ng double-stranded DNA (dsDNA) or 2 μM 90 bp oligo containing 45–77 bp genomic homology, depending on the desired recombination. Electroporated cells were recovered for 1–6 h and plated on either LB or minimal media. Single colonies from selection plates were re-streaked and subjected to colony PCR using KAPA-2G polymerase (Kapa Biosystems), following manufacturer's instructions, to verify desired recombination.

2.4.5 Building ATMW1

EcNR1 was transformed with pUltraG_ScW40CCA. To remove the *E. coli* tryptophanyl–tRNA synthetase (trpS) 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 TrpRS.Z.ab-F and TrpRS.Z.ab-R to generate the PCR product trpS::ZeoR. 50 ng of the trpS::ZeoR PCR cassette was transformed in the recombination following the aforementioned protocol, and the resulting strains were plated on LB–Agar plates supplemented with Zeocin. The resulting colonies were screened via colony PCR using TrpRS150F + TrpRS150R, TrpRS150F + Zeo-iR, and TrpRS150F + TrpRS_sqiR, as well as sequencing these colony PCR products. This strain was named EZ4.

To replace the E. coli tryptophanyl-tRNACCA (*trpT*), the *trpT*::GentR PCR cassette was amplified using trpTKO.Gent-F and trpTKO.Gent-R (965 bp). 50 ng of *trpT*::GentR PCR cassette was transformed into EZ4, induced as previously described. Resulting gentamycin-resistant colonies were screened for the desired recombination using colony PCR primers *trpT* GsqF and *trpT* GsqR, as well as sequencing of the PCR product. The resulting strain was named EZG4.

2 μM 90 bp oligo, *galK*.90 del, was used to delete *galK* from the endogenous genomic location. Following transformation, cells were recovered for 6 h and washed twice with M9 minimal media at 5,000 × g for 5 min. 100 μl of a 10^4 dilution was plated on M63 minimal media supplemented with glycerol and 2-deoxygalactose to select for successful galK deletion1. Colony PCR was used to verify the deletion of galK using galK_KO_verf-F/R. The resulting strain was named G4.

galK.PCR cassette containing the endogenous promoter dlambda.*GalK*-F and dlambda.*galK* dterm-R (1,348 bp) was amplified and was used to remove the λ-RED genes

from the 4 strain. Following transformation of 50 ng of this PCR product into strain 4, successful insertion of the *galK*.PCR cassette into the λ-RED site was selected by plating the cells on M63 plates containing galactose as the sole carbon source. Surviving colonies were screened by colony PCR using dlambda.sqF with dlambda.sqR or dLambda.sqiR for the desired deletion. This final strain was named ATMW1 (EcNR1 *trpS*::ZeoR *trpT*::GentR ⊿*galK* λ-RED::*galK*).

2.4.6 Growth comparison

5 ml starter cultures of EcNR1G, EcNR1G + pUltraG_ScW40CCA, and ATMW1 strains were grown for 16 h in LB with all strain-dependent antibiotics. For each strain, the starter culture was diluted to an initial OD600 of 0.01 in three identical cultures of 80 mL LB with no antibiotics and allowed to grow in 250 ml sterile Erlenmeyer flasks at 30 °C, with shaking (250 r.p.m.). Growth was monitored every 30 min by measuring OD600 in a 10-mm cuvette.

2.4.7 Assessment of aaRS-tRNA activity using a chloramphenicol reporter

Overnight cultures of ATMW1 harboring pRepAC-EcW-TAG or pRepJI-EcW-TGA, with or without pBK–EcWRSwt, were diluted to an OD600 of 0.1, and 3 µl was spot plated on LB agar plates supplemented with kanamycin (+pBK plates), spectinomycin, tetracycline, and varying chloramphenicol concentrations. Growth was analyzed after 48 h of incubation at 37 °C.

2.4.8 Assessment of aaRS-tRNA activity using a sfGFP151 reporter

EcNR1 or ATMW1 harboring pEvolT5-EcW sfGFP151 (TAG or TGA) with or without pBK-EcWRSwt, pBK-EcWRS-h14, or pBK-EcWRS-h9 were grown overnight in LB.

The starter cultures were diluted in LB supplemented with required antibiotics to 0.05 OD600. Cultures were grown at 30 °C or 37 °C (30 °C when comparing to progenitor strain EcNR1) until 0.55 OD600, at which point the sfGFP expression was induced with a final concentration of 1 mM IPTG. Unnatural amino acids (UAA) were added during induction to a final concentration of 1 mM. Cultures were grown for an additional 17–20 h at 37 °C with shaking. To evaluate sfGFP expression, cells from 150 µl of the cultures were pelleted at 5,000 × g, resuspended in 150 µl PBS, and transferred to a 96-well clear-bottom assay plate. Fluorescence was measured by using a SpectraMAX M5 (Molecular Devices) (Ex. 488 nm; Em. 534 nm). Fluorescence for each sample was normalized using its OD600.

2.4.9 Protein purification

To maximize the yield of UAA-modified protein expression, a different plasmid combination was used: EcTrpRS-h14 and EcTrpRS-h9 were cloned into a pEvoltac plasmid that expresses them from a strong *tacI* promoter, while the $tRNA^{EcTrp}_{UCA}$ is expressed from the *proK* promoter. The sfGFP reporter gene (sfGFP-151-TGA or wild type sfGFP) was expressed from pET22b-T5lac plasmid driven by the strong *t5.lac* promoter. Overnight expression cultures were centrifuged and resuspended in lysis buffer: B-PER Bacterial Protein Extraction Reagent (Thermo Scientific) + 1× Halt Protease Inhibitor Cocktail (Thermo Scientific) + 0.01% Pierce Universal Nuclease (Thermo Scientific). After 30 min incubation at room temperature, the lysate was clarified by centrifuging at 22,000 × *g* for 5 min. The C-terminally polyhistidine tagged soluble sfGFP in the supernatant was purified using a HisPur Ni-NTA resin (Thermo Scientific) following manufacturer's protocol. Protein purity was confirmed by SDS–PAGE and purified protein molecular weight was confirmed by ESI–MS (Agilent Technologies, 1260 Inifinity ESI–TOF).

2.4.10 Construction of the EcWRS-5HTP pBK library

Overlap extension was used to introduce degenerate codons, creating the five-residue tryptophanyl tRNA-synthetase library pBK-EcWRS1.5 (786,432 diversity): F7-NBT, S8-NST, V144-NNK, P145-NST, V146-NNK. Using Phusion HSII (Fisher Scientific) and manufacturer's protocol, EcWRS1_mut7-8-F + EcWRS1_mut-VPViR and EcWRS1_mut144-6_F + EcWRS_NcoI_PstI_termR were used to PCR amplify the N-terminal and C-terminal of the EcWRS PCR product, respectively. The N-terminal and C-terminal PCR products were joined together by overlap extension PCR using the following terminal primers: libEcWRS-NdeI-F and EcWRS_NcoI_PstI_termR. These inserts were digested with NdeI/NcoI (NEB) and ligated by T4 DNA Ligase (NEB) into pBK vector cut with the same restriction enzymes. The ligation mixture was ethanol precipitated with Yeast-tRNA (Ambion) and transformed into electrocompetent DH10b cells. The library was covered using ~10^7distinct colony forming units (CFU).

2.4.11 Selection for synthetase charging 5HTP

ATMW1 was co-transformed with the pBK-EcWRS1.5b library and the positive selection reporter plasmid pRepJI EcW TGA. The reporter plasmid harbors a *lpp*-promoted *E. coli* $tRNA^{Trp}_{UCA}$, a CAT reporter modified to contain TGA codons (Q98TGA), an arabinose-inducible T7 RNA polymerase harboring two TGA nonsense codons (at positions 8 and 114), and a wild-type GFPuv reporter expressed from a *t7* promoter. Suppression of CAT leads to chloramphenicol resistance, and suppression of T7 RNA polymerase drives expression of a *t7*-promoted GFPuv. 9.2×10^7 CFU were plated on LB + 0.5× Spec/0.5× Tet/Kan/0.02%

arabinose + chloramphenicol (25, 35, 45 μg/ml) in the presence of 1 mM 5HTP for 36 h at 37 °C.

Colonies from the 35 and 45 μg/ml chloramphenicol positive selection plates were harvested, and the pBK plasmids harboring mutant EcTrpRS were purified by miniprep and gel purification. These were co-transformed into ATMW1 harboring the negative selection plasmid pNegJ2-EcW (araBad-Barnase harboring two TGA codons at 3 and 45). 3×10^7 CFU were plated on LB + Amp/0.5×Kan/0.02% arabinose and incubated for 12 h at 37 °C. Cells were harvested and library pBK plasmid was purified by miniprep/gel purification.

Isolated pBK plasmids from the negative selection were transformed again into ATMW1 pRepJI-EcW TGA, and 10^6 CFU were plated on LB + 0.5×Spec/Tet/Kan/0.02% arabinose + chloramphenicol (30, 40, 50 μg/ml) in the presence or absence of 1 mM 5HTP for 18 h, 37 °C, which revealed substantially higher number of colonies in the presence of the UAA. 96 colonies were picked into a 1 ml LB supplemented with Spec/Tet/Kan in deep 96 well polypropylene plates and grown overnight. The resulting cultures were diluted to ~0.01 OD_{600} and 3 μl of each was spot plated on LB/Agar plates supplemented with Spec/Tet/Kan, and chloramphenicol (50, 60 μg/ml) in the presence or absence of 5HTP. Four clones exhibiting the most prominent UAA dependent growth were picked and sequenced.

2.4.12 Assessing activity of $tRNA^{EcTrp}_{CUA}$ and synthetase hits in HEK293T

Dulbecco's modified Eagle's medium (high glucose DMEM) supplemented with 10% FBS (FBS) and penicillin–streptomycin (0.5×) was used to culture HEK293T cells. Cells were incubated in a humidified incubator at 37 °C in the presence of 5% CO_2. HEK293T were seeded at a density of 600,000 cells per well for a 12-well plate 1 d before transfection and transfected at ~70% confluence. Polyethylenimine (PEI, Sigma) and DNA were mixed at a

ratio of 4 µl PEI (1 mg/ml) to 1 µg DNA in DMEM. After 20 min incubation, 100 µl of this mixture was used to transfect one single well in a 12-well plate. For these transfections, 500 ng of pAcBac EGFP39* U6-EcWtR TAG was transfected in the presence or absence of pAcBac-TrpRS (wt, h14, or h9) U6-EcWtR TAG. UAAs were added to the culture medium to a final concentration of 1 mM at the time of transfection. Fluorescence images were taken at 48 h post-transfection using a Zeiss Axio Observer fluorescence microscope.

To obtain EGFP39* expression data, cells were harvested, washed once with PBS buffer (5,000 × *g*), and lysed with CelLytic M lysis buffer (Sigma) supplemented with 1× Halt protease inhibitor and 0.01% Pierce universal nuclease. 50 µl lysis buffer was used for each well of a 12-well plate, and was allowed to incubate for 20 min. After incubation, the lysate was clarified by centrifuging at 22,000 × *g* for 5 min and was transferred to a clear bottom 96-well assay plate. Fluorescence was measured using a SpectraMAX M5 (Molecular Devices) (Ex. 488 nm; Em. 534 nm).

2.4.13 Isolation of EGFP-39-TAG reporters from HEK293T

HEK293T cells were cultured as previously described. One day before transfection, cells were seeded at a density of 8 million cells/10 cm dish. 50 µl PEI MAX (Polysciences) was mixed with 10 µg total DNA (5 µg reporter, 5 µg tRNA–aaRS plasmid) with 180 µl DMEM (no FBS), incubated for 20 min, and added evenly to the dish at 90% confluence. Desired UAAs were supplemented at a final concentration of 1 mM and cells were allowed to express the desired protein for 48 h.

Cells from a 10 cm dish were harvested and lysed with 600 µl CelLytic M lysis buffer (Sigma, 1× Halt protease inhibitor, 0.01% Pierce universal nuclease). Lysate was clarified as described and purified via Ni^{2+}-NTA, following manufacturer's protocol. Purified protein was

analyzed by SDS–PAGE and molecular weight was confirmed by ESI–MS (Agilent Technologies 1260 Infinity ESI–TOF).

2.4.14 Click-labeling of 5AzW and 5PrW residues

Purified proteins containing 5AzW were incubated with or without 20 μM DBCO–Cy5 (Sigma) for 1 h in the dark at room temp. Proteins samples were resolved by SDS–PAGE gel and imaged using Cy5 specific settings on a ChemiDoc MP Imaging System (Bio-Rad). The SDS–PAGE gels were then Coomassie stained and imaged.

5PrW containing proteins were labeled using the Click-iT Plus Alexa Fluor Picolyl Azide kit (Life Technologies) with a modified protocol. The following were mixed in order: 1 μg protein (4 μl), 3 μl 10× buffer additive, 0.3 μl 100 mM $CuSO_4$, 0.3 μl copper protectant, 1.2 μl Alexa 488 PCA (50 μM, 2 μM final), 21.2 μl 1× Click-iT reaction buffer, to a final volume of 30 μL. Samples were incubated for 40 min in dark and subsequently resolved by SDS–PAGE and imaged using the ChemiDoc MP Imaging System (Bio-Rad) with Alexa 488 specific settings.

2.4.15 Tryptic digestion and LC-MS/MS analysis of reporter proteins

12 μg purified reporter protein was treated with DMSO (0.2 μM) and precipitated with 5 μl 100% solution of trichloracetic acid. Sample was frozen at −80 °C overnight. Thawed sample was centrifuged at 15,000 r.p.m. for 10 min at room temperature. Supernatant was removed and pellets were vortexed to resuspend in 500 μl cold acetone. Samples were then centrifuged at 5,000 r.p.m. for 10 min. Acetone was then removed and pellet was allowed to air dry. Pellet was resuspended in 30 μl 8 M urea in PBS followed by 70 μl 100 mM ammonium bicarbonate, and then 1.5 μl 1 M DTT was added. Samples were incubated at 65 °C for 15

min. After incubation, 2.5 µl of 500 mM iodoacetamide in PBS was added and the sample was left at room temperature for 30 min. Following incubation, 120 µl PBS was added to each sample and vortexed rapidly. 4 µl of trypsin was added to the samples followed by 2.5 µl 100 mM $CaCl_2$. Samples were then agitated for 37 °C overnight. Trypsin was then quenched with 10 µl formic acid and pelleted at 15,000 r.p.m for 20 min. Supernatants were stored at −20 °C.

Samples were subsequently analyzed by LC–MS/MS using a LTQ Orbitrap XL mass spectrometer (Thermo Fisher) coupled to an EASY-nLC 1000 nanoLC (Thermo Fisher). 10 µl samples were loaded onto 100 µm fused silica column with a 5 µm tip packed with 10 cm of Aqua C18 reverse-phase resin (Phenomenex) using the EASY-nLC 1000 autosampler. The digests were eluted using a gradient 0–100% buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B; 20% water, 80% acetonitrile, 0.1% formic acid). The flow rate through the column was set to 400 nl/min and the spray voltage was set to 3.5 kV. One full MS scan (FTMS) (400–1800 MW) was followed by seven data-dependent scans (ITMS) of the nth most intense ion from the imported mass list with dynamic exclusion. The tandem MS data were searched using the SEQUEST algorithm using a concatenated target/decoy variant of the human IPI databases. A static modification of +57.02146 on cysteine was specified to account for iodoacetamide alkylation. SEQUEST output files were filtered using DTASelect 2.0.

2.4.16 ATMW plasmid construction and additional commentary

Complementation plasmid pUltraG_ScW40$_{CCA}$. Previously reported pUltra Hit14, containing a *tacI* promoted wild-type yeast tryptophanyl–tRNA synthetase and *proK* promoted evolved yeast tRNA amber suppressor,[12] was used to generate additional pUltra derivatives. To generate pUltra_ScW40$_{CUA}$, a previously evolved tRNA 40A was amplified from 40A

gBlock (IDT) using primers proK-F and TrpH NcoI-R.[13] The full-length prok-tRNA PCR product was cloned into pUltra Hit14 via SbfI/NcoI (NEB), producing pUltra ScW40$_{CUA}$.

To build pUltraG_ScW40$_{CCA}$, a complementation plasmid with a constitutively active synthetase, the *tacI* promoter region was removed by digesting the pUltra_ScW40$_{CUA}$ vector with XbaI/SbfI, and a glnS'-ScWRS cassette was amplified from an pEVOL[8]-derived plasmid, pEvol ScW, using glnS-XbaI-F + glnS-SbfI-R and subsequently cloned using the same set of restriction sites. The 40A tRNA anticodon was then mutated to CCA using overlap extension PCR. The two PCR products were amplified with proK-F + Trp40CCA-iR and Trp40CCA-iF + SmR-R, overlap extended, and cloned into pUltraG_ScW40$_{CUA}$ using SbfI/NcoI, producing the final complementation plasmid pUltraG_ScW40$_{CCA}$.

Construction of selection plasmids. To generate the TAG-positive selection plasmid, pRep-ScW14[12] was digested with SpeI/BglII. The *lpp*-promoted *E. coli* tRNA$^{Trp}_{CUA}$ was amplified from gBlock I-EcWtR TAG with pNP-SpeI-F + pNP-BAB-R. The PCR product was digested with SpeI/BamHI, and cloned into the SpeI/BglII sites of the pRep vector creating pRepAC-EcWtR-TAG.

In order to create a selection plasmid based on TGA suppression, a smaller pRep-Cm3b was used as a starting template. To delete the chloramphenicol–UPP fusion protein, chloramphenicol acetyl transferase was amplified from pEvol with pEvol CmR SpeI-F + pEvol CmR PstI-R, digested with SpeI/PstI, and ligated into the same sites of pRep-Cm3b, creating pRep-Cm3J-wt. TGA sites were introduced into the CAT gene by replacing G7, I84, or Q98 via site-directed mutagenesis using primers CmR-TGA-7, CmR-TGA-84, or CmR-TGAT-98, building three different reporters. The *lpp*-promoted *E. coli* tRNA$^{Trp}_{CUA}$ was amplified from pRepAC-EcWtR-TAG with pRep-KpnI-tR-F and pRep-EcWtR-NdeI/AvrII-R and cloned via KpnI/AvrII into these pRep-Cm3J plasmids, creating three different pRep-

Cm3J-#TGA-EcWtR TAG. The anticodon of the tRNA was mutated to TGA using site-directed mutagenesis with primer EcWtR-TGA-MSDM. While all three reporter plasmids exhibited desired phenotypes upon TGA suppression, pRep-Cm3J-98TGA-EcWtR was used for the subsequent cloning steps.

To add the T7 RNA polymerase-GFPuv reporter system into pRep-Cm3J-98TGA-EcWtR, these elements were amplified from pRepAC-EcWtR-TAG using multiple overlap extension PCR reactions to change the two TAG nonsense codons in the T7RP gene to TGA: Three PCR products were first amplified using T7F1 + T7R3, T7F3 + T7R4b, T7F4 + T7R5, which were overlapped to produce T7-araC PCR cassette. This cassette was first cloned back into pRepAC-EcWtR-TAG via KpnI/NsiI, and then the entire GFP-T7araC cassette was amplified from the resulting plasmid with T7_mut-SbfI-F + T7_mut-NotI/SbfI-R, digested with SbfI, and ligated into the PstI site of pRep-Cm3J-98TGA-EcWtR vector producing pRepJI-EcW.

The negative selection plasmid was built by modifying an existing pNeg plasmid.[7,86] The two barnase suppression sites (3TAG, 45TAG) were mutated to TGA with site-directed mutagenesis using primers barnase MSDM 3 TGA and barnase MSDM 45 TGA, creating plasmid pNegJ2. The *E. coli* $tRNA^{Trp}_{CUA}$ was amplified from pRep-Cm3J-98TGA-EcWtR with pRep_NegtR-EcoRI-R and pRep_KpnI-tR-F and then cloned into pNegJ2, replacing the preexisting tRNA, creating pNegJ2-EcW.

Construction of bacterial suppression plasmids. Since ATMW1 uses the pUltraG plasmid to express the yeast tryptophanyl pair, the suppression plasmids cannot use the CloDF13 origin of replication, or the spectinomycin marker. The previously described pEvoltac MjY plasmid,[8] which harbors a compatible p15a origin of replication and a chloramphenicol resistance marker, was used as the template to generate these plasmids. The plasmids pEvolT5 EcW

sfGFP151 (TAG or TGA) were built to allow the rapid evaluation of various EcTrpRS mutants isolated from the selection scheme (encoded in pBK vector). Initially, a *t5.lac*-promoted sfGFP-151-TAG was amplified from pET22b-T5-sfGFP151TAG using GFPflip-NotI-F + sfGFP-T7+lam-PstI-R and subsequently cloned into pEvoltac MjY using EcoNI/PstI, to generate pEvolT5 MjY sfGFP151TAG.

A *proK*-promoted *E. coli* $tRNA^{Trp}_{CUA}$ was assembled by overlap extension as follows. The *proK* promoter was amplified from pUltraG ScW40 using EcWtR-PstI-F + EcWtR proK-oR, which was overlapped with $tRNA^{EcTrp}_{CUA}$ amplified with EcWtR-proK-oF and EcWtR-KpnI-R. This PCR product was cloned into pEvoltac MjY sfGFP151TAG using PstI/KpnI, producing pEvoltac MjY EcWtR sfGFP151TAG. The MjY tRNA was removed by using Polymerase Incomplete Primer Extension (PIPE) cloning with primers MjYtR-del-oF and MjYtR-del-oR, producing the final plasmid pEvolT5-EcW-sfGFP151TAG. Additionally, site-directed mutagenesis was used to build the TGA reporter pEvolT5-EcW-sfGFP151TGA with primers sfGFP(pEvol)TGA151 and EcWtR TGA MSDM.

For more efficient protein expression using evolved aaRS–tRNA pairs, plasmids containing the *proK* promoted $tRNA^{EcTrp}_{UCA}$ and *tacI* promoted EcWRS-h14 or -h9 were assembled. The tRNA in pEvoltac MjY was first replaced by amplifying the EcTrp-$tRNA_{UCA}$ from pEvoltac-EcW-sfGFP151TGA with EcWtR PstI-F and pUltraII-tRsqR, and subsequently cloning into PstI/SphI to generate pEvoltac-EcW-MjYRS. EcWRS-h14 and -h9 were then amplified from their respective pBK plasmids using EcWRS1.FA.NotI-F and EcWRS-NotI-R, digested with NotI, and cloned into NotI-digested pEvoltac-EcW-MjYRS-pAcF, producing pEvoltac-EcW-TGA-h14 or -h9.

pBK MjYRS[86] was used as a template to introduce tryptophanyl-tRNA synthetase variants. Top10 genomic DNA was purified using previously described protocols, and used as

the template to amplify the EcTrpRS using EcTrpRS-NdeI-F and EcTrpRS-PstI-R. The PCR product was digested with NdeI/PstI, producing pBK EcWRS.

Construction of mammalian suppression plasmids. Previously reported pAcBac1 was used to generate mammalian reporter and suppression plasmids.[100] To build pAcBac1 TrpRS, EcTrpRS was amplified from Top10 genomic DNA with TrpRS-F-NheI + TrpRS-R-EcoRI and subsequently cloned into pAcBac1 via NheI/EcoRI. U6-EcWtR gBlock was used as a template and amplified with tRNAtrp*-NheI-R + tRNAtrp* NheI-R to produce the human U6 promoted *E. coli* $tRNA^{EcTrp}_{CUA}$ PCR product. This PCR product was then digested with AvrII/NheI and cloned into the AvrII site in the pAcBac variants, resulting in pAcBac1-TrpRS-U6EcWtR-TAG and pAcBac1-EGFP39*-U6EcWtR-TAG. EcWRS-h14 and -h9 synthetase variants were cloned into the pAcBac1-TrpRS-U6EcWtR-TAG plasmid via NheI/EcoRI digestion after amplification with EcWRS_mamNheI-F + EcWRS_mamEcoRI-R to make pAcBac1-EcWRS-h14-U6EcWtR-TAG and pAcBac1-EcWRS-h9-U6EcWtR-TAG.

2.5 References

1. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550,** 53–60 (2017).
2. Dumas, A., Lercher, L., Spicer, C. & Davis, B. G. This journal is © The Royal Society of Chemistry 2014 Designing logical codon reassignment – expanding the chemistry in biology. 1–36 (2014). doi:10.1126/scienc
3. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79,** 413–444 (2010).
4. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science (80-.).* **292,** 498–500 (2001).
5. Chin, J. W. An Expanded Eukaryotic Genetic Code. *Science (80-.).* **301,** 964–967 (2003).
6. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 1059–1070 (2014).

7. Niu, W., Schultz, P. G. & Guo, J. An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **8,** 1640–1645 (2013).
8. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry* **52,** 1828–1837 (2013).
9. Wan, W. *et al.* A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chemie - Int. Ed.* **49,** 3211–3214 (2010).
10. Xiao, H. *et al.* Genetic incorporation of multiple unnatural amino acids into proteins in Mammalian cells. *Angew. Chem. Int. Ed. Engl.* **52,** 14080–14083 (2013).
11. Iraha, F. *et al.* Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion. *Nucleic Acids Res.* **38,** 3682–3691 (2010).
12. Chatterjee, A., Xiao, H., Yang, P.-Y., Soundararajan, G. & Schultz, P. G. A Tryptophanyl-tRNA Synthetase/tRNA Pair for Unnatural Amino Acid Mutagenesis in *E. coli*. *Angew. Chemie Int. Ed.* **52,** 5106–5109 (2013).
13. Ellefson, J. W. *et al.* Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **32,** 97–101 (2014).
14. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–8 (2009).
15. Warming, S., Costantino, N., Court, D. L., Jenkins, N. A. & Copeland, N. G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res.* **33,** 1–12 (2005).
16. Soll, L. & Paul, B. Recessive Lethal Nonsense Suppressor in Escherichia coli which inserts Glutamine. *Nature* **223,** 1340–1342 (1969).
17. Jahn, M., Rogers, M. J. & Söll, D. Anticodon and acceptor stem nucleotides in tRNA(Gln) are major recognition elements for E. coli glutaminyl-tRNA synthetase. *Nature* **352,** 258–60 (1991).
18. Rogers, M. J., Adachi, T., Inokuchi, H. & Söll, D. Switching tRNA(Gln) identity from glutamine to tryptophan. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3463–7 (1992).
19. Kopelowitz, J., Hampe, C., Goldman, R., Reches, M. & Engelberg-Kulka, H. Influence of codon context on UGA suppression and readthrough. *J. Mol. Biol.* **225,** 261–269 (1992).
20. O'Donoghue, P. *et al.* Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. *FEBS Lett.* **586,** 3931–3937 (2012).
21. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20,** 1044–1048 (2002).
22. Zhang, Z. *et al.* Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 8882–8887 (2004).
23. Antonczak, A. K. *et al.* Importance of single molecular determinants in the fidelity of expanded genetic codes. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 1320–1325 (2011).
24. Cooley, R. B., Karplus, P. A. & Mehl, R. A. Gleaning unexpected fruits from hard-won synthetases: Probing principles of permissivity in non-canonical amino acid-tRNA synthetases. *ChemBioChem* **15,** 1810–1819 (2014).
25. Young, D. D. *et al.* An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50,** 1894–1900 (2011).

26. Raftery, L. A., Egan, J. B., Cline, S. W. & Yarus, M. Designed set of cloned termination suppressors: In Vivo activity of isogenetic UAG, UAA, and UGA suppressor tRNAs. *Mol. Cell. Ther.* **158,** 849–859 (1984).
27. Chatterjee, A., Xiao, H., Bollong, M., Ai, H.-W. & Schultz, P. G. Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl. Acad. Sci.* **110,** 11803–11808 (2013).
28. Chatterjee, A., Xiao, H. & Schultz, P. G. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. *Proc. Natl. Acad. Sci.* **109,** 14841–14846 (2012).
29. Guo, J., Melancon, C. E., Lee, H. S., Groff, D. & Schultz, P. G. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chemie - Int. Ed.* **48,** 9148–9151 (2009).

Chapter 3

Resurrecting the bacterial tyrosyl-tRNA synthetase/tRNA pair for expanding the genetic code of both *E. coli* and eukaryotes

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

Italia, J.S., Latour, C., Wrobel, C.J., Chatterjee, A. Resurrecting the bacterial tyrosyl-tRNA synthetase/tRNA pair for expanding the genetic code of both *E. coli* and eukaryotes. *Cell. Chem. Biol.,* **2018,** 25, 1304-1312.

3.1 Introduction

3.1.1 Overview

Site-specific incorporation of unnatural amino acids (UAAs) into proteins in living cells provide exciting new ways to understand and engineer their function.[1–5] To co-translationally incorporate a UAA, it is encoded by a repurposed nonsense codon, which is suppressed by an orthogonal (i.e., does not cross-react with its host counterparts) UAA selective aminoacyl-tRNA synthetase (aaRS)/tRNA pair.[1–4] Eukaryote- or archaea derived aaRS/tRNA pairs are orthogonal in bacteria and used for UAA incorporation, while those derived from bacteria are typically orthogonal in eukaryotes (Figure 3-1).[1–5] To create UAA-specific variants of aaRS/tRNA pairs through directed evolution, two cell-based selection systems have been developed, using *Escherichia coli*[6,7] or *Saccharomyces cerevisiae* (yeast)[8,9] as selection hosts, for engineering pairs that are orthogonal in bacteria or eukaryotes, respectively. Due to its facile nature, the E. coli-based selection system has been significantly more successful in creating novel UAA-selective aaRS/tRNA pairs than its yeast counterpart.[2,3,10] As a result, genetically encoding new UAAs in eukaryotes in the past decade has overwhelmingly relied on the unique pyrrolysyl pair—the only pair that can be engineered using the facile *E. coli* selection system followed by application in eukaryotes (Figure 3-1).[1–3,5,11] However, such excessive dependence on a single platform limits the structural diversity of UAAs that can be genetically encoded in eukaryotes.[2,3,10] It also limits the scope of the nascent technology for concurrently incorporating multiple different UAAs into proteins in eukaryotes, each of which must be charged by a distinct orthogonal pair.[3,12–14]

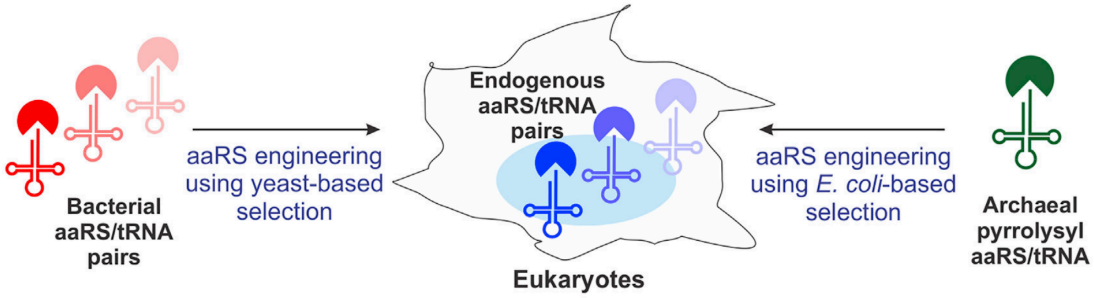


Figure 3-1. Two GCE platforms. Bacteria-derived aaRS/tRNA pairs are typically suitable for expanding the genetic code of eukaryotes, but engineering their substrate specificity must be performed using a yeast-based selection system. The unique pyrrolysyl pair from archaea can be used for Uaa mutagenesis in both bacteria and eukaryotes. Thus, it can be engineered to charge desired UAAs using the E. coli selection system, followed by application in eukaryotes.

3.1.2 Genetic code expansion and previous selection system paradigms

The first demonstrations of site-specific UAA incorporation in eukaryotic cells was achieved over 15 years ago using a bacteria-derived tyrosyl-tRNA synthetase (TyrRS)/tRNA pair.[8,15] However, this pair has been engineered to charge only a small collection of simple tyrosine analogs so far, with no new engineered variants reported in the last decade.[2,3,10] This can be largely attributed to the limitations associated with the aforementioned yeast-based selection system, which is currently needed to engineer its substrate specificity. In contrast, the analogous *Methanococcus jannaschii* (archaea)-derived TyrRS (MjTyrRS)/tRNA pair, which is orthogonal in bacteria (but not in eukaryotes), has been engineered using the *E. coli* selection system to charge over 70 UAAs, including a variety of bioconjugation handles,[16–18] photoaffinity probes,[16,19] fluorophores,[20] metal-binding groups,[21] light-responsive functionalities,[22] posttranslational modifications of tyrosine[23,24] etc., many of which cannot currently be incorporated into proteins expressed in eukaryotic cells.[2,5] The ability to similarly

engineer the bacterial TyrRS/tRNA pair using the facile *E. coli* selections system can facilitate the introduction of such useful UAAs to the eukaryotic genetic code.

3.1.3 Altered translational machinery of the tyrosyl system

Directed evolution of bacteria-derived aaRS/tRNA pairs in *E. coli* is forbidden due to their cross-reactivity in the host cell. However, functionally replacing an endogenous aaRS/tRNA pair of E. coli with a eukaryotic/archaeal counterpart can enable its reintroduction in the resulting "altered translational machinery" (ATM) strain as an orthogonal nonsense suppressor (Figure 3-2).[10,25,26] Recently, we demonstrated that such an ATM strain can serve as the selection host for altering the substrate specificity of the "liberated" bacterial tryptophanyl-tRNA synthetase/tRNA pair.[10] The resulting engineered variants of this pair then can be used for UAA mutagenesis both in eukaryotes and in the engineered *E. coli* strain.[10] Here, we extend this strategy to liberate the endogenous TyrRS (EcTyrRS)/tRNA pair from *E. coli*, complemented by an archaeal counterpart (Figure 3-2). Furthermore, we demonstrate the feasibility of using the resulting ATMY strains to readily evolve variants of the EcTyrRS/tRNA pair, which facilitate the incorporation of several UAAs in both the ATM *E. coli* strain, as well as in mammalian cells.

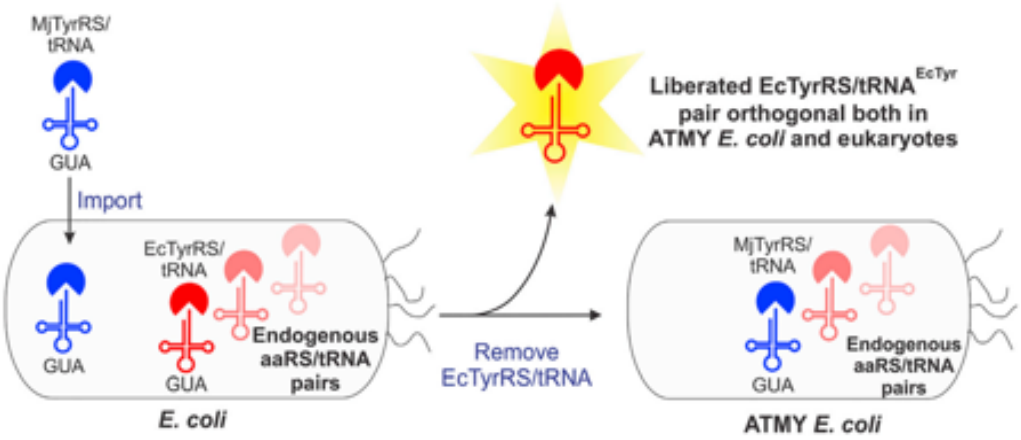


Figure 3-2. ATMY scheme. Functionally substituting the endogenous EcTyrRS/tRNA pair of E. coli with an archaeal counterpart liberates it for reintroduction into the resulting ATMY strain as an orthogonal nonsense suppressor, where its substrate specificity can be engineered using the facile E. coli-based selection system.

3.2 Results and discussion

3.2.1 Functionally substituting the endogenous EcTyRS/tRNA pair of E. coli

Iraha *et al.*[26] have previously substituted the endogenous tyrosyl pair of *E. coli* with eukaryotic and archaeal counterparts, corroborating the feasibility of our approach. However, the resulting strains suffer from growth defects of unknown origin.[26] In addition, whether the resulting *E. coli* strains can be used as the selection host to alter the substrate specificity of the liberated EcTyrRS/tRNA pair has not been explored. Such efforts are not trivial, as the nonsense-suppressing variant of the liberated endogenous pair can cross-react with other endogenous pairs.[10,27,28] To better understand the logic of efficiently substituting the endogenous EcTyrRS/tRNA pair with an orthogonal counterpart from a different domain life, we took a systematic approach that involved the removal of the genes for EcTyrRS (*tyrS*)

and $tRNA^{EcTyr}$ (*tyrT*, *tyrU*, and *tyrV*) from the genome using several different strategies, and characterization of the resulting strains.

The archaea-derived MjTyrRS/tRNA pair represents a good candidate for substituting its *E. coli* counterpart, as suggested by its extensive use for efficient nonsense suppression in bacteria.[5,29,30] We encoded the substituting MjTyrRS/ tRNA pair in a pUltra vector (pUltraBR-MjY), which has an unusual CloDF13 origin of replication that is compatible with most commonly used plasmids. The K12-derived EcNR1 strain of *E. coli*, harboring a temperature-inducible λ-Red recombination system, and optimized for efficient genome recombineering, was used as the host. [31] To enable the removal of the genes encoding the EcTyrRS/tRNA pair, we either used gene cassettes selectable using antibiotics (gentamycin or zeocin), or the dual-selectable marker *tolC*, which can be repeatedly used to edit multiple loci in the genome (Figure 3-3).[32]

a)

Strain	*tyrS*	*tyrTV*	*tyrU*	λ-Red
EcNR1Z	WT	WT	WT	ZeoR
ATMY1	-	*tolC*	GentR	ZeoR
ATMY2	-	GentR	*tolC*	ZeoR
ATMY3	-	$tRNA^{EcTyr}_{CUA}$	$tRNA^{EcTyr}_{Inact}$	*tolC*
ATMY4	-	$tRNA^{EcTyr}_{CUA}$	$tRNA^{EcTyr}_{CUA}$	*tolC*
ATMY5	-	$tRNA^{EcTyr}_{Inact}$	$tRNA^{EcTyr}_{Inact}$	*tolC*

Strain	Genotype
ATMY1	EcNR1GT pUltraBR-MjY dtyrS tyrTV::tolC tyrU::gentR lambda::ZeoR
ATMY2	EcNR1GT pUltraBR-MjY dtyrS tyrTV::gentR tyrU::tolC lambda::ZeoR
ATMY3	EcNR1GT pUltraBR-MjY dtyrS tyrTV::$tRNA^{EcTyr}_{CUA}$ tyrU::$tRNA^{EcTyr}_{Inact}$ lambda::tolC
ATMY4	EcNR1GT pUltraBR-MjY dtyrS tyrTV::$tRNA^{EcTyr}_{CUA}$ tyrU::$tRNA^{EcTyr}_{CUA}$ lambda::tolC
ATMY5	EcNR1GT pUltraBR-MjY dtyrS tyrTV::$tRNA^{EcTyr}_{Inact}$ tyrU::$tRNA^{EcTyr}_{Inact}$ lambda::tolC

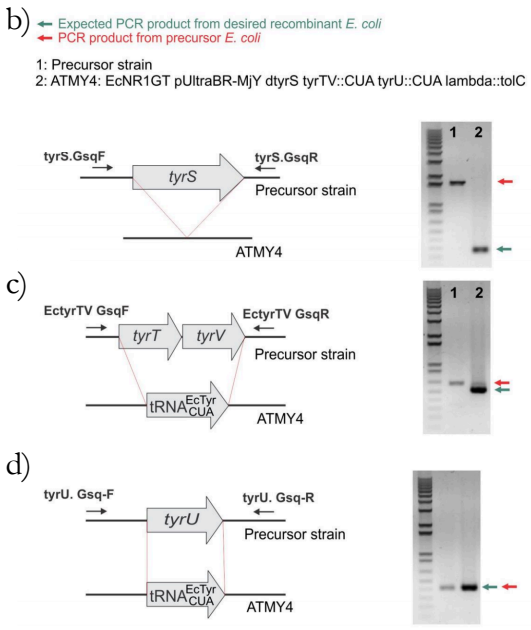


Figure 3-3. ATMY strain recombinations overview. a) Shorthand and full genotypes of strains ATMY1-5, created by first scarlessly deleting tyrS in the presence of pUltraBR-MjY, followed

by the removal of tyrTV, tyrU, and the λ-Red.machinery using the indicated cassettes. b-d elucidate the genomic recombinations of ATMY4 as a representative example. Verification of genomic alterations of the ATMY4 strain using colony PCR coupled with DNA sequencing. The expected size of the PCR products for the precursor and ATMY4 strain is shown in red and green, respectively. In each case, oligonucleotides used to verify each genomic region are shown in the context of a representative map of the corresponding genomic locus. b) Deletion of tyrS in ATMY4 is confirmed by the appearance of the expected shorter PCR product. c) Replacement of tyrTV in ATMY4 by $tRNA^{EcTyr}_{CUA}$ is confirmed by the appearance of the expected slightly shorter PCR product. DNA sequencing confirms the expected change. d) Replacement of tyrU with $tRNA^{EcTyr}_{CUA}$ does not alter the size of the expected PCR product, but DNA sequencing confirms the expected change.

We started by knocking out *tyrS* from the genome in the presence of the pUltraBR-MjY plasmid, by first replacing it with *tolC*, followed by the scarless removal of this dual-selectable marker. From here, we created five different final strains, named ATMY1-5 (Figure 3-3a), by knocking out the three $tRNA^{EcTyr}$ genes either with different selectable markers, or with its variants that suppress the TAG nonsense codon ($tRNA^{EcTyr}_{CUA}$) or inactivated by the truncation of the anticodon loop ($tRNA^{EcTyr}_{Inact}$). The λ-Red machinery was also removed from the genome of each strain in the final step. Analysis of their growth (Figure 3-4) revealed that replacing the $tRNA^{EcTyr}$ genes with their inactivated or TAG-suppressor variants (ATMY3-5) is tolerated better than the use of selectable markers (ATMY1-2), likely by better preserving the endogenous sequence architecture of these loci, and minimizing perturbation to the expression/processing of genes that these tRNAs are clustered with. Strains ATMY1 and 3–5 exhibited similar but somewhat slower growth rate relative to the progenitor strain. It is

possible that the heterologous MjTyrRS/tRNA pair does not interface as efficiently with the *E. coli* translation system as the native pair, which leads to reduced fitness. However, the viability of these strains was sufficiently robust for their seamless use as selection and expression hosts as described below.

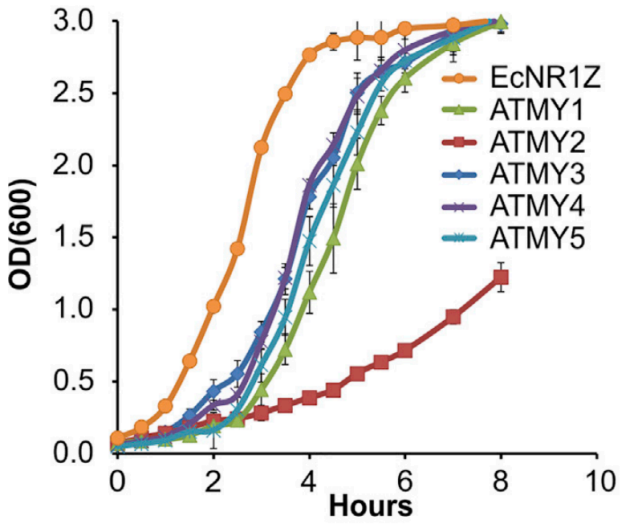


Figure 3-4. Growth rates of resulting strains. Data represents means +/- SD (n=3).

3.2.2 Establishing the EcTyrRS/tRNAEcTyrCUA pair in the ATMY strains as an orthogonal pair

When we attempted to reintroduce the liberated tyrosyl pair as a TAG-suppressor (EcTyrRS/$tRNA^{EcTyr}_{CUA}$) pair in the ATMY1 strain, we encountered two challenges. (1) Co-expression of this pair from plasmids established for other orthogonal pairs (pBK and pRep for aaRS and tRNA, respectively)[6,7] led to a significant level of toxicity. It is likely that the overexpression of the EcTyrRS/$tRNA^{EcTyr}_{CUA}$ pair, which is highly active in *E. coli* due to its native origin, leads to efficient suppression of endogenous TAG stop codons, resulting in the loss of viability. (2) Expression of the $tRNA^{EcTyr}_{CUA}$ alone (in the absence of the cognate EcTyrRS) resulted in a significant level of TAG suppression activity, observed using both a chloramphenicol-acetyl transferase (CAT-TAG) and a superfolder GFP (sfGFP-151TAG)

reporter, each harboring a TAG codon at a permissive site, suggesting that it cross-reacts with another *E. coli* aaRS (Figures 3-5). In the presence of the $tRNA^{EcTyr}_{CUA}$, the sfGFP expression assay showed a small but significant enhancement in cellular fluorescence (Figure 3-5), while in the CAT expression assay, the cells survived up to 60 mg/mL of chloramphenicol (Figure 3-5).

Purification of the resulting sfGFP protein followed by its mass spectrometry (MS) analysis (Figure 3-6) indicated that the $tRNA^{EcTyr}_{CUA}$ is charged by the glutaminyl-tRNA synthetase (GlnRS). Indeed, Söll and coworkers have previously observed that upon overexpression, *E. coli* GlnRS is capable of misacylating the $tRNA^{EcTyr}_{CUA}$.[33] Analogous cross-reactivity of *E. coli* GlnRS has also been reported with TAG-suppressor variants of several other *E. coli*-derived tRNAs, driven by its recognition of the non-native CUA anticodon.[10,27,28,34]

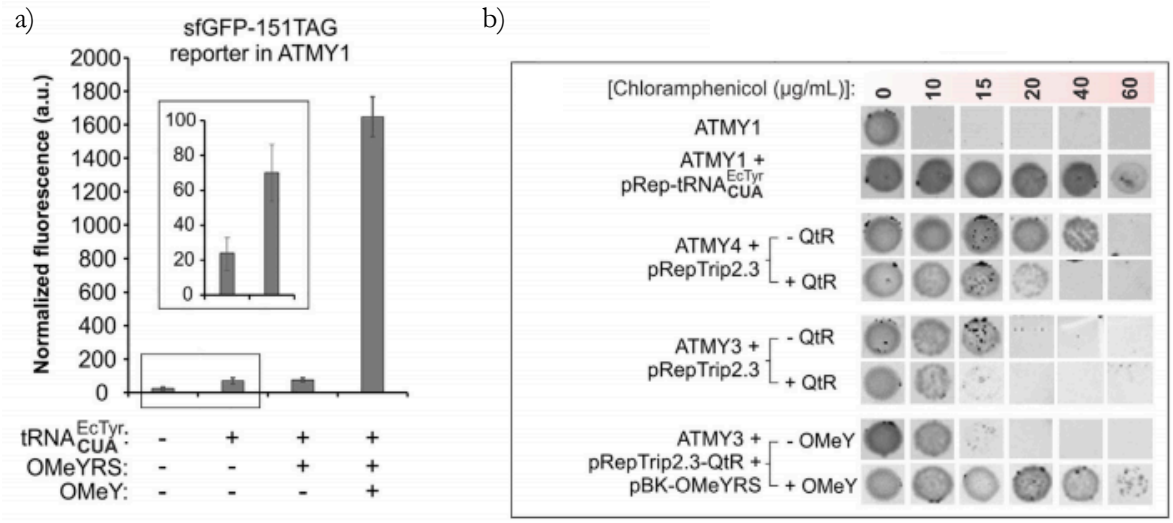


Figure 3-5. Reintroduction of $tRNA^{EcTyr}_{CUA}$ into ATMY strains. a) $tRNA^{EcTyr}_{CUA}$ is weakly active in ATMY1 cells in the absence of EcTyrRS, as indicated by a small but significant increase in sfGFP-151TAG reporter expression (magnified in the inset). Co-expression of the

OMeY-selective EcTyrRS (OMeYRS) significantly increases TAG suppression levels, leading to strong OMeY dependent sfGFP-151TAG expression (see Figure 3-6). Expression of full-length sfGFP was measured using an optical density at 600 nm (OD600)-normalized fluorescence of cells suspended in PBS (the normalized fluorescence of ATMY1 cells not harboring a GFP reporter was subtracted from each). Data represent means ± SD (n = 3). (b) The cross-reactivity of $tRNA^{EcTyr}_{CUA}$ in ATMY1 is further confirmed using a CAT-TAG reporter, which leads to cellular survival at up to 60 μg/mL of chloramphenicol. Lowering the $tRNA^{EcTyr}_{CUA}$ levels by expressing it from the genome in ATMY4 and ATMY3 strains reduces this cross-reactivity (survival up to 40 and 20 μg/mL, respectively), which can be further attenuated by overexpressing the native $tRNA^{Gln}$ from the pRepTrip2.3-QtR plasmid in ATMY4 and ATMY3. Coexpressing OMeYRS in ATMY3 enables OMeY-dependent survival at up to 60 μg/mL, revealing a range where Uaa-selective aaRS variants can be efficiently selected.

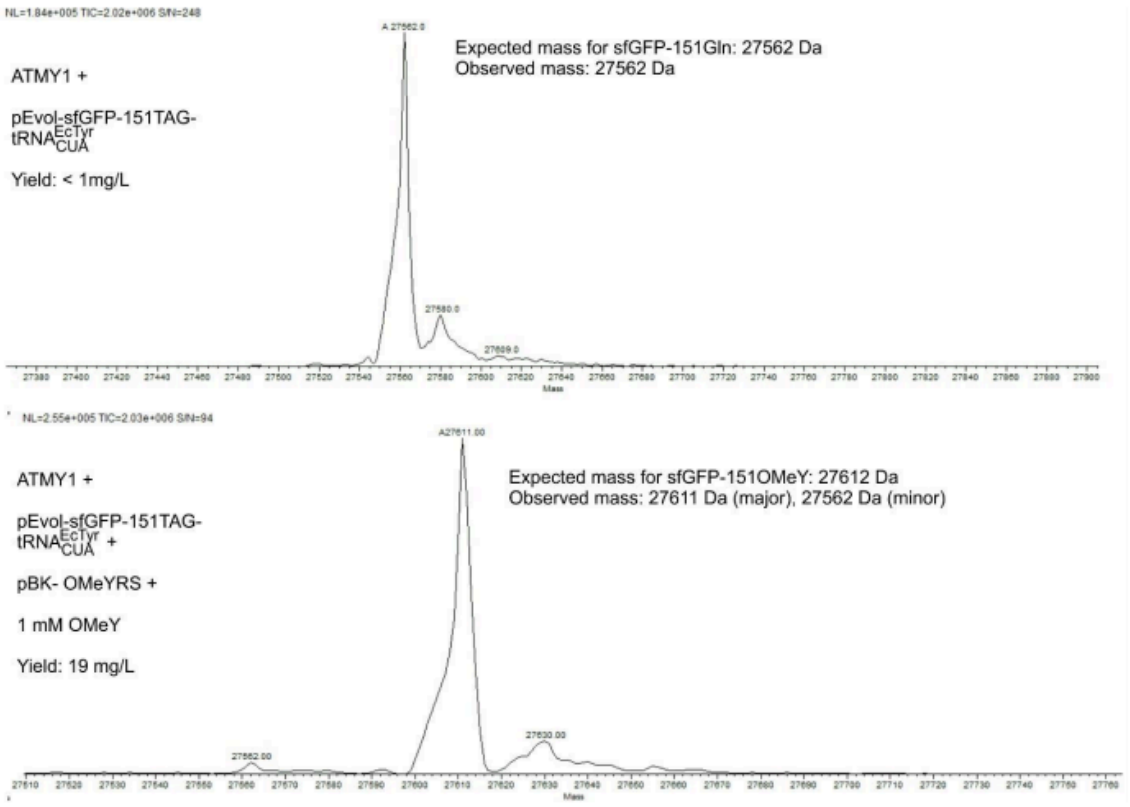


Figure 3-6. Expressing $tRNA^{EcTyr}_{CUA}$ in ATMY1 strain in the absence of EcTyrRS leads to low levels of sfGFP-151-TAG expression, and the mass (top panel) of the isolated protein is consistent with the incorporation of glutamine at this site. It has been previously shown that upon overexpression, *E. coli* GlnRS can charge $tRNA^{EcTyr}_{CUA}$. Co-expression of OMeYRS from a pBK plasmid (driven by the *glnS* promoter) largely outcompetes the misacylation of $tRNA^{EcTyr}_{CUA}$ by GlnRS, confirmed by the mass of the isolated sfGFP-151-TAG reporter (bottom panel).

In this particular case, the degree of the cross-reactivity was weak, and, when a previously developed O-methyltyrosine (OMeY) (Figure 3-7a) selective EcTyrRS mutant (OMeYRS)[8,35] was co-expressed with $tRNA^{EcTyr}_{CUA}$, a significantly higher level of reporter expression was observed in the presence of OMeY (Figure 3-5a), and MS analysis of the

isolated sfGFP showed a mass consistent with the incorporation of OMeY (Figure 3-6). Despite its weak nature, the extent of this cross-reactivity was sufficient to jeopardize the selection scheme that is required to engineer the substrate specificity of EcTyrRS in this strain.

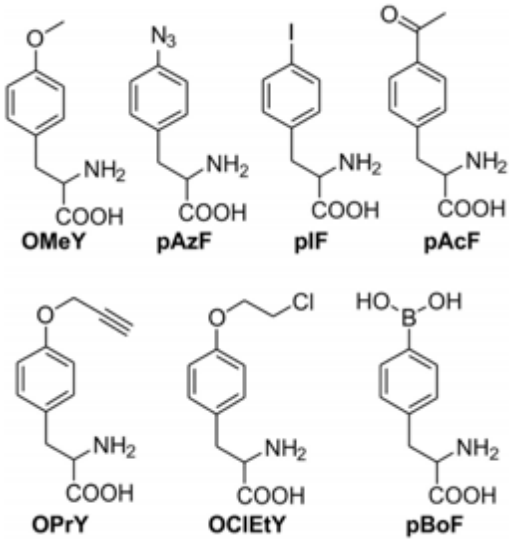


Figure 3-7. UAA used in this study.

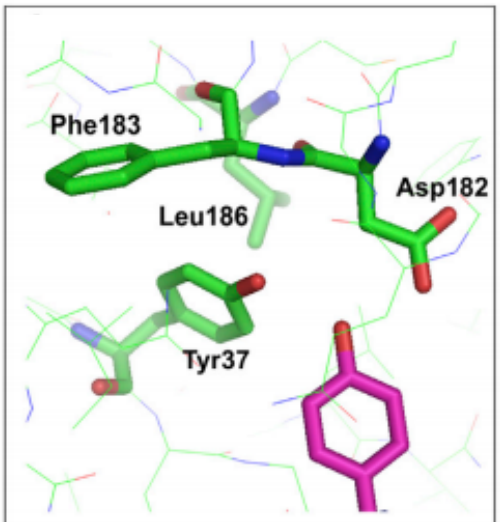
We envisioned that lowering the expression of the highly active $tRNA^{EcTyr}_{CUA}$, by encoding it in the genome instead of a multicopy plasmid, may alleviate both the associated toxicity and the extent of the cross-reactivity. In addition, we thought that the unwanted cross-reactivity could be further attenuated by overexpressing the endogenous $tRNA^{Gln}$, which would presumably outcompete the non-cognate $tRNA^{EcTyr}_{CUA}$ for interaction with GlnRS. We had already created strains ATMY3 and ATMY4, which encode one and two copies of $tRNA^{EcTyr}_{CUA}$ in the genome, respectively (Figure 3A). When subjected to the aforementioned CAT-TAG expression assay, ATMY4 and ATMY3 indeed showed reduced levels of cross-reactivity, surviving up to 40 and 20 μg/mL of chloramphenicol (Figure 3D), respectively. To overexpress the native $tRNA^{Gln}$ in these strains, its gene was amplified from the *E. coli* genome and was incorporated into the multi-copy plasmid pRep to generate pRepTrip2.3-QtR, which

also harbors the CAT-TAG reporter. ATMY4 and ATMY3 strains transformed with the pRepTrip2.3- QtR plasmid survived only up to 20 and 15 μg/mL of chloramphenicol (Figure 3-5b), confirming a significantly attenuated cross-reactivity between GlnRS and $tRNA^{EcTyr}_{CUA}$ in response to the altered expression patterns of these tRNAs. It was also possible to express EcTyrRS from the pBK plasmid in the ATMY3 and ATMY4 strains without compromising their fitness. Owing to the lower degree of background TAG suppression observed in the ATMY3 strain, it represents the most suitable host for the evolution of EcTyrRS. Expression of the aforementioned OMeYRS from a pBK plasmid enabled ATMY3 to survive up to 60 μg/mL of chloramphenicol in the presence 1 mM OMeY, but only up to 15 μg/mL in its absence (Figure 3-5b), confirming that the UAA-selective EcTyrRS mutants can be enriched from their inactive counterparts using this facile antibiotic-based selection in this strain.

3.2.3 Development of a polyspecific EcTyrRS mutant through the directed evolution in ATMY3

The ability to establish the EcTyrRS/$tRNA^{EcTyr}_{CUA}$ pair as an adequately orthogonal nonsense suppressor in ATMY3 opens the door to allow its directed evolution for generating novel UAA-selective variants, using the facile double-sieve selection system hosted in *E. coli*. We have shown that the pRepTrip2.3-QtR plasmid enables the survival of the UAA-selective EcTyrRS variants in the presence of its cognate UAA (Figure 3-5b), providing an efficient mechanism for positive selection. To enable the removal of EcTyrRS variants that charge canonical amino acids (the negative selection),[7] we also created the pNeg-QtR plasmid, encoding the TAG-inactivated toxic gene barnase, as well as the $tRNA^{Gln}$ expression cassettes. To verify the efficacy of the ATMY3-based selection system, we first sought to develop an OMeY-selective EcTyrRS variant, the feasibility of which has been previously demonstrated using the yeast-based selection system.[8]

Based on the available crystal structure (PDB: 1X8X), four key amino acid residues (Tyr37, Asp182, Phe183, and Leu186) in the active site of EcTyrRS were randomized by site-saturation mutagenesis (Figure 3-8) to create library EcYRS1a (as stored in -80°C). This library of mutants was then subjected to alternating positive and negative selection steps in the ATMY3 strain in the presence and absence of 1 mM OMeY, respectively. After just one round of each selection, the mutant library exhibited a significant OMeY-dependent survival in the subsequent round of positive selection, indicating successful enrichment of OMeY-selective EcTyrRS variants. Individual colonies were then screened for OMeY-dependent survival under positive selection conditions, and ten successful clones were characterized by DNA sequencing. Five unique EcTyrRS sequences were identified, which were homologous to each other, as well as to the previously identified OMeYRS (Figure 4A): Tyr37 mutated to hydrophobic residues (Val or Ile), Asp182 to small residues (Ser or Cys), Phe183 to Met, and Leu186 to small residues (Ala or Ser).



Clones	Tyr37	Asp182	Phe183	Leu186
ISMA	Ile	Ser	Met	Ala
VCMA	Val	Cys	Met	Ala
VSMS	Val	Ser	Met	Ser
VSMA	Val	Ser	Met	Ala
VSMA*	Val	Ser	Met	Ala
OMeYRS (previous)	Val	Ser	Met	Leu

*Additional D165G mutation

Figure 3-8. Active site of EcTyrRS (PDB: 1X8X). The bound tyrosine is shown in magenta and the active site residues subjected to randomization are highlighted and sequence of mutants that show OMeY-depdendent survival under positive selection conditions.

To evaluate and compare the OMeY-charging efficiency of these mutants relative to the previously developed OMeYRS, we employed the sfGFP-151TAG reporter expression assay in the ATMY4 strain, which was expected to facilitate higher reporter expression (two copies of the $tRNA^{EcTyr}_{CUA}$ in the genome). Each EcTyrRS mutant facilitated robust full-length reporter expression, observed using its characteristic fluorescence, only in the presence of 1 mM OMeY (Figure 3-9a). Intriguingly, the most efficient mutant (OMeYRS-VSMA*), which facilitated improved reporter expression relative to the original OMeYRS, had a fortuitous Asp165Gly mutation that was found to be beneficial. This Asp residue sits in close proximity

with the $tRNA^{EcTyr}$ backbone and its mutation to Gly may enhance the EcTyrRS-tRNA interaction (Figure 3-9b). Successful incorporation of OMeY was confirmed by MS analysis (Figure 3-10) of the reporter protein following its isolation using a C-terminal polyhistidine tag by immobilized metal-ion chromatography.

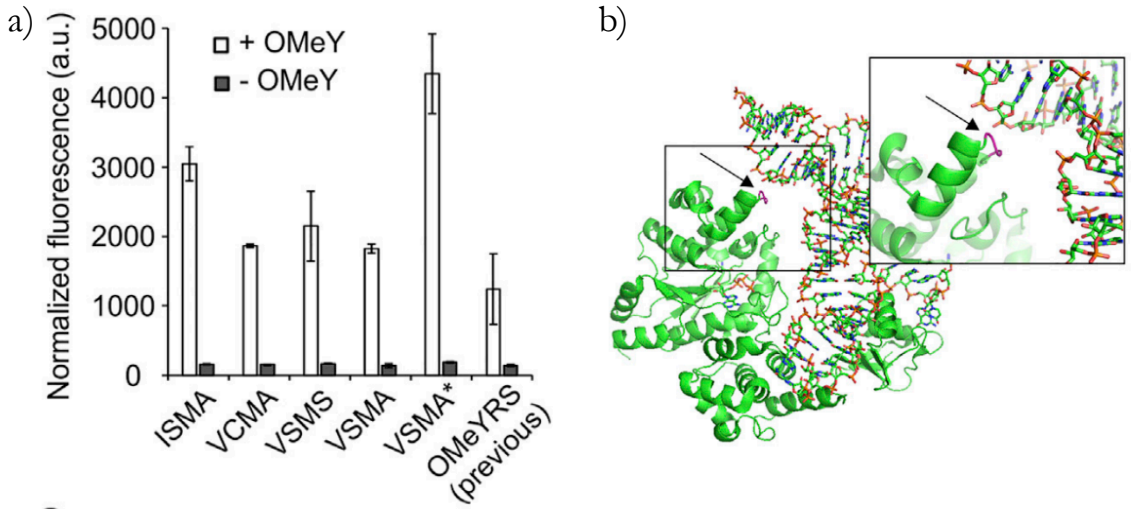


Figure 3-9. Characterization of OMeY-Selective EcTyrRS mutants in *E. coli*. a) OMeY-selective EcTyrRS mutants show OMeY-dependent sfGFP-151TAG reporter expression in ATMY4 (OD600-normalized fluorescence of suspended cells). The activity of the previously developed OMeYRS is shown for comparison (Figure 3-10). Data represent means ± SD (n = 3). b) The location of the D165G mutation mapped to *T. thermophilus* TyrRS/tRNA co-crystal structure (PDB: 1H3E). The black arrows point arrows point at the location of this residue, which is colored magenta.

	Uaa	Yield (μg/ 10^7 cells)	Yield (mg/L)	Expected Mass (Da)	Observed mass (Da)
EGFP-39TAG reporter expressed in HEK293T cells	OMeY	127	-	29698	29696
	OPrY	165	-	29722	29720
	OClEtY	107	-	29746	29745
	pAzF	77*	-	29709	29707, 29680 (minor, from azide reduction)
	pAcF	176		29710	29708
	pIF	54	-	29794	29792
	pBoF	164	-	29710	29708, 29691
	Wild-type (Tyr)	240	-	29684	27596
sfGFP-151TAG reporter expressed in ATMY4	OMeY	-	19	27612	27611
	OPrY	-	7	27636	27634
	OClEtY	-	3	27660	27659
	pAzF	-	25	27623	27621
	pBoF	-	18	27626	27624, 27605
	Wild-type (Tyr)	-	46	27598	27596

Figure 3-10. Yield and MS analysis for reporter proteins. The EGFP39TAG reporter was expressed in HEK293T cells in the presence of engineered EcTyrRS/$tRNA^{EcTyr}_{CUA}$ pairs and its UAA substrates. VSMA* (Figure 3-8) EcTyrRS mutant was used to incorporate OMeY, OPrY, OClEtY, pAzF, pAcF and pIF while the GSIE mutant was used charge pBoF. The sfGFP-151TAG reporter was expressed in ATMY4 in the presence of engineered EcTyrRS/$tRNA^{EcTyr}_{CUA}$ pairs and its UAA substrates. VSMA* (Figure 3-8) EcTyrRS mutant was used to incorporate OMeY, OPrY, OClEtY and pAzF, while the GSIE mutant was used charge pBoF. Overexpression of the EcTyrRS/$tRNA^{EcTyr}_{CUA}$ pair from optimized UAA mutagenesis plasmids (e.g., pEVOL) led to severe toxicity (due to high efficiency of the native pair). In this case, $tRNA^{EcTyr}_{CUA}$ was expressed from the genome (two copies) and the EcTyrRS variants were expressed from the weak glnS promoter encoded in the pBK plasmid. Incorporation efficiencies of these UAAs can be further improved by optimizing the

expression levels of the EcTyrRS/$tRNA^{EcTyr}_{CUA}$ pair, to enhance suppression efficiency without compromising the cellular fitness.

To demonstrate that the EcTyrRS mutants developed here would be beneficial for UAA mutagenesis in mammalian cells, we cloned OMeYRS-VSMA* into the previously described mammalian expression plasmid pAcBac1-16xtRY,[35,36] also encoding 16 copies of its cognate tRNA, and its activity was compared with an otherwise identical plasmid, encoding the previously developed OMeYRS. In addition, we incorporated the Asp265Arg mutation into each EcTyrRS variant, which confers improved interaction with the $tRNA^{EcTyr}_{CUA}$.[37] When co-transfected into HEK293T cells with a plasmid expressing EGFP39TAG, both plasmids facilitated OMeY-dependent EGFP expression (Figures 3-11). It was previously found that the original OMeYRS can charge several other para- substituted phenylalanine analogs,[35] (including those with bioorthogonal conjugation handles, significantly expanding its utility. We showed that our OMeYRS-VSMA* is also capable of charging these useful UAAs, in some cases (e.g., O-propargyl-tyrosine) with improved efficiency (Figures 3-11). Successful incorporation of each UAA was confirmed by the MS analysis (Figure 3-10) of the reporter protein following isolation with excellent yields (60–150 μg per 10-cm dish; 25%–60% relative to the wild-type EGFP reporter). The ability to enrich active OMeY-selective variants from a naive EcTyrRS library validates the utility of the ATMY3-based selection scheme. Furthermore, the development of a polyspecific EcTyrRS variant with improved efficiency will be useful for single and dual UAA mutagenesis in mammalian cells.

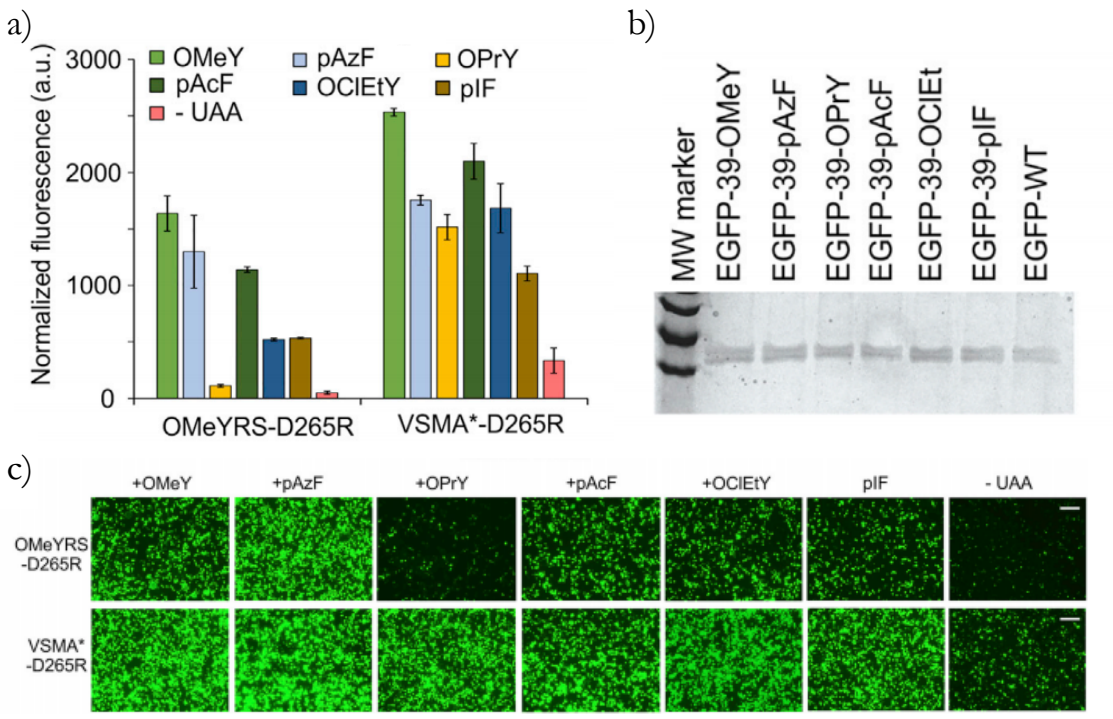


Figure 3-11. Characterization of OMeY-Selective EcTyrRS mutant VSMA* in HEK293T. The VSMA* mutant also facilitates improved OMeY incorporation into the EGFP-39TAG reporter expressed in HEK293T cells, relative to the previously reported OMeYRS, observed by a) measuring EGFP fluorescence in clarified cell-free extract c) fluorescence microscopy. b) SDS-PAGE analysis of EGFP-39TAG reporters incorporating different UAAs, charged by the VSMA* mutant in HEK293T cells.

3.2.4 Engineering EcTyrRS to charge p-Boronophenylalanine

The MjTyrRS/tRNA pair has been engineered to genetically encode p-boronophenylalanine (pBoF) (Figure 3-7) in *E. coli*,[38] where it has facilitated several enabling applications including the development of engineered carbohydrate-binding proteins,[38,39] genetically encoded fluorescent sensors for physiologically relevant reactive oxygen species (e.g., hydrogen peroxide, peroxynitrite),[40–42] novel bioconjugation strategies,[43] etc. The ability

to engineer EcTyrRS to charge pBoF with high fidelity and efficiency will extend the utility of this versatile chemical handle for application in eukaryotic cells. We subjected the aforementioned EcTyrRS mutant library in the ATMY3-based selection system to enrich potential mutants capable of charging pBoF. Numerous colonies were identified after just two rounds of positive and a round of negative selection, which exhibited pBoF-dependent survival under positive selection conditions. Characterization by DNA sequencing identified three distinct but convergent clones, where Tyr37 mutates to Gly or Ser, Asp182 to Ser, Phe183 to Ala or Ile, and Leu186 to Asp (Figure 3-12).

Clones	Tyr37	Asp182	Phe183	Leu186
SSAE	Ser	Ser	Ala	Glu
SSIE	Ser	Ser	Ile	Glu
GSIE	Gly	Ser	Ile	Glu

Figure 3-12. Sequence of mutants that show pBoF-dependent survival under positive conditions.

The hydrophilic nature of these engineered active sites are in agreement with the need to accommodate the polar boronic acid functionality. The aforementioned sfGFP-151TAG reporter expression assay in ATMY4 was used to confirm that all three mutants charge pBoF with high fidelity and efficiency (Figure 3-13a). MS analysis of the reporter protein confirmed the expected mass along with a dehydrated peak, characteristic of phenyl boronic acids (Figure 3-13b; Figure 3-10).[30,38,44,45]

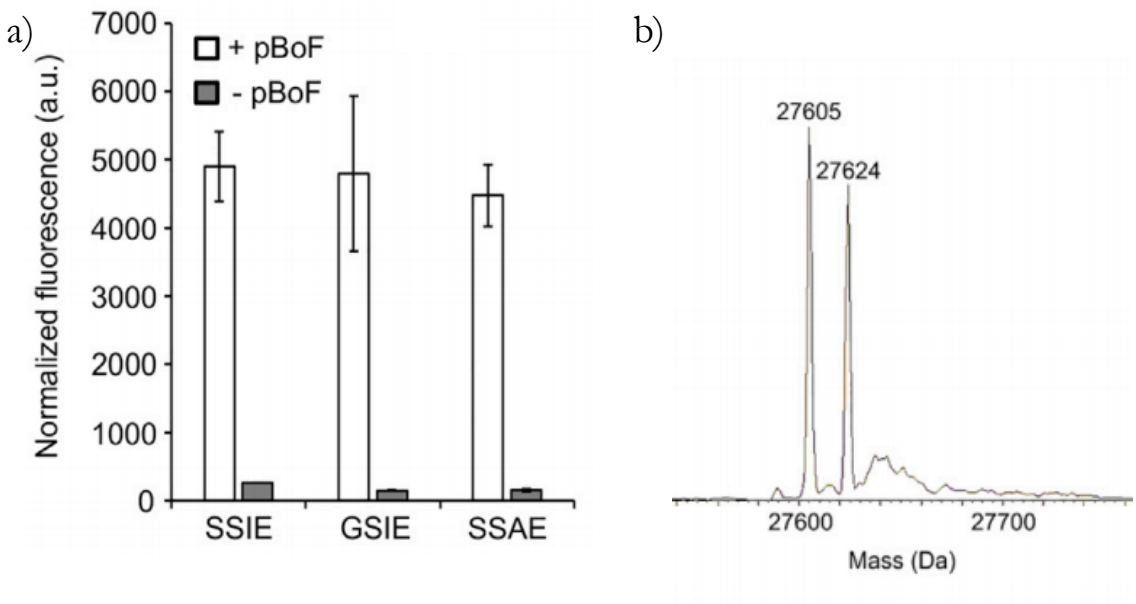


Figure 3-13. Characterization of pBoF-Selective EcTyrRS mutants in *E. coli*. a) Each of three mutants facilitate pBoF-dependent sfGFP-151TAG reporter expression in ATMY4. b) MS analysis of the reporter protein confirm pBoF incorporation (expected mass: 27,626 Da, additional dehydration peak is an established behavior of phenylboronates).

All three mutants were cloned into the mammalian expression vector described above, further incorporating the beneficial Asp265Arg mutation. Co-transfection of HEK293T cells with the EGFP-39TAG reporter resulted in efficient reporter expression only in the presence of 1 mM pBoF (Figure 3-14). The reporter protein in each case was isolated in good yield (140–160 μg per 10-cm dish; 58%–65% relative to the wild-type EGFP reporter; Figure 3-10). Taking further advantage of this excellent efficiency, we demonstrated the ability to incorporate pBoF into two different sites in the EGFP-39TAG-151TAG reporter expressed in HEK293T cells with only a minor loss in the yield (Figure 3-15). Such efficient multi-site incorporation of pBoF can facilitate the development of novel carbohydrate-binding proteins (e.g., antibodies), strategically employing multiple boronate handles.

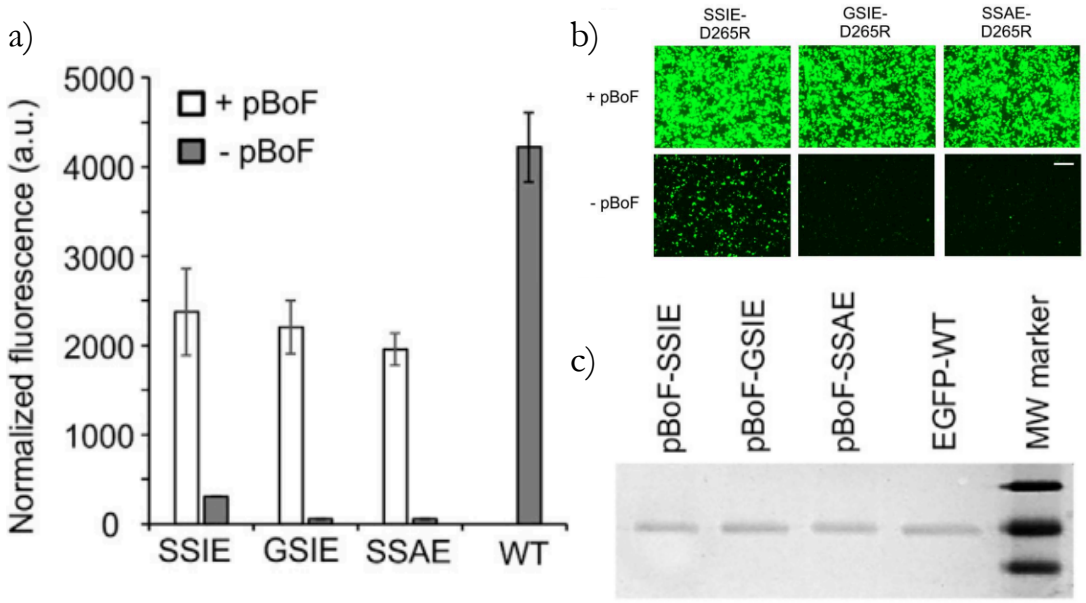


Figure 3-14. Characterization of pBoF-Selective EcTyrRS mutants in HEK293T. a) These mutatns charge pBoF into the EGFP-39TAG reporter expressed in HEK293T cells monitored by EGFP fluorescence in cell-free extract. Expression of the wild-type EGFP reporter is also shown for reference, demonstrating high efficiency of pBoF incorporation. Data represent means +/- SD (n=3). b) Fluorescence microspy images of HEK293T cells showing pBoF-dependent EGFP-39TAG expression. Scale bar, 200 μm. c) SDS-PAGE of pBoF-incorporating reporter proteins.

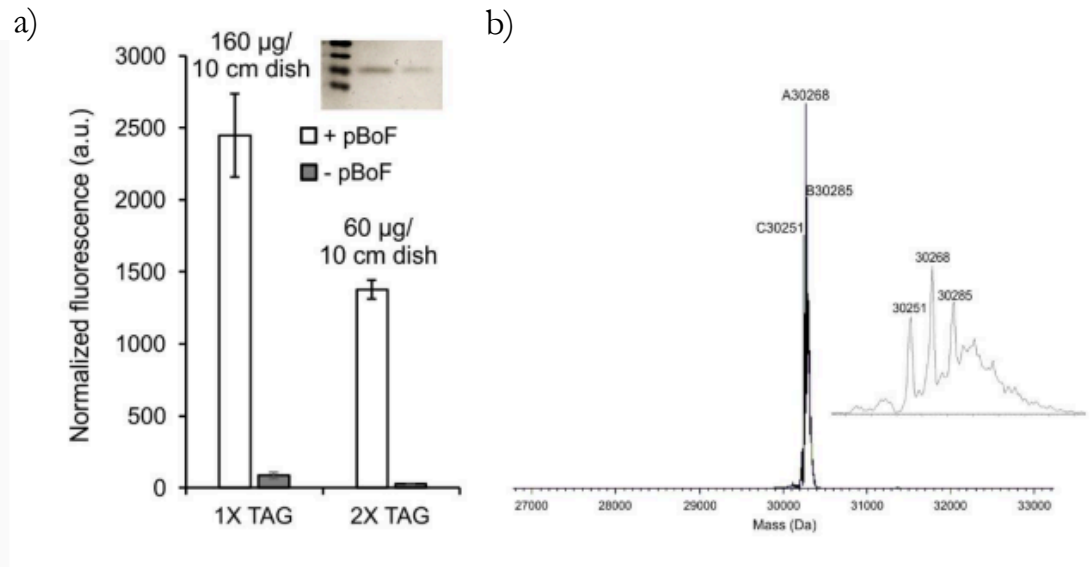


Figure 3-15. Incorporation of multiple pBoF into a single EGFP protein in HEK293T. a) Expression of the EGFP-39TAG-151TAG reporter in HEK293T cells incorporating pBoF using the GSIE-EcTyrRS/tRNA$^{EcTyr}_{CUA}$ pair. Expression of the EGFP-39TAG under same conditions is also shown for reference. b) The MS analysis of the double-pBoF mutant show the presence of the correct mass (expected: 30284 Da), as well as corresponding single and double dehydration products

3.3 Conclusions

The great potential of the TyrRS/tRNA pair for driving the genetic code expansion technology has been highlighted by the remarkable success of the MjTyrRS/tRNA pair in E. coli.[2,5] However, the use of the EcTyrRS/tRNA pair to similarly expand the eukaryotic genetic code has been significantly limited by the challenges associated with the eukaryote-hosted selection system required to engineer its substrate specificity.[2,3,10] Recently, the universally orthogonal pyrrolysyl pair has been engineered using the E. coli selection system to charge several hydrophobic phenylalanine derivatives, providing an alternative platform for

genetically encoding such UAAs in eukaryotes.[2,11] However, it has been particularly difficult to use this platform to genetically encode hydrophilic amino acids such as pBoF. The ability to readily engineer the EcTyrRS/tRNA pair using the *E. coli*-based selection scheme overcomes these limitations, paving the way for potentially introducing numerous other highly sought-after UAAs, such as those mimicking post-translational modifications of tyrosine (sulfation, phosphorylation, etc.),[23,24] into the genetic code of eukaryotes. In addition, the EcTyrRS/tRNA pair have been used together with the pyrrolysyl pair to site-specifically incorporate two distinct UAAs into proteins expressed in mammalian cells.[12,13] The scope of this technology would be further expanded with the ability to incorporate a broader set of useful UAAs using this pair.

Our work also reveals several important insights for successful functional replacement of aaRS/tRNA pairs in *E. coli* with evolutionarily distant counterparts. We found that removing endogenous tRNA genes using large selectable markers may lead to growth defects. The tRNA genes are often clustered with other tRNAs and other genes, which are co-transcribed, and the resulting transcript is subsequently processed. It is possible that the insertion of large genetic elements may perturb these processes. Replacing the tRNA genes with their inactivated or nonsense-suppressing counterpart can alleviate these concerns. It is possible that the minor growth defects of the ATMY strains are a consequence of the suboptimal performance of the substituting MjTyrRS/tRNA pair. This heterologous pair—the tRNA in particular—did not evolve to optimally interface with the *E. coli* translation machinery. Furthermore, this pair may not be fully orthogonal in *E. coli*. Indeed, MjTyrRS was recently shown to be capable of charging the endogenous prolyl-tRNA of *E. coli* at very low levels.[46] Consequently, it may be possible to further overcome the minor residual growth

defects of ATMY strains through optimizing the performance of the substituting MjTyrRS/tRNA pair.

The cross-reactivity of the nonsense-suppressing variants of endogenous tRNAs with other host synthetases represents a major challenge in our strategy to enable directed evolution of endogenous aaRS/tRNA pairs in *E. coli*. GlnRS appears to be a prevalent source of such cross-reactivity, particularly with TAG-suppressor tRNAs.[10,27,28,33,34] We have previously shown that misacylation by GlnRS can be avoided by using a TGA-suppressor $tRNA^{EcTrp}$ instead of its TAG-suppressing counterpart.[10] This strategy did not work for $tRNA^{EcTyr}$ due to its poor TGA suppression efficiency. However, lowering its expression, coupled with overexpression of endogenous tRNAGln, sufficiently attenuated the degree of this misacylation. To further eliminate the residual cross-reactivity, a more orthogonal $tRNA^{EcTyr}_{CUA}$ can be engineered in the future through its directed evolution in the ATMY strains.

Although the MjTyrRS/tRNA pair has already been well established for Uaa mutagenesis in *E. coli*, the EcTyrRS/ $tRNA^{EcTyr}_{CUA}$ pair provides an alternative platform in ATMY strains. The very high efficiency of this native pair in *E. coli*, which enables high level of TAG suppression from just one or two $tRNA^{EcTyr}_{CUA}$ genes encoded in the *E. coli* genome, represents a unique advantage of this system. In contrast, the MjTyrRS/tRNA pair must be heavily overexpressed from optimized multi-copy vectors to provide good suppression efficiency.[29,30,47]

In conclusion, here we further demonstrate the utility of "liberating" the endogenous aaRS/tRNA pairs of *E. coli*, and their use for genetic code expansion of both *E. coli* and eukaryotes. The ability to readily engineer the substrate specificity of the EcTyrRS/tRNA pair using the facile *E. coli* selection system resurrects its untapped potential to further expand the catalog of useful UAAs genetically encoded in eukaryotes.

3.4 Experimental procedures

3.4.1 Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
ATMY1	This paper	N/A
ATMY2	This paper	N/A
ATMY3	This paper	N/A
ATMY4	This paper	N/A
ATMY5	This paper	N/A
EcNR1Z	This paper	N/A
EcNR1GT	This paper	N/A
EcNR1	George Church/Addgene	Cat#26930
E.coli DH10B	ThermoFisher Scientific	Cat#18297010
Chemicals, Peptides, and Recombinant Proteins		
O-methyltyrosine	Fisher Scientific	Cat#AAH6309606
4-azido-L-phenylalanine	Chem-Impex	CAT#06162
o-propargyltyrosine	Syntides	CAT#11252961
p-acetylphenylalanine	Chem-Impex	CAT#24756
p-iodophenylalanine	Chem-Impex	CAT#03376
p-boronophenylalanine	Fisher Scientific	Cat#AC358932500
Experimental Models: Cell Lines		
HEK293T	ATCC	Cat#CRL-1573

3.4.2 Bacterial Strains and Mammalian Cell Lines

The *E. coli* strain DH10b (Life Technologies) was used for cloning and propagation. The bacterial EcNR1 progenitor strain was a gift from Prof. George M. Church. Strains were propagated at 37°C or 30°C for desired time while shaking at 250 RPM in sterile culture tubes. The strains were grown on LB-agar plates and liquid LB medium with the following antibiotic concentrations unless noted otherwise: 95 µg/mL spectinomycin, 50 µg/mL chloramphenicol, 10 µg/mL gentamycin, 100 µg/mL ampicillin, 15 µg/mL zeocin, 12 µg/mL tetracycline, 30 µg/mL kanamycin.

The HEK293T cell line (Sex: Female) was purchased from ATCC (ATCC CRL-3216) and was used without further authentication. Cells were maintained in DMEM, high glucose media, supplemented with 10% FBS, in the presence of Penicillin/Streptomycin, at 37°C, 100% humidity, 5% CO2.

3.4.3 Lambda-Red recombination

All strains were derived from EcNR1 which encodes the temperature inducible λ-Red recombinase genes (exo, beta, gam). Cultures of EcNR1, or strains derived from it, were inoculated in 10 mL LB and grown to 0.5 OD_{600} at 30°C. Upon reaching 0.5 OD_{600}, the λ-Red recombinase genes were induced for 15 minutes in a 42°C water bath. After induction, cells were pelleted at 4,500 x g and resuspended in residual ddH2O (200-300 mL). 50 µl aliquots were electroporated with 50 ng dsDNA or 2 µM DNA oligos containing flanking regions of 40-80 bp of genomic homology. After 1-6 hr of recovery, cells were plated on LB-agar plates containing suitable selection agents (e.g., antibiotics, SDS to select for *tolC* insertion, etc). Colonies from selection plates were individually re-streaked and analyzed by colony PCR, followed by sequencing, using KAPA-2G polymerase (Kapa Biosystems) to verify desired genotypes. EcNR1 was additionally engineered to create a new precursor strain EcNR1GT (EcNR1 Δ*galK* Δ*tolC*) to take advantage of dual selectable markers *galK* and *tolC*. All following recombinations were performed with the above protocol. A list of oligonucleotides is listed in the appendix.

Strategy common to ATMY1-5: EcNR1GT was transformed with pUltraBR-MjY. To genomically remove the *E. coli* tyrosyl-tRNA synthetase (*tyrS*), the gene encoding *tolC* was PCR amplified using primers EcYS.tolC-F2 and EcYS.tolC-R2 to generate the PCR product *tyrS*::*tolC*. 50 ng of the *tyrS*::*tolC* PCR cassette was transformed to facilitate recombination following the aforementioned protocol,[10] and the resulting strain was plated on LB-Agar plates supplemented with 0.005% SDS. The resulting colonies were characterized via colony PCR, as well as DNA sequencing the colony PCR products. The intermediate strain was stocked with genotype EcNR1GT pUltraBR-MjY *tyrS*::*tolC*. Subsequently, dual selectable *tolC* was

deleted from the genome using the 90 bp oligo tyrS.90 del and plating on LB-Agar plates supplemented with colicin and 30 mg/mL Vancomycin, resulting in intermediate strain EcNR1GT pUltraBR MjY Δ*tyrS*. This intermediate strain was used to build the five variants of tyrosyl ATM strains.

3.4.4 Genomic recombination to build ATMY1

Wild-type *E. coli* contains three copies of the $tRNA^{Tyr}_{GUA}$, localized in two genomic locations. The two tRNAs *tyrT* and *tyrV* are clustered in the same cassette, while *tyrU* is in a different genomic location. To build ATMY1, *tyrTV* was deleted by recombination using the *tolC* selectable marker. The *tolC* marker was amplified with EcY_TV.tolC-F and EcY_TV.tolC-R, resulting in PCR cassette tyrTV::tolC. The tyrTV::tolC PCR product was transformed in the recombination following the aforementioned protocol and plated on 0.005% SDS. Colonies were verified, resulting in the intermediate strain EcNR1GT pUltraBR MjY Δ*tyrS tyrTV*::*tolC*. To replace the *E. coli* tRNATyr (*tyrU*), antibiotic resistance gene $Gent^R$ was amplified using tyrU.KO.gentR-RBS-F and tyrU.KO.gentR-R to create *tyrU*::$Gent^R$ and used with the recombination protocol, creating intermediate EcNR1GT pUltraBR MjY Δ*tyrS tyrTV*::*tolC tyrU*::$Gent^R$. In order to delete the λ-Red genes from the final strain, a second antibiotic resistance gene ZeoR was amplified with primers dlambda-ZtattB-F and dlambda-ZtattB-R to create PCR product λ-Red::ZeoR. The PCR product was transformed and recombinant hits were selected on LB-Agar supplemented with Zeocin. Confirmed screening resulted in final strain ATMY1 (EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV*::*tolC tyrU*::$Gent^R$ λ-Red::ZeoR).

3.4.5 Genomic recombination to build ATMY2

To build ATMY2, *tyrTV* was deleted using a Gent^{R} selectable marker. The Gent^{R} marker was amplified with tyrTV.KOgent-F and tyrTV.KOgent-R, resulting in PCR cassette *tyrTV*::Gent^{R}. The *tyrTV*::Gent^{R} PCR product was transformed in the recombination following the aforementioned protocol and plated on gentamycin. Colonies were verified, resulting in intermediate strain EcNR1GT pUltraBR MjY ⊿tyrS *tyrTV*::Gent^{R}. To replace the E. coli tRNA$^{Tyr}_{CCA}$ (*tyrU*), the *tyrU*::*tolC* cassette was amplified using tyrU.KO.tolCRBS-F and tyrU-tolC.R and used with the recombination protocol with cells plated on LB supplemented with 0.005% SDS, creating intermediate EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV*::Gent^{R} *tyrU*::*tolC*. The λ-Red genes were removed from the final strain using a ZeoR cassette as described above. Colony screening by PCR and DNA sequencing confirmed the generation of the final strain ATMY2 (EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV*::Gent^{R} *tyrU*::*tolC* λ-Red::Zeo^{R}).

The second generation of strains, ATMY3-5, were generated from EcNR1GT pUltraBR MjY ⊿*tyrS* using the dual-selectable marker *tolC*. Cells were plated on either LB-Agar supplemented with 0.005% SDS to select for the presence of tolC or LB-Agar supplemented with colicin+vancomycin to select for the absence of *tolC*. In the second generation of strains, the tyrosyl-tRNA genes encoded in the *tyrTV* and *tyrU* locus are replaced with either tRNA$^{Tyr}_{CUA}$, its TAG-suppressing counterpart, or tRNA^{Tyr} (Inact), which contains an inactivated tRNA that was generated by deleting two residues in the anticodon (GUA/XXA).

3.4.6 Genomic recombination to build ATMY3

The previously built intermediate strain (EcNR1GT pUltraBR-MjY ⊿*tyrS*) was used to build ATMY3. The *tolC* cassette was amplified with primers EcY_TV.tolC-F and EcY_TV.tolC-R to create tyrTV::tolC. Next, *tolC* was replaced from the *tyrTV* location with PCR cassette tyrTV::CUA, which was amplified from a gblock (IDT) using tyrTV.lhm-F and tyrTV.rhm-F, and was designed to introduce the $tRNA^{Tyr}_{CUA}$ into this locus expressed from the endogenous *tyrTV* promoter. This resulted in the intermediate strain EcNR1GT pUltraBR MjY DtyrS tyrTV::CUA. Next, tyrU was replaced by tolC, using the tyrU::tolC cassette (amplified with primers tyrU.KO. tolC-RBS-F and tyrU-tolC.R), to develop strain EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::CUA tyrU::tolC*. Afterwards, the *tolC* cassette at the *tyrU* locus was replaced with the $tRNA^{Tyr}_{Inact}$ cassette amplified from a tyrU.inact-gblock with primers tyrU.lh-F and tyrU.rh-R via tolC negative selection, resulting in strain EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::CUA tyrU::Inact*. Finally, the λ-RED genes were deleted using the PCR cassette λ-RED::tolC, amplified with primers dlambda.tolC-F and dlambda.tolC-R (tolC positive selection). Final screening resulted in ATMY3 strain: EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::CUA tyrU::Inact* λ-RED::tolC.

3.4.7 Genomic recombination to build ATMY4

The recombinations to build ATMY4 are identical to ATMY3, but only differ in the *tyrU* genomic location. Instead of replacing the *tyrU* locus with a $tRNA^{Tyr}_{Inact}$, a $tRNA^{Tyr}_{CUA}$ cassette was used for constructing ATMY4, resulting in two copies of $tRNA^{Tyr}_{CUA}$ in its genome. To replace the *tolC* cassette at the *tyrU* locus, the PCR cassette tyrU::CUA was amplified with primers tyrU.lh-F and tyrU.rh-R from tyrU.CUA-gblock. Subsequent λ-RED

recombinations were performed as described with ATMY3, resulting in a final ATMY4 strain with genotype: EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::CUA tyrU*::CUA λ-RED::*tolC*.

3.4.8 Genomic recombination to build ATMY5

We also built a strain where both *tyrU* and *tyrTV* loci were replaced with $tRNA^{Tyr}_{Inact}$, thus creating a strain with no amber suppressor tRNAs. This strain was generated in the same order as ATMY3 and ATMY4, but used inactivated tRNA gblocks for tolC deletion at *tyrU* and *tyrTV* loci. From EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::tolC*, *tolC* was removed using tyrTV::Inact PCR cassette, containing one copy of the *E.coli* $tRNA^{Tyr}_{Inact}$, amplified from tyrTV.inact-gblock.orig with primers tyrTV.lhm-F and tyrTV.rhm-R, resulting in strain EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::Inact*. Next *tyrU* was replaced with *tolC* to build EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::Inact tyrU::tolC*, and the *tolC* was subsequently removed using tyrU::Inact PCR cassette (containing one copy of the *E. coli* $tRNA^{Tyr}_{Inact}$) from tyrU.inact-gblock with primers tyrU.lh-F and tyrU.rh-R, resulting in strain EcNR1GT pUltraBR MjY pUltraBR MjY ⊿*tyrS tyrTV::Inact tyrU::tolC*. The λ-RED recombination was performed as described with ATMY3, resulting in a final ATMY5 strain containing two genomically expressed inactviated tyrosyl tRNAs with genotype: EcNR1GT pUltraBR MjY ⊿*tyrS tyrTV::Inact tyrU::inact* λ-RED::*tolC*.

3.4.9 Growth comparison of ATMY strains

5 mL starter cultures of EcNR1Z, ATMY1-5 strains were grown to saturation overnight (16 hrs) in LB with all strain-dependent antibiotics. The starter cultures were diluted to an initial OD_{600} of 0.02 in three identical cultures of 50 mL LB with Spectinomycin and

grown in 250 mL sterile Erlenmeyer flasks at 37°C, with shaking (250 rpm). A 10 mm cuvette was used to monitor OD_{600} every 30 minutes.

3.4.10 Assessment of aaRS/tRNA activity using the CAT-reporter

The reporter plasmid used for chloramphenicol acetyltransferase (CAT) assays varied based on the strain used (wild-type or ATMY1-5). Refer to Supplementary sequences for plasmid sequences. *Strains Containing No Genomically Incorporated TAG-suppressor tRNA (ATMY1, ATMY2, ATMY5):* These strains were transformed with any of the following variants of pRep: pRepTrip2.3-EcY-TAG (expresses $tRNA^{Tyr}_{CUA}$) or pRepTrip2.3-TAG (does not express $tRNA^{Tyr}_{CUA}$). To co-express the EcTyrRS, a pBK plasmid encoding the appropriate EcTyrRS mutant was also co-transformed into these strains. As a control for no-aaRS co-expression, empty pBK plasmids were used. Cells containing intended plasmids were grown to saturation overnight in LB media supplemented with spectinomycin, tetracyclin and kanamycin (if pBK is present), diluted to an OD_{600} of 0.1, and 3 mL was spot-plated on LB-agar plates supplemented with kanamycin (+pBK plates), spectinomycin, tetracycline, ampicillin, and varying concentrations of chloramphenicol. Growth was recorded after 24 or 48 hrs of incubation at 37°C.

Strains Containing Genomically Incorporated TAG-suppressor tRNA (ATMY3 and ATMY4): The strains were transformed with any of the following variants of pRep: pRepTrip2.3-TAG (does not express $tRNA^{Tyr}_{CUA}$) or pRepTrip2.3P-EcQtR-2x TAG (does not express $tRNA^{Tyr}_{CUA}$, but contains two copies of the *E. coli* $tRNA^{Gln}$ expressed from the proK promoter). Strains containing pRep variants were further co-transformed with pBK-EcTyrRS variants (if EcTyrRS co-expression is desired). The chloramphenicol resistance of these strains were assayed as described above.

3.4.11 Assessment of tRNA/aaRS activity using the sfGFP-151TAG reporter

Strains Containing No Genomically Incorporated TAG-suppressor tRNA (ATMY1): ATMY1 harboring pEvolT5-EcY-sfGFP151TAG or pEvolT5-sfGFP151TAG (to assess background nonsense suppression in the absence of the suppressor tRNA) with or without pBK-EcTyrRSwt, pBK-EcOMeYRS, pBK-Empty, or newly evolved pBK-EcTyrRS variants were grown to saturation overnight in LB supplemented with spectinomycin, kanamycin and chloramphenicol. The starter cultures were diluted 100x in LB supplemented with required antibiotics. Cultures were grown at 37°C until 0.55 OD_{600} and induced with a final concentration of 1 mM IPTG. Unnatural amino acids (UAA) were also added during induction to a final concentration of 1 mM. Cultures were grown for an additional 16 hours at 37°C. For fluorescence measurements, cells were pelleted at 5000 x g, resuspended in PBS, diluted 10-fold, and transferred to a 96-well clear-bottom plate. Fluorescence was measured by using a SpectraMAX M5 (Molecular Devices) (Ex. 488 nm; Em. 534 nm) and normalized by OD_{600}. A culture of ATMY1 with no sfGFP reporter was also grown similarly and used to assess the background fluorescence, which was subtracted from all measurements.

Strains Containing Genomically Incorporated TAG-suppressor tRNA: For ATMY4, assessment of the tRNA/aaRS activity was accomplished with the same methods as above except strains harbored pEvolT5-sfGFP151TAG instead of its counterpart that encodes the suppressor tRNA.

3.4.12 Protein purification

Protein expression was performed using the same strain, plasmid combinations, and methods as described above (Assessment of tRNA/aaRS activity using a sfGFP151 reporter). Overnight expression cultures (10 mL) were centrifuged and resuspended in lysis buffer: B-

PER Bacterial Protein Extraction Reagent (Thermo Scientific) + 1X Halt Protease Inhibitor Cocktail (Thermo Scientific) + 0.01% Pierce Universal Nuclease (Thermo Scientific). After 30 min incubation at room temperature, the lysate was clarified by centrifuging at 22,000 xg for 5 min. The soluble C-terminally polyhistidine-tagged sfGFP was purified from the supernatant using HisPur Ni-NTA resin (Thermo Scientific) following manufacturer's protocol. Protein purity was confirmed by SDS-PAGE and protein molecular weight was confirmed by ESI-MS (Agilent Technologies 1260 Inifinity ESI-TOF).

3.4.13 Construction of the EcTyrRS library

To construct the library, four residues were randomized in tyrosyl tRNA-synthetase (pBK-EcYRS1a): Y37-FLIMVSTAYHCG, D182-NST, F183-NNK, L186-NNK. The EcYRS1a library was built via sequential overlap extension of three gel-purified PCR products. N-terminal piece A was amplified from the endogenous EcYRS using primers pBK seqT-F and EcYRS-35oR. Piece B was amplified with primers EcYRS-Y37mut-F and EcYRS-181oR, which was subsequently overlap extended to join piece AB using the terminal primers: pBK seqT-F and EcYRS-181oR. Lastly, C-terminal piece C was amplified using EcYRS-D182.F193.L186- NST.NNK.NNK-F and pBK MCS JIsqR. Piece AB was joined with piece C via overlap extension with terminal primers pBK MCS JIsqF and pBK MCS JIsqR. These inserts were digested with NdeI/NcoI (NEB) and ligated by T4 DNA Ligase (NEB) into the pBK vector, cut with the same restriction enzymes. The ligation mixture was ethanol precipitated with Yeast-tRNA (Ambion) and transformed into electrocompetent DH10b cells. The library was covered using $>10^7$ transformants.

3.4.14 OMeY and pBoF synthetase selection in ATMY3

ATMY3 was co-transformed with the pBK-EcYRS1a library and the positive selection reporter plasmid pRepTrip2.3P-EcQtR-2x. The reporter plasmid harbors two copies of proK-promoted *E. coli* $tRNA^{Gln}$ in its genomic context, a CAT reporter containing TAG codon (Q98TAG), an ampicillin reporter containing one TAG codon (3TAG) an arabinose-inducible T7 RNA polymerase containing two TAG nonsense codons (at positions 8 and 114), and a wild-type GFPuv reporter expressed from the t7 promoter. Suppression of CAT leads to chloramphenicol resistance, and suppression of T7 RNA polymerase drives expression of a t7-promoted GFPuv. 9.2x107 cfu (colony forming units) were plated on LB + 0.5x each of spectinomycin, tetracyclin, and kanamycin + 0.02% arabinose + chloramphenicol (30, 50 μg/mL) + 60 μg/mL Amp in the presence of 1 μM OMeY or pBoF for 36 hrs at 37°C.

Colonies from the 30 and 50 mg/mL chloramphenicol positive selection plates were harvested, and the pBK plasmids harboring mutant EcTyrRS were isolated. These were co-transformed into ATMY1 harboring the negative selection plasmid pNeg2-2xQtR, which contains an arabinose induced barnase gene with two stop TAG codons (3 and 45). 3x107 cfu were plated on LB + ampicillin + 0.5x kanamycin and incubated for 12 hrs at 37°C. Cells were harvested and library pBK plasmid was purified.

Isolated pBK plasmids from the negative selection were transformed back into ATMY3 pRepTrip2.3P-EcQtR-2x, and 10^6 cfu were plated on LB + 0.5x each of spectinomycin, tetracyclin, and kanamycin + 0.02% arabinose + chloramphenicol (30, 50 mg/mL) + 60 mg/mL Amp in the presence or absence of 1 mM OMeY or pBoF for 18 hrs at 37°C. 96 colonies from the +UAA plates were picked into 1 mL LB supplemented with Spec/Tet/Kan in deep 96-well polypropylene plates and grown overnight. Overnight cultures were diluted to 100x and 3 mL of each was spot plated on LB/Agar plates supplemented with

spectinomycin/tetracyclin/kanamycin, and chloramphenicol (30 or 50 μg/mL) in the presence or absence of the UAA. 10 clones showing the most prominent UAA-dependent growth were picked, sequenced, and spot-plated again to confirm the growth phenotype.

3.4.15 Assessment of aaRS/tRNA activity of HEK293T cells

HEK293T cells were maintained in Dulbecco's modified Eagle's medium (high glucose DMEM) supplemented with 10% FBS and penicillin-streptomycin (0.5x) in a humidified incubator at 37°C with 5% CO_2. $8.5x10^6$ cells were seeded one day before transfection in a 10 cm dish or distributed amongst a 24-well plate. Cells were transfected the following day at 70% confluency. For small-scale expression and imaging analysis, PEI (Sigma) was used to co-transfect pAcBac-EGFP39* and pAcBac-EcYRS (WT, OMeYRS, pBoFRS variants)-16xtRY-TAG into 24-well dishes. 250 ng each of the two plasmids were incubated in 18 μL DMEM (no FBS) with 3 μL of PEI (1 mg/mL) for 20 min, then added to the cells dropwise and mixed gently. UAAs were added to the culture medium to a final concentration of 1 mM at this time. Fluorescence images were taken at 48 hr post-transfection using a Zeiss Axio Observer fluorescence microscope. In order to obtain quantitative fluorescence measurements, cells were harvested by centrifugation at 5000 x g and lysed with CelLytic M lysis buffer containing 1x Halt protease inhibitor and 0.01% Pierce universal nuclease. 50 μL of this lysis buffer was used for cells from each well and incubated for 10 minutes. The resulting lysate was subsequently clarified and the supernatant was transferred to a clear bottom 96-well plate and fluorescence was recorded using a SpectraMAX M5 (Molecular Devices); Ex. 488 nm; Em. 534 nm.

To isolate EGFP39* from HEK293T cells in a larger scale expression, 5 μg each of pAcBac-EGFP-TAG (or pAcBac-GFP-2xTAG) and pAcBac-EcYRS (WT, OmeYRS,

pBoFRS variants)-16xtRY-TAG were incubated for 20 min in 180 µL DMEM (no FBS) with 40 µL PEImax (Polysciences; 1 mg/mL), and added evenly to 10 cm dishes seeded with $8.5x10^6$ cells one day prior to transfection. UAA was also added at this time at a concentration of 1 mM. Protein expression was allowed to continue for 48 hours.

After 48 hours, cells were harvested at 5000 x g and lysed with 600 µl CelLytic M lysis buffer (supplemented with 1x Halt protease inhibitor, 0.01% Pierce universal nuclease). Lysate was clarified and purified by Ni-NTA affinity chromatography. 0.5 mL Ni-NTA columns were equilibrated with equilibrium buffer (20 mM Na2HPO4, 300 mM NaCl (PBS), 10 mM imidazole pH 7.4). Clarified lysate was diluted with 2x volume of equilibrium buffer and loaded on the equilibrated Ni-NTA column. The column was washed with 20x column volumes of wash buffer (PBS, 25 mM imidazole pH 7.4). Protein was eluted in fractions with 3 CV elution buffer (PBS, 300 mM imidazole pH 7.4) and fluorescent fractions were pooled. Purified protein was then analyzed by SDS-PAGE and the molecular weight was confirmed with ESI-MS (Agilent Technologies 1260 Infinity ESI-TOF).

3.5 References

1. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550,** 53–60 (2017).
2. Dumas, a., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical reassignment - expanding the chemistry in biology. *Chem. Sci.* **6,** 50–69 (2014).
3. Italia, J. S. *et al.* Expanding the genetic code of mammalian cells. *Biochem. Soc. Trans.* **45,** 555–562 (2017).
4. Mukai, T., Lajoie, M. J., Englert, M. & Söll, D. Rewriting the Genetic Code. *Annu. Rev. Microbiol.* **71,** annurev-micro-090816-093247 (2017).
5. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13,** 854–870 (2018).
6. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20,** 1044–1048 (2002).
7. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science (80-.).* **292,** 498–500 (2001).
8. Chin, J. W. An Expanded Eukaryotic Genetic Code. *Science (80-.).* **301,** 964–967

(2003).
9. Chin, J. W., Cropp, T. A., Chu, S., Meggers, E. & Schultz, P. G. Progress Toward an Expanded Eukaryotic Genetic Code. *Chem. Biol.* **10,** 511–519 (2003).
10. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat Chem Biol* **13,** 446–450 (2017).
11. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 1059–1070 (2014).
12. Xiao, H. *et al.* Genetic incorporation of multiple unnatural amino acids into proteins in Mammalian cells. *Angew. Chem. Int. Ed. Engl.* **52,** 14080–14083 (2013).
13. Zheng, Y., Addy, P. S., Mukherjee, R. & Chatterjee, A. Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells. *Chem. Sci.* **8,** 7211–7217 (2017).
14. 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* **57,** 441–445 (2018).
15. Sakamoto, K. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* **30,** 4692–4699 (2002).
16. Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **124,** 9026–9027 (2002).
17. Seitchik, J. L. *et al.* Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. *J. Am. Chem. Soc.* **134,** 2898–2901 (2012).
18. Wang, J. *et al.* A biosynthetic route to photoclick chemistry on proteins. *J. Am. Chem. Soc.* **132,** 14812–14818 (2010).
19. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci.* **99,** 11020–11024 (2002).
20. Wang, J., Xie, J. & Schultz, P. G. A genetically encoded fluorescent amino acid SI. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 9785–9789 (2006).
21. Lee, H. S., Spraggon, G., Schultz, P. G. & Wang, F. Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. *J. Am. Chem. Soc.* **131,** 2481–2483 (2009).
22. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126,** 14306–14307 (2004).
23. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24,** 1436–1440 (2006).
24. Luo, X. *et al.* Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* 1–7 (2017). doi:10.1038/nchembio.2405
25. Englert, M. *et al.* A genomically modified Escherichia coli strain carrying an orthogonal E. coli histidyl-tRNA synthetase•tRNAHis pair. *Biochim. Biophys. Acta - Gen. Subj.* **1861,** 3009–3015 (2017).
26. Iraha, F. *et al.* Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion. *Nucleic Acids Res.* **38,** 3682–3691 (2010).
27. Kleina, L. G., Masson, J. M., Normanly, J., Abelson, J. & Miller, J. H. Construction of Escherichia coli amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J. Mol. Biol.* **213,** 705–717 (1990).

28. Normanly, J., Kleina, L. G., Masson, J. M., Abelson, J. & Miller, J. H. Construction of Escherichia coli amber suppressor tRNA genes. III. Determination of tRNA specificity. *J. Mol. Biol.* **213,** 719–726 (1990).
29. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry* **52,** 1828–1837 (2013).
30. Zheng, Y. *et al.* Molecular BioSystems Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1 †. *Mol. Biosyst.* (2016). doi:10.1039/C6MB00070C
31. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–8 (2009).
32. DeVito, J. A. Recombineering with tolC as a Selectable/Counter-selectable Marker: remodeling the rRNA Operons of Escherichia coli. *Nucleic Acids Res.* **36,** e4–e4 (2007).
33. Swanson, R. *et al.* Accuracy of *in vivo* aminoacylation reguires proper balance of tRNA and aminoacyl-tRNA synthetase. *Science (80-.).* **242,** 1548–1551 (1988).
34. Jahn, M., Rogers, M. J. & Söll, D. Anticodon and acceptor stem nucleotides in tRNA(Gln) are major recognition elements for E. coli glutaminyl-tRNA synthetase. *Nature* **352,** 258–60 (1991).
35. Chatterjee, A., Xiao, H., Bollong, M., Ai, H.-W. & Schultz, P. G. Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. *Proc. Natl. Acad. Sci.* **110,** 11803–11808 (2013).
36. Zheng, Y., Lewis, T. L., Igo, P., Polleux, F. & Chatterjee, A. Virus-enabled optimization and delivery of the genetic machinery for efficient unnatural amino acid mutagenesis in mammalian cells and tissues. *ACS Synth. Biol.* **6,** 13–18 (2017).
37. Takimoto, J. K., Adams, K. L., Xiang, Z. & Wang, L. Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol. Biosyst.* **5,** 931 (2009).
38. Brustad, E. *et al.* A genetically encoded boronate-containing amino acid. *Angew. Chemie - Int. Ed.* **47,** 8220–8223 (2008).
39. Liu, C. C. *et al.* Evolution of proteins with genetically encoded 'chemical warheads'. *J. Am. Chem. Soc.* **131,** 9616–9617 (2009).
40. Chen, Z. J., Ren, W., Wright, Q. E. & Ai, H. W. Genetically encoded fluorescent probe for the selective detection of peroxynitrite. *J. Am. Chem. Soc.* **135,** 14940–14943 (2013).
41. Chen, Z. *et al.* The N-B Interaction through a Water Bridge: Understanding the Chemoselectivity of a Fluorescent Protein Based Probe for Peroxynitrite. *J. Am. Chem. Soc.* jacs.6b01285 (2016). doi:10.1021/jacs.6b01285
42. Wang, F., Niu, W., Guo, J. & Schultz, P. G. Unnatural amino acid mutagenesis of fluorescent proteins. *Angew. Chemie - Int. Ed.* **51,** 10132–10135 (2012).
43. Yang, M., Li, J. & Chen, P. R. Transition metal-mediated bioorthogonal protein chemistry in living cells. *Chem. Soc. Rev.* **43,** 6511–6526 (2014).
44. Bandyopadhyay, A. & Gao, J. Iminoboronate-Based Peptide Cyclization That Responds to pH, Oxidation, and Small Molecule Modulators. *J. Am. Chem. Soc.* **138,** 2098–2101 (2016).
45. Cal, P. M. S. D. *et al.* Iminoboronates: A new strategy for reversible protein modification. *J. Am. Chem. Soc.* **134,** 10299–10305 (2012).
46. Javahishvili, T. *et al.* Role of tRNA orthogonality in an expanded genetic code. *ACS*

Chem. Biol. **9,** 874–879 (2014).
47. Amiram, M. *et al.* Evolution of translation components in recoded organisms enables multi-site nonstandard amino acid incorporation in proteins at high yield and purity. *Nat. Biotechnol.* (2015). doi:10.1038/nbt.3372

Chapter 4

Genetically encoded protein sulfation in eukaryotes

A significant portion of the work described in this chapter is currently in revision:

Italia, J.S., Peeler, J.C., Hillenbrand, C., Latour, C., Weerapana, E., Chatterjee, A. Genetically encoded protein sulfation in eukaryotes. *Manuscript in revision*.

We thank Prof. Dougald M. Monroe III (UNC) for helpful discussions on the kinetic assay for thrombin inhibition by HCII.

4.1 Introduction

4.1.1 Tyrosine sulfation

Sulfation of tyrosine residues is a post-translational modification (PTM) that occur exclusively in multicellular eukaryotes.[1–6] Golgi-resident tyrosylprotein sulfotransferases (TPST1 and TPST2) use 3'-phosphoadenosine-5'-phosphosulfate (PAPS) as the sulfate donor to install this PTM (Figure 4-1). Consequently, only secreted and membrane-associated proteins, which are processed through the trans-Golgi network, are subjected to this modification.[1–6] Tyrosine sulfation is believed to be irreversible and it facilitates numerous protein-protein and protein-ligand interactions that are important in diverse physiological processes such as immunity, hormone function, and blood-coagulation. Additionally, pathogens such as HIV rely on tyrosine-sulfated cell-surface receptors to gain access to human cells.[7] In turn, our immune system employs tyrosine-sulfated antibodies to target such pathogens,[8,9] attesting to the importance of this PTM in our biology. It is estimated that ~1% of all tyrosine residues in the eukaryotic proteome are sulfated, but the physiological roles for the most remain poorly understood.[3–5] A major factor contributing to this knowledge gap is the difficulty of expressing target proteins in a homogeneously sulfated state. Recombinant expression in common eukaryotic hosts often leads to incomplete sulfation of native sites.[9,10] Additionally, many proteins are sulfated at multiple tyrosine residues, and the difficulty of only modifying a chosen subset of these makes it challenging to evaluate the roles of individual sulfations.

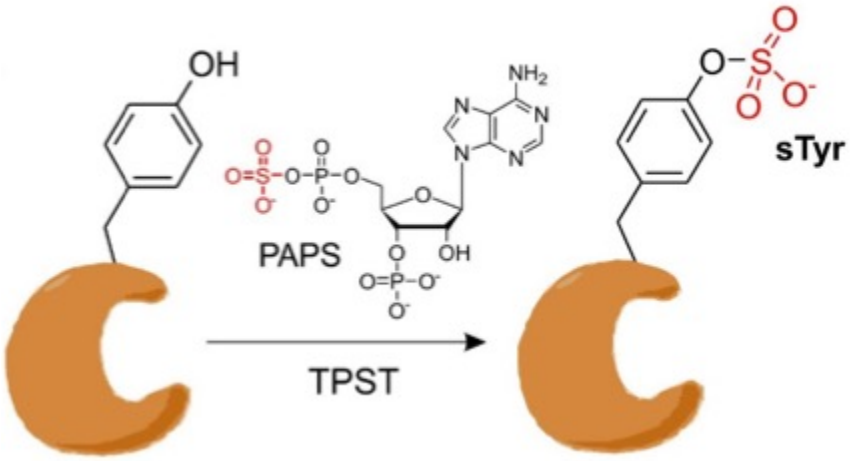


Figure 4-1. Tyrosylprotein sulfotransferases. Proteins processed through the trans-Golgi network in multicellular eukaryotes are subjected to tyrosine sulfation by TPST enzymes that use PAPS as a cofactor.

4.1.2 General strategy for genetically encoded sulfotyrosine

The genetic code expansion technology offers an elegant solution to these challenges by enabling co-translation site-specific incorporation of modified amino acid residues such as O-sulfotyrosine (sTyr) in response to a repurposed nonsense codon (Figure 4-2).[11–14] Indeed, the *M. jannaschii* tyrosyl-tRNA synthetase (MjTyrRS)/tRNA pair has been engineered to site-specifically incorporate sTyr into proteins expressed in *E. coli*, which has been useful for investigating the roles of tyrosine sulfation.[9,15–18] However, the MjTyrRS/tRNA pair is cross-reactive with its eukaryotic counterparts and cannot be used for non-canonical amino acid (ncAA) mutagenesis in eukaryotic cells.[19] This significantly limits the utility of this platform, given that tyrosine sulfation is only found in proteins from multicellular eukaryotes, and that the class of eukaryotic proteins that are subjected to sulfation (secreted and membrane-associated proteins) are frequently incompatible with recombinant expression in *E. coli*, as they require specialized processing through the ER-Golgi network. Furthermore, the ability to

express a eukaryotic protein in its native host is indispensable for investigating how its sulfation affects the cellular pathways it participates in (e.g., how sulfation of GPCRs affect their signaling). Genetically encoding sTyr in eukaryotic cells would overcome these limitations.

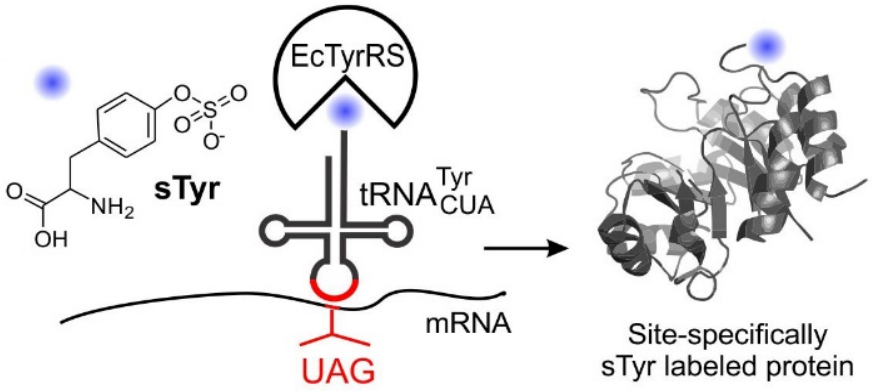


Figure 4-2. Co-translational incorporation of sulfotyrosine into proteins expressed in living cells in response to a nonsense codon using an engineered TyrRS/tRNA pair.

4.1.3 Heparin Cofactor II (HCII), a sulfated protein

Our platform should allow facile expression of native eukaryotic proteins homogeneously sulfated at native sites. In this work, we sought to demonstrate this using human heparin cofactor II (HCII) as a model system. HCII, a large secreted glycoprotein, is a serine protease inhibitor (serpin) that irreversibly inhibits thrombin, a key player in executing blood coagulation.[20,21] This anticoagulant activity of HCII is triggered by glycosaminoglycans (GAGs) such as heparin.[20–22] In the absence of GAGs, the acidic N-terminal domain (AND) of HCII binds its glycosoaminoglycan binding domain (GBD), resulting in an auto-inhibited state (Figure 4-3). GAGs activate HCII by binding its GBD and displacing the AND, which then recruits thrombin by binding its exosite 1 (Figure 4-3). The AND of HCII, which can bind both thrombin exosite I and GBD, is sulfated at two distinct sites (Tyr60 and Tyr73)

whose roles in HCII activity is poorly understood.[23] The absence of ER-Golgi processing precludes bacterial expression of HCII in its native glycosylated state, while overexpression in eukaryotic hosts can result in incomplete sulfation. We sought to demonstrate our ability to incorporate sulfotyrosine in GFP in bacteria and eukaryotes, and then move on to a more challenging protein, such as HCII.

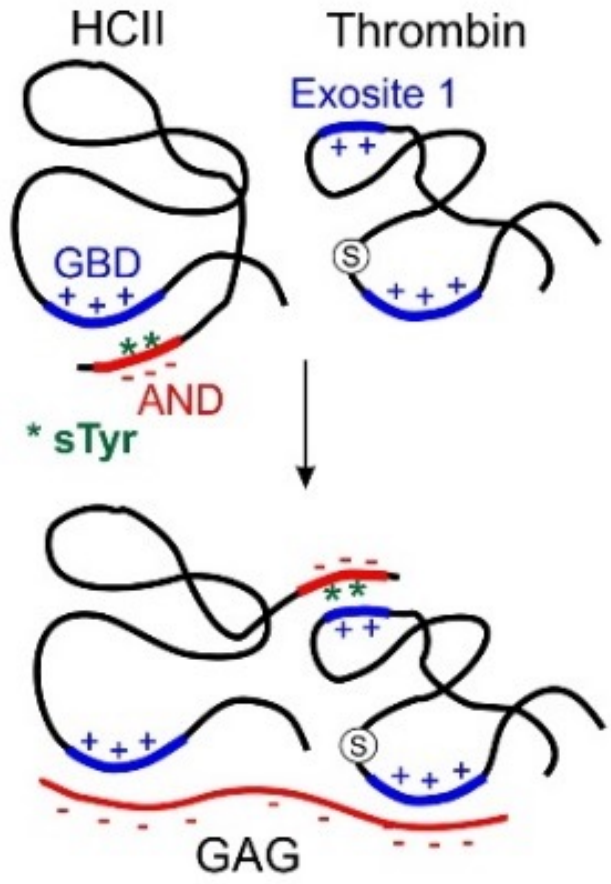
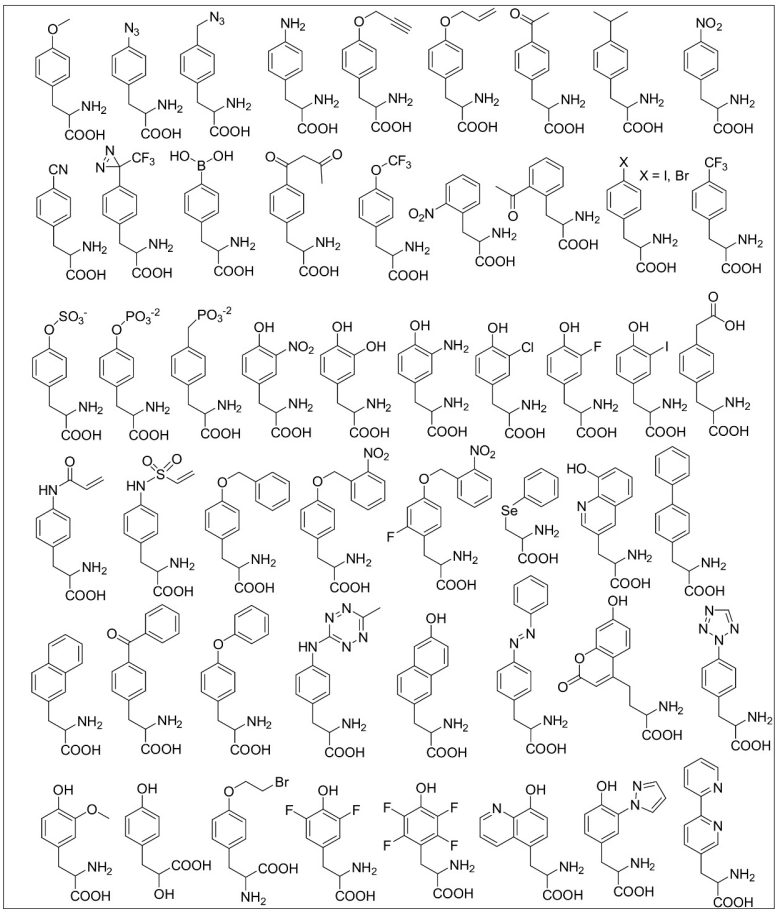


Figure 4-3. The model for GAG-activated thrombin inhibition of HCII, which is sulfated at Tyr60 and Tyr73 (shown as green stars).

M. jannaschii pair derived in *E. coli*



E. coli pair derived in yeast

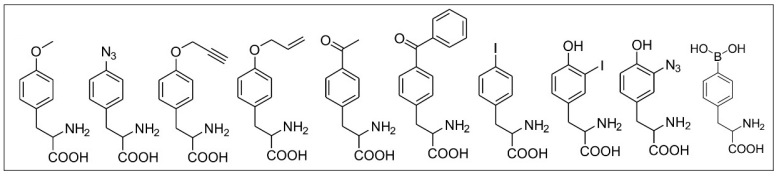


Figure 4-4. Comparison of ncAA diversity for tyrosyl pairs developed with the *E. coli* or yeast based selection platforms.

4.2 Results and Discussion

4.2.1 Functionally substituting the endogenous EcTyrRS/tRNA pair of E. coli

The *E. coli* derived tyrosyl-tRNA synthetase (EcTyrRS)/tRNA pair represents a promising platform to this end, as it has already been established for ncAA mutagenesis in eukaryotes.[19,24,25] Bacterially derived pairs are generally orthogonal in eukaryotes and can be used for genetic code expansion (Figure 4-5). However, the repertoire of ncAAs genetically encoded using this platform has been significantly limited relative to its *M. jannaschii* derived counterpart (Figure 4-4). [14,19] While the substrate specificity of MjTyrRS can be engineered using a facile *E. coli* based directed evolution system, the engineering of EcTyrRS relies on a cumbersome yeast-based system, which has experienced much less success.[12,14,19,26] Recently, we have established a novel approach to facilitate the directed evolution of *E. coli* derived aminoacyl-tRNA synthetase (aaRS)/tRNA pairs in *E. coli* (Figure 4-6).[12,19,26] First, one of the endogenous aaRS/tRNA pairs of *E. coli* is functionally substituted by an orthogonal counterpart from archaea/eukaryote. Next, the liberated endogenous pair is reintroduced in the resulting 'altered translational machinery' (ATM) *E. coli* strain as an orthogonal nonsense suppressor, where it can be engineered using the *E. coli* based directed evolution platform. We have used this strategy to create ATMY strains of *E. coli*, in which the endogenous EcTyrRS/tRNA pair is functionally replaced by an archaea-derived TyrRS/tRNA pair (Figure 4-6).[19] We have further demonstrated the feasibility of engineering the EcTyrRS/tRNA pair in such an ATMY *E. coli* strains to genetically encode ncAAs in both eukaryotes and ATMY *E. coli* strains. This platform provides an exciting opportunity to genetically encode sTyr in eukaryotic cells.

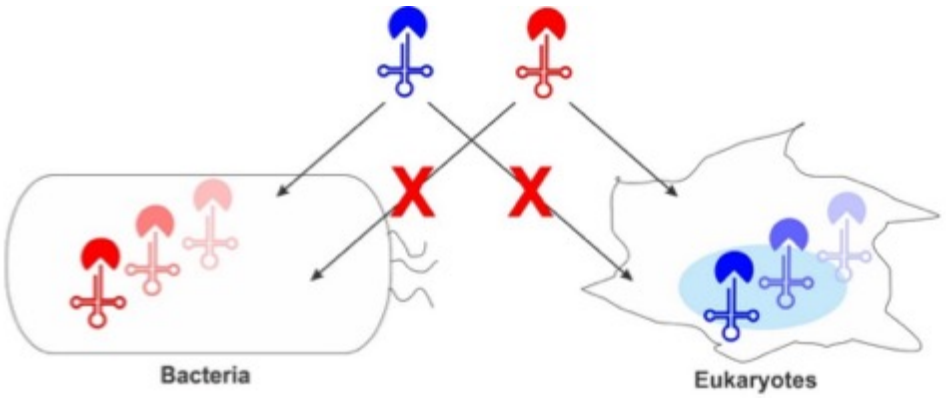


Figure 4-5. Typical orthogonality of bacterial vs eukaryotic pairs and their use in the GCE field. Typically, the bacteria-derived aaRS/tRNA pairs (color-coded red) are orthogonal in eukaryotes and can be used for eukaryotic genetic code expansion, while eukaryote or archaea derived pairs (color-coded blue) are orthogonal in bacteria and are useful for bacterial genetic code expansion.

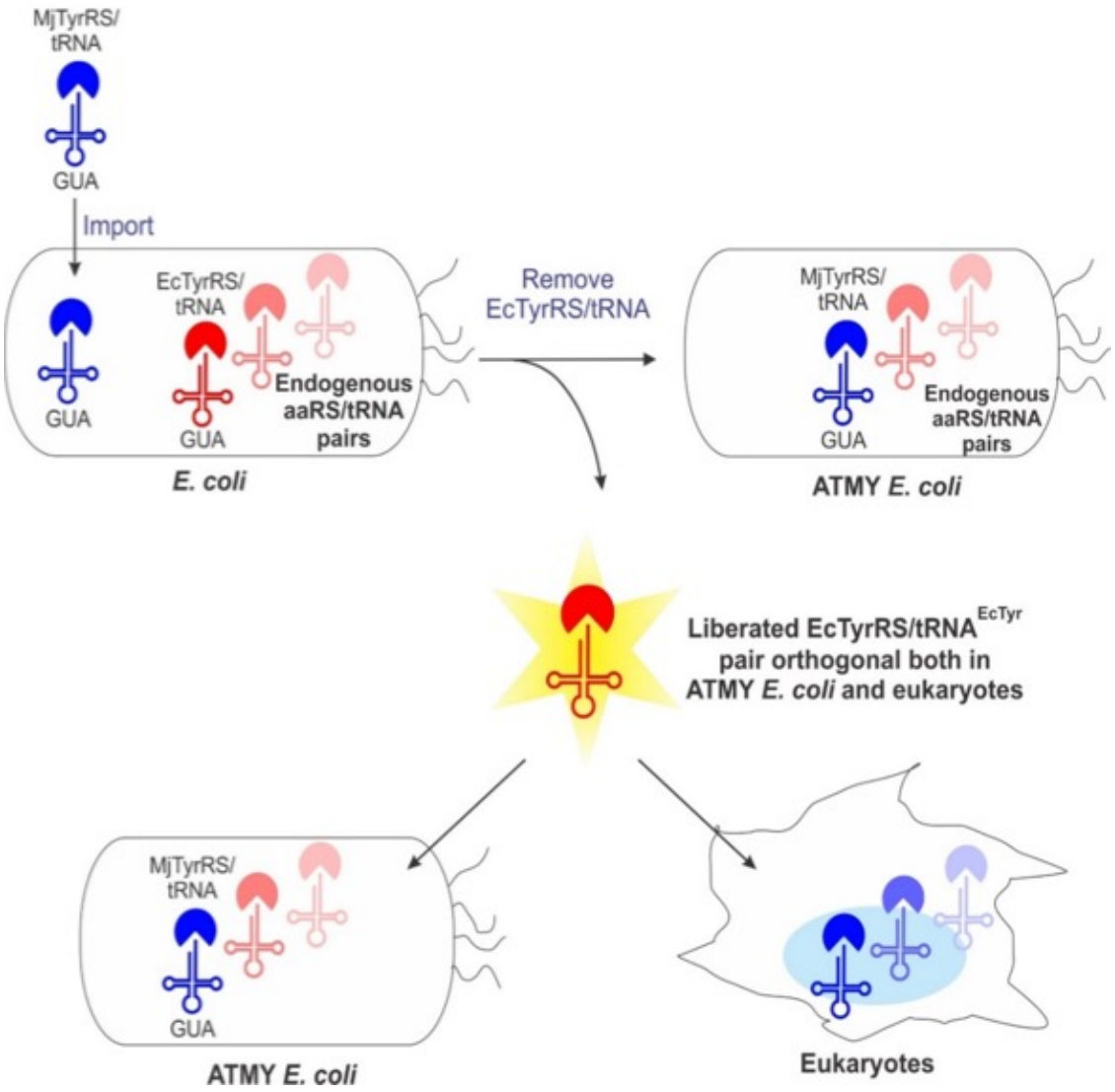


Figure 4-6. General scheme for the creation of ATMY strains. Functionally substituting the EcTyrRS/tRNA pair in E. coli with the archaea derived MjTyrRS/tRNA pair creates an engineered ATMY strain. The 'liberated' EcTyrRS/tRNA pair can be established as an orthogonal nonsense suppressor in ATMY *E. coli*, and engineered in this strain for altering its substrate specificity.

4.2.2 EcTyrRS library construction and directed evolution for the incorporation of sTyr

We started by constructing a library of EcTyrRS mutants encoded in the pBK vector (pBK-EcYRS1) by randomizing six active site residues (Y37 to FLIMVSTAYHCG, L71 to NBT, N126 to NSPTACGDH, D182 to NST, F183 to NNK, L186 to NNK) surrounding the phenolic hydroxyl of the bound tyrosine substrate (Figure 4-7). The pBK-EcYRS1 library was subjected to a previously developed double-sieve selection system in ATMY *E. coli*.[19] The positive selection enriches active aaRS mutants using a TAG-inactivated chloramphenicol acetyltransferase reporter, while the negative selection removes mutants that charge canonical amino acids using a TAG-inactivated toxic barnase gene. Just after a single round of positive and negative selection each in ATMY3 *E. coli*, the library demonstrated highly sTyr-dependent survival in the presence of chloramphenicol, indicating the enrichment of sTyr-selective EcTyrRS mutants (Figure 4-8a). Several individual clones from this selected pool of mutants also replicated the same sTyr-dependent phenotype (Figure 4-8b). DNA sequencing of eight such clones revealed the presence of two distinct but highly convergent clones, where Y37, N126, and F183 are conserved, L71 and D182 are mutated to V and G, respectively, and L186 is either conserved (VGL) or is mutated to M (VGM)(Figure 4-7). The enlarged active site of these mutants were consistent with the need to accommodate the additional sulfate group of sTyr.

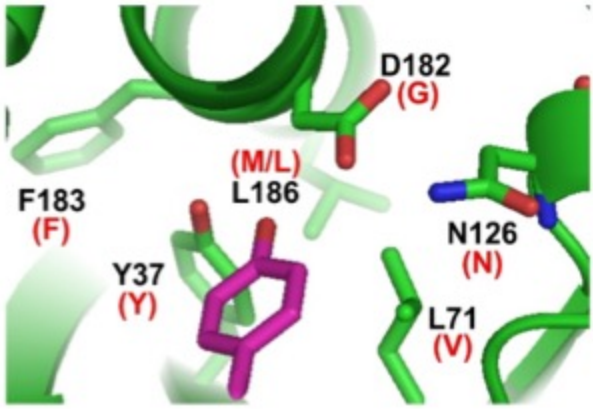


Figure 4-7. EcTyrRS active site and with EcYRS1 library mutation hits annotated in red.

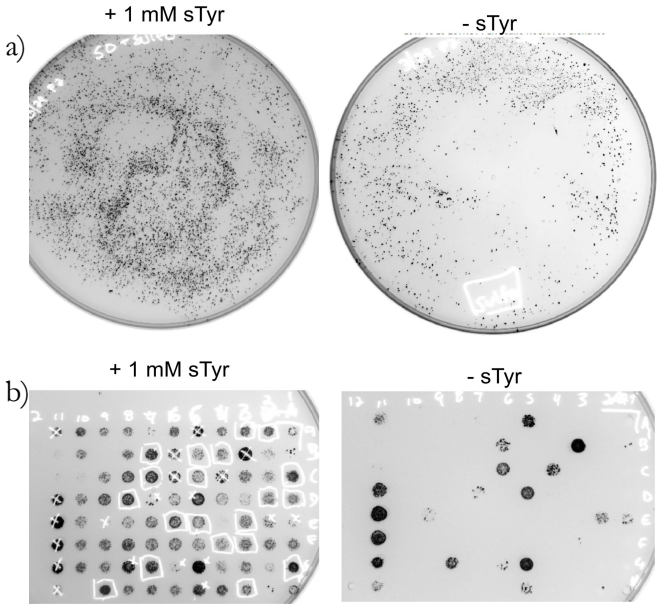


Figure 4-8. Selection of SulfoRS mutants with EcYRS1 library. a) Positive selection plates (second round) containing library mutants previously subjected to one positive and one negative round of selection. b) Individual colonies spot plated from P2N1 (two positive, one negative) selection plates.

4.2.3 Evaluation of sulfotyrosine incorporation efficiency into sfGFP in ATMY4 E. coli

To evaluate the sulfotyrosine incorporation efficiency of these two EcTyrRS mutants (VGL and VGM), we individually co-transformed them (encoded in the pBK plasmid) with a pEvolT5-sfGFP-151-TAG reporter plasmid in ATMY4 *E. coli* strain (encodes two genomic copies of $tRNA^{EcTyr}_{CUA}$).[19] These cells expressed sfGFP only in the presence of sTyr upon induction with IPTG (Figure 4-9a). Purification of this reporter protein using a C-terminal polyhistidine tag (8-10 mg/L) followed by ESI-MS analysis showed a mass consistent with the incorporation of sTyr (Figure 4-9a). Western-blot analysis using an anti-sTyr monoclonal antibody further corroborated the presence of sTyr in this protein (Figure 4-9b). These observations confirm that we have generated an engineered EcTyrRS/tRNA pair that selectively incorporates sTyr in response to UAG.

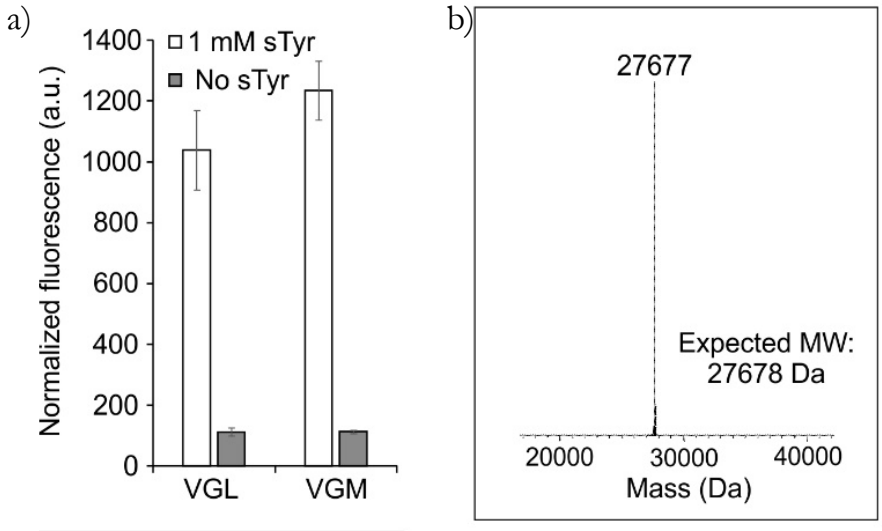


Figure 4-9. Characterization of EcTyrRS mutants in *E. coli*. a) Two EcTyrRS mutants facilitate sfGFP-151-TAG reporter expression in ATMY4 *E. coli* in the presence of sTyr (fluorescence in resuspended cells). b) ESI-MS analysis of the purified sfGFP-151-sTyr show expected mass.

4.2.4 Evaluation of sulfotyrosine incorporation efficiency into EGFP in HEK293T

Next, we explored if these mutant EcTyrRS/tRNA pairs can be used in mammalian cells for co-translational sTyr incorporation. We created a mammalian expression plasmid pB1U-Sulfo-16xtYR-TAG that expresses the VGL or the VGM EcTyrRS as well as 16 copies of the $tRNA^{EcTyr}_{CUA}$. Co-transfection of this plasmids with pAcBac1-EGFP-39-TAG led to robust expression of EGFP in the presence of sTyr, while significantly reduced reporter expression was observed in its absence (Figure 4-10). The VGM mutant exhibited lower levels of UAG suppression in the absence of sTyr (Figure 4-10). See Chapter 6 for results of the background of the VGL mutant (tyrosine is charged). The reporter protein expressed in the presence of sTyr was isolated from HEK293T cells (100-120 μg from ~10^7 cells) using a C-terminal polyhistidine tag and analyzed by ESI-MS, which showed a mass consistent with sTyr incorporation (Figure 4-10c). Western-blot using an anti-sTyr antibody further confirmed the presence of sTyr in EGFP-39-sTyr, but not in wild-type EGFP (Figure 4-11) that was expressed and purified in a similar manner.

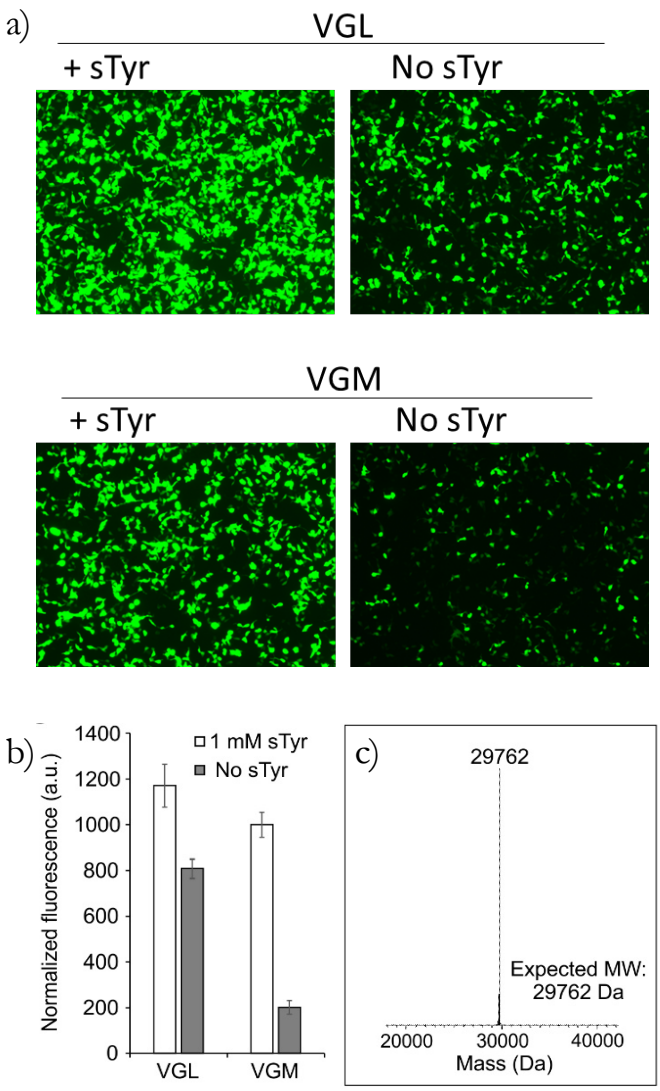


Figure 4-10. Characterization of EcTyrRS mutants in HEK293T. a) Fluorescence images of HEK293T cells expressing EGFP-39-TAG reporter using VGL- or VGM-EcTyrRS mutant in the presence or absence of sTyr (1 mM). b) EGFP-39-TAG expression via fluorescence in cell free extract. c) ESI-MS analysis of purified EGFP-39-sTyr.

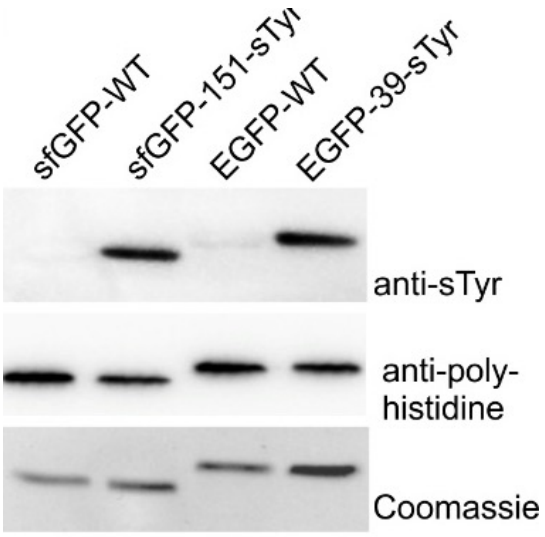


Figure 4-11. Western blot analysis of purified sfGFPwt/151-sTyr or EGFPwt/39-sTyr with either anti-sTyr or anti-polyhistidine antibodies.

4.2.5 Expression and purification of recombinant Heparin Cofactor II (HCII) with varying levels of sulfation

We introduced UAG codons at 60 and 73 positions of full-length human HCII and overexpressed it in HEK293T cells in the presence of our sTyr incorporation system. Full-length HCII was successfully isolated from the culture medium using a C-terminal polyhistidine tag (Figure 4-12a). Whole-protein ESI-MS of this large protein was challenging, but we confirmed the presence of sTyr at both sites through protease digestion followed by LC-MS analysis (Figure 4-13, 4-14). Glycosylase (PNGase) treatment significantly reduced the molecular weight of the protein (Figure 4-12b), suggesting the presence of robust N-linked glycosylation. These results confirm that our platform can be used to express endogenous eukaryotic proteins precisely sulfated at multiple sites.

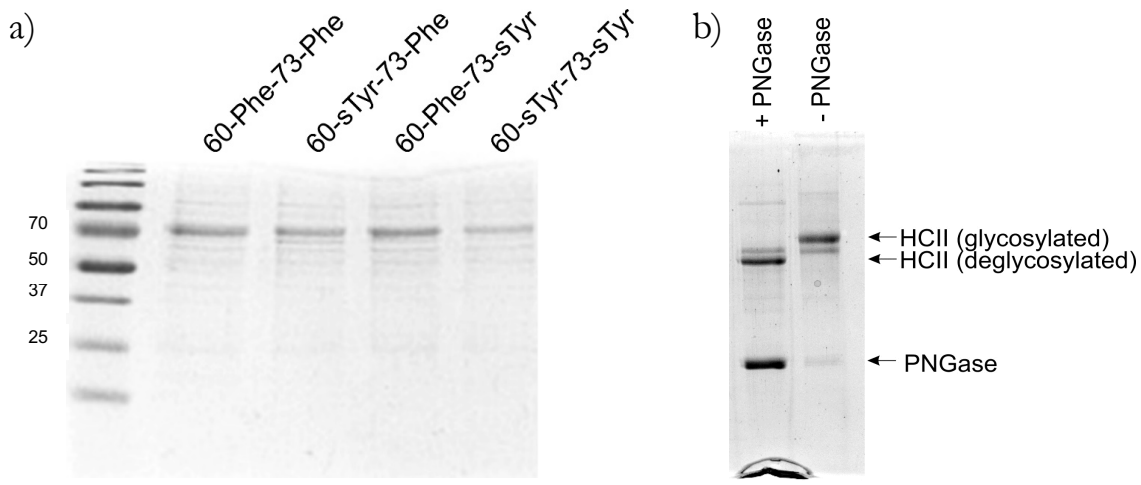


Figure 4-12. SDS-PAGE analysis of secreted HCII mutants expressed in HEK293T. a) Due to glycosylation, the observed molecular weight is higher than the primary sequence prediction (~57 kDa). b) PNGase F treatment (cleaves N-linked glycosylation) of purified HCII reduces the molecular weight.

Peptide harboring 60-sTyr
ENTVTNDWIPEGEEDDDY*LDLEK
Expected m/z (+2): 1410.07

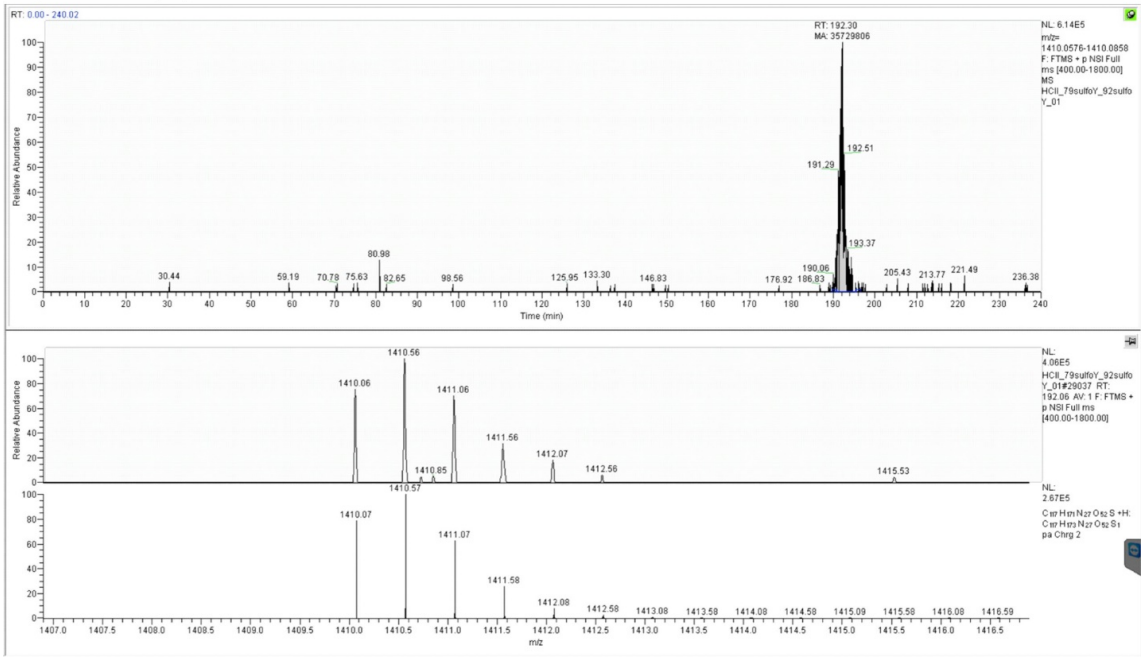


Figure 4-13. Trypsin digestion followed by LC-MS analysis of HCII-60-sTyr-73-sTyr isolated from HEK293T cells identifies the presence of the peptide harboring 60-sTyr. HCII-60-Tyr-73-Tyr was 89% sulfated at this 60 site, while the above double mutant was 94% sulfated (MS not shown). The predicted and observed fragmentation patterns match (bottom two panels). Note: While resolution in this thesis is poor, one can look into the published paper to find the exact mass spec data.

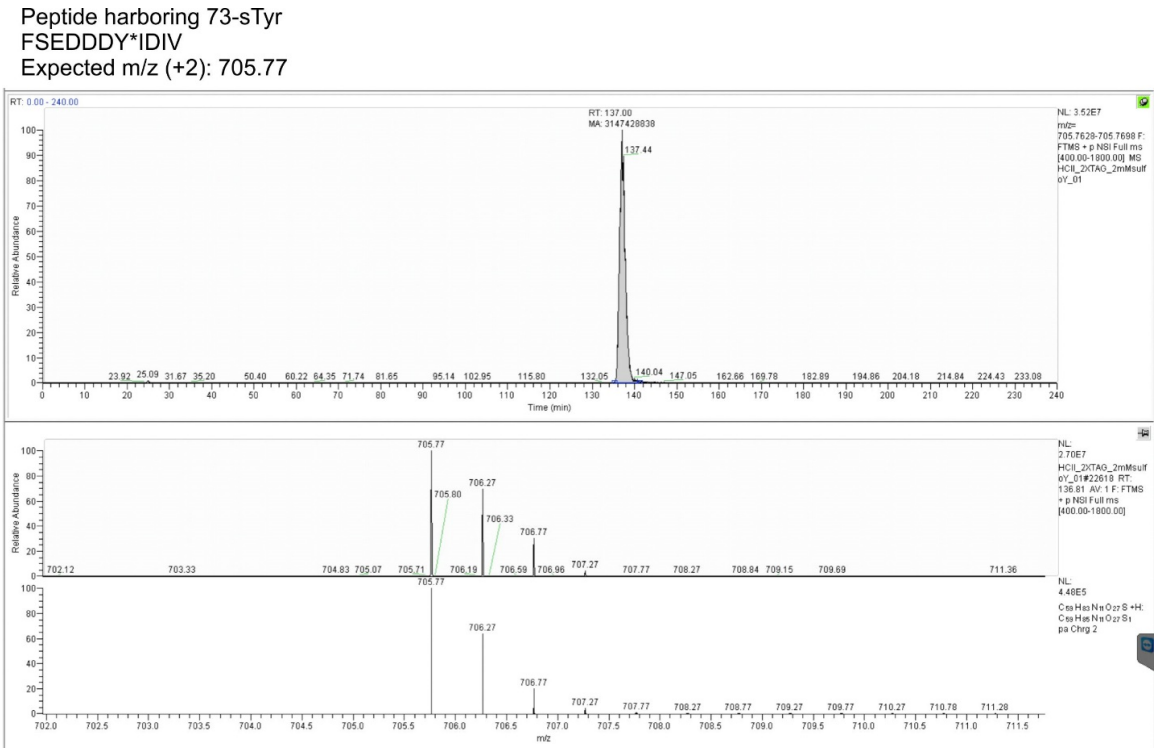


Figure 4-14. Trypsin + elastase double digestion followed by LC-MS analysis of HCII-60-sTyr-73-sTyr isolated from HEK293T cells identifies the presence of the peptide harboring 73-sTyr. The predicted and observed fragmentation patterns match (bottom two panels). We were unable to find the HCII fragment harboring the 73 residue through trypsin digestion alone, likely due to its large predicted size. HCII-60-Tyr-73-Tyr was 95% sulfated at this 73 site, while the above double mutant was >95% sulfated (MS not shown).

4.2.6 Exploration of the effect of sulfation on HCII's natural thrombin inhibition

In addition to 60-sTyr-73-sTyr, we also expressed and purified HCII mutants 60-Phe-73-Phe (to prevent sulfation), 60-sTyr-73-Phe, and 60-Phe-73-sTyr (Figure 4-12a) and evaluated their thrombin inhibition activities using an established biochemical assay.[27–29] For each HCII mutant, second-order rate constants (k_2) of thrombin inhibition were measured at different heparin concentrations to find the optimal [heparin], at which maximal inhibition rate is observed (Figure 4-15). 60-sTyr-73-sTyr exhibited a maximal rate constant of 3×10^8 M^-

$^{1}min^{-1}$ at ~20 μg/mL heparin, which is in close agreement with previously reported data.[29] Interestingly, the absence of sTyr at site 73 (60-sTyr-73-Phe) led to a slightly lower maximal k_2 but a substantially reduced (~3 fold) optimal [heparin], whereas the 60-Phe-73-sTyr mutant (no sTyr at site 60) had an unchanged optimal [heparin] but a significantly lower maximal k_2 (Figure 4-15). The 60-Phe-73-Phe mutant showed both a low maximal k_2, and a reduced optimal [heparin]. The preliminary biochemical evaluation of precisely sulfated HCII mutants suggests important – yet distinct – roles the two sulfation PTMs play in fine-tuning its GAG-triggered thrombin inhibition activity: while the 73-sTyr appears to contribute more to AND-GBD association, the 60-sTyr might be more important for thrombin recruitment.

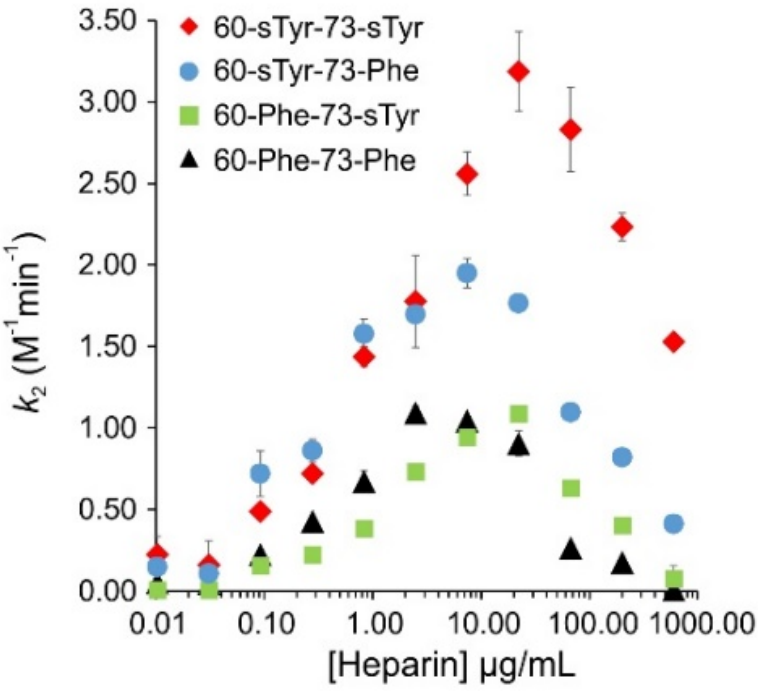


Figure 4-15. Second-order rate constant of thrombin inhibition by different HCII mutants at various heparin concentrations. Note: The 60-sTyr-73-sTyr has a comparable curve to the 60-Tyr-73-Tyr (due to almost equivalent sulfation but the data is not shown in this graph.

4.3 Conclusions

In summary, we have developed a platform for site-specific incorporation of sTyr into proteins expressed in eukaryotic cells with high fidelity and efficiency, which would be a valuable tool for investigating the consequences of tyrosine sulfations found in the eukaryotic proteome. This platform can also be used to express therapeutically relevant proteins homogeneously modified with functionally important sulfations.[9,10,30] Additionally, the ability to incorporate sTyr into virtually any site of any protein in eukaryotic cells offers intriguing opportunities for novel synthetic biology applications.

4.4 Experimental procedures

4.4.1 General biological reagents, strains, and protocols

E. coli strain DH10B (Life Technologies) was used for plasmid propagation and cloning. *E. coli* strains were cultured on LB-agar plates with appropriate antibiotic concentrations as follows: 95 μg/mL spectinomycin, 50 μg/mL chloramphenicol, 30 μg/mL kanamycin. Phusion high fidelity DNA polymerase (Thermo-Fischer) was used for PCR amplifications and restriction enzymes were obtained from New England Biolabs. DNA oligonucleotides were purchased from Integrated DNA Technologies, while Sanger sequencing was performed by Eton Bio. Engineered *E. coli* strain ATMY3 (contains one genomic copy of $tRNA^{EcTyr}_{CUA}$; no genomic EcTyrRS) was used as the selection host for the directed evolution of EcTyrRS.[1] Engineered *E. coli* strain ATMY4 (contains two genomic copies of $tRNA^{EcTyr}_{CUA}$; no genomic EcTyrRS) was used as the expression host for expressing recombinant proteins incorporating sTyr.

HEK293T cell line was purchased from ATCC (ATCC CRL-3216) and maintained in DMEM (high glucose) supplemented with 10% FBS and Penicillin/Streptomycin. Cells were grown in a 37 °C 100% humidity, 5% CO_2.

4.4.2 EcTyrRS library construction

In the EcYRS1 library (pBK-EcYRS1), six residues were randomized as follows: Y37-FLIMVSTAYHCG, L71-NBT, N126-NSPTACGDH, D182-NST, F183-NNK, L186-NNK. We used a previously reported library (pBK-EcYRS1a), which contains the desired Y37, D182, F183, and L186 randomizations, as the template to generate pBK-EcYRS1 by sequential overlap of extension PCR. Piece A was amplified with primers pBK seqT-F and EcYRS-L71-oR. Piece B was amplified with EcYRS-L71-NBT-F and EcYRS-N126-oR, and subsequently overlapped with piece A using terminal primers pBK seqT-F and EcYRS-N126oR to create piece AB. Lastly, piece C was amplified with EcYRS-N126x-F (x corresponds to nine different codons) and pBK MCS JIsqR for all desired N126 variants. Piece C variants were combined in equal distribution and were subsequently overlapped with piece AB to form the full length aaRS PCR product.

After amplification, the aaRS PCR product was digested with NdeI/NcoI (NEB) and ligated by T4 DNA Ligase (NEB) into the pBK vector digested with the same restriction enzymes. The ligation mixture was ethanol precipitated with Yeast-tRNA (Ambion) and transformed into electrocompetent DH10B cells. Greater than 10^8 transformants were obtained to ensure library coverage.

4.4.3 Directed evolution of EcTyrRS-SulfoY variant in ATMY3

Positive selection 1: The pBK-EcYRS1 library was transformed into ATMY3 containing the positive selection plasmid pRepTrip2.3P-EcQtR-2x. The pRep plasmid expresses a

chloramphenicol acetyl transferase (CAT) reporter containing a Q98TAG, an ampicillin resistance gene containing a 3TAG, an arabinose inducible T7 RNA polyermase containing two TAG codons (site 8 and 14), a T7 promoted GFPuv, and two copies of the *E. coli* $tRNA^{Gln}$ expressed from its endogenous promoter. Approximately $9x10^8$ colony forming units were plated on LB + 0.5x spectinomycin, tetracyclin, and kanamycin + 0.02% arabinose + 30 μg/mL ampicillin + 30 or 50 μg/mL chloramphenicol in the presence of 1 mM sTyr for 18 h at 37 °C. After 18 h, colonies from plates were harvested with 15 mL LB, centrifuged and selected pBK plasmid pool (pBK-EcYRS1a-P1) was purified via miniprep and isolated via gel purifcation.

Negative selection: The isolated plasmid was subsequently transformed into ATMY3 containing pNeg2-2xQtR (contains arabinose dependent barnase with 3TAG, 45TAG, and two copies of the *E. coli* $tRNA^{Gln}$). Approximately 10^8 cells were plated on LB-agar plates containing 0.5x spectinomycin, ampicillin, and kanamycin + 0.02% arabinose in the absence of sTyr for 12 h at 37 °C. After 12 h, colonies from plates were harvested with 15 mL LB, centrifuged and the pBK library subjected to one positive selection and one negative selection (pBK-EcYRS1a-P1N1) was purified via miniprep.

Positive selection 2: pBK-EcYRS1a-P2N1 was subjected to a second round of positive selection (10^6 cfu plated). 96 single colonies from the second round of positive selection plates containing 1 mM sTyr were picked into 500 μL LB supplemented with spectinomycin, tetracyclin, kanamycin in a 96 deep-well plate and grown to confluence overnight. These overnight cultures were diluted 100 fold and 3 μL were individually spot plated on LB-Agar plates containing spectinomycin, tetracyclin, kanamycin + 0.02% arabinose and 30 or 50 μg/mL chloramphenicol in the presence or absence of 1 mM sTyr. Eight pBK variants showing the most sTyr-dependent survival were picked for further characterization.

4.4.4 Characterization of tRNA/aaRS activity in E. coli via sfGFP reporter

ATMY4 containing pEvolT5-sfGFP151TAG was transformed with pBK-EcTyrRS variants. Overnight starter cultures were diluted 100 fold in 10 mL LB containing required antibiotics and grown at 37 °C while shaking at 250 rpm in 50 mL flasks. Upon reaching 0.55 OD_{600}, 1 mM final IPTG was added to induce protein expression. 1 mL aliquots of induced cultures were placed in 15 mL culture tubes with or without 1 mM sulfotyrosine and grown for 18-20 h at 30 °C. Afterwards, cells were pelleted, resuspended in PBS, and diluted 10 fold. Dilutions were transferred to a 96-well clear bottom plate. Expression of full-length sfGFP was measured using the associated characteristic fluorescence by a SpectraMAX M5 (Molecular Devices) multimode plate reader (Ex. 488 nm; Em. 534 nm; 515 cutoff) and normalized with respect to OD_{600}.

4.4.5 Purification of sfGFP-TAG from bacterial expression

Protein expression was performed in 10 mL culture as described above (sfGFP151-TAG reporter assay). Afterwards, the cells were pelleted at 5000 x g, resuspended in lysis buffer [B-PER Bacterial Protein Extraction Reagent (Thermo Scientific), 1x Halt Protease Inhibitor Cocktail (Thermo Scientific), 0.01% Pierce Universal Nuclease (Thermo Scientific), and incubated for 10 min on ice. After incubation, the crude lysate was clarified at 22,000 x g. The full-length sfGFP containing a C-terminal 6x HisTag was purified using HisPur Ni-NTA resin (Thermo Scientific) according to the manufacturers protocol. SDS-PAGE and Bradford analysis were used to assess protein purity, while the molecular weight was confirmed by ESI-MS (Agilent Technologies 1260 Infinity ESI-TOF).

4.4.6 Site-specific incorporation of sTyr into proteins expressed in mammalian cells

HEK293T cells were maintained as described above. pB1U-SulfoA1-16xtYR-TAG (VGL) or pB1U-SulfoB7-16xtYR-TAG (VGM) contain 16 copies of alternating U6/H1 promoted *E. coli* $tRNA^{Tyr}_{CUA}$ and UbiC promoted EcTyrRS mutants. pAcBac1-EGFP-39TAG was used as a reporter plasmid. $0.7x10^6$ cells per well were seeded one day prior to transfection in a 12 well plate. At 70% confluence, the transfection mixture (500 ng each of suppressor and reporter plasmid, 18 µL DMEM, 3.5 µL Sigma PEI (1 mg/mL), 10 min incubation prior to addition) was added to each well and gently mixed. A final concentration of 2 mM sTyr was added to the wells at the time of transfection. After 48 h, cells were harvested by centrifugation at 5000 x g and residual media was removed. 50 µL lysis buffer (10 mL CellLytic M, 1x Halt Protease inhibitor, 0.01% Pierce universal nuclease) was added per well and incubated for 10 min. After incubation, cells were clarified by centrifugation and lysate was analyzed for fluorescence in the SpectraMAX M5 (Molecular Devices) under the same conditions as sfGFP.

For purification and further charectrization, EGFP-39-sTyr was expressed in 10 cm dishes ($8.5x10^6$ seeded 24 h prior to transfection). 5 µg suppressor plasmid and 5 µg reporter plasmid were incubated with 180 µL DMEM (no FBS) and 40 µL PEImax (Polysciences; 1 mg/mL). 2 mM sTyr and 2 mM Sodium Butyrate was added at the time of transfection. After 48 hr, cells were harvested at 5000 x g. 600 µL lysis buffer (CellLytic M, 1x Halt protease inhibitor, 0.01% Pierce universal nuclease) was used to lyse the cells. After 10 min incubation, the lysate was clarified by centrifugation and the protein was purified using HisPur Ni-NTA resin (Thermo-Scientific). Purity and the molecular weight of the expressed protein was analyzed by SDS-PAGE and ESI-MS (Agilent Technologies 1260 Inifinity ESI-TOF).

4.4.7 Anti-His and anti-Sulfotyrosine western blot of GFP reporters

Western blot was used to confirm the presence of a polyhistidine tag (via anti-HisTag blot) and the presence of sulfotyrosine (via anti-sTyr blot) in reporter proteins expressed above. 500 ng each of purified wild-type or sTyr-incorporated mutant of sfGFP or EGFP reporter proteins were resolved by SDS-PAGE, and transferred to a PVDF membrane (Life Technologies) using a Trans-Blot Turbo Transfer Sytem 15 (BioRad) in Towbin Transfer Buffer (at 12V for 30min, twice). After complete transfer, membrane was blocked in 10 mL 5% milk in TBST (HisBlot) or 10 mL Pierce Superblocker (Fisher Scientific) at 4 °C overnight with constant agitation. Membranes were subsequently incubated in 1:3000 anti-HisTag mouse mAb (Invitrogen, MA121315, in 5% milk TBST) or 1:6000 anti-Sulfotyrosine mouse mAb (Millipore Sigma, Clone: 1CA2, in Pierce Superblocker) overnight. Next, the membrane was washed six times, 10 min per wash, using TBST at room temperature. Afterwards, 1:6000 dilution of chicken anti-mouse secondary antibody (Invitrogen, SA1-72021, in 5% milk TBST) was incubated for 2 h at room temperature. The membrane was washed and activated using SuperSignal West Dura Kit (Fisher Scientific). The activated blot was imaged on the ChemiDoc MP imaging system (BioRad).

4.4.8 Expression and purification of Heparin Cofactor II (HCII)

HEK293T cells were maintained as described above. pB3-SulfoRS-16xYtR-TAG-HCII contains the following three components: 16 copies of alternatingly H1/U6 promoted *E. coli* tyrosine $tRNA_{CUA}$, a UbiC promoted EcTyrRS mutant, and HCII mutants under a CAG promoter. 10 cm dishes were seeded with 8.5 x 10^6 cells 24 h prior to transfection. Afterwards, DMEM +FBS media was aspirated and replaced with DMEM without FBS. A transfection mixture (10 μg pB3 plasmid, 180 μL DMEM, 50 μL PEI Max) was incubated for 10 min prior

to addition. 2 mM sTyr and 2 mM sodium butyrate were added at the time of transfection. Since HCII is a secreted protein, the media was harvested on days 2 and 3 post transfection, stored at 4 °C for up to 2 days, and adherent HCII expressing cells were re-supplemented with DMEM (no FBS) + 2 mM sTyr + 2 mM sodium butyrate. Collected media containing overexpressed HCII (20 mL total per 10 cm plate) were pooled and subjected to purification.

HCII containing media was centrifuged at 5,000 x g at 4 °C for 30 min to remove any residual debris. The supernatant was concentrated with Amicon 30 kDa MWCO centrifugal filters to approximately 2 mL. For concentrated media harvested from five 10 cm dishes, 1 mL Ni-NTA (Thermo-Scientific) resin was used for protein purification. Bound protein was washed with 50 mL of wash buffer containing PBS + 45 mM imidazole. HCII was eluted with 10 mL elution buffer, concentrated down to 1 mL using a 30 kDa MWCO filter, and buffer exchanged into HNPN –PEG buffer (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% NaN_3). Protein yields and purity were analyzed by Bradford, SDS-PAGE, anti-His tag dot blot, and tryptic/elastase mass spectrometry.

4.4.9 Deglycosylation assay of HCII

PNGase F was purchased from Promega (V4831). 18 µL (10 µg, in 0.5 mM Tris-HCl, pH 7.8) purified recombinant HCII was incubated at 37 °C with or without 2 µL PNGase for 18 hrs. After incubation, mixtures were resolved by SDS-PAGE and imaged via ChemiDoc imaging.

4.4.10 Tryptic & Elastase mass spectrometry characterization of HCII

An in gel digestion was performed to prepare peptides for MS analysis. 1000-2000 ng HCII was resolved by SDS-PAGE. Gel was stained for 1 hr, and destained overnight. After

destain, HCII bands were sliced and cut into approximately 1 mm^2 pieces. Pieces were placed in microcentrifuge tubes containing 500 µL 100 mM ammonium bicarbonate. Gel bands were frozen at -80°C overnight in the 500 µL ammonium bicarbonate. Gel bands were thawed, supernatant was removed, and gel bands were washed 1-2x for 15 min with 500 µL 100 mM ammonium bicarbonate. After washes, supernatant was removed and 200 µL 10 mM TCEP was added to completely cover gel bands. Samples were placed in a 60 °C water bath for 30 min. Samples were quickly spun and TCEP was aspirated. 200 µL 55 mM iodoacetamide was added to cover the gel bands. Tubes were placed in the dark for 30 min at RT. Supernatant was removed and gel bands were washed 3x for 15 min in 500 µL 50:50 acetonitrile:100 mM ammonium bicarbonate. After washes, supernatant was removed and 50 µL acetonitrile was added to completely dehydrate the gel bands (turned opaque). Acetonitrile was removed and residual solvent was removed using a SpeedVac for 5 min.

Sequencing grade trypsin (V5111) and neutrophil elastase (V1891) was purchased from Promega. Sample was resuspended in either trypsin (for 60 site) or trypsin+elastase (for 73 site). For trypsin, 200 ng trypsin (20 µL, resuspended in 25 mM ammonium bicarbonate) was added to dehydrated gel slices. For trypsin+elastase, 300 ng (30 µL, 25 mM ammonium bicarbonate) trypsin was added, immediately followed by 35 µL elastase (30 ng, resuspended in double distilled water according to manufacturer protocol) + 50 µL 50 mM Tris-HCl. In both cases, enzymes were incubated with gel sample for 10 min before 200 µL 50 mM ammonium bicarbonate was added and placed at 37 °C incubator overnight. Next, the supernatant was transferred to a clean tube and 100 µL formic acid was added to the gel bands followed by a 15 min incubation at RT. The supernatant was aspirated and combined with the supernatant from the last step. Formic acid washes of the gel slices were repeated two more times. Next, 150 µL acetonitrile was added to cover the gel slices, incubated at RT for 15 min,

and combined with all previous washes. Acetonitrile washes were repeated two more times until bands became opaque. Lastly, the peptide sample (~500 μL consisting of the overnight incubation supernatant, formic acid washes, and acetonitrile washes) was evaporated down to 10 μL using SpeedVac and stored at -80 °C until subjected to HPLC-MS analysis.

Digested peptides were analyzed by LC-MS using an LTQ Orbitrap XL mass spectrometer (Thermo Fisher) coupled to an EASY-nLC 1000 nanoLC (Thermo Fisher). 18 μL of sample was loaded onto 100 μm fused silica column with a 5 μm tip packed with 10 cm of Aqua C18 reverse-phase resin (Phenomenex) using the EASY-nLC 1000 autosampler. Peptides were eluted with a gradient 0–55% buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B; 20% water, 80% acetonitrile, 0.1% formic acid). The flow rate through the column was set to 400 nl/min and the spray voltage was set to 3.5 kV. One full MS scan (400–1800 MW) was followed by seven data dependent scans. For the data dependent scans, a mass list was used to target the predicted peptides with sTyr at residues 60 and 73. In the absence of a targeted peptide, data dependent scans were performed on the nth most intense ions in the MS1. MS1 spectra and total ion chromatograms were manually analyzed for peptide identification and presence of sulfation at each residue.

4.4.11 HCII-Thrombin activity assay

To calculate the second order rate constant of thrombin inhibition by HCII, thrombin was incubated with excess HCII under pseudo-first order conditions in the presence of different heparin concentrations (details below). The reaction was quenched after 1 minute and the residual thrombin activity ($k_{inhibited}$) was measured using a chromogenic substrate. The pseudo-first order rate constant (k_1) was calculated from this using the equation $k_1 = -\ln(k_{inhib}/k_{uninhib})/t$, where $k_{uninhib}$ is the activity of thrombin in the absence of HCII inhibition

under identical treatment. The second order inhibition rate constant (k_2) was calculated from $k_1$using the equation $k_2 = k_1/[HCII]$ with units of $M^{-1}min^{-1}$. The second order rate constant at each heparin concentration was plotted against the corresponding heparin concentration.

Concentrations of different HCII protein were measured by Bradford and normalized by anti-His dot-blot assay (blot intensities quantified via ChemiDoc imaging). Clear plastic 96 well plates were coated with 2 mg/mL ovalbumin (Fisher) for 1 hr at 37 °C. Ovalbumin was removed by tapping the plate on a paper towel. A master mix of 2 mg/mL ovalbumin, 0-2 mg/mL heparin (Fisher), 0.6 nM α-thrombin (Fisher) in HNPN (20 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% NaN_3) were incubated in the treated 96 wells for 1 min with 10 nM HCII. After 1 min, 10 μL of a solution of 1 mg/mL polybrene was directly added to all wells using a multichannel pipet to quench the heparin-dependent inhibition of thrombin by HCII. The plates were spun down in a bucket centrifuge for 10 min at 3,500 rpm to precipitate the heparin/polybrene complex. 100 μL supernatant was removed and 50 μL 450 μM ChromozymeTH (Sigma) substrate was added to measure the amount of residual thrombin activity by monitoring the absorbance on the SpectroMax plate reader at 405 nm for 1 hr in triplicate.

4.5 References

1. Huttner, W. B. Tyrosine sulfation and the secretory pathway. *Annu. Rev. Physiol.* **50,** 363–376 (1988).
2. Kehoe, J. W. & Bertozzi, C. R. Tyrosine sulfation: A modulator of extracellular protein-protein interactions. *Chem. Biol.* **7,** 57–61 (2000).
3. Moore, K. L. Protein tyrosine sulfation: A critical posttranslation modification in plants and animals. *Proc. Natl. Acad. Sci.* **106,** 14741–14742 (2009).
4. Seibert, C. & Sakmar, T. P. Toward a framework for sulfoproteomics: Synthesis and characterization of sulfotyrosine-containing peptides. *Biopolym. - Pept. Sci. Sect.* **90,** 459–477 (2008).
5. Stone, M. J., Chuang, S., Hou, X., Shoham, M. & Zhu, J. Z. Tyrosine sulfation: an increasingly recognised post-translational modification of secreted proteins. *N. Biotechnol.* **25,** 299–317 (2009).

6. Yang, Y. *et al.* Tyrosine sulfation as a protein post-translational modification. *Molecules* **20,** 2138–2164 (2015).
7. Farzan, M. *et al.* Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* **96,** 667–676 (1999).
8. Huang, C. *et al.* Structural basis of tyrosine sulfation and V. (2004).
9. Li, X. *et al.* Characterization of a sulfated anti-HIV antibody using an expanded genetic code Characterization of a sulfated anti-HIV antibody using an expanded genetic code. (2018). doi:10.1021/acs.biochem.8b00374
10. Mikkelsen, J., Thomsen, J. & Ezban, M. Heterogeneity in the Tyrosine Sulfation of Chinese Hamster Ovary Cell Produced Recombinant FVIII. *Biochemistry* **30,** 1533–1537 (1991).
11. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550,** 53–60 (2017).
12. Italia, J. S. *et al.* Expanding the genetic code of mammalian cells. *Biochem. Soc. Trans.* **45,** 555–562 (2017).
13. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13,** 854–870 (2018).
14. Dumas, A., Lercher, L., Spicer, C. & Davis, B. G. This journal is © The Royal Society of Chemistry 2014 Designing logical codon reassignment – expanding the chemistry in biology. 1–36 (2014). doi:10.1126/scienc
15. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24,** 1436–1440 (2006).
16. Liu, C. C., Cellitti, S. E., Geierstanger, B. H. & Schultz, P. G. Efficient expression of tyrosine-sulfated proteins in E. coli using an expanded genetic code. *Nat. Protoc.* **4,** 1784–1789 (2009).
17. Liu, C. C., Choe, H., Farzan, M., Smider, V. V. & Schultz, P. G. Mutagenesis and evolution of sulfated antibodies using an expanded genetic code. *Biochemistry* **48,** 8891–8898 (2009).
18. Liu, C. C., Brustad, E., Liu, W. & Schultz, P. G. Crystal structure of a biosynthetic sulfo-hirudin complexed to thrombin. *J. Am. Chem. Soc.* **129,** 10648–10649 (2007).
19. Italia, J. S., Latour, C., Wrobel, C. J. J. & Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. *Cell Chem. Biol.* 1–9 (2018). doi:10.1016/j.chembiol.2018.07.002
20. Rau, J. C., Mitchell, J. W., Fortenberry, Y. M. & Church, F. C. Heparin cofactor II: Discovery, properties, and role in controlling vascular homeostasis. *Semin. Thromb. Hemost.* **37,** 339–348 (2011).
21. Tollefsen, D. M. Regulation of Hemostasis and Thrombosis : Insights From Murine Models Heparin Cofactor II Modulates the Response to Vascular Injury. *Arterioscler. Thromb.* 454–460 (2007). doi:10.1161/01.ATV.0000256471.22437.8
22. Blinder, M. A., Marasa, J. C., Reynolds, C. H., Deaven, L. L. & Tollefsen, D. M. Heparin Cofactor II: cDNA Sequence, Chromosome Localization, Restriction Fragment Length Polymorphism, and Expression in Escherichia Coli. *Biochemistry* **27,** 752–759 (1988).
23. Hortin, G., Tollefsen, D. M. & Strauss, A. W. Identification of two sites of sulfation of human heparin cofactor II. *J. Biol. Chem.* **261,** 15827–15830 (1986).
24. Chin, J. W. An Expanded Eukaryotic Genetic Code. *Science (80-.).* **301,** 964–967

(2003).
25. Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods* **4,** 239–244 (2007).
26. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat. Chem. Biol.* 1–8 (2017). doi:10.1038/nchembio.2312
27. Ciaccia, A. V., Willemze, A. J. & Church, F. C. Heparin promotes proteolytic inactivation by thrombin of a reactive site mutant (L444R) of recombinant heparin cofactor II. *J. Biol. Chem.* **272,** 888–893 (1997).
28. Ciaccia, A. V., Monroe, D. M. & Church, F. C. Arginine 200 of heparin cofactor II promotes intramolecular interactions of the acidic domain: Implication for thrombin inhibition. *J. Biol. Chem.* **272,** 14074–14079 (1997).
29. Mitchell, J. W. & Church, F. C. Aspartic acid residues 72 and 75 and tyrosine-sulfate 73 of heparin cofactor II promote intramolecular interactions during glycosaminoglycan binding and thrombin inhibition. *J. Biol. Chem.* **277,** 19823–19830 (2002).
30. Tollefsen, D. M., Majerus, D. W. & Blank, M. K. Heparin Cofactor II. *Arch Pathol Lab Med* **257,** 2162–2169 (1982).

Chapter 5

Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites

A significant portion of the work described in this chapter has been accepted for publication: Italia, J.S., Peeler, Addy, P.S., Erickson, S.B., Peeler, J.C., Weerapana, E., Chatterjee, A. Mutually orthogonal nonsense-suppression systems and conjugation chemistries for precise protein labeling at up to three distinct sites. *J. Am. Chem. Soc.* 2019. Just accepted.

5.1 Introduction

5.1.1 Co-translational labeling of recombinantly expressed proteins

The ability to precisely label proteins with multiple distinct entities at predefined locations is highly desirable for numerous cross-disciplinary applications. A variety of different strategies have been developed for labeling proteins that target native residues, the termini, as well as peptide and protein tags appended to a recombinant protein.[1–17] Co-translational site-specific incorporation of noncanonical amino acids (ncAA) enables precise placement of uniquely reactive chemical functionalities into proteins, which can be subsequently labeled using chemoselective bioconjugation reactions.[18–24] Significant advantages of this strategy include the small size of the ncAAs relative to other genetically encoded tags, the flexibility of incorporating them into virtually any site of any target protein, and the large variety of unique chemical groups that can be genetically encoded using this technology. Incorporation of ncAAs into proteins in living cells is performed using engineered aminoacyl-tRNA synthetase (aaRS)/tRNA pairs that suppress nonsense or frameshift codons and do not cross-react with their counterparts from the host cell (i.e., orthogonal).[18–22] Several different pairs have been developed for ncAA incorporation into proteins expressed in different domains of life.[18–22,25] Using this technology, many chemoselective bioconjugation handles have been genetically encoded, including azides, alkynes, ketones, strained alkenes and alkynes, tetrazines, etc., which has enabled novel approaches to understand and manipulate protein function.[18–24,26,27]

5.1.2 UAA incorporation systems for multi-site suppression and site-specific protein labeling

So far, the large majority of work using this technology has been restricted to the incorporation of a single ncAA into a polypeptide. However, the ability to incorporate multiple distinct bioconjugation handles into proteins, which can be independently functionalized with

different entities, has the potential to facilitate many powerful applications, including the attachment of different probes for sophisticated biophysical studies, and the synthesis of advanced protein-based therapeutics and diagnostics. Such expression systems would require multiple different aaRS/tRNA pairs that do not cross-react with each other or with their host counterparts. It is important to remember that two pairs may cross-react at three different levels: ncAA-aaRS interaction, aaRS-tRNA interaction, and codon-anticodon interaction.[28] These two pairs must be able to efficiently suppress two distinct nonsense or frameshift codons, and incorporate two different ncAAs that encode mutually compatible bioconjugation chemistries for their independent labeling. Using the *M. jannaschii* derived tyrosyl pair (MjTyr) and the methanosarcina-derived pyrrolysyl pair (Pyl), which are mutually orthogonal, it has indeed been possible to site-specifically incorporate two distinct ncAAs into proteins expressed in *E. coli*.[29–34] MjTyr is typically used to suppress the UAG nonsense codon, while the more anticodon-permissive Pyl is assigned to UAA or AGGA codons.[29–33] Taking advantage of bioconjugation handles that have been genetically encoded using these pairs, particularly those for strain-promoted azide-alkyne cycloaddition (SPAAC)[35,36] and strain-promoted inverse electron-demand Diels-Alder cycloaddition (SPIEDAC),[37–41] it has been possible to achieve site-specific dual labeling of proteins targeted to distinct ncAA residues.[29,31–33]

5.1.3 Use of ATM platforms for double and triple suppression with three compatible chemistries

Despite these exciting and empowering advances, further expansion of this technology is impeded by the unavailability of additional mutually orthogonal aaRS/tRNA pairs, as well as genetically encodable novel bioconjugation chemistries that are compatible with SPAAC and SPIEDAC. Recently, we developed a unique *E. coli* derived tryptophanyl-tRNA synthetase

(EcTrpRS)/tRNA pair which can be used for ncAA incorporation in both *E. coli* and eukaryotes.[42] We functionally substituted the endogenous EcTrp pair of *E. coli* with its yeast counterpart to generate the novel strain ATMW1. The "liberated" EcTrpRS/tRNA pair was then established as an orthogonal UGA suppressor in ATMW1, and was engineered to selectively charge a variety of tryptophan analogs, including 5-hydroxytryptophan (5HTP; Figure 5-1).[42] Furthermore, we have developed chemoselective bioconjugation strategies that take advantage of the highly electron rich 5-hydroxyindole group of 5HTP.[43,44] For example, based on the remarkably rapid coupling between 5-hydroxyindoles and various aromatic diazonium ions, we have developed the chemoselective rapid azo-coupling reaction (CRACR) for selective labeling of proteins at 5HTP residues. In this report, we show that 5HTP-directed CRACR is compatible with both SPAAC and SPIEDAC for chemoselective labeling of proteins at different sites. Next, we demonstrate that our UGA-suppressing EcTrpRS/tRNA pair can be orthogonally used with UAG-suppressing MjTyr or Pyl to site-specifically incorporate two distinct ncAAs into proteins expressed in ATMW1 *E. coli*. We used these new dual-ncAA mutagenesis platforms to express proteins co-incorporating 5HTP with *p*-azidophenylalanine (pAzF) or cyclopropene-lysine (CpK) (Figure 5-1), which were subsequently site-specifically double labeled using CRACR and SPAAC or SPIEDAC, respectively. Finally, we combined all three suppressors to site-specifically co-incorporate 5HTP,[42] pAzF,[45,46] and CpK[47] into the same protein in response to three different stop codons and demonstrated their independent labeling with three distinct entities.

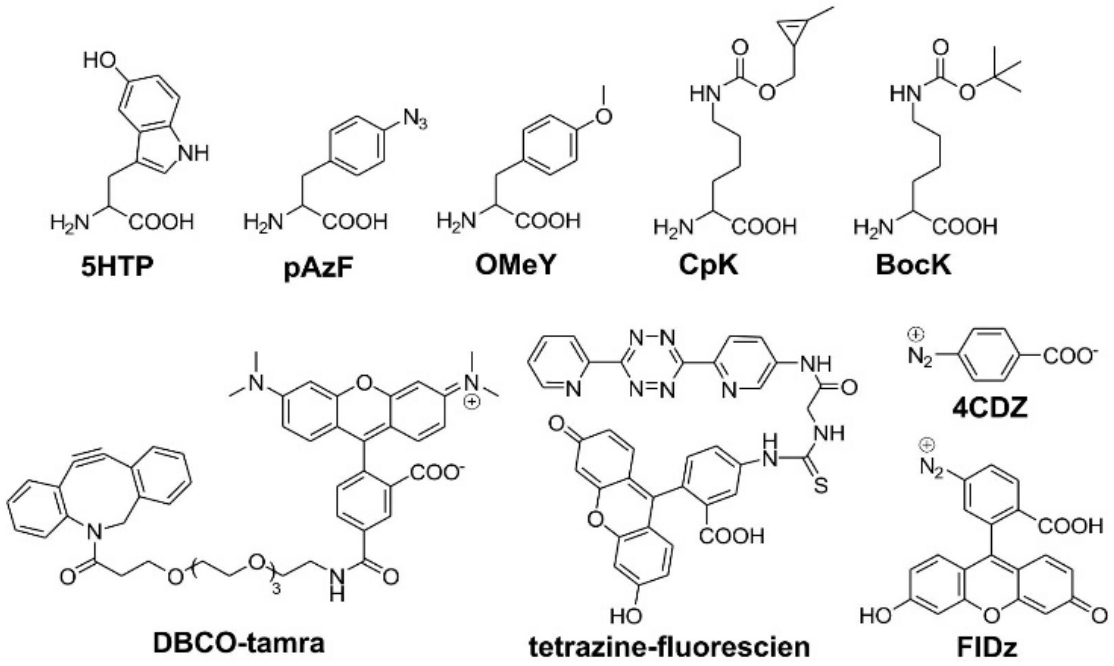


Figure 5-1. Structures of the ncAAs and labeling agents used in this report.

5.2 Results and Discussion

5.2.1 CRACR is compatible with SPAAC and SPIEDAC

For site-specifically labeling proteins with different entities, distinct ncAAs with bioconjugation handles are needed which can be functionalized by chemistries that do not interfere with each other. So far, this has been achieved most successfully by combining ncAAs that can be labeled by SPAAC and SPIEDAC.[28,31,33,48] For example, proteins containing an azide and a cyclopropene sidechain have been sequentially labeled in one pot by a cyclooctyne and an electron-deficient tetrazine, respectively.[28,48] Distinctly reactive SPIEDAC handles have also been used to this end.[74,78] Additionally, Cu(I)-catalyzed azide-alkyne cycloaddition and oxime conjugation have been shown to be compatible with SPIEDAC and SPAAC,[29,30,49] respectively, but their scope is somewhat limited due to the requirement of toxic catalysts or low pH.

CRACR enables rapid labeling of proteins (>100 $M^{-1}s^{-1}$ for 4-carboxydiazonium; 4CDZ) at 5HTP residues under ambient conditions.[43] To evaluate if CRACR is compatible with SPAAC and SPIEDAC mediated protein labeling, we expressed superfolder green fluorescent protein (sfGFP) incorporating 5HTP, CpK or pAzF at position 151. A previously reported polyspecific MjTyrRS/tRNA pair[46] and the wild-type MbPylRS/tRNA pair[47] were used to incorporate pAzF and CpK, respectively, in response to the UAG stop codons. 5HTP was incorporated using our recently reported EcTrpRS/$tRNA_{UCA}$ pair in the ATMW1 strain.[42] To investigate whether the labeling conditions for each of these bioconjugation handles are compatible with each other, we subjected each of the three different sfGFP mutants to three different labeling conditions using DBCO-tamra, tetrazine-fluorescein, and 4CDZ (or 4CDZ-biotin)(Figure 5-1). Subsequent analysis by whole-protein ESI-MS analysis revealed that these reagents facilitate complete and selective labeling of sfGFP mutants harboring their corresponding bioconjugation partners (DBCO-pAzF, tetrazine-CpK, 4CDZ-5HTP), while leaving the other two mutants untouched (Figure 5-2). We further resolved these proteins by SDS-PAGE and analyzed the labeling reaction by fluorescence imaging (for DBCO-TAMRA, and tetrazine-fluorescein) or Western blot (for 4CDZ-biotin, using streptavidin-HRP conjugate), further confirming selective labeling of each of the three mutants only with their intended reagents (Figure 5-3). These observations establish that CRACR-mediated labeling of 5HTP, SPAAC-labeling of pAzF and SPIEDAC-labeling of CpK are mutually compatible. Therefore, it should be possible to distinctly functionalize these ncAA sidechains with multiple different entities if they could be simultaneously incorporated into one protein.

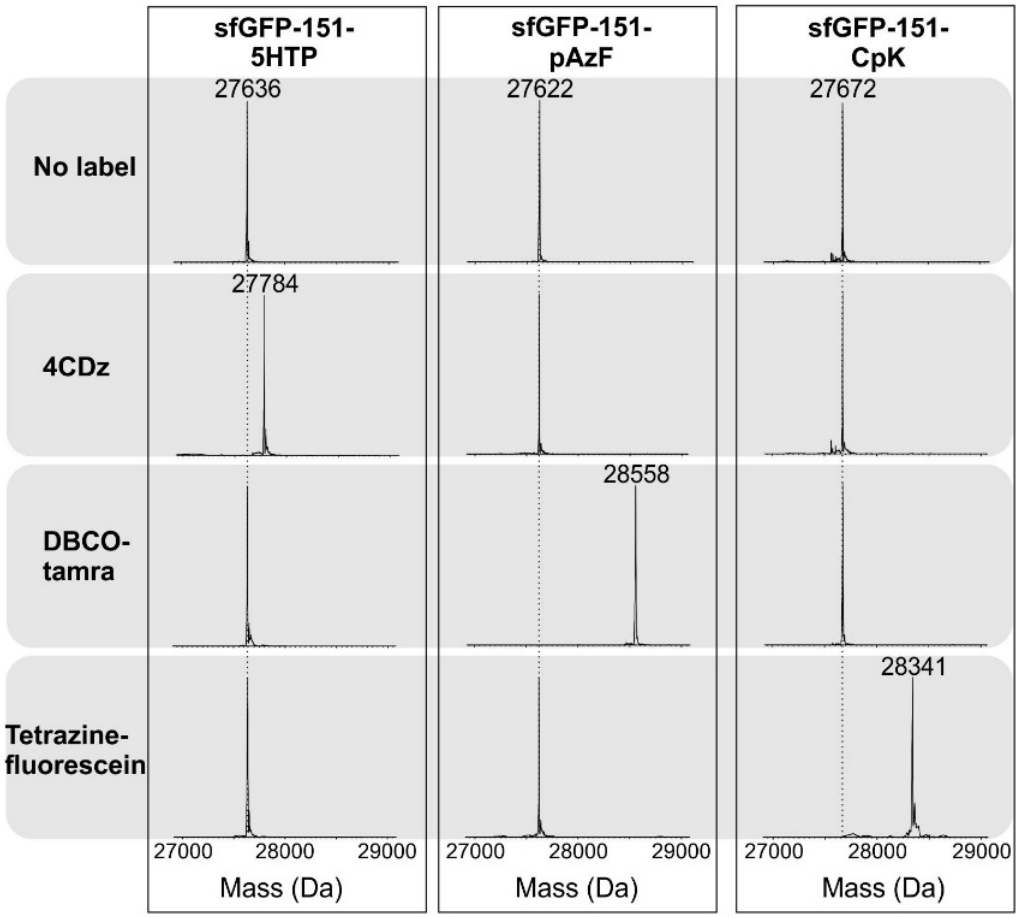


Figure 5-2. Selective labeling and orthogonality of CRACR, SPAAC, and SPIEDAC via MS. Treatment of three sfGFP mutants (incorporating 5HTP, pAzF, or CpK at the 151 position) with 4CDZ, DBCO-tamra, and tetrazine-fluoresceine followed by ESI-MS analysis.

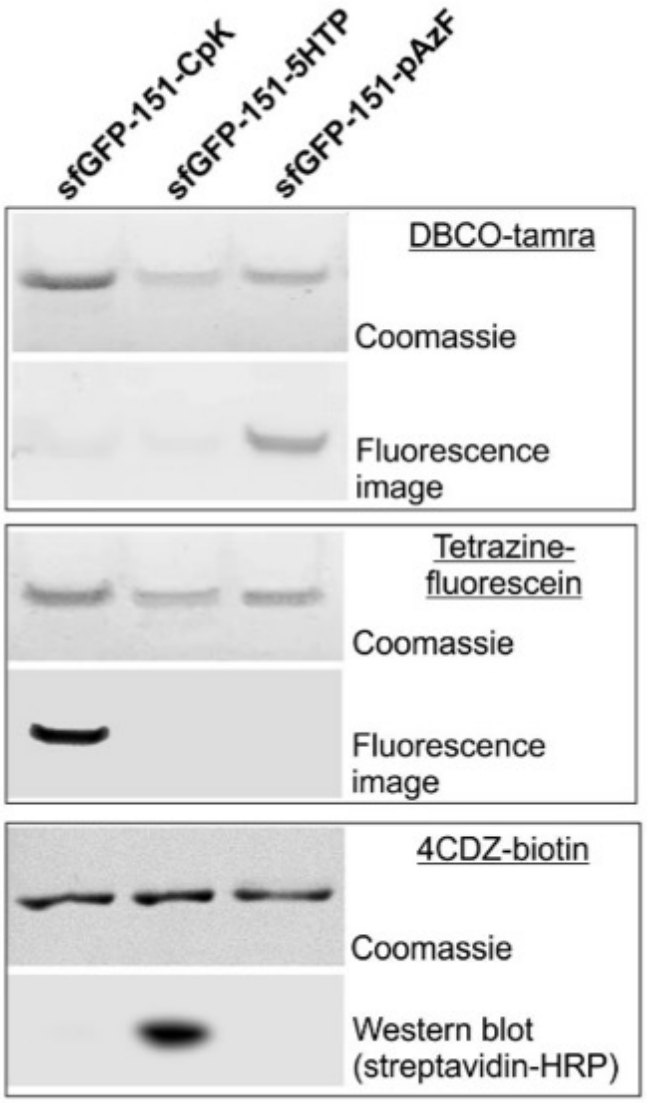


Figure 5-3. Selective labeling and orthogonality of CRACR, SPAAC, and SPIEDAC via fluorescent probe conjugation. Analysis of protein samples from a similar experiement 94CDZ-biotin used instead of 4CDZ) by SDS-PAGE followed by fluorescence imaging or Western blot (for 4CDZ-Biotin).

To facilitate the co-incorporation of 5HTP into proteins with pAzF and CpK, we next evaluated if the EcTrpRS/$tRNA_{UCA}$ pair can be combined with MjTyr or Pyl to suppress two distinct nonsense codons in a mutually orthogonal manner in our ATMW1 strain of *E. coli*. Unlike all other pairs (including MjTyr and Pyl) that typically use a UAG nonsense codon for ncAA incorporation, our EcTrpRS/$tRNA_{UCA}$ pair was developed as an efficient UGA-suppressor,[49] making it intrinsically easier to combine it with other pairs for suppressing two distinct nonsense codons. Additionally, both MjTyr and Pyl have been used in regular *E. coli* strains as orthogonal UAG suppressors, suggesting that these do not cross-react with any of the endogenous pairs (including the EcTrp, which is used as an orthogonal pair in our engineered ATMW1 strain). However, it is important to note that in our expression system, the anticodon of $tRNA^{EcTrp}$ is altered to perform UGA suppression, and the pair is expressed at a significantly higher level relative to endogenous expression (to achieve high levels of nonsense suppression efficiency), which may compromise its mutual orthogonality with MjTyr and Pyl pairs.

5.2.2 A dual ncAA incorporation system using EcTrpRS/$tRNA^{EcTrp}_{UCA}$ and PylRS/$tRNA^{Pyl}_{CUA}$

Due to its unique structural features, the pyrrolysyl (Pyl) pair is orthogonal in all domains of life. Its mutual orthogonality with other suppressor pairs has been the basis of all dual-ncAA incorporation systems developed so far.[28–33,48,49] Consequently, we anticipated that it would also be orthogonal to our EcTrp suppressor pair in ATMW1 *E. coli*. To ensure that the EcTrp and the Pyl pairs can be simultaneously assigned to UGA and UAG codons with high fidelity, we constructed a dual suppressor plasmid expressing both of these pairs. Our previously reported polyspecific EcTrpRS (h14) and the wild-type MbPylRS were each expressed from *tacI* promoters, while the $tRNA^{EcTrp}_{UCA}$ and $tRNA^{Pyl}_{CUA}$ were expressed using

separate *proK* promoters. This plasmid was co-transformed into ATMW1 *E. coli* strain with a sfGFP reporter that harbors a UGA codon at position 3 and a UAG codon at position 151 (Figure 5-4). The expression of this sfGFP double mutant was monitored in rich media in the presence or absence of substrate ncAAs (BocK and CpK for Pyl, 5HTP for EcTrp) for both pairs. Robust reporter expression was observed when both substrates were present, while no sfGFP fluorescence was detected when both substrates were absent, or when 5HTP alone was present (Figure 5-4). The latter observation is consistent with the fact that the Pyl pair is essentially silent in the absence of a substrate ncAA, causing no "leaky" UAG suppression. When only a Pyl-substrate (BocK or CpK) was present, weak reporter expression was observed due to the low levels of tryptophan-charging activity of the EcTrpRS mutant (h14) in the absence of 5HTP. We have previously shown that this mutant exclusively charges 5HTP when it is supplemented in the expression medium.[49] We isolated full-length sfGFP-3-5HTP-151-BocK/CpK by immobilized metal ion chromatography (IMAC) using a C-terminal polyhistidine tag (17-20 mg/L) and analyzed the protein by SDS-PAGE and whole-protein ESI-MS (Figure 5-5), confirming successful incorporation of the intended ncAAs.

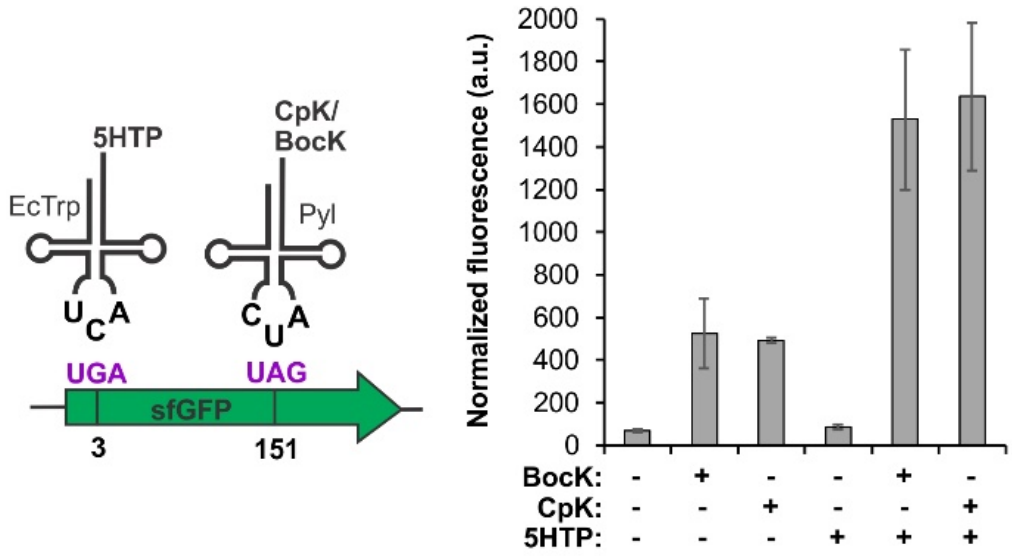


Figure 5-4. Dual ncAA incorporation system using EcTrp(UCA) and Pyl(CUA) pairs in ATMW1 *E. coli*. A graphical scheme is depicted on the left, with the in-cell fluorescence expression of indicated ncAA substrates shown on the right.

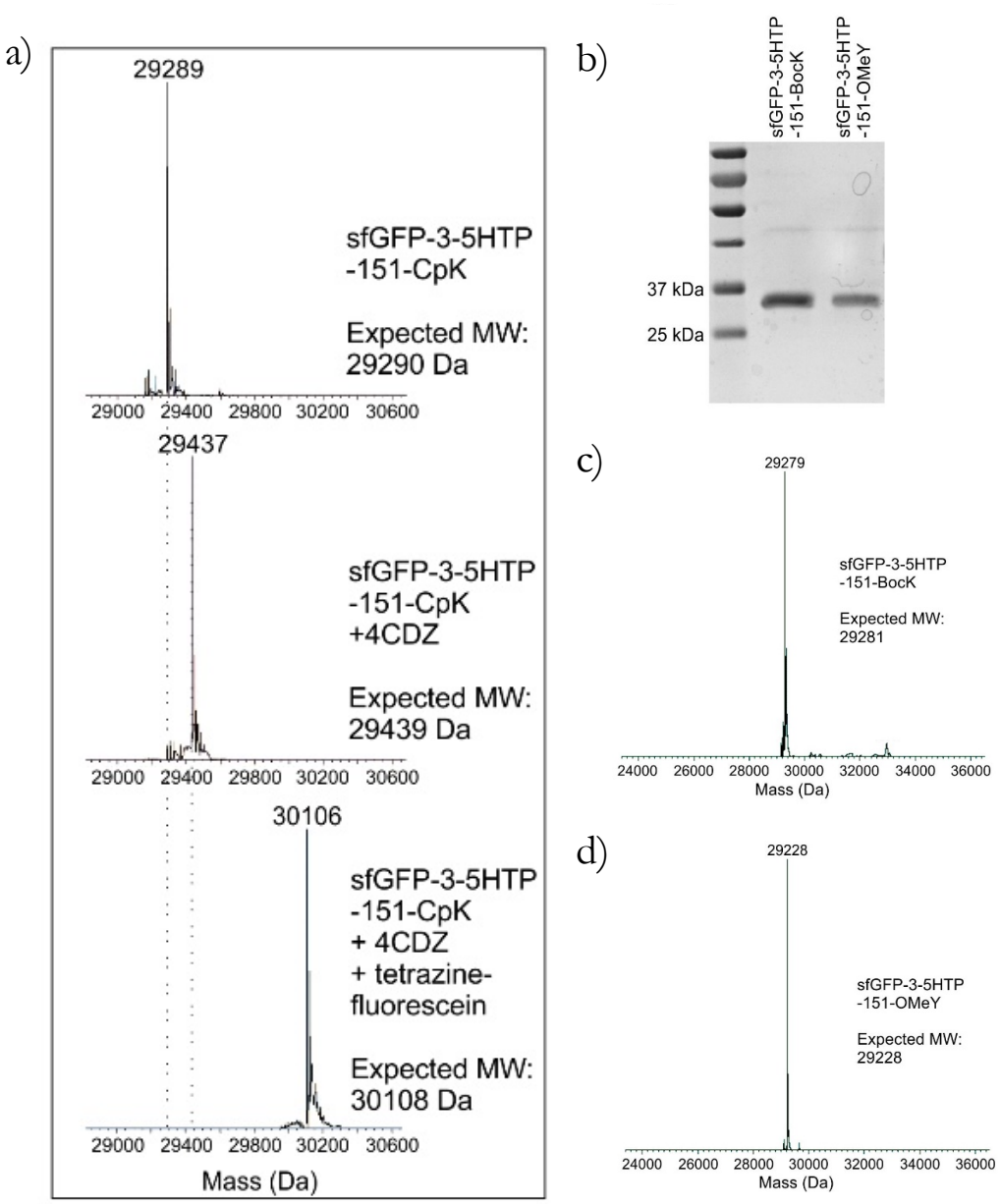


Figure 5-5. Mass spec and SDS-PAGE analysis of double suppressed protein. a) Purified sfGFP-3-5HTP-151-CpK has the expected mass (Top panel). Treatment with 4CDZ and tetrazine-fluorescein results in its one-pot quantitative and selective double labeling of CRACR with SPIEDAC. b) SDS-PAGE analysis of sfGFP-3-5HTP-151-BocK and sfGFP-3-5HTP-151-OMeY. c) and d) correspond to the mass spec analysis of the BocK or OMeY containing double suppressed proteins from the SDS-PAGE analysis.

The ability to site-specifically incorporate CpK and 5HTP into a protein with high fidelity and efficiency using the above dual ncAA incorporation system makes it possible to test whether these two bioconjugation handles can be used to attach two distinct labels using SPIEDAC and CRACR. Purified sfGFP-3-5HTP-151-CpK was subjected to sequential labeling first with 4CDZ, and then with tetrazine-fluorescein in the same pot under ambient conditions. ESI-MS analysis of the protein after each step revealed complete labeling with expected mass changes, confirming the suitability of these two chemoselective labeling strategies for facile site-specific dual labeling of recombinant proteins (Figure 5-5a).

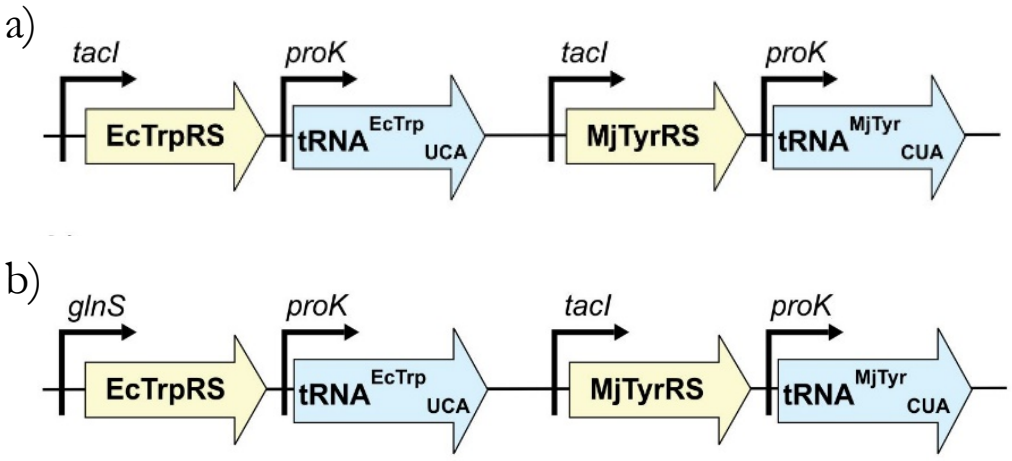


Figure 5-6. Map of double suppressor plasmid systems for EcTrp/MjTyr systems. a) Cross-reactive scheme. b) Scheme with reduced cross-reactivity containing the weaker glnS promoting EcTrpRS.

5.2.3 Cross-reactivity identification and circumvention between EcTrpRS/tRNA$^{EcTrp}_{UCA}$ and MjTyrRS/tRNA$^{MjTyr}_{CUA}$

Next, we investigated if the EcTrp pair can be combined with the MjTyr pair in a mutually orthogonal fashion. We built a dual suppressor plasmid, similar to the one described in the previous section, that expresses the EcTrpRS(h14)/tRNA$^{EcTrp}_{UCA}$ and the MjTyrRS(polyspecific)/tRNA$^{MjTyr}_{CUA}$ pairs (Figure 5-6a). However, a detailed characterization of its suppression behavior revealed a significant level of cross-reactivity between EcTrpRS and tRNA$^{MjTyr}_{CUA}$ (Figure 5-7, 5-8, 5-9). When this plasmid was used to express the sfGFP-3-UAG reporter, significant protein expression was observed in the absence of a substrate for MjTyrRS, and MS analysis of the resulting protein revealed the incorporation of tryptophan at the UAG codon (Figure 5-7). However, OMeY was selectively incorporated at the UAG codon when it was supplemented in the growth medium, which shows that MjTyrRS can outcompete EcTrpRS for charging tRNA$^{MjTyr}_{CUA}$ in the presence of its substrate ncAA. This cross-reactivity was surprising, given that the MjTyrRS/tRNA$^{MjTyr}_{CUA}$ pair has been extensively used for ncAA mutagenesis in various *E. coli* strains with no evidence of cross-reactivity with the endogenous EcTrpRS.

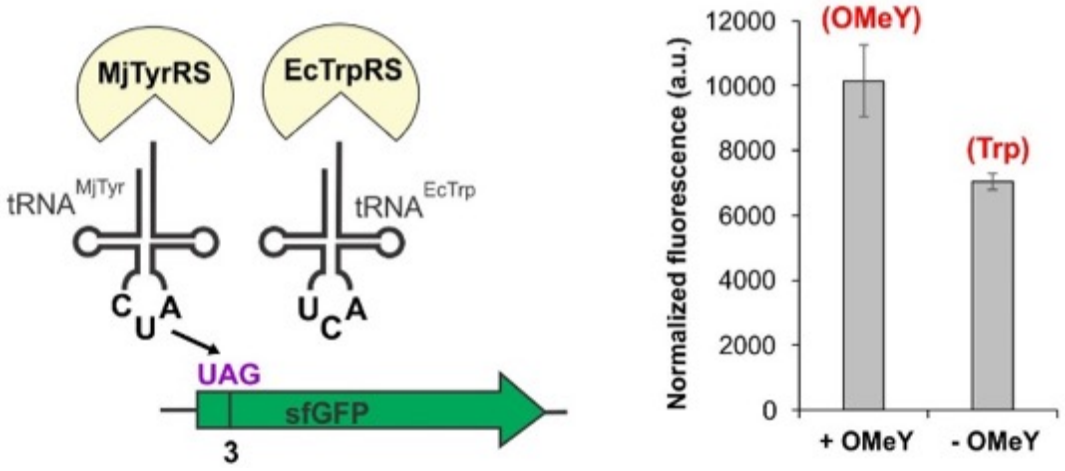


Figure 5-7. Cross-reactive incorporation of tryptophan at the UAG codon in the presence of both EcTrp(UCA) and MjTyr(CUA) suppressors. Expressing sfGFP-3-UAG using this double suppressor plasmid (scheme on the left) show significant level of protein expression both in the presence and absence of OMeY (substrate for the MjTyr pair). The amino acid incorporated at the UAG codon in each case is identified by ESI-MS analysis of the resulting protein, and shown in red in parenthesis. Incorporation of Trp in the absence of OMeY reveals an underlying cross-reactivity between the EcTrp and the MjTyr system.

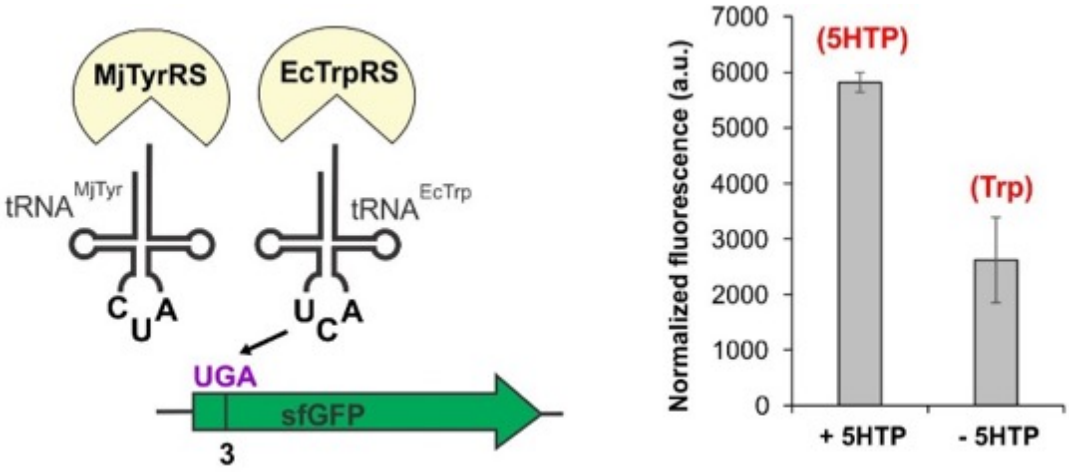


Figure 5-8. Cross-reactive incorporation of tryptophan at the UGA codon in the presence of both EcTrp(UCA) and MjTyr(CUA) suppressors. Expressing sfGFP-3-UGA reporter in the presence of the double suppressor plasmid (scheme on the left) show expected incorporation of 5HTP at the UGA codon when it is added to the media, and tryptophan in its absence.

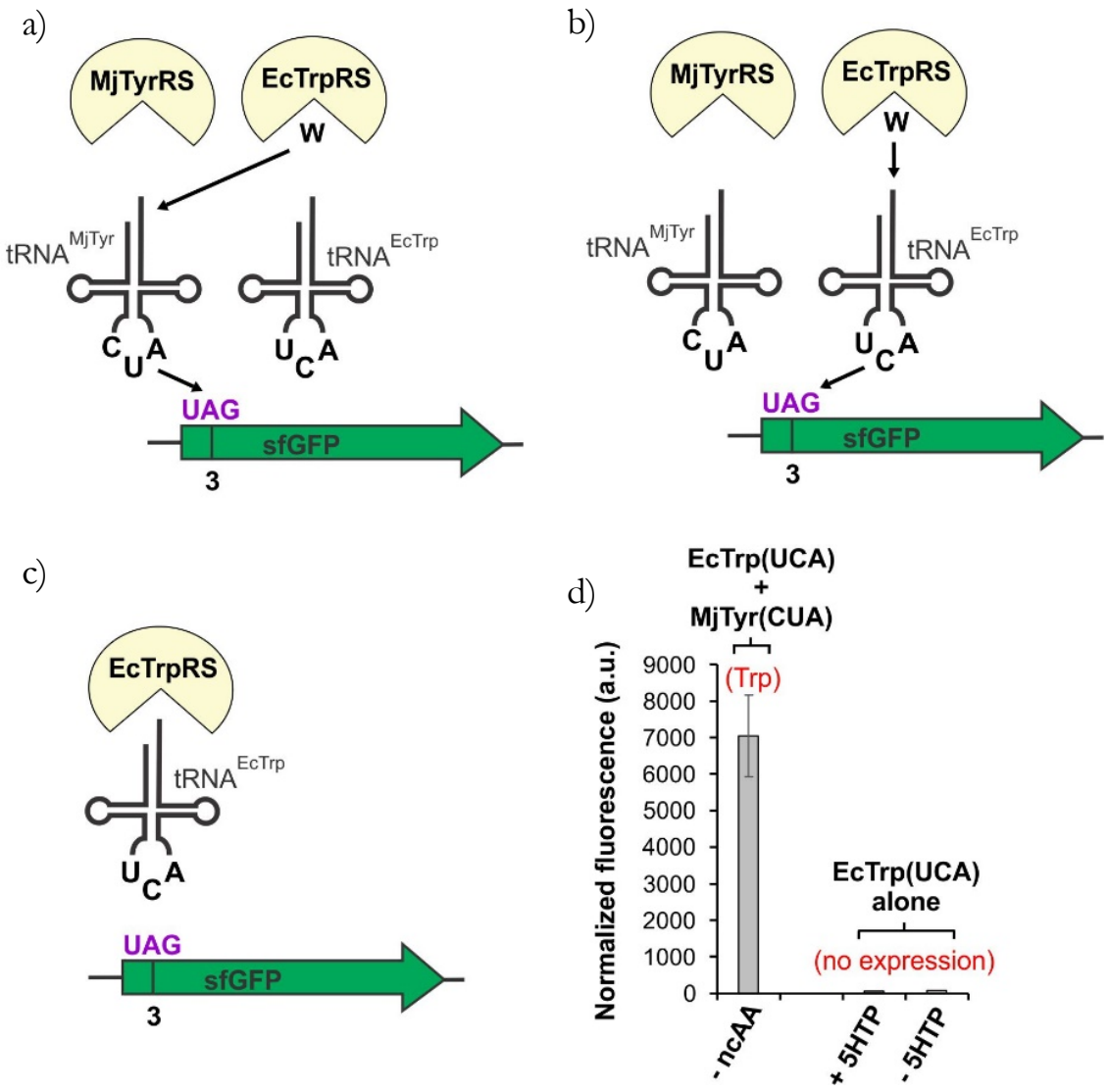


Figure 5-9. Investigating the mechanism of cross reactivity between EcTrp and MjTyr suppression systems. The $tRNA^{EcTrp}_{UCA}$ cannot suppress the UAG codon, and does not contribute to cross-reactivity. Two possible mechanisms for cross-reactive incorporation of Trp into UAG codon observed in Figure 5-5d: a) EcTrpRS charges non-cognate $tRNA^{MjTyr}_{CUA}$ leading to UAG suppression, and b) EcTrpRS charges Trp to $tRNA^{EcTrp}_{UCA}$, which misincorporates it at the UAG codon. c) Scheme of the expression of sfGFP-3-UAG reporter in the presence of EcTrp(UCA) pair alone. d) Result of EcTrp(UCA) alone in cell fluorescence demonstrates no protein expression in the presence or absence of 5HTP. Expression level of

the same reporter using the double suppressor plasmid EcTrp(UCA)+MjTyr(CUA) in the absence of any added ncAA is also shown for reference (left bar). This indirectly confirms that the mechanism A is responsible for incorporation of Trp in response to UAG.

We hypothesized that at its low endogenous expression levels in regular *E. coli* strains, EcTrpRS does not exhibit detectable levels of cross-reactivity toward $tRNA^{MjTyr}_{CUA}$. In contrast, the overexpression of EcTrpRS using a strong promoter (*tacI*) from a multi-copy plasmid in our system causes it to significantly cross-react with this non-cognate tRNA. Indeed, it has been previously reported that overexpression of an aaRS can result in its cross-reaction with non-cognate tRNAs.[50] If correct, this hypothesis would predict a relief from cross-reactivity if we lower the expression level of EcTrpRS. To this end, we created an alternative double suppression plasmid, where the strong inducible *tacI* promoter was replaced with a much weaker constitutively active *glnS* promoter to drive EcTrpRS expression (Figure 5-6b). This plasmid facilitated the expression of the sfGFP-3-UAG reporter only in the presence of OMeY, the MjTyrRS substrate (Figure 5-10).

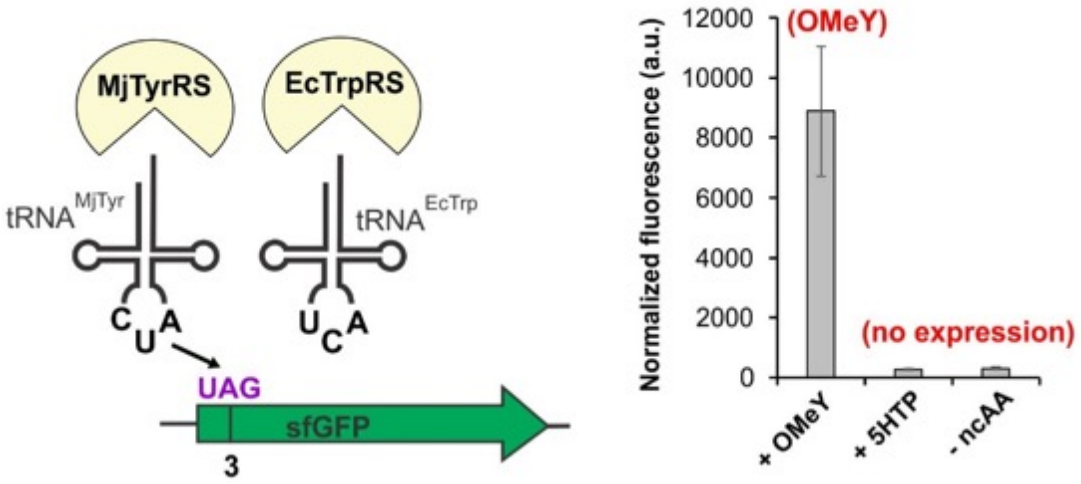


Figure 5-10. Redesigned EcTrp(UCA) and MjTyr(CUA) suppressor system eliminates cross-reactive incorporation of tryptophan at the UAG codon. Genetic scheme is outlined in Figure 5-6b. Expressing sfGFP-3-UAG using this double suppressor plasmid (Figure 5-6b) show reporter expression only in the presence of OMeY, while the cross-reactive tryptophan incorporation is completely eliminated.

The lack of UAG suppression in the absence of a MjTyrRS-substrate confirms a relief from the aforementioned cross-reactivity, corroborating our hypothesis. We were also gratified to find that the reduced expression EcTrpRS did not significantly compromise its efficiency of 5HTP incorporation in response to UGA codons, but it drastically attenuated "leaky" incorporation of tryptophan in the absence of its substrate 5HTP: while the new double suppressor plasmid facilitated comparable levels of the sfGFP-3-UGA expression relative to its earlier counterpart in the presence of 5HTP, little reporter expression was observed in the absence of the ncAA (Figure 5-8 and Figure 5-10). These observations underscore the importance of expressing the aaRS/tRNA pairs at an optimal level for incorporating ncAAs with high fidelity and efficiency.

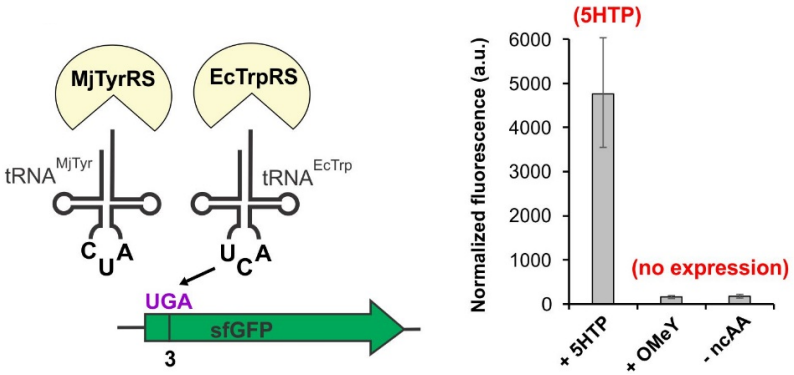


Figure 5-11. Expressing sfGFP-3-UGA reporter in the presence of the optimized double suppressor plasmid (Figure 5-6b) also show reporter expression only in the presence of 5HTP. It is notable that while the reduced expression of EcTrpRS eliminates cross-charging of $tRNA^{MjTyr}_{CUA}$ and alleviates tryptophan incorporation at UGA in the absence of 5HTP, incorporation efficiency of 5HTP is not significantly affected (relative to previous plasmid, Figure 5-8).

5.2.4 An optimized dual ncAA incorporation system using EcTrpRS/$tRNA^{EcTrp}_{UCA}$ and MjTyrRS/$tRNA^{MjTyr}_{CUA}$

We then tested the incorporation of two distinct ncAAs into the sfGFP-3-UGA-151-UAG reporter using this optimized EcTrp(UCA)+MjTyr(CUA) double suppression system in ATMW1 *E. coli*. 5HTP was used as the substrate for EcTrpRS(h14), while OMeY or pAzF were tested for the polyspecific MjTyrRS mutant. We observed high levels of reporter expression only in the presence of both 5HTP and OMeY/pAzF (Figure 5-12). Purification of the full-length double mutant using a C-terminal polyhistidine tag by IMAC (yields of 25-30 mg/L) followed by whole-protein ESI-MS analysis revealed the expected mass (Figure 5-13, Figure 5-5b, Figure 5-5d), validating incorporation of the two ncAAs at desired sites. The sfGFP-3-5HTP-151-pAzF also provides the opportunity to test if CRACR and SPAAC can

be used together to attach two different labels on the same protein. Purified sfGFP-3-5HTP-151-pAzF was labeled first with 4CDZ, followed by DBCO-tamra under ambient conditions and analyzed by ESI-MS to demonstrate quantitative labeling of each ncAA residue (Figure 5-13).

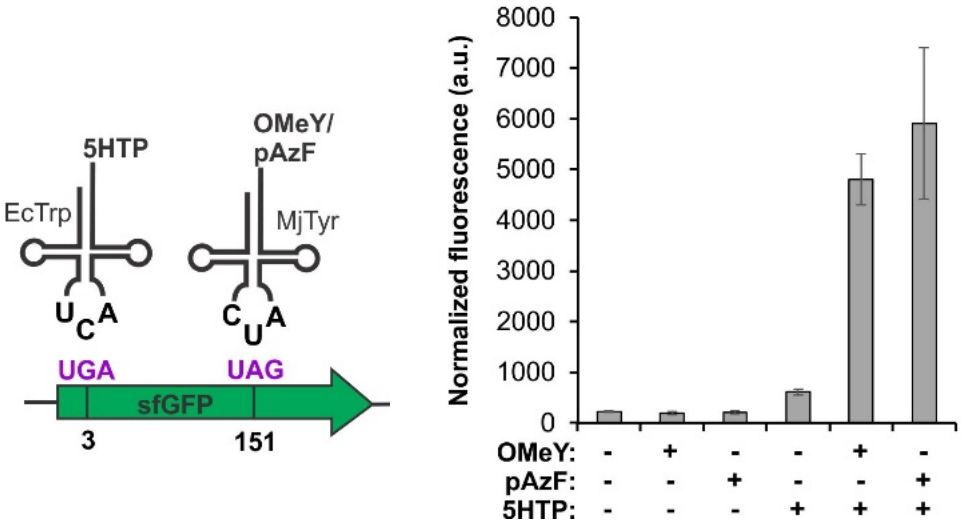


Figure 5-12. Dual ncAA incorporation system using the optimized EcTrp(UCA) and MjY(CUA) pairs in ATMW1 *E. coli*. A graphical scheme is depicted on the left, with the in-cell fluorescence expression of indicated ncAA substrates shown on the right.

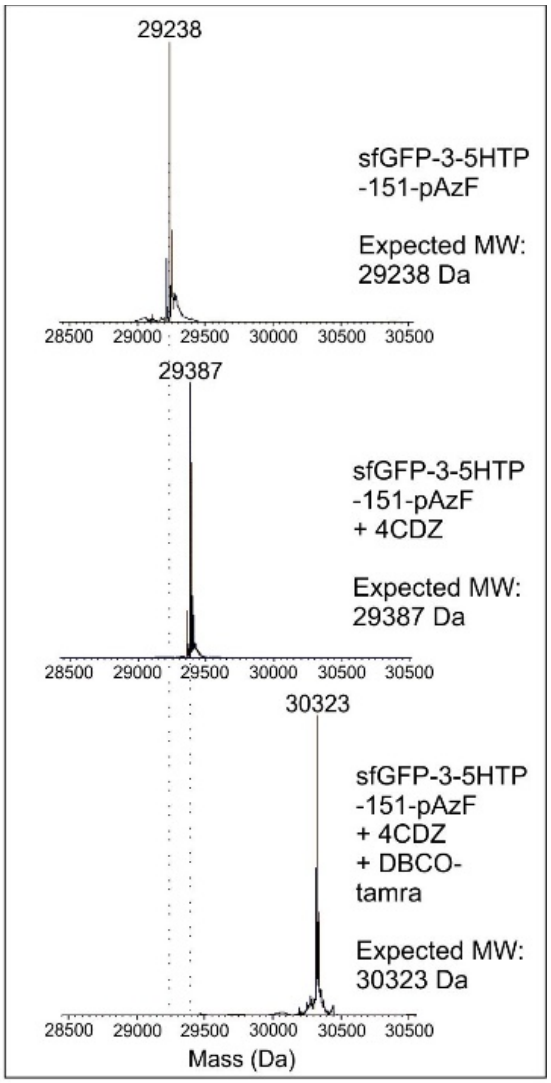


Figure 5-13. Mass spec analysis of EcTrp/MjTyr double suppressed protein. Purified sfGFP-3-5HTP-151-pAzF has the expected mass (Top panel). Treatment with 4CDZ and DBCO-tamra results in its one-pot quantitative and selective double labeling of CRACR with SPAAC. SDS-PAGE and MS analysis of sfGFP-3-5HTP-151-OMeY is shown in Figure 5-5b and Figure 5-5d, respectively.

The ability to precisely functionalize antibodies and antibody fragments is highly important for numerous applications. Efficient site-specific dual modification of antibodies can facilitate the development of sophisticated diagnostics and therapeutics.[1,26,27,51] To explore the possibility of extending our dual labeling strategies described above to this important class of proteins, we selected the 5HTP+pAzF incorporation system using the EcTrpRS/$tRNA^{EcTrp}_{UCA}$ + MjTyrRS/$tRNA^{MjTyr}_{CUA}$ pairs, which was somewhat more efficient than the 5HTP+CpK incorporation system. The optimized double suppression plasmid harboring both pairs was co-transformed into ATMW1 *E. coli* with a plasmid expressing the Fab fragment of the anti-HER2 antibody Herceptin, where the heavy chain encoded two mutations: K169UGA and S202UAG (Figure 5-14).[52,53] We deliberately chose to incorporate both ncAAs into the same subunit of the antibody to highlight the robustness of our expression system. The wild-type Fab fragment was also expressed as a control. The double mutant Fab was expressed in the presence of 5HTP and pAzF and was purified from the periplasmic fraction by protein-A affinity chromatography.

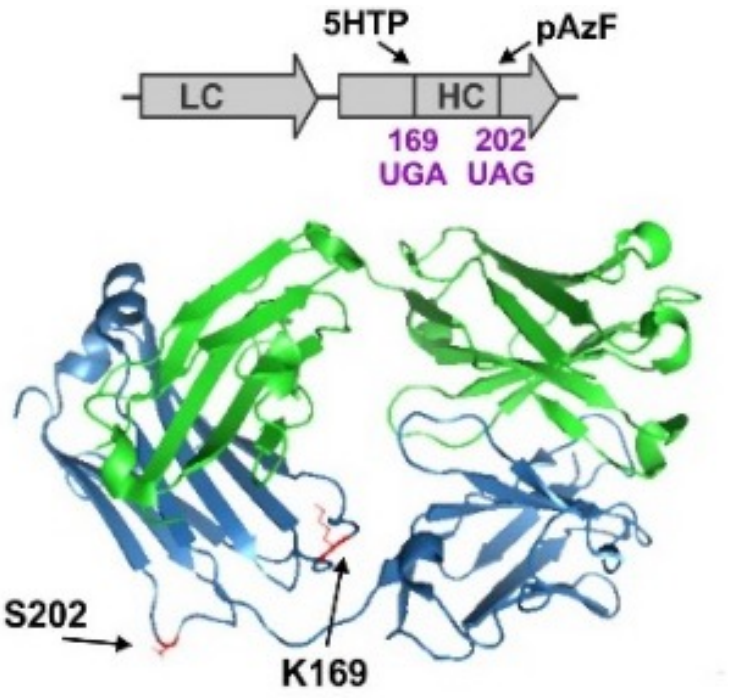


Figure 5-14. Scheme for the site-specific double labeling of anti-HER2 Fab using CRACR and SPAAC. The Fab was expressed incorporating 5HTP and pAzF into 169 and 202 sites of the heavy chain using the dual-ncAA incorporation system described in Figure 5-12.

ESI-MS analysis of the whole antibody confirmed the incorporation of both ncAAs (Figure 5-15). When the purified double mutant Fab was treated with 4CDZ, or DBCO-tamra or both reagents, clean labeling of the corresponding sites was observed by whole-protein ESI-MS analysis (Figure 5-15a). Identical treatment of the wild-type Fab resulted in no modification (Figure 5-15b and Figure 5-15c), underscoring the selectivity of this double labeling strategy. We also showed that the Fab double mutant can be labeled with DBCO-tamra and fluorescein-diazonium (FlDZ) reagents to install two distinct fluorescent labels at two different sites (Figure 5-16).

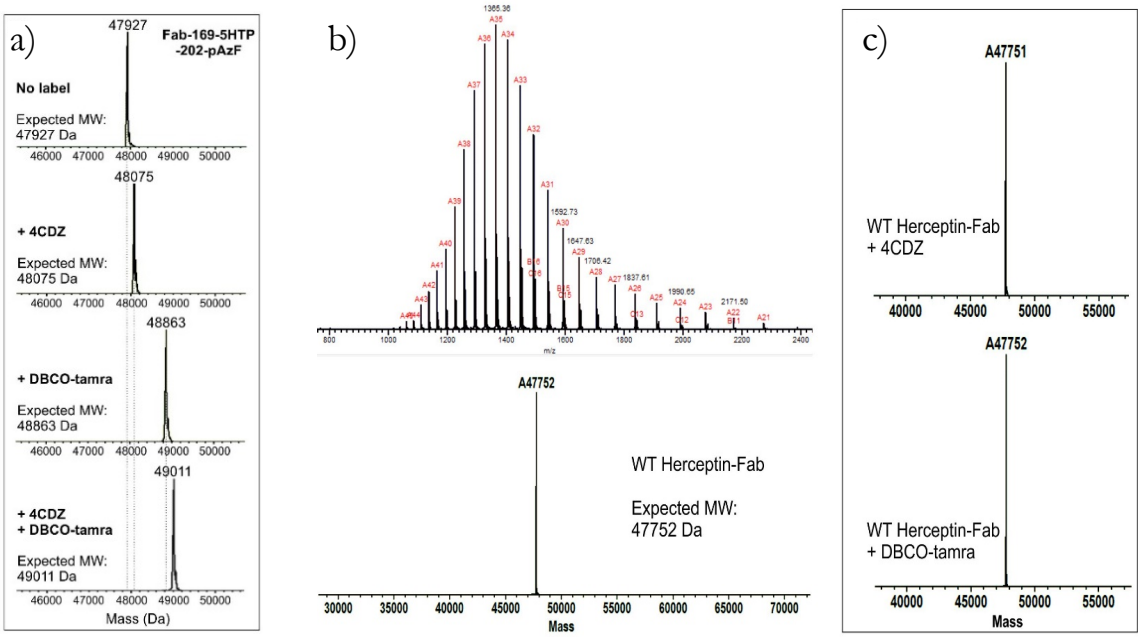


Figure 5-15. Mass spectrometry analysis of anti-HER2-Fab double suppression and labeling. a) ESI-MS analysis of Fab-169-5HTP-202-pAzF (not reduced; HC and LC connected by a disulfide) shows the expected mass (top panel). Treatment of this Fab double mutant with 4CDZ or DBCO-tamra produces expected single modification, while quantitative double modification is observed when both are used. b) Full protein mas of WT Herceptin-Fab demonstrates expected molecular weight. c) Identical treatment of the wild-type Fab leads to no modification.

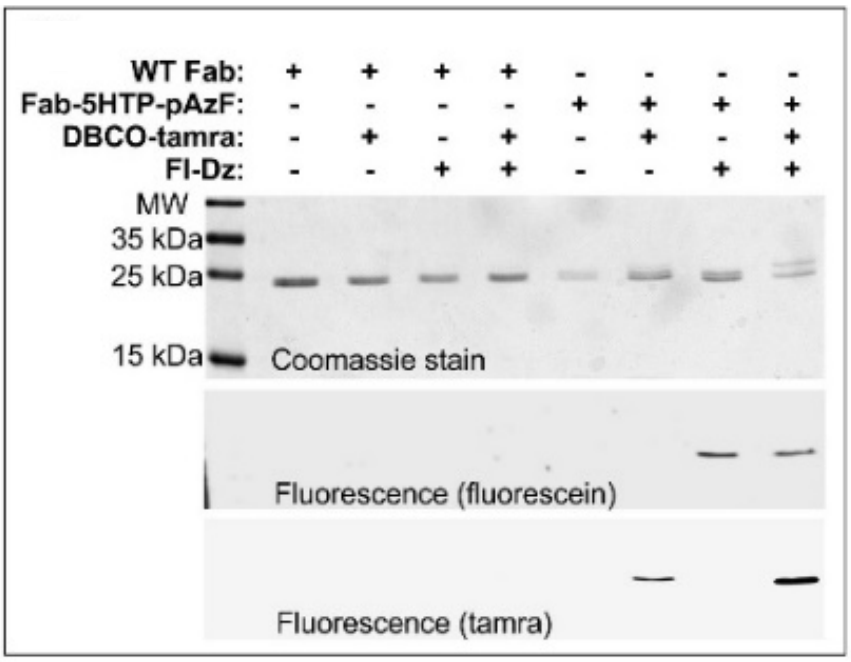


Figure 5-16. SDS-PAGE and gel fluorescence analysis of anti-HER2-Fab double suppression and labeling. Labeling Fab-169-5HTP-202-pAzF or wild-type Fab with DBCO-tamra or fluorescein-diazonium or both leads to expected labeling of the double mutant but not the wild-type, as revealed by reducing SDS-PAGE (HC and LC separates; each ~24 kDa) followed by fluorescence imaging.

5.2.5 Site-specific incorporation of three different ncAAs into one protein, and its chemoselective labeling with three distinct entities

Our work establishes EcTrp, MjTyr and Pyl as a set of three mutually orthogonal pairs in ATMW1 *E. coli*. We wondered if it would be possible to simultaneously use all three pairs to site-specifically incorporate 5HTP, pAzF and CpK into one protein. The compatible bioconjugation chemistries these encode should then enable the labeling of the resulting protein with three distinct entities. To our knowledge, site-specific incorporation of three different ncAAs into one protein in living cells has not yet been achieved, and the ability to

do so would mark a major milestone for the ncAA mutagenesis technology. However, this would require the assignment of these three pairs to three distinct nonsense/frameshift codons. Since the suppression efficiency of nonsense codons are generally higher, we chose to simultaneously use the three nonsense codons to encode three distinct ncAAs. We envisioned assigning EcTrp, MjTyr and the Pyl pairs to suppress UGA, UAG, and UAA, respectively, to achieve site-specific incorporation of three different ncAAs, since Pyl has been previously used as a UAA-suppressor together with a UAG-suppressing MjTyr pair for dual ncAA incorporation.[29,32] We confirmed that each of these three nonsense suppressors can facilitate efficient expression of sfGFP reporters encoding the appropriate nonsense codon at the 151 site in ATMW1 *E. coli* (Figure 5-17).

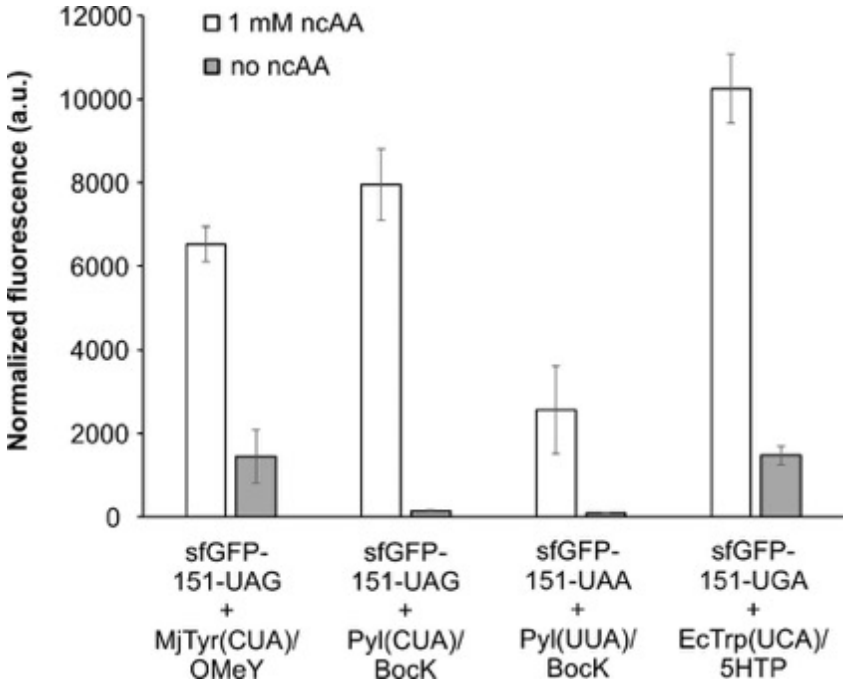


Figure 5-17. Single suppression efficiency of ncAA incorporation by MjTyr(CUA), Pyl(CUA), Pyl(UUA), and EcTrp(UCA) into sfGFP mutants harboring the appropriate nonsense codon. The ncAA used in each case is shown below.

However, simultaneous reassignment of all three stop codons raises the important question of how to terminate translation. It is important to remember that at a reassigned nonsense codon, incorporation of the designated ncAA is not the only outcome; significant levels of termination is also observed. In particular, multiple consecutive stop codons have been shown to effectively terminate translation even in the presence of highly efficient suppressors.[54] Although placing multiple consecutive stop codons at the end of a target gene offers an effective way to terminate its translation, the resulting protein would likely have a heterogeneous C-terminus from partial ncAA incorporation at these sites. To address this concern, we designed a novel expression system GTEV, where the desired recombinant protein is appended with a purification tag (polyhistidine in this case) at the C-terminus, followed by a tobacco-etch virus (TEV) protease cleavage sequence, the TEV protease,[19] and three consecutive UAA stop codons (Figure 5-18). We envisioned that upon expression, the C-terminal TEV protease will cleave itself out, leaving the desired recombinant protein with a clean C-terminus. We expressed wild-type sfGFP from the GTEV vector in ATMW1 *E. coli* and isolated it using the C-terminal polyhistidine tag by IMAC in high yield (140 mg/L; comparable with a non-GTEV expression system). Its SDS-PAGE and ESI-MS analysis showed only the desired protein, confirming efficient self-cleavage of the TEV protease (Figure 5-19).

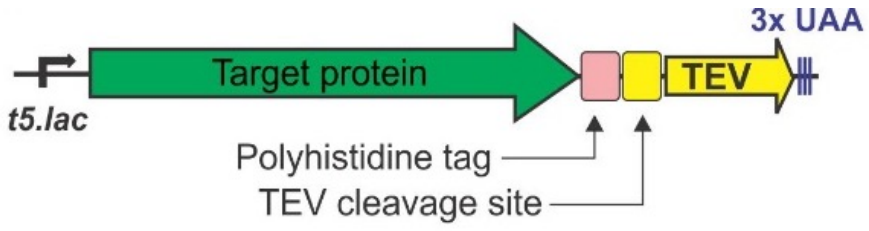


Figure 5-18. Map of the GTEV expression system.

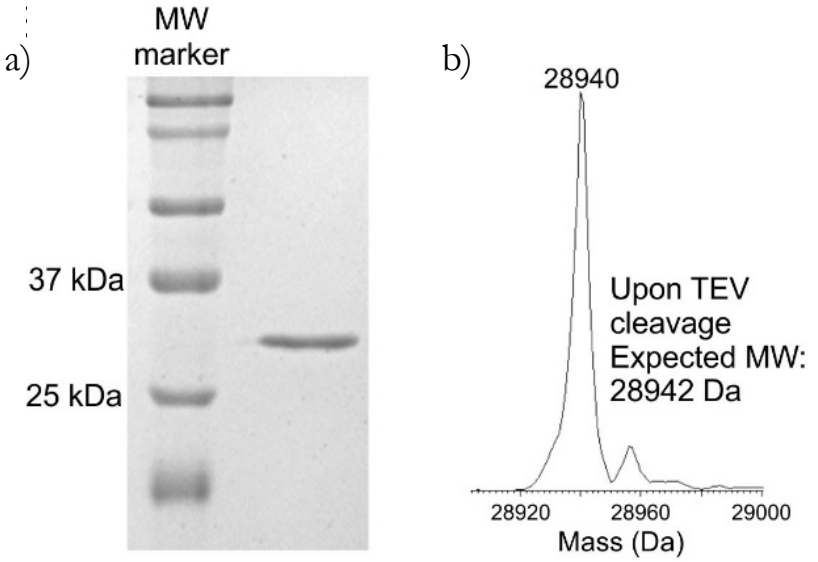


Figure 5-19. Characterization of GFP purified from the GTEV expression system. a) SDS-PAGE analysis of purified (IMAC) wild-type sfGFP protein expressed from the GTEV system show only fully processed product with TEV self-cleavage. b) ESI-MS analysis of the resulting protein reveals the expected mass.

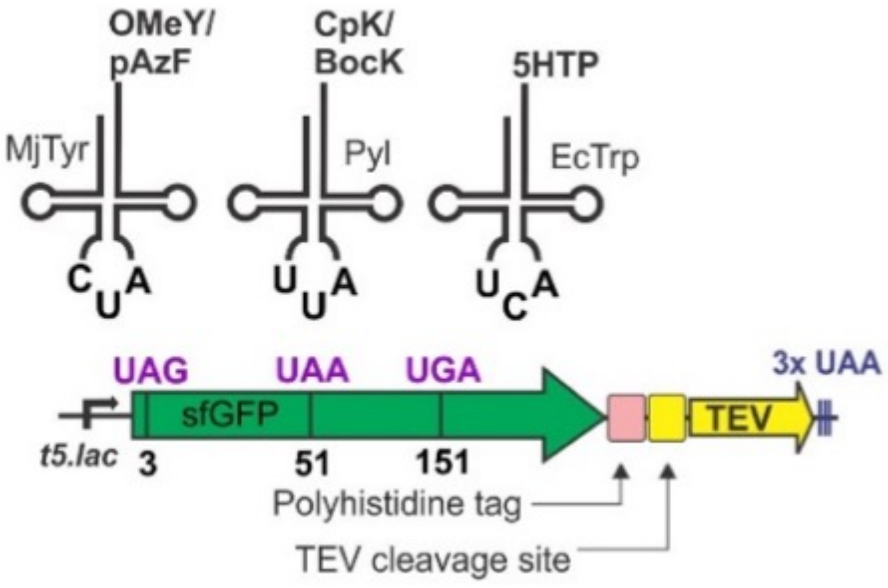


Figure 5-20. Scheme describing the triple ncAA incorporation experiment.

To achieve simultaneous incorporation of three different ncAAs, we needed a plasmid system for co-expressing all three suppressor pairs as well as a reporter gene in the same cell. To this end, we designed a two-plasmid expression system: 1) The aforementioned optimized dual suppressor plasmid encoding EcTrpRS(h14)/$tRNA^{EcTrp}_{UCA}$ and the MjTyrRS(polyspecific)/$tRNA^{MjTyr}_{CUA}$ pairs. 2) A GTEV vector for expressing a triple mutant sfGFP reporter (sfGFP-3UAG-51UAA-151UGA) that also encodes the MbPylRS/$tRNA^{Pyl}_{UUA}$ pair. These two plasmids were co-transformed into ATMW1 *E. coli* (Figure 5-20) and the expression of the full-length reporter was monitored in the presence or absence of ncAAs (5HTP for EcTrp, OMeY or pAzF for MjTyr, BocK or CpK for Pyl) by its characteristic fluorescence. Strong sfGFP expression was observed only when ncAA substrates for each of the three pairs were present in the medium, indicating successful incorporation of three ncAAs at desired sites (Figure 5-21). The full-length sfGFP-3-OMeY-51-BocK-151-5HTP reporter was isolated by IMAC (yield 3 mg/L) and was analyzed by SDS-PAGE (Figure 5-22a) whole-protein ESI-MS (Figure 5-22b) to reveal a mass consistent with the incorporation of the intended ncAAs at targeted sites. We also subjected this protein to tryptic digestion and analyzed the resulting peptides by HPLC-MS/MS analysis, to unambiguously confirm the incorporation of each ncAA at the desired sites (Figure 5-24, Figure 5-25, Figure 5-26). The same expression system was also used to incorporate pAzF+CpK+5HTP into the sfGFP triple mutant reporter with similar yield (2 mg/L). When we treated this protein with either 4CDz, DBCO-tamra, or tetrazine-fluorescein, separately, ESI-MS revealed complete single labeling in each case, confirming the presence of the three different bioconjugation handles which can be independently functionalized (Figure 5-23). We then showed that all three ncAA residues can be labeled sequentially in one pot using CRACR, SPAAC and SPIEDAC to yield a triply functionalized protein (Figure 5-23). The set of

mutually compatible bioconjugation handles and the technology for their site-specific co-incorporation that we develop here should be valuable for numerous applications in chemical biology.

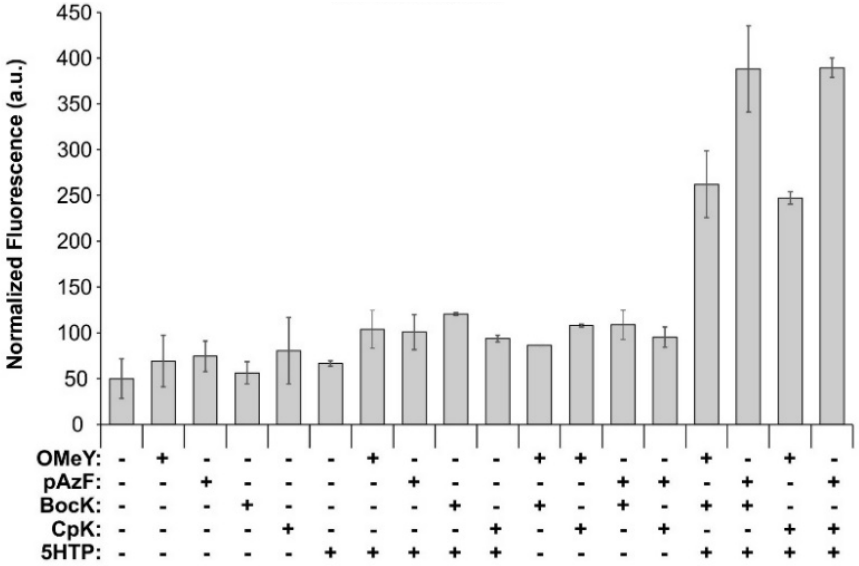


Figure 5-21. Expression of sfGFP-3UAG-51-UAA-151UGA using this system in the presence of indicated ncAA substrates, measured by its characteristic in-cell fluorescence.

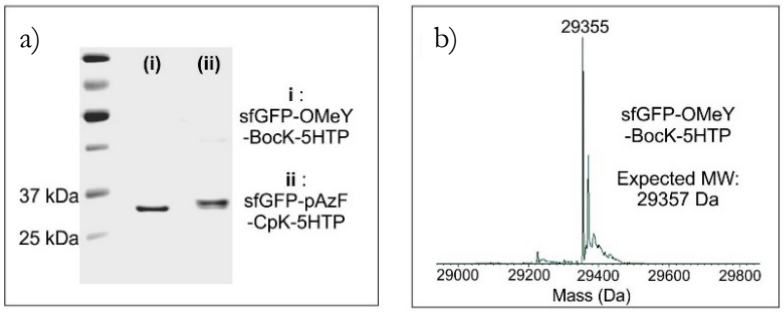


Figure 5-22. Characterization of purified unlabeled triple suppressed protein. a) SDS-PAGE analysis of purified sfGFP triple mutants incorporating either OMeY+BocK+5HTP or pAzF+CpK+5HTP. b) Whole-protein ESI-MS analysis of isolated sfGFP-OMeY-BocK-5HTP shows a mass consistent with the incorporation of the three different ncAAs at intended sites.

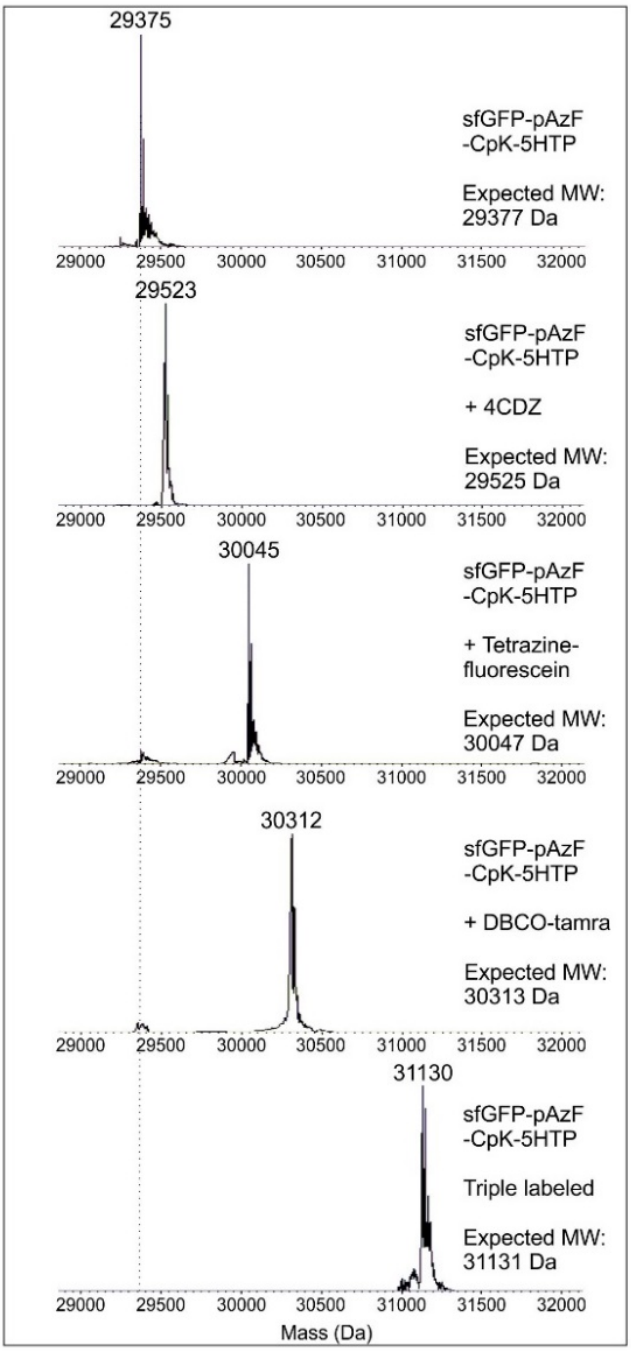


Figure 5-23. One pot but sequential labeling of triple suppressed protein. Purified sfGFP-pAzF-CpK-5HTP also shows the expected mass (top panel). Its treatment with 4CDZ, tetrazine-fluorescein, and DBCO-tamra results in expected single labeling, while one-pot triple labeling is observed when all three labeling reagents are used.

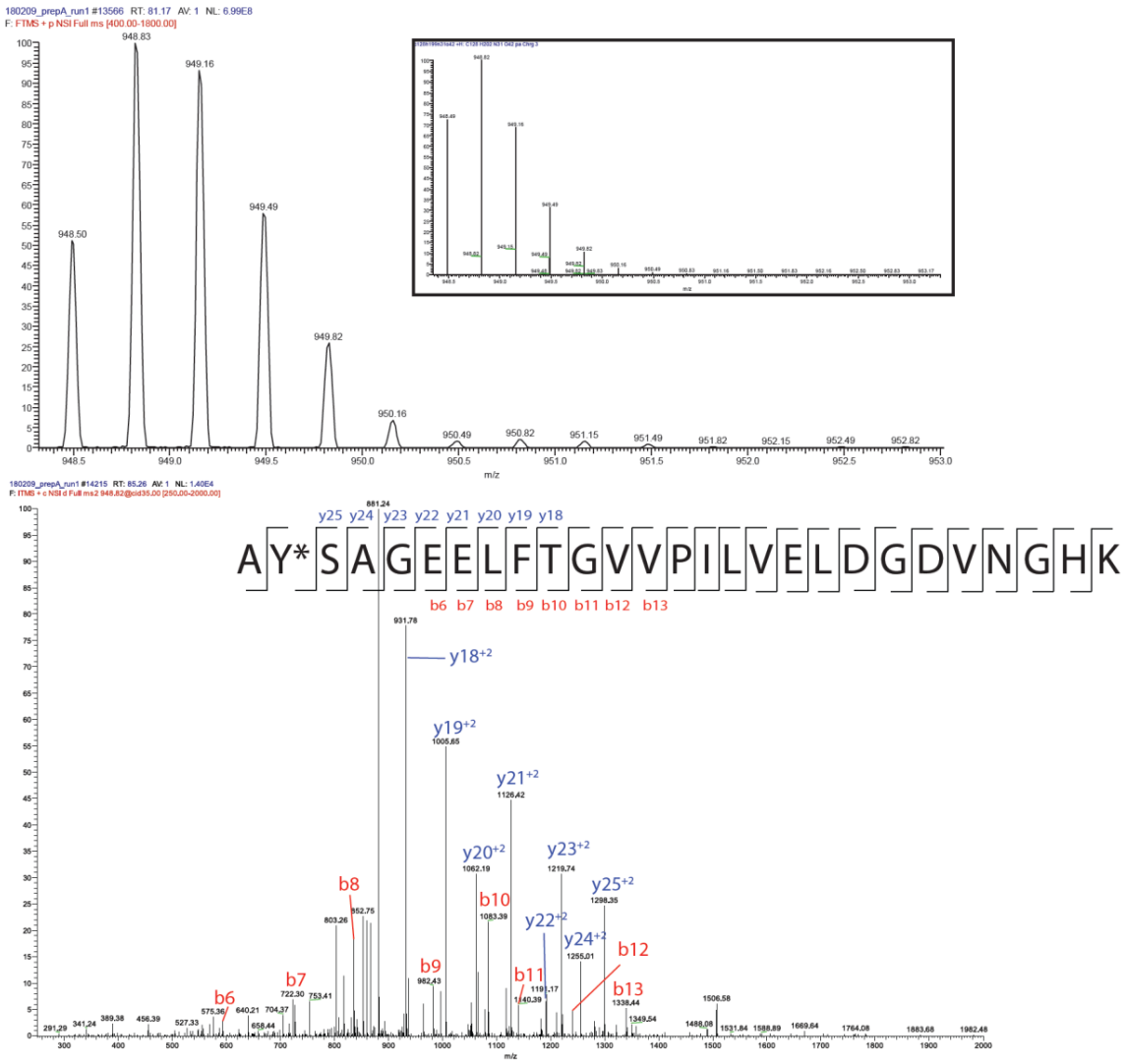


Figure 5-24. Mass and fragmentation of the peptide containing the OMeY residue produced by the tryptic digestion of sfGFP-3-OMeY-51-BocK-151-5HTP (Lys5Ala). Tryptic digestion at Lys5 of original sfGFP produces a very short peptide containing the OMeY residue that was challenging to characterize using our LC-MS set up. Mutating Lys5 to Ala produces a longer OMeY-containing tryptic fragment (shown in the bottom panel) that was readily characterized.

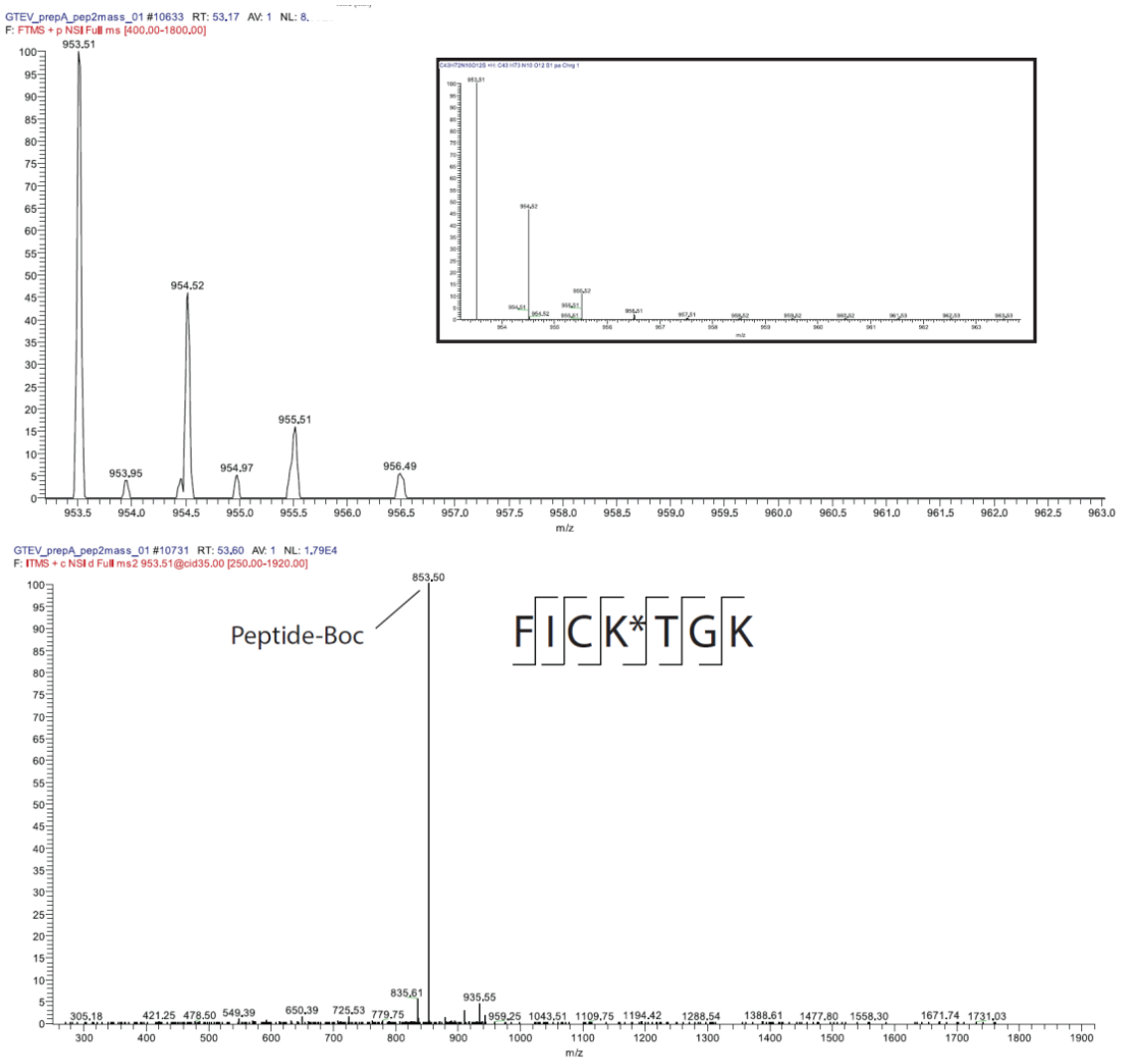


Figure 5-25. Mass and fragmentation of the peptide containing the BocK residue produced by the tryptic digestion of sfGFP-3-OMeY-51-BocK-151-5HTP (Lys5Ala). The fragmentation of the parent peak (m/z=953.5; top panel) produces neutral loss of the Boc group to give a peak at m/z=853.5 (bottom panel).

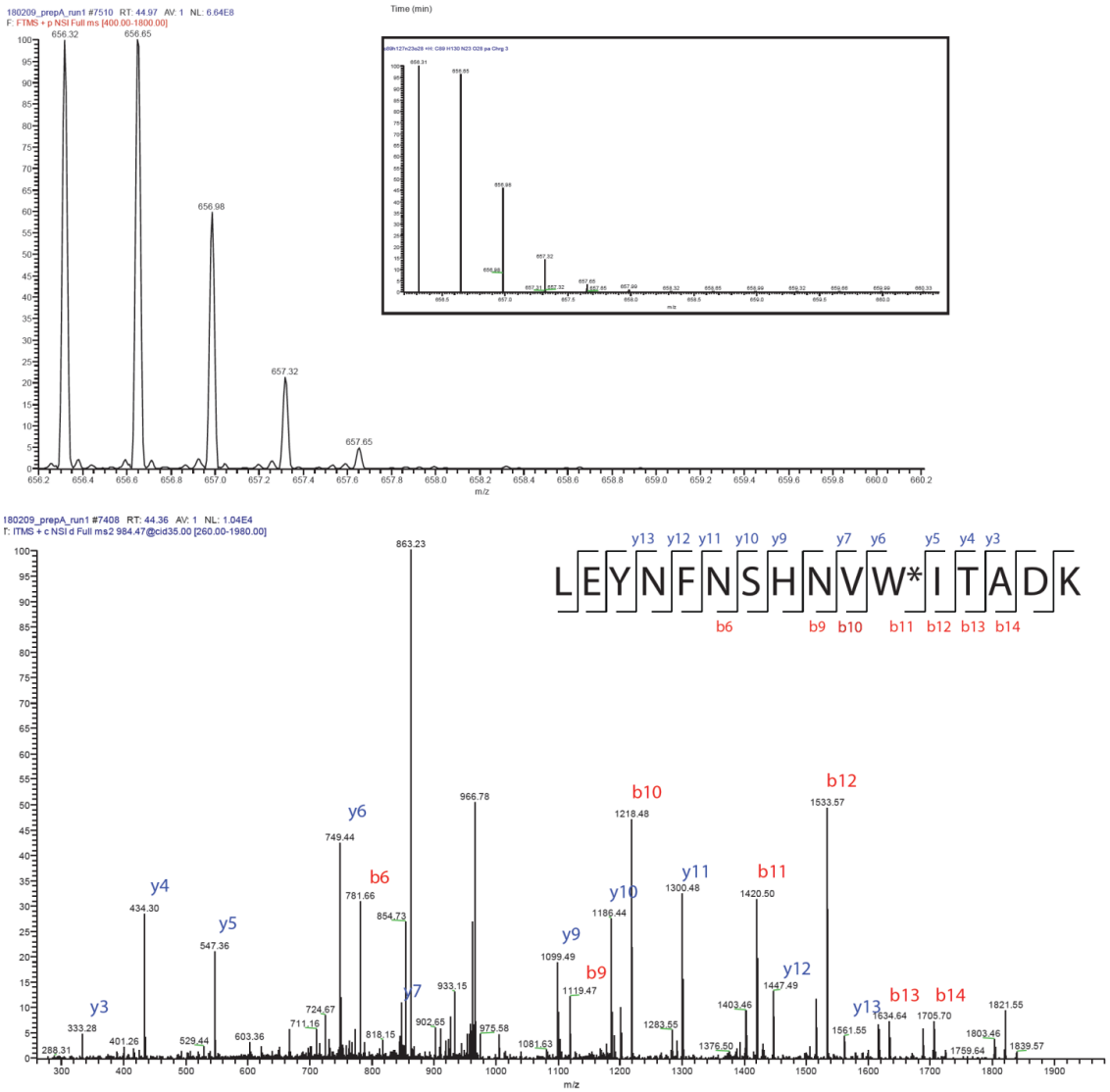


Figure 5-26. Mass and fragmentation of the peptide containing the 5HTP residue produced by the tryptic digestion of sfGFP-3-OMeY-51-BocK-151-5HTP (Lys5Ala).

5.3 Conclusions

Site-specific incorporation of ncAAs with bioconjugation handles provides a facile route to covalently attach a diverse variety of molecules onto proteins with remarkable precision. Extending this strategy to enable protein labeling at multiple chosen sites poses a multifaceted challenge that demand the development of both mutually compatible conjugation chemistries, and new routes for their site-directed incorporation into proteins. It has been possible to incorporate two different bioconjugation-ncAAs into one protein expressed in *E. coli* using MjTyr and Pyl pairs, and double label the resulting protein.[29–33] Mutually compatible SPAAC and SPIEDAC chemistries have emerged as a facile strategy for such dual labeling under ambient catalyst-free conditions. Here we expand this toolbox by first demonstrating that our recently developed 5HTP-directed protein conjugation strategy CRACR is compatible with both SPAAC and SPIEDAC, owing to its fundamentally different labeling chemistry (Figure 5-2 and Figure 5-3). We further show that the *E. coli* derived EcTrp pair, which we established as an orthogonal UGA-suppressor in the engineered *E. coli* strain ATMW1, can interface with MjTyr and Pyl pairs in a mutually orthogonal fashion to provide powerful new routes for incorporating multiple different bioconjugation-ncAAs into the same protein. We have previously demonstrated that the EcTrp pair can also be used in eukaryotic cells for site-specific incorporation of ncAAs in response to UAG o r UGA codons.[42] Consequently, it can be potentially combined with other orthogonal pairs in eukaryotic cells to site-specifically co-incorporate 5HTP and another bioconjugation-ncAA (such as CpK using Pyl, or pAzF using the *E. coli* derived tyrosyl pair),[28,48] enabling new routes for double labeling recombinant eukaryotic proteins – a direction that we are actively pursuing.

We also report the first example of site-specific incorporation of three distinct ncAAs into one protein expressed in a living cell. Development of EcTrp as a new orthogonal

suppressor in the ATMW1 strain and its mutual orthogonality with MjTyr and Pyl pairs set the stage for their simultaneous use for triple ncAA incorporation. However, to achieve this, two key challenges needed addressing: 1) assigning these pairs to three distinct "blank" codons, and 2) developing a plasmid system that co-expresses all three orthogonal pairs. MjTyr and Pyl have been previously together in *E. coli* to suppress UAG and UAA, respectively.[29,32] It appeared logical to combine the UGA-suppressing EcTrp pair with this dual-suppression system to build the first triple-suppression platform. To facilitate efficient termination of translation in an expression system where all three nonsense codons are reassigned, we took advantage of multiple consecutive stop codons. The UAA nonsense codon was used for this purpose, as the PylRS/$tRNA^{Pyl}_{UUA}$ pair was the least efficient among the three different suppressors. To get around partial undesired ncAA incorporation at the C-terminal UAA codons, we also created a novel self-cleaving tag that leaves a clean C-terminus on the target protein. We also developed a plasmid system that co-expresses all three suppressor pairs to facilitate triple ncAA incorporation, without causing significant toxicity. Although the efficiency of incorporating three distinct ncAAs is not high (2-3 mg/L; ~2% of wild-type reporter), it can be improved by further optimization of this first-generation platform. The specific MjTyr, Pyl and the EcTrp pairs used here can be used to incorporate a variety of additional ncAAs with useful properties,[18,19,22] further expanding the scope of the double and triple suppression systems described here.

Our work also highlight the need for carefully investigating multiple ncAA incorporation systems for any underlying cross-reactivity that can compromise its fidelity. Nearly all dual-ncAA incorporation systems developed to date have taken advantage of the unique Pyl pair, which does not cross react with other pairs owing to its distinctive structure. However, more caution must be taken when venturing beyond Pyl for creating new multi-

ncAA incorporation systems. In addition to this work, we have also previously shown that in mammalian cells, an *E. coli* derive tyrosyl-tRNA synthetase can charge an *E. coli* derived leucyl-$tRNA_{CUA}$,[28] which was unexpected given both pairs were derived from the same species. Here, we were able to overcome the mild cross-reactivity of EcTrpRS toward $tRNA^{MjTyr}_{CUA}$ by simply reducing the expression level of EcTrpRS. This also underscores the importance of fine-tuning the expression levels of various components in complex multi-suppression systems, akin to the fine balance of endogenous aaRS/tRNA pairs that underlie the high fidelity of natural systems.

In this work, we have simultaneously reassigned all three nonsense codons and have thus reached the limit of the maximum number of distinct ncAAs that can be incorporated within the framework of the canonical triplet genetic code. However, exciting progress is being made to overcome this barrier and significantly expand the coding capacity for ncAAs in *E. coli*. For example, efforts are under way to reduce the number of triplet codons *E. coli* uses by global engineering of its genome.[55,56] Such efforts can liberate additional triplet codons that can be simultaneously reassigned to encode more ncAAs. In an alternative approach, the Romesberg group have demonstrated the feasibility of using unnatural base pairs to encode ncAAs in *E. coli*.[57,58] Use of an unnatural base pair can dramatically increase the number of available triplet codons that can be reassigned for ncAA incorporation. However, as these approaches overcome the limited selection of codons that are currently available for reassignment to ncAAs, an arsenal of mutually orthogonal ncAA incorporation systems would be needed to facilitate their simultaneous use to build a dramatically expanded genetic code. Our work offers a primer on how to approach this complex endeavor, reveals some of the underlying challenges, and provides novel strategies to overcome them.

5.4 Experimental procedures

5.4.1 General cloning methods

All cloning and plasmid propagation were performed in *E. coli* strain DH10b (Life Technologies). DNA was amplified using Phusion Hot Start II DNA polymerase (Fisher Scientific) and purified with spin columns (Epoch Life Science). DNA gblocks and oligos were purchased from Integrated DNA Technologies (IDT). Restriction enzymes for cloning and plasmid verification were purchased from New England Biolabs (NEB). Sanger equencing of DNA was performed by Eton Biosciences. All other materials were purchased through Sigma-Aldrich or Fisher Scientific unless otherwise noted.

5.4.2 Antibiotics, concentrations, and strains

95 μg/mL spectinomycin, 50 μg/mL chloramphenicol, 10 μg/mL gentamycin, 100 μg/mL ampicillin, 15 μg/mL zeocin, 12 μg/mL tetracycline, 30 μg/mL kanamycin were used to supplement LB or LB-Agar containing plasmids with the corresponding antibiotic resistance gene unless otherwise noted. ATMW1 was used for all bacterial protein expressions for fluorescence analysis or protein purification.

Statistical methods. Error bars represent standard deviation of three independent replicates expressed in *E. coli* or HEK293T cells, which has been a reliable interpretation of error in past work.

5.4.3 Analysis of sfGFP reporter expression in ATMW1 by fluorescence spectroscopy

All expression and protein production were performed in previously developed *E. coli* strain ATMW1. ATMW1 co-transformed with various pGTEV/pEvoltac reporter/suppression plasmid combinations were grown in a 1 mL overnight starter culture,

inoculated into 20 mL LB+antibiotics and grown at 37 °C, 250 rpm until reaching a 0.55 OD_{600}. Expression was induced with 1 mM IPTG and 1 mM ncAA (OMeY, HTP, BocK, CpK, 1.5 mM pAzF) and grown for 20 hr at 30 °C in 15 mL culture tubes in 1 mL aliquots. For fluorescence measurements, cells were pelleted at 5000 x g, resuspended in PBS, diluted 10-fold, and transferred to a 96-well clear-bottom assay plate. Fluorescence was measured by using a SpectraMAX M5 (Molecular Devices) (Ex. 488nm; Em. 534nm) and normalized by OD_{600}.

Single suppression of sfGFP reporter in ATMW1: Our newly developed pGTEV reporter plasmid (described in Plasmid Construction Section) was transformed into ATMW1 with a suppressor pEvoltac plasmid. Our GTEV construct contains an sfGFP-10xHisTag-TEVcleavage-TEVprotease fusion protein expressed from a *T5-lac* promoter. Expression of the TEV fusion results in the autocleavage of the TEV protease *in vivo*, which allows expression of a protein containing defined C-terminus when all three stop codons are used to direct ncAA incorporation. The sfGFP portion of the fusion protein contains one UAG, UAA, or UGA stop codon as described in each experiment. For tyrosine, pyrrolysine, and tryptophan based ncAA incorporation, pEvoltac MjY TAG, pEvoltac Pyl TAG or pEvoltac Pyl TAA, or pEvoltac EcW TGA were co-transformed with the respective pGTEV repoter (e.g., pGTEV-TAG with pEvoltac MjY TAG, etc) into ATMW1. Single stop codon suppression of the sfGFP reporter was performed as described above.

Double suppression of sfGFP reporter in ATMW1: ATMW1 was cotransformed with pGTEV-3TGA-151TAG and either pEvoltac EcW.h14-MbPylRS TAG, pEvoltac MjY.CNF-EcW.h14, or pEvolG EcW.h14-MjY.CNF(tac). GFP was expressed and analyzed as described above.

Triple suppression of sfGFP reporter in ATMW1: ATMW1 was cotransformed with pGTEV-3TAG-K5A-51TAA-151TGA-MbPylTAA (contains GTEV reporter and MbPyl tRNA/aaRS pair) and pEvolG EcW.h14-MjY.CNF(tac). sfGFP was expressed and analyzed as described above.

5.4.4 Reporter sfGFP protein purification and mass spectroscopy analysis

sfGFP was expressed with the same temperature, shaking conditions, and amino acid concentrations as described in the single, double, and triple expression analysis. In order to obtain more protein for LCMS analysis, larger culture volumes (20 mL or 1 L cultures in Erlenmeyer flasks) were used. The expression culture was centrifuged and resuspended in incubated in lysis buffer [B-PER Bacterial Protein Extraction Reagent (Thermo Scientific) + 1X Halt Protease Inhibitor Cocktail (Thermo Scientific) + 0.01% Pierce Universal Nuclease (Thermo Scientific)] for 30 minutes on ice. The resulting cell lysate was subsequently clarified at 22,000 x g for 10 min. The supernatant was removed and His-tagged GFP was purified on a HisPur Ni-NTA resin (Thermo Scientific) following manufacturer's protocols. To buffer exchange and concentrate protein after purification, protein was diluted 100 fold in PBS buffer and concentrated back using a 3 kDa MWCO spin column (Millipore). Protein concentration was determined using Bradford assay. All SDS-PAGE and blot images were recorded by a ChemiDoc MP imaging system (BioRad). HPLC-coupled ESI-MS analysis of intact proteins were performed using a 1260 Agilent Infinity Series HPLC/6230 Agilent TOF Mass Spectrometer.

5.4.5 Analysis of single labeling cross reactivity via fluorescent gel imaging

Three different sfGFP-151 mutants containing either 5HTP, pAzF, or CpK were expressed and purified as described above. All three sfGFP reporters were individually subjected to three bioconjugation reagents (details below), and analyzed by LCMS or SDS-PAGE/fluorescence imaging: a) 10 μM of protein was treated with 50 μM of 4CDZ for 30 min at room temperature and quenched with excess 5HTP (1 mM final concentration), b) 10 μM of protein was treated with 50 μM of DBCO-tamra for 3 h at room temperature and quenched with excess pAzF (1 mM final concentration), c) 10 μM of protein was treated with 200 μM of tetrazine-fluorescein for 3 h at room temperature.

5.4.6 Biotinylation and anti-biotin western blot of sfGFP-151 variants

10 μM of sfGFP-151-5HTP, AzF, or CyK were used to test the cross reactivity of the diazo coupling reaction. Biotinylation of modified protein was carried out using a biotin-diazonium conjugate **2**. Biotin diazonium salt was generated by irradiation of compound **1** at 365 nm using 120 W LED-array (Larson Electronics) for 10 s. 50 μM of **2** was added to the individual solution of 10 μM sfGFP containing 151 HTP, AzF, or CyK in 0.1 M phosphate buffer (pH 7) and incubated for 30 min at room temperature. The reaction was quenched by the addition of excess HTP (1 mM). The protein was subsequently desalted by dilution with 0.1 M phosphate buffer and centrifugation using a concentrator (10 kDa molecular weight cutoff, Amicon Ultra-0.5 mL, centrifugal filters). The protein was resuspended in 100 mM phosphate buffer adjusting the concentration of the protein to be 5 μM.

Anti-biotin western blot: sfGFP mutants incubated with **2** were resolved using a 12 % SDS-PAGE gels and transferred onto a PVDF membrane (Life Technologies) using a Trans-Blot Turbo Transfer System 15 (Bio-Rad). After transferring, the membrane was incubated

overnight in BSA in PBS (Thermo scientific) blocking buffer at 4 °C with gentle agitation. After overnight incubation, the blocking solution was removed and the membrane was incubated with Streptavidin-HRP (Pierce; at 1:2500 dilution) for 2 hr in fresh blocking solution with gentle shaking at room temperature. The membrane was washed 7 times (10 min incubation with agitation) with wash solution (0.1% Tween 20 in TBS). Then, the membrane was developed for 2 min using SuperSignal West Dura Kit (Fisher Scientific) and the signal was detected by the ChemiDoc MP imaging system (BioRad).

5.4.7 Labeling parameters for double and triple suppressed protein

Labeling of double and triple suppressed protein was performed in one pot under the concentrations and conditions as described in the single labeling cross reactivity tests. All proteins before and after labeling were analyzed by LCMS.

Double suppression labeling: Using 10 μM sfGFP-3-pAzF-151-5HTP, the 5HTP residue was first labeled with 50 μM 4CDz for 30 min at RT and quenched with 1 mM free 5HTP. pAzF was subsequently labeled in the same pot using 200 μM DBCO-tamra for 3 h at RT and quenched with 1 mM free pAzF. sfGFP-3-5HTP-151-CyK was first labeled with 50 μM 4CDZ for 30 min at RT and quenched with 1 mM free 5HTP, followed by CpK labeling using 200 μM tetrazine-fluorescine for 3 hr at RT. Proteins were directly analyzed by LCMS with no further processing.

Triple suppression labeling: Site specific single labeling of triple suppressed protein was first verified using 10 μM of the unlabeled triple suppressed protein sfGFP-3-pAzF-51-CpK-151-5HTP. The unlabeled protein was subjected to labeling by solely 4CDZ, DBCO-tamra, or tetrazine-fluorescine with the aforementioned conditions to reaffirm single-site labeling of the triply suppressed protein. Once single labeling of the triple suppressed protein was

confirmed by LCMS, the reporter protein was triply labeled at RT first with 50 μM 4CDZ for 30 min, followed by 50 μM DBCO-tamra for 3 h, and lastly 200 μM tetrazine-fluorescine for 3 h in a one pot reaction. This triply labeled mass was analyzed by whole protein LCMS and tryptic MS without further processing.

5.4.8 Tryptic mass spec analysis of triple suppressed protein

Purified sfGFP-3-OMeY-51-BocK-151-5HTP(Lys5Ala) protein (100 μg) was precipitated with TCA, and the protein pellet was washed with ice-cold acetone. The protein was resuspended in 15 μL 8M urea in PBS and 20 μL 0.2% ProteaseMax Surfactant (Promega). The protein solution was diluted with 65 μL 100 mM ammonium bicarbonate. The protein was reduced with the addition of 1.5 μL of 1M DTT, with heating at 65°C for 15 minutes. The protein was then alkylated with the addition of 2.5 μL 500 mM iodoacetamide, with room temperature incubation for 45 minutes. The protein was diluted with 120 μL PBS and digested overnight at 37 °C with the addition of 4 μL of 0.5 μg/μL sequencing-grade trypsin (Promega), and 2.5 μL 100 mM $CaCl_2$. Peptides were acidified with 10 μL formic acid and centrifuged at 15,000 rpm for 20 min. The supernatant was then analyzed by LC-MS/MS. Peptides were analyzed on an LTQ Orbitrap XL mass spectrometer (Thermo Fisher) coupled to an EASY-nLC 1000 nanoLC (Thermo Fisher). 10 μL of sample was loaded onto 100 μm fused silica column with a 5 μm tip packed with 10 cm of Aqua C18 reverse-phase resin (Phenomenex) using the EASY-nLC 1000 autosampler. Peptides were eluted with a gradient 0–55% buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B; 20% water, 80% acetonitrile, 0.1% formic acid). The flow rate through the column was set to 400 nl/min and the spray voltage was set to 3.5 kV. One full MS scan (400–1800 MW) was followed by seven

data dependent scans of the nth most intense ions with dynamic exclusion enabled. MS1 and MS2 spectra were manually analyzed.

5.4.9 Antibody double suppression and purification

pBK-aHer2-K169TGA-S202TAG was co-transformed with pEvolG-EcW.h14-MjY-CNF into ATMW1. The cells were grown in LB supplemented with 35 μg/mL chloramphenicol, 10 μg/mL kanamycin, and 100 μg/mL spectinomycin. Upon reaching an OD_{600} of 0.6, cells were induced with 0.02% arabinose, 1 mM IPTG, 1 mM pAzF, and 1 mM HTP (Chem Impex Int'l. Inc). Cells were grown for 16 h at room temperature. Cells were harvested using a periplasmic-lysis buffer (20% sucrose, 30 mM Tris, pH8, 1mM EDTA, 0.2 mg/mL lysosyme, and Halt protease inhibitor) for 30 min at 37 °C. Lysate was then diluted 1:1 with binding buffer (50 mM NaOAc, pH 5.2) and clarified by centrifugation at 1700 rpm for 30 min. Herceptin-FAB was then purified using Pierce Protein G agarose according to the manufacturer's instructions. After purification, protein was buffer exchanged into 100 mM phosphate buffer, pH 7.

5.4.10 Antibody double labeling and SDS-PAGE fluorescence analysis

Wild-type or double suppressed antibody was labeled with 200 μM diazo-fluorescein for 20 min on ice and the reaction was quenched with 1 mM HTP. 100 μM DBCO-tamra was then added and the reaction proceeded for 30 min at room temperature, followed by quenching with 10 mM azido-lysine. Mixture was dialyzed overnight at 4 °C into PBS. Concentration after dialysis was measured by Bradford assay, and equal amounts of each protein was resolved using a 15% SDS-PAGE gel.

5.5 References

1. Agarwal, P. & Bertozzi, C. R. Site-specific antibody-drug conjugates: The nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug. Chem.* **26,** 176–192 (2015).
2. Bloom, S. *et al.* Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **10,** 205–211 (2018).
3. Boutureira, O. & Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **115,** 2174–2195 (2015).
4. DeGruyter, J. N., Malins, L. R. & Baran, P. S. Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* **56,** 3863–3873 (2017).
5. Koniev, O. & Wagner, A. Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev.* **44,** 5495–5551 (2015).
6. Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science (80-.).* **355,** 597–602 (2017).
7. Shih, H. W., Kamber, D. N. & Prescher, J. A. Building better bioorthogonal reactions. *Curr. Opin. Chem. Biol.* **21,** 103–111 (2014).
8. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chemie - Int. Ed.* **48,** 6974–6998 (2009).
9. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5,** 1–14 (2014).
10. Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **7,** 876–884 (2011).
11. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21,** 86–89 (2003).
12. Gautier, A. *et al.* An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **15,** 128–136 (2008).
13. Martin, B. R., Giepmans, B. N. G., Adams, S. R. & Tsien, R. Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat. Biotechnol.* **23,** 1308–1314 (2005).
14. Fernández-Suárez, M. *et al.* Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **25,** 1483–1487 (2007).
15. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2,** 99–104 (2005).
16. Zhang, C. *et al.* π-Clamp-mediated cysteine conjugation. *Nat. Chem.* **8,** 120–128 (2016).
17. McKay, C. S. & Finn, M. G. Click chemistry in complex mixtures: Bioorthogonal bioconjugation. *Chem. Biol.* **21,** 1075–1101 (2014).
18. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550,** 53–60 (2017).
19. Dumas, A., Lercher, L., Spicer, C. & Davis, B. G. This journal is © The Royal Society of Chemistry 2014 Designing logical codon reassignment – expanding the chemistry

in biology. 1–36 (2014). doi:10.1126/scienc
20. Italia, J. S. *et al.* Expanding the genetic code of mammalian cells. *Biochem. Soc. Trans.* **45,** 555–562 (2017).
21. Mukai, T., Lajoie, M. J., Englert, M. & Söll, D. Rewriting the Genetic Code. *Annu. Rev. Microbiol.* **71,** annurev-micro-090816-093247 (2017).
22. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13,** 854–870 (2018).
23. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal Labeling of Proteins. *Chem. Rev.* **114,** 4764–4806 (2014).
24. Lang, K. & Chin, J. W. Bioorthogonal reactions for labeling proteins. *ACS Chem. Biol.* **9,** 16–20 (2014).
25. Vargas-Rodriguez, O., Sevostyanova, A., Söll, D. & Crnković, A. Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Curr. Opin. Chem. Biol.* **46,** 115–122 (2018).
26. Kim, C. H., Axup, J. Y. & Schultz, P. G. Protein conjugation with genetically encoded unnatural amino acids. *Curr. Opin. Chem. Biol.* **17,** 412–419 (2013).
27. Sun, S. B., Schultz, P. G. & Kim, C. H. Therapeutic applications of an expanded genetic code. *ChemBioChem* **15,** 1721–1729 (2014).
28. Zheng, Y., Addy, P. S., Mukherjee, R. & Chatterjee, A. Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells. *Chem. Sci.* **8,** 7211–7217 (2017).
29. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli. *Biochemistry* **52,** 1828–1837 (2013).
30. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464,** 441–444 (2010).
31. Sachdeva, A., Wang, K., Elliott, T. & Chin, J. W. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *J. Am. Chem. Soc.* **136,** 7785–7788 (2014).
32. Wan, W. *et al.* A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew. Chemie - Int. Ed.* **49,** 3211–3214 (2010).
33. Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* **6,** 393–403 (2014).
34. Zheng, Y., Gilgenast, M. J., Hauc, S. & Chatterjee, A. Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis. *ACS Chem. Biol.* **13,** 1137–1141 (2018).
35. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126,** 15046–15047 (2004).
36. Jewett, J. C. & Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical

biology. *Chem. Soc. Rev.* **39,** 1272–1279 (2010).
37. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine Ligation : Fast Bioconjugation Based on Tetrazine Ligation : Fast Bioconjugation Based on Inverse-Electron-Demand. *J. Am. Chem. Soc.* **130,** 13518–13519 (2008).
38. Devaraj, N. K. & Weissleder, R. Biomedical applications of tetrazine cycloadditions. *Acc. Chem. Res.* **44,** 816–827 (2011).
39. Kamber, D. N. *et al.* 1,2,4-Triazines Are Versatile Bioorthogonal Reagents. *J. Am. Chem. Soc.* **137,** 8388–8391 (2015).
40. Lang, K. *et al.* Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic diels-alder reactions. *J. Am. Chem. Soc.* **134,** 10317–10320 (2012).
41. Seitchik, J. L. *et al.* Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. *J. Am. Chem. Soc.* **134,** 2898–2901 (2012).
42. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat. Chem. Biol.* 1–8 (2017). doi:10.1038/nchembio.2312
43. Addy, P. S., Erickson, S. B., Italia, J. S. & Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **139,** 11670–11673 (2017).
44. Addy, P. S., Italia, J. S. & Chatterjee, A. An oxidative bioconjugation strategy targeted to a genetically encoded 5-hydroxytryptophan. *ChemBioChem* **19,** 1375–1378 (2018).
45. Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **124,** 9026–9027 (2002).
46. Young, D. D. *et al.* An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50,** 1894–1900 (2011).
47. Elliott, T. S. *et al.* Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat. Biotechnol.* **32,** 465–472 (2014).
48. 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* **57,** 441–445 (2018).
49. Xiao, H. *et al.* Genetic incorporation of multiple unnatural amino acids into proteins in Mammalian cells. *Angew. Chem. Int. Ed. Engl.* **52,** 14080–14083 (2013).
50. Swanson, R. *et al.* Accuracy of *in vivo* aminoacylation reguires proper balance of tRNA and aminoacyl-tRNA synthetase. *Science (80-.).* **242,** 1548–1551 (1988).
51. Beck, A., Goetsch, L., Dumontet, C. & Corvaïa, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat. Rev. Drug Discov.* **16,** 315–337 (2017).
52. Hudis, C. A. Trastuzumab — Mechanism of Action and Use in Clinical Practice. *N. Engl. J. Med.* **357,** 39–51 (2007).
53. Hutchins, B. M. *et al.* Site-specific coupling and sterically controlled formation of multimeric antibody Fab fragments with unnatural amino acids. *J. Mol. Biol.* **406,** 595–

603 (2011).
54. Zheng, Y. *et al.* Molecular BioSystems Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1 †. *Mol. Biosyst.* (2016). doi:10.1039/C6MB00070C
55. Ostrov, N. *et al.* Design, synthesis, and testing toward a 57-codon genome. *Science (80-.).* **353,** 819–822 (2016).
56. Wang, K. *et al.* Defining synonymous codon compression schemes by genome recoding. *Nature* **539,** 59–64 (2016).
57. Dien, V. T., Morris, S. E., Karadeema, R. J. & Romesberg, F. E. Expansion of the genetic code via expansion of the genetic alphabet. *Curr. Opin. Chem. Biol.* **46,** 196–202 (2018).
58. Zhang, Y. *et al.* A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551,** 644–647 (2017).

Chapter 6

Collection of unpublished short stories

Acknowledgements

Chester Wrobel, now in Cornell's PhD program, had a significant role in developing the ATM-BL21 strain and the preliminary expression data. He also developed the NitroW libraries, amongst many other contributions to the team as a whole.

Christen Hillenbrand, who will be a senior at BC at the time of publication, completely optimized the malachite green assay for analysis of EcYRS variants. She had a significant contribution to section 6.4 and also has made large contributions to the ATM team as a whole.

Elise Ficaretta, a first year graduate student, has begun to take over the ATMW-BL21 project and some of her fluorescence experiments were used in section 6-2.

Many of the common methods and protocols used in this chapter from the Chatterjee lab have been detailed extensively in previous chapters and publications. Therefore, only new protocols and signficiant modifications to existing protocols have been mentioned in Chapter 6. Much of this work is still preliminary, but we believe the results are useful for analysis and direction for future efforts.

6.1 Applications of the ATM system for biocontainment and proteome-wide incorporation of UAAs

6.1.1 Brief Overview

While development of new tRNA/aaRS pairs for the incorporation of novel chemical functionalities has been the major goal of the ATM systems,[1,2] we wanted to explore other potential applications of this technology. Two distinct areas in *E. coli*-based synthetic biology have recently focused on biocontainment[3,4] and proteome-wide incorporation of UAAs.[5–8] We sought to address these two topical issues with the ATM tryptophan platform. While this preliminary work was carried out in bacteria, it is possible that these two preliminary applications could eventually be expanded from bacteria to eukaryotic organisms.

Genetically modified organisms (GMO) are becoming increasingly popular from a therapeutic, consumption, and agricultural standpoint.[3,4] While GMOs can be beneficial, their use can risk the homeostasis of human health or the ecosystem and their application has therefore been limited to reduce the potential dangers. One such way to increase the widespread use of safe GMOs is to introduce biocontainment strategies in order to circumvent escape mechanisms: mutagenic drift, environmental supplementation, and horizontal gene transfer.[3] Many biocontainment mechanisms are being developed, but some of the more popular methods can vary between construction of kill-switch genetic circuits,[4] organism modification with survival dependent on proteins engineered to be functional only in the presence of UAA,[3] or cellular survival linked to a specific metabolite.[9]

We previously showed there is a low level of background charging of tryptophan by EcTrpRS hits -9 and -14 in the absence of UAA and a complete selective "switch" of specificity in the presence of UAA.[1] We envisioned deleting the *E. coli* TrpRSwt from the genome using the *tolC* recombination system[10] using the *E. coli* TrpRS –h9 or –h14 as the

complementary pair and leaving the endogenous *E. coli* $tRNA^{Trp}_{CCA}$ intact in the genome (rather than the *Saccharomyces cerevisiae* TrpRS). This strain would hypothetically be able to survive in the absence of UAA due to the low basal incorporation of tryptophan by the mutant 5HTP synthetases. However, in the presence of UAA, the specificity would shift from tryptophan to an UAA and result in proteome-wide incorporation of the UAA and the death of the genetically modified organism due to perturbation of protein function. We serendipitously discovered that 5HTP is commercially available as an over the counter therapeutic and is also naturally produced in the body (at low levels). It is unclear how the physiological levels of this type of amino acid would affect our new strains. This type of biocontainment strategy could be used in an application such as a probiotic or bacterially delivered therapeutic whose survival can be modulated with the addition of a benign UAA, 5HTP.

The new ATM-Trp strain would enable the ability to proteomically incorporate the UAA of interest at low levels, avoiding a toxic phenotype while allowing for the changes in protein expression on a global level. While GMO biocontainment is one exciting avenue of this new strain, this strain could also be developed for proteomic applications. There has been significant research in using selective pressure incorporation[7,8] via auxotrophic strains or ribosome engineering[6] to globally incorporate an UAA of interest. These methodologies can be used to study the effects of an amino acid analog on global protein function,[8] changes in protein levels as cells undergo stimuli (apoptosis, development, etc),[6] or even create an "inventory" of photocrosslinked proteins.[5] The development of the new bioconjuction strategy CRACR[13] as well as existing SPAAC opens new possibilities for the global replacement of tryptophan in the proteome with 5HTP and/or azide containing UAAs. Development and characterization of this new ATM-Trp strain would explore the exciting avenues of biocontainment and proteome studies.

6.1.2 Results and Discussion

We first investigated the potential of cell killing by transforming pBK-EcWRS–h14 (Figure 6-1a), which contains the EcWRS-h14 under the constitutively active glnS' promoter, into DH10b, a standard *E. coli* lab strain which contains the wild-type EcWRS and $EcWtR_{CCA}$. After 1 hr incubation of the transformation at 37°C, 100 µL of the transformation ($1/10^{th}$ competent cell, exact cell # unknown) was plated on LB agar plates with or without 1 mM 5HTP (Figure 6-1b). The striking difference in colony counts between the 1 mM 5HTP and no 5HTP plates shows that 5HTP is toxic in the presence of EcWRS-h14 and corresponding TGG-suppressing tRNA. This result suggested the synthetase can disrupt cell survival as well while incorporating 5HTP, even though optimization of the lethality of this system needed to be improved. Since these results are preliminary, it is possible this low level of escape frequency in the 1 mM 5HTP plate is due to contamination, which was an issue we saw at the time of these experiments resolved by sterile filtration.

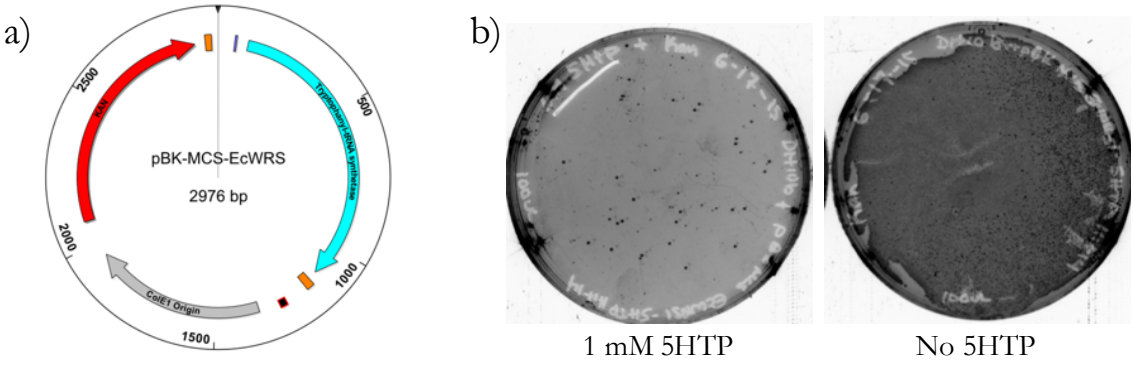


Figure 6-1. DH10b cell death due to proteomic 5HTP incorporation. a) pBK plasmid map. b) DH10b transformation plated +/- 5HTP.

We decided to build an EcWRSwt deleted strain, ATMW2, in order to reduce the presence of background tryptophan charging to the endogenous EcTrp-tRNA. ATMW2 would contain a plasmid harboring EcWRS-h9 or –h14, rather than the pUltraG ScW40 as employed in ATMW1.[1] This pBK plasmid does not contain a TGA-suppressing $tRNA^{Trp}$, and we hoped the resulting strain would be able to survive due to background tryptophan charging in the absence of UAA.

This recombination strategy was made easier by the existence of a precursor strain previously developed in the lab named EcNR1GT (EcNR1 ⊿*GalK* ⊿*tolC*), which allows for the use of *galK and tolC* as selectable markers for λ-Red recombineering. We employed the use of λ-Red recombineering (see Chapter 2 Experimental Procedures)[1,12] to replace *trpS* with *tolC*. The *tolC* cassette was amplified with two oligos containing the same homologous regions as the zeocin cassette in ATMW1 and transformed to remove *trpS* (Figure 6-2a).[1] Recombinants containing either –h9 or –h14 transformed with or without (control) the *trpS::tolC* cassette were plated on LB + 0.005% SDS to test for the presence of *tolC* (Figure 6-2b), as only recombinants containing *tolC* should survive. While the difference between the plates transformed with the *trpS::tolC* cassette versus the control plate (no DNA) is minimal, we decided to proceed with the colony PCR (Figure 6-2c). The colony PCR confirmed the deletion of *trpS* (Figure 6-2c), and further preliminary phenotypic screening demonstrated improved cell killing when plated in the presence of 5HTP (Figure 6-2d). The best hits for –h9 and –h14 were named JSI-W9 (EcNR1 pBK-EcWRS-h9 ⊿*GalK* ⊿*tolC trpS::tolC*) and JSI-W10 (EcNR1 pBK-EcWRS-h14 ⊿*GalK* ⊿*tolC trpS::tolC*), respectively. The λ-RED genes were removed later with their replacement by a Zeocin resistance cassette to create ATMW2.h14 (EcNR1 ⊿*GalK* ⊿*tolC trpS::tolC* λ-RED::Zeo^R), but follow up biocontainment experiments were performed with JSI-W9 or JSI-W10. (Data not shown)

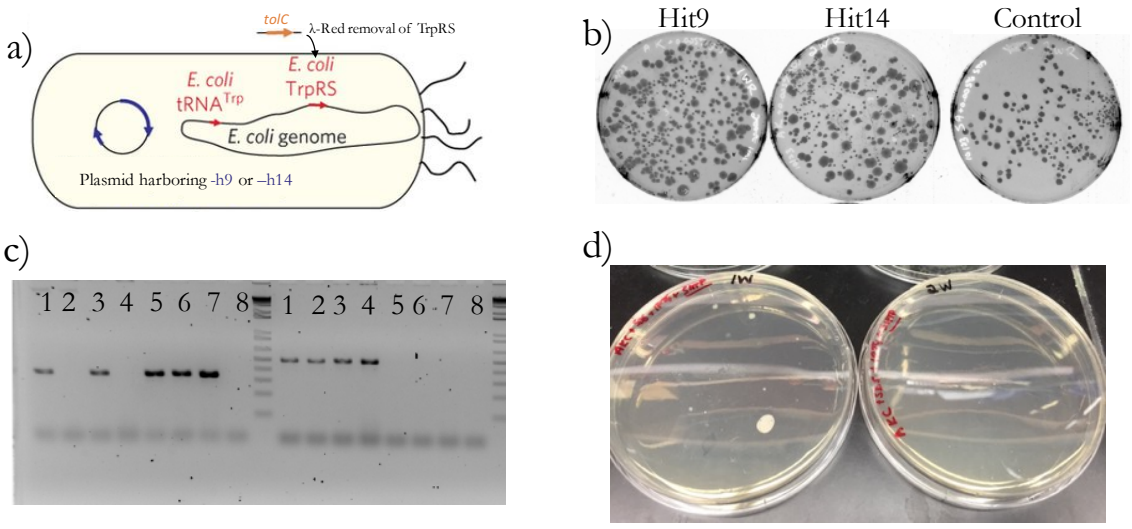


Figure 6-2. ATMW2 strain construction. a) TrpRS removal scheme. b) TrpRS recombination. –h9 and –h14 plates contain the respective plasmid and were transformed with the *trpS::tolC* cassette. The control plate contains no DNA and should therefore be empty. c) Colony PCR of *trpS* locus. Lanes 1-4 are hits, Lane 5-7 controls, Lane 8 PBS. (Left) band indicates trpS present at trpS location, (Right) band indicates *tolC* present at *trpS* location. d) Successful hits 2 and 4 were plated on LB + 5HTP to test for cell killing.

With JSI-W9 and JSI-W10 in hand, we wanted to investigate the amount of UAA necessary to induce cell death. $1x10^9$ cells were plated on LB + agar plates containing a gradient of either 5HTP or 5MTP (Figure 6-3). We found that in the best case system (JSI-W10, contains h14), cell death was achieved for all $1x10^9$ cells in either 1 mM or 3 mM UAA. The improved cell killing may be due to the lower activity of -h14 relative to -h9 and decreased background charging of tryptophan in the absence of UAA. Our preliminary data does not indicate the minimum amount of UAA necessary to eliminate escape frequency, but this would be an area of interest for future work. Regardless, this data suggests the constitutive expression of these mutants allows for the survival in the absence of UAA and $>1x10^{-9}$ frequency of cell

killing in the presence of 1 mM UAA. Because 5HTP is known to be safe for human consumption, this functions as an innocuous killswitch which could be used to reduce the dangers of GMOs in our society.

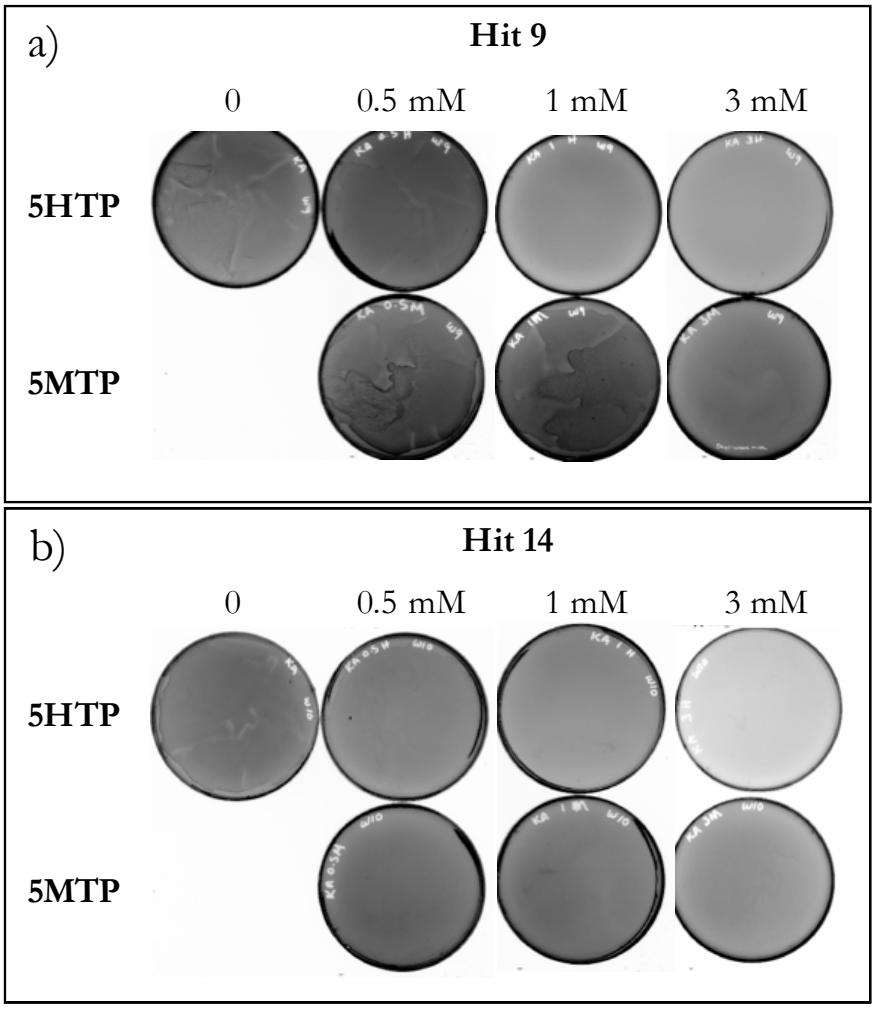


Figure 6-3. Biocontainment test with JSI-W9 and JSI-W10. a) JSI-W9 complemented with –h9 or b) JSI-W10 complemented with –h14 is plated (~$1x10^9$ cells) on a gradient of either 5HTP or 5MTP. a) For JSI-W9, we see strong growth on 0.5 mM UAA, and a haze on 1 & 3 mM 5MTP. There is no growth on >1 mM HTP. b) For –h14, we only see a haze of bacterial growth on the 0.5 mM plates. (Description due to the image quality).

While ATMW2.h14 (JSI-W10 without prophage) is effective as a killswitch/biocontainment GMO, we have also mentioned the strain's potential to be used for proteomic studies. For proof of concept, we wanted to demonstrate the potential of global replacement of TGG codons (tryptophan) with UAA analogs, similarly to Budisa *et. al.*[8] To do this, we used the plasimd pEvolSPI-sfGFPwt, which contains an ITPG inducible sfGFPwt with one natural TGG codon, to explore the incorporation rates of tryptophan versus UAA. An ATMW2.h14 pEvolSPI-sfGFPwt starter culture was used to inoculate 15 mL LB (Kan/Chlor). At 0.5 OD_{600}, cells were induced with 1 mM IPTG and 1 mM UAA was added. Interestingly, the fluorescence in the presence of UAA were similar to the no UAA control, except for 5MTP (Figure 6-4a). We then purified the sfGFPwt (contains 6x-HisTag) via Ni-NTA purification and analyzed the protein by SDS-PAGE (Figure 6-4b) and LCMS (Figure 6-4c). While yields of sfGFP were proportionally similar to the no UAA control (96, 38, 74, 102 mg/L respectively), the mass did not correspond to the expected UAA-containing mass but instead the WT mass. It is possible that the presence of residual $tRNA^{EcTrp}_{CCA}$-Tryptophan is in such an abundance at 0.5 OD600 and is outcompeting the newly formed $tRNA^{EcTrp}_{CCA}$-UAA for incorporation by the ribosome. Further studies could be performed to investigate this proportion of different tRNA charging rates but we determined optimization of the induction variables (time of induction, time of UAA) could be beneficial.

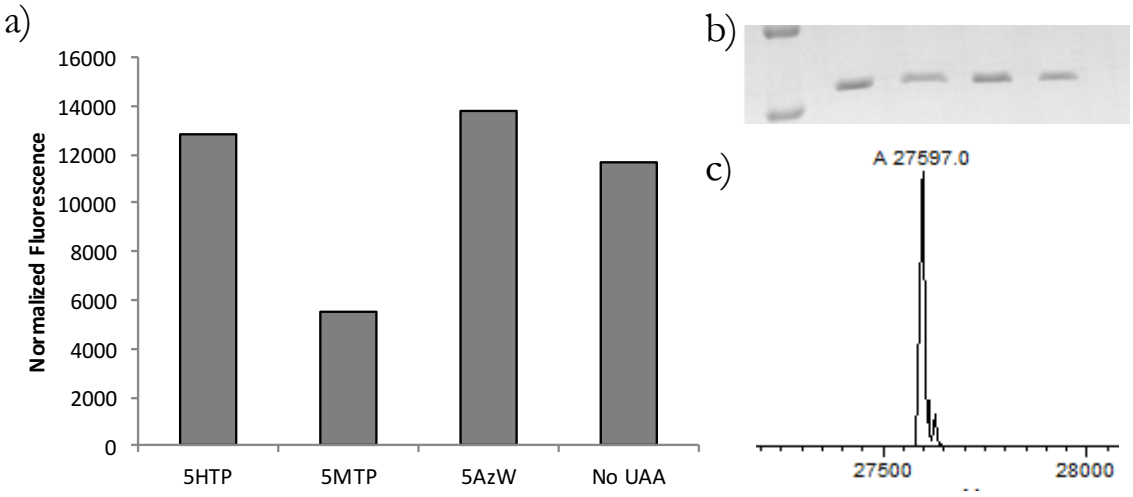


Figure 6-4. Selective pressure incorporation of sfGFPwt. a) Normalized fluorescence in the presence of UAA. b) SDS-PAGE of purified sfGFPwt. c) Representative LCMS of all purified sfGFPwt. All LCMS spectra display the same mass (Expected/Observed: 29597).

Two expression variables seemed like the most obvious areas for optimization: OD_{600} at induction and OD_{600} at UAA addition. We hypothesized that we could first add 5HTP without IPTG which would allow the cell to use up the $tRNA^{EcTrp}_{CCA}$-Tryptophan over an optimizeable period of time. When $tRNA^{EcTrp}_{CCA}$-Tryptophan is consumed, IPTG could then be added to induce the expression of sfGFP as a model protein, as the proportion of $tRNA^{EcTrp}_{CCA}$-UAA to $tRNA^{EcTrp}_{CCA}$-Tryptophan would be at its greatest level. Fortunately, only a minor amount of incorporation is needed. Additionally, previous work has suggested the –h14 synthetase can function at similar incorporation levels with 0.25 mM UAA, which would decrease the toxicity effects of global 5HTP incorporation (data not shown).

We first explored the optimal OD_{600} at which UAA should be added. Starter cultures of ATMW2.14 pEvolSPI-sfGFPwt were used to inoculate a 50 mL LB (Kan/Chlor) culture. The OD_{600} was measured every half hour and 1 mM or 5 mM 5HTP was added with 1 mM

IPTG (Figure 6-5a). Final growth OD_{600} were measured after ~19 hrs at 37°C. We hypothesized that the best UAA addition point would be at a starting OD_{600} which results in a proportionally low final OD_{600} (Figure 6-5a, 1mM and 5mM bars are final OD_{600}). If the culture grows much higher than the OD_{600} at induction, it would indicate there is a high residual amount of $tRNA^{EcTrp}_{CCA}$-Tryptophan resulting in cell viability for a longer period of time. If $tRNA^{EcTrp}_{CCA}$-UAA is the only tRNA species present, the cell should grow poorly. While it was difficult to identify the most optimal UAA addition point, addition at 0.181-0.200 OD_{600} only doubled after overnight growth while still maintaining a reasonable cell population for protein expression (Figure 6-5a) and a target OD_{600} of 0.2 was used for subsequent experiments. A more direct method of quantification could be pursued in the future.

We next investigated the optimal amount of time to wait for the depletion of $tRNA^{EcTrp}_{CCA}$-Tryptophan. We performed a similar experiment (Figure 6-5b, description in legend) to determine best time increment which would result in reasonable fluorescence while shifting the incorporated tryptophan analog from the endogenous tryptophan to the UAA of interest. IPTG addition between 1-1.5 hours resulted in minimal dropoff in protein expression levels (Figure 6-5b) and a near complete shift to the presence of 5HTP in the purified sfGFPwt after 1 hr (Figure 6-5c). We additionally noticed that increasing the concentration to 5 mM HTP (Figure 6-5b) or only 2 mM HTP (Figure 6-5d) results in significantly decreased protein expression, likely due to cell death. These data suggest the effective use of ATMW2.h14 as a selective pressure incorporation strain with optimized conditions to incorporate >90% 5HTP in many proteomic TGG codons.

In conclusion, the ATMW2.h14 strain has broad applications as a biocontainment mechanism for GMOs as well as a platform for the study of the bacterial proteome. It is exciting to think about expanding this technology toward eukaryotic systems with analogous

translational machinery. Future efforts should revolve around optimizing the bacterial system, demonstrating an application, and possibly expanding this type of technology to a eukaryotic system. Envisioning this type of GMO biocontainment would be exciting for something like CAR-T or other biological therapeutics. For SPI, the use of this system for the study of proteomic changes in response to stimuli seems like one of the more exciting avenues of research.

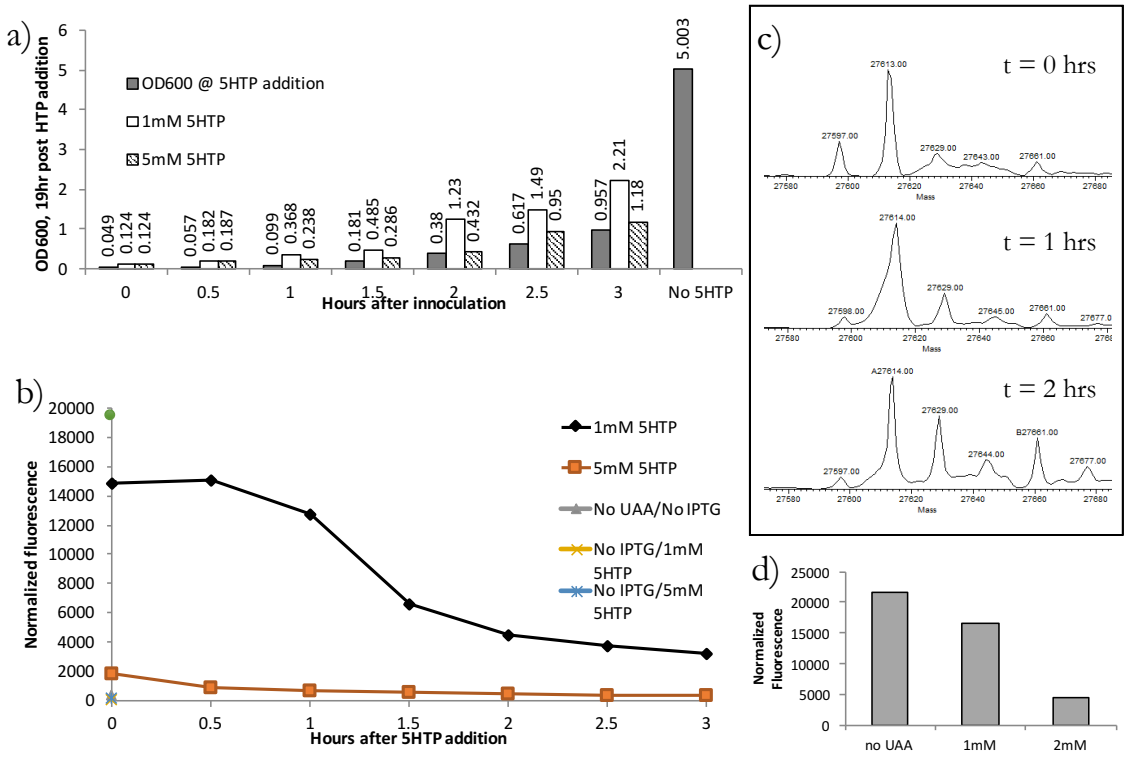


Figure 6-5. Selective pressure incorporation parameter optimization. a) 50 mL cultures were inoculated with 1:100 ratio of starter culture ATMW2.h14 pEvolSPI-sfGFPwt. OD_{600} was measured every half hour, aliquoted to 1 mL, and 1 or 5 mM 5HTP was added with 1 mM IPTG. Measurements are OD_{600} after ~19 hrs growth at 37°C unless otherwise stated. b) Normalized fluorescence of sfGFPwt expression. 25 mL cultures were inoculated with 1:100 ratio of start culture ATMW2.h14 pEvolSPI-sfGFPwt. They were grown at 37°C until 0.2 OD_{600} (~1.5 hrs) and 1 mM or 5 mM 5HTP was added. Cultures were returned to the shaker

and 1 mM 5HTP was added 0-3 hrs after at 30 min increments. After IPTG addition, cultures were transferred to 30°C overnight for ~19 hrs. c) LCMS analysis of purified sfGFPwt produced from 1 mM 5HTP cultures from Figure 6-5b at various time intervals. (Masses-sfGFPwt: 27597, sfGFP-5HTP: 27614). Multiple peaks are due to oxidation and the scale of the experiment. d) Repeat of Figure 6-5b at 1 hrs IPTG induction with 1 mM, 2 mM, and 5 mM 5HTP. 5 mM resulted in complete cell death, while 2 mM is visibily much less fluorescent.

6.2 BL21-ized ATMW strains

6.2.1 Brief Overview

A great question of bacterial based protein expression, now occasionally taken for granted, is "Which host should be used?" While there are many specialty strains of *E. coli*, one of the most commonly used strains is the K12 derived BL21 (DE3).[13,14] This strain is exceptionally efficient as an expression host, producing great yields of high quality protein. This advantage is due to significant genetic engineering, such as deletion of the Lon protease, OmpT outer membrane protein (involved in protein degradation), *hsd*SB mutation (prevents plasmid loss), and the insertion of the lambda prophage containing T7 RNAP under a lac inducible promoter (in the DE3 version of the strain).[13]

Over the last few years, we have investigated variations of the ATM-tryptophan[73] strain and their potential applications to the scientific community. While wild-type *E. coli* derived ATMW1 (Chapter 2) based protein expression was reasonable,[1] the strain lacks the safeguards that a BL21(DE3) strain can provide for more sensitive or low yielding protein expression. Additionally, ATMW1 is incapable of expressing protein from the commonly used T7 RNAP promoters. These lack of safeguards and compatibility with many common expression vectors caused our lab to take efforts to create a BL21-ized version of ATM-tryptophan.[1]

6.2.2 Results and Discussion

We decided to use λ-Red recombination to mimic the deletion of *E. coli* TrpRS/tRNA in the commonly used BL21(DE3) strain. Since BL21 does not contain a genomically incorporated set of the temperature inducible λ-Red genes like EcNR1,[1] we needed to pursue an alternative route of expression. Instead, we used plasmid pKD46, which contains the λ-Red genes under an arabinose inducible promoter.[15] pKD46 also contains a useful feature where growth at higher temperatures results in the strain's loss of the plasmid, thereby removing the λ-Red genes from the final strain after recombination is complete. As mentioned, we envisioned deleting the *E. coli* TrpRS and $tRNA^{Trp}$ from the genome by sequential λ-Red recombination with a Gentamycin resistance and Zeocin resistance cassette, respectively. Deletion of the *E. coli* translational machinery would be complemented with the same pUltra ScW40 as used previously.[1] pKD46, containing the λ-Red recombination genes, would be removed from the final strain by heat incubation (Figure 6-6).

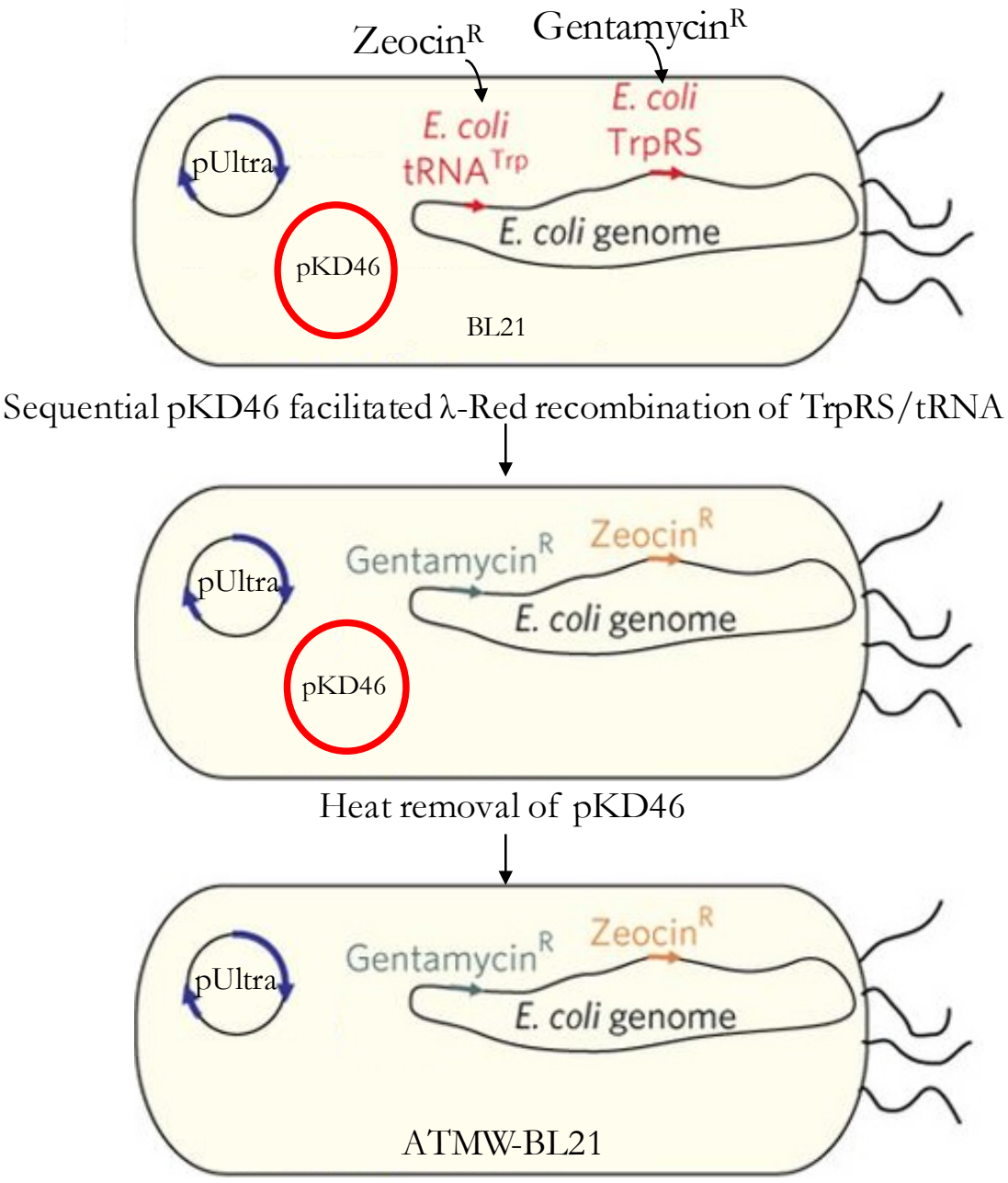


Figure 6-6. Scheme of pKD46 facilitated λ-Red recombination of BL21 to create ATMW-BL21.

While the protocol for recombination with pKD46 has been extensively studied, it did require some optimization by a great undergraduate Chester Wrobel. The protocol can hypothetically be scaled for larger *E. coli* preps as needed. First, the complementation plasmid (pUltra ScW40) and pKD46 were transformed into BL21. Due to temperature sensitivity, *E. coli* containing pKD46 must be grown at 30°C to maintain the plasmid. Overnight starter cultures were used to inoculate cultures at a 1:100 ratio in 2 mL LB + antibiotics. At 0.1 OD_{600}, 10 mM final L-arabinose was added to induce pKD46 λ-Red expression (include a no arabinose control). Allow the cultures to grow at 30°C to 0.4 OD_{600}. Aliquot 1 mL of each culture into two 1.5 mL centrifuge tubes and chill in ice-water bath for 10 min. After bath, centrifuge for 10 min at 4000 x g and 4°C. Aspirate supernatant and resuspend pellets in 1 mL ice-cold sterile ddH_2O. Repeat wash, centrifuging for 10 min at 4000 x g and 4°C. Resuspend pellet in 50 μL sterile ddH_2O. These 50 μL aliquots are now ready for electroporation with dsDNA or ssDNA oligo cassettes for λ-Red recombination. Add 5 pg to 0.5 μg DNA to cells and electroporate/recover via standard protocol with a minor modification to incubate for 120 min prior to plating on selection plates. If a subsequent round of recombination needs to be performed, recover and plate at 30°C. However, if this is the final round of recombination, recover and plate at 37°C to remove pKD46. If pKD46 removal is proving tricky, one can try growth at up to 42°C for extended lengths of time. The absence of pKD46 can be confirmed by plating on ampicillin. All colony PCR confirmation and plating experiments for this strain was completed by Chester identically to ATMW1,[73] creating the ATMW-BL21 strain with genotype BL21 pUltra ScW40 *trpS*::Zeo^R *trpT*::$Gent^R$ (Figure 6-6).

Our work on the ATMW-BL21 strain had paused after confirmation of expression with a T7 promoter, but Elise Ficaretta, a newer member of the lab, has recently revived this story. I'd highlight her findings and Chester's preliminary data here. One concern we've had

with some ATM variants is their often lower growth rates compared to their progenitor strains. ATMW1 had no significant growth defect[1] compared to EcNR1Z or EcNR1Z+pUltraG, and the same could be said for ATMW-BL21 versus BL21 or BL21+pUltraG (Figure 6-7). It is interesting to see the growth comparison between wildtypes EcNR1Z and BL21, as they do not vary significantly in the exponential growth phase but BL21 has a lower stationary OD_{600}. Regardless, both ATM versions performed similarly to their progenitors.

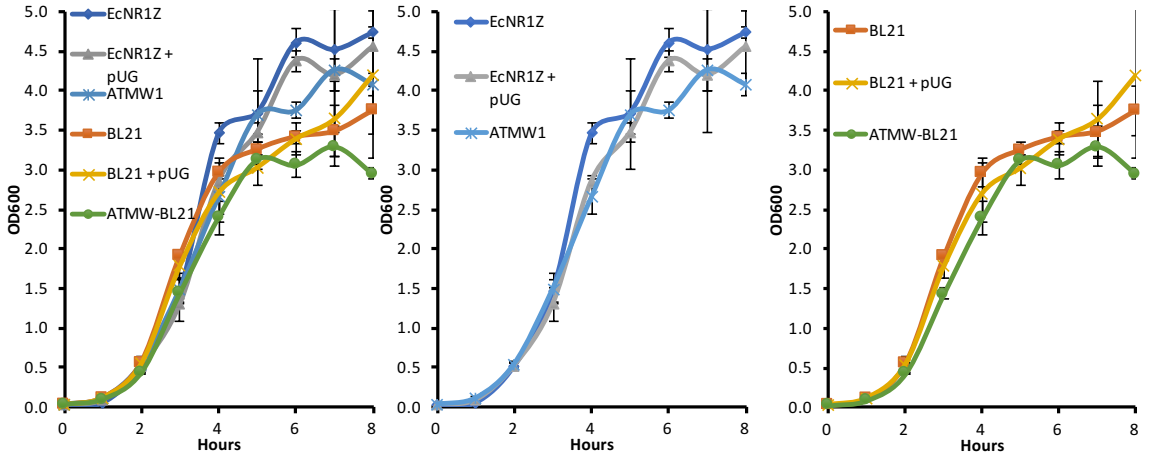


Figure 6-7. Growth comparisons between ATM and their progenitor strains. Left panel: consolidation of all strains of interest. Middle panel: Repeat of ATMW1 generations from Italia *et. al.*[1] Right panel: ATMW-BL21 growth comparison with BL21 progenitors. Overnight starter cultures were used to inoculate 15 mL cultures at a starting OD_{600} of 0.03 and grown in 50 mL flasks at 37°C, 250 rpm. 1 mL of cultures were removed every hour to measure OD_{600}.

We next wanted to investigate the protein expression capabilities of the ATMW-BL21 strain compared to parent BL21, as well as a comparison between ATMW1 and ATMW-BL21. Protein expression of sfGFPwt or sfGFP-151TGA was performed as described[1] unless

otherwise specified. We first investigated the protein expression of sfGFPwt. We were relatively unsurprised to see the BL21 strains outperforming their wild-type (EcNR1Z/ATMW1) counterparts (Figure 6-8a). It was also promising to see the modification of the wild-type strains to ATMW strains resulted in no dropoff of protein expression measured by normalized fluorescence (Figure 6-8a). The resultant proteins were purified via Ni-NTA chromatography and analyzed by SDS-PAGE (Figure 6-8b). Yields were proportional to the fluorescence data (48-69 mg/L for EcNR1Z/ATMW1; 76-121 mg/L for BL21/ATMW-BL21). However, this result indicates the modification of a strain to an ATMW strain does not result in a decrease in protein yield relative to parent strains.

We next wanted to investigate the UAA incorporation efficiency between ATMW1 and ATMW-BL21. Surprisingly, the UAA incorporation efficiency of sfGFP under the lac promoter was much worse in the ATMW-BL21 strain (Figure 6-8c), despite the sfGFPwt expression being significantly better (Figure 6-8a). This result should be repeated, but does raise possibly the most important question: which strain is truly better for UAA incorporation? Obviously, BL21(DE3) based strains are necessary for the common T7 RNAP promoted expressions, as seen by the effective expression of T7 RNAP promoted GFPuv (Figure 6-8d), but it is possible that the BL21 strain is worse at UAA incorporation. Future experiments should include a detailed comparison of UAA incorporation between these four strains of both lac promoted or T7 RNAP promoted GFP as well as other proteins of interest. We originally believed the ATMW-BL21 strain would supercede all use for ATMW1, but these results indicate that may not necessarily be the case as some applications (non-T7 RNAP) may benefit from the use of ATMW1!

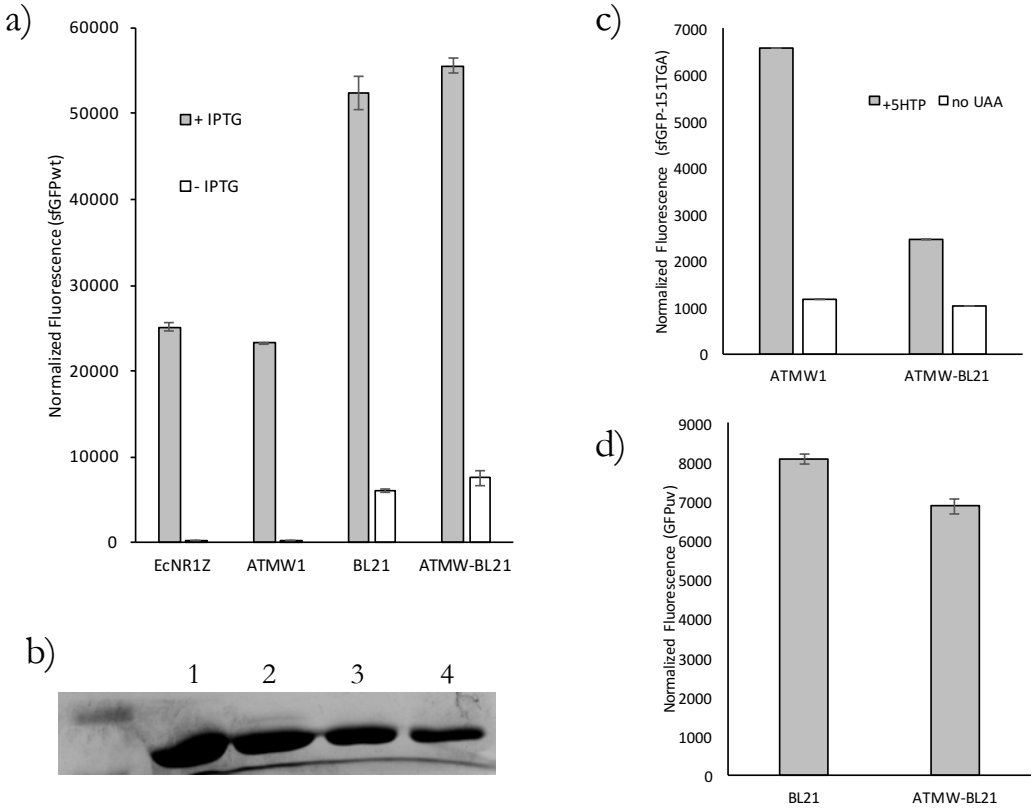


Figure 6-8. Protein expression by ATMW strains. All expressions were performed by diluting an overnight starter culture into 20 mL cultures at a 1:100 ratio. Cultures were induced with 1 mM IPTG or 0.2% arabinose and 1mM UAA was added at 0.5 OD_{600}. Cultures were expressed after induction for ~19 hrs at 30°C. Fluorescence (sfGFPwt: 488-534, 515co; GFPuv: 395-510) and OD_{600} were measured via plate reader. Fluorescence (PBS measurement subtracted) was divided by OD_{600} (PBS measurement subtracted) to achieve a normalized fluorescence. a) Normalized fluorescence of sfGFPwt (Elise Ficaretta). b) SDS-PAGE analysis of protein expressed in Figure 6-8a. (Elise Ficaretta) c) 5HTP dependent expression of sfGFP-151TGA in ATMW1 vs ATMW1-BL21, both transformed with pEvolT5 EcW sfGFP-151TGA + pBK EcWRS1.h14 (Chester Wrobel). d) GFPuv-wt expression in BL21 vs ATMW-BL21, both transformed with pET101-GFPuv, which contains GFPuv under an arabinose induced T7 RNAP promoter (Chester Wrobel).

6.3 Selection of NitroW charging tryptophan synthetase mutants

6.3.1 Brief Overview

Reactive nitrogen and oxygen species are found in a variety of diseases, such as heart disease and Alzheimer's. These reactive species can modify proteins, possibly resulting in a change of protein function.[16] Aromatic & electron rich residues such as tyrosine and tryptophan are prime candidates for post translational modification by these species.[17] These hydrophobic and aromatic residues have been shown to be largely involved in protein-protein interactions[18] due to their unique structure, hydrogen bonding, and ability to interact via cation-π stacking.[18]

It has been shown that peroxynitrite or peroxidases, when incubated with tyrosine or tryptophan, can catalyze a slew of nitrated products *in vitro*.[17] While great efforts have been employed to study tyrosine nitration (3-nitrotyrosine),[17] tryptophan nitration is less understood. Recent literature indicates that in physiological conditions (neutral pH), the major nitration product of L-Trp when exposed to nitrative reactive species is 6-nitrotryptophan (6NTP).[19] Advances in proteomics has found this species of 6NTP in proteomes of PC12 cells,[20] rat hippocampus and cerebellum,[21] and cardiac tissue.[16] This 6NTP species has also been identified as a post-translational modification in PC12 cells by analysis with an anti-6NTP antibody.[19] These modifications of tryptophan have varied consequences, either enhancing[22] or retarding[23] protein function. Therefore, site-specific tryptophan nitration/oxidation are relevant to protein structure in function and have significant consequence to human health.

While nitrotryptophan has been shown to be present in the proteome, the function of these modified residues can be difficult to understand due to challenges in purifying homogenously modified protein. We envisioned using genetic code expansion as a method to

incorporate the physiologically relevant 6NTP or other tryptophan species such as 5-hydroxy-6-nitrotryptophan[24] to characterize the effects of these nitrated residues on protein function. Other PTMs have been genetically encoded in *E. coli*,[25,26] but our ATM[1] technology allows us to develop new pairs for the use in eukaryotes.

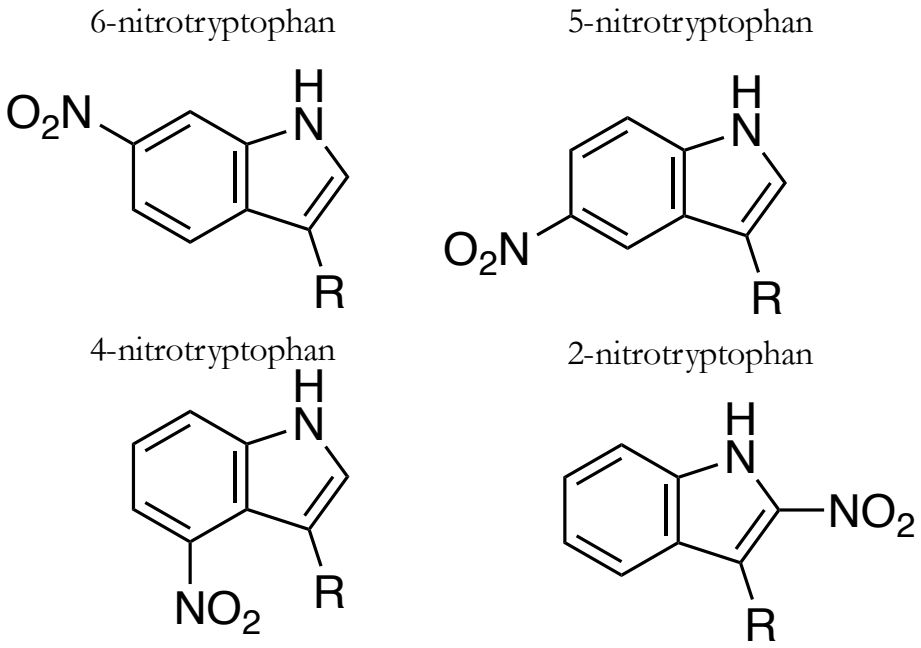


Figure 6-9. Nitrotryptophan analogs used in this study.

6.3.2 Results and Discussion

We first attempted to incorporate the less physiologically relevant 5-nitrotryptophan (Figure 6-9) due to our belief that a 5-position substituted tryptophan analog may be accepted by our previously developed EcWRS1 library.[1] This library was developed for the selection of 5HTP and contains mutations of residues: Phe7(NBT), Ser8(NST), Val144(NNK), Pro145(NNK), and Val146(NNK). After a round of positive (P1) and negative (N1) selection in ATMW1,[1] the second positive (P2) selection plates show the an unexpected outcome

(Figure 6-10a). Increased amounts of 5NTP resulted in worsened growth (Data not shown, measured by eye). This result may suggest that 5NTP or successful EcWRS1 hits are toxic to *E. coli,* with the former being the more likely scenario. Hits were picked and spot plated from the +5NTP selection (P2) plate but showed no enhancement of survival in the presence of 1 mM 5NTP over the absence of UAA (Figure 6-10b).

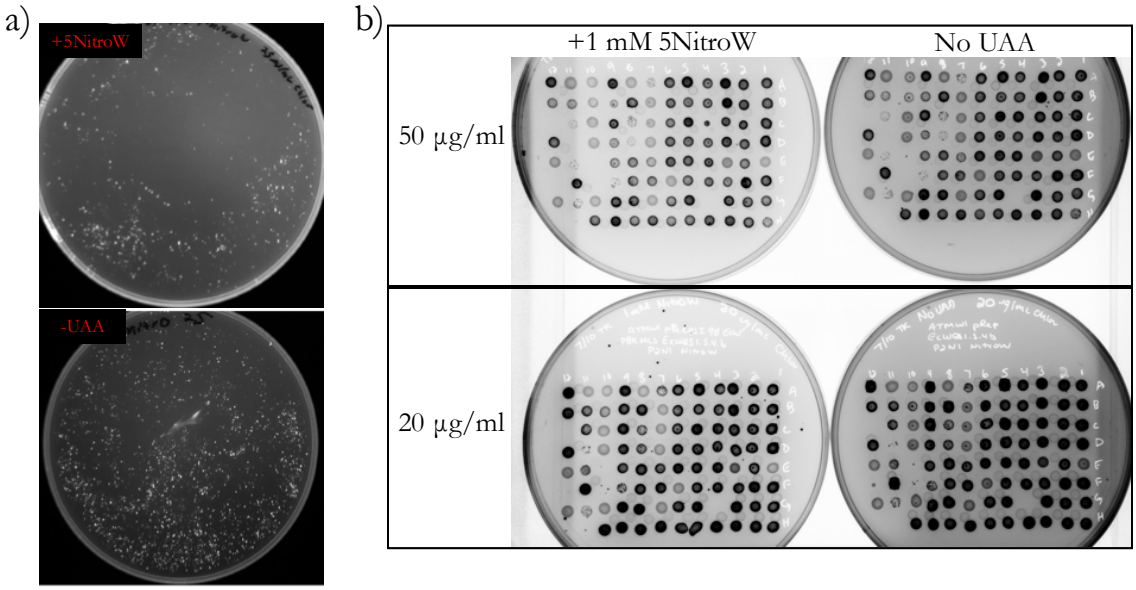


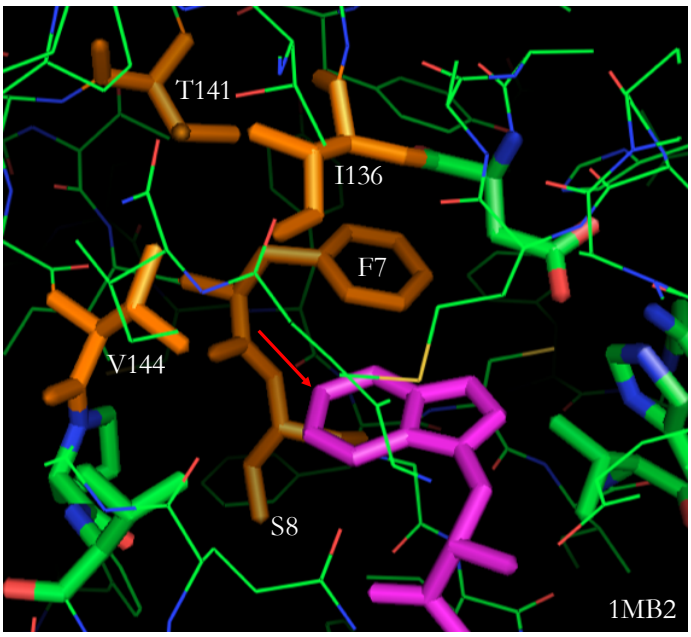
Figure 6-10. EcWRS1 selection for 5NTP. a) P2 (second round of positive selection) plates (See Italia *et. al*[4] for more detailed selection protocol). 10^9 cfu of ATMW1 expressing pRepJI-EcW-TGA and pBK-EcWRS1 was plated on LB + 0.5x Spec/Tet/Kan + 25 and 35 µg/mL Chlor and 50 µg/mL Amp, and 1 mM 5NTP. After 24 hrs incubation, all plates were harvested and plasmid was purified as previously described to make library EcWRS1-5NTP.P1.[73] For negative selection, 10^8 cfu of ATMW1 expressing pNegJ2-EcW and pBK-EcWRS1-5NTP.P1 were plated on LB + 2xAmp, 0.02% arabinose. After 10 hrs incubation, pBK-EcWRS1-5NTP.P1.N2 was purified and transformed for another round of positive selection. Under the same positive selection conditions unless otherwise specified, 10^6 cfu were plated on 25

μg/mL chloramphenicol with or without 1 mM 5NTP. b) Spot plates of colonies picked from Figure 6-10a +5NTP plate. See Italia *et. al.*[73] or Chapter 2 for more detail on spot plating protocol.

While the 5NTP selection was unsuccessful, we were more interested in the ability to incorporate the physiologically relevant 6NTP. To do this, new libraries would need to be developed in order to target the 6-position of the tryptophan indole ring. The EcWRS4 library was cloned using the library cloning strategy used in EcWRS1,[1] albeit with different mutagenic primers. EcWRS4 consists of mutations: F7(NNT), S8(S or A), I136(NNT), T141(NNT), and V144(NNT) (Figure 6-11a). For the selections of 6NTP with both EcWRS4 and EcWRS5, a crude mixture of 6NTP and 5NTP was used. This crude mixture was synthesized by Partha Sarathi Addy and was used in the interest of time while he developed an optimized method to isolate pure 6NTP and can be noted as 6NTP (contains 6 and 5NTP). The P2N1 (two rounds of positive, one round of negative) selection of 6NTPmix with the EcWRS4 library did result in three hits which were highly conserved (Figure 6-11b). It was even more exciting to see these hits had a reasonable level of specificity in the presence of 6NTPmix, as measured by expression of sfGFP-151TGA in ATMW1 (Figure 6-11c). However, the purified sfGFP from the +6NTPmix plates did not have the expected nitrotryptophan mass by whole protein LCMS, but instead showed the mass for wild type tryptophan for all three hits (Figure 6-11d). It is possible that the Nitro- group falls off the indole ring during mass spec, but the more likely conclusion is the EcWRS4 library selection was unsuccessful despite such a large sequence convergence.

a)

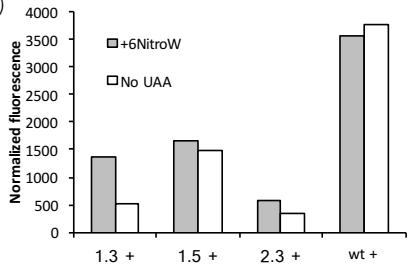
	F7	S8	I126	T141	V144
EcWRS4 Library	NNT	S or A	NNT	NNT	NNT



b)

	F7	S8	I126	T141	V144
Hit 1.3	I	S	I	G	I
Hit 1.5	I	S	I	G	V
Hit 2.3	I	S	I	G	C

c)



d)

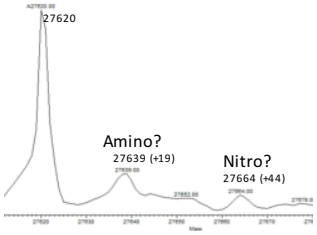


Figure 6-11. EcWRS4 library and selection results for 6NTPmix. a) EcWRS4 library design and crystal structure of active site. Mutated residues are in orange, tryptophan is in magenta. b) EcWRS4 hits after two rounds of positive and one round of negative selection for 6NTPmix. c) Expression of sfGFP-151TGA in the ATMW1 pEvol-EcW-sfGFP-151TGA pBK-EcWRS4-hit system. Expression and measurement of GFP fluorescence was performed using our lab's standard expression protocol as previously described.[1] d) Whole protein LCMS of sfGFP-151-"6NTPmix." Mass does not indicate 6NTPmix incorporation as expected.

Library	F7	S8	V43	H45	D135	I136	V144
EcWRS5.1A	NNT	A	NNT	NNT	NST	NBT	NST
EcWRS5.1S	NNT	S	NNT	NNT	NST	NBT	NST

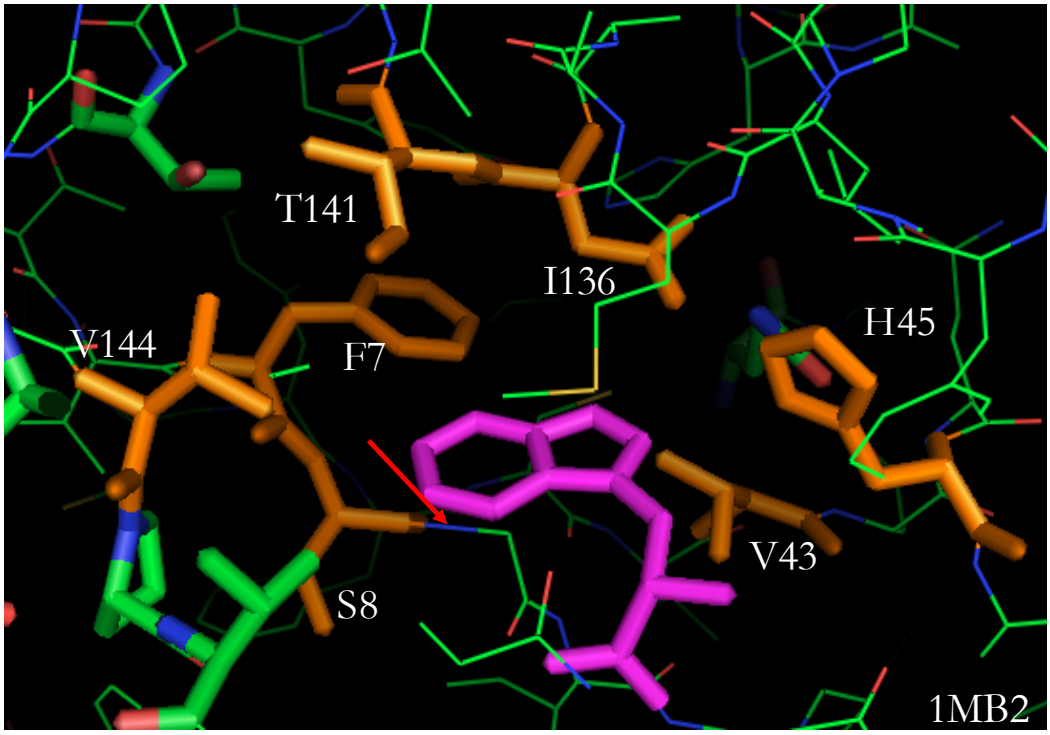


Figure 6-12. EcWRS5 library design and crystal structure.

Undeterred, we decided to try another library of synthetase mutants to incorporate the continually elusive 6NTP. We took a much larger library approach with EcWRS5, which contains mutations: F7(NNT), S8(S or A), V43(NNT), H45(NNT), D135(NST), I136(NBT), V144(NST). These mutations were introduced through sequential overlap extension by Chester Wrobel (Figure 6-12). After EcWRS5 selection (performed identically to EcWRS1 and EcWRS4), we found a series of hits displaying 6NTP survival dependence (Figure 6-13a). Hits were highly convergent in some residues: V43F, conservation of H45, D135, I136. S8 could either stay as serine or would also change to alanine, possibly irrelevant to the synthetase

solutions. The most interesting residues are F7(VHLTS) and V144 to cysteine. It is surprising that V43 turns into the larger hydrophobic phenylalanine, but possibly this structurally beneficial when stacking with indole of 6NTP in the active site. Expression of an sfGFP-151TGA reporter with the EcWRS5 hits displayed minimal 6NTPmix dependence (Figure 6-13b). However, when magnified, hit 2 does display a 2:1 ratio of UAA dependence (Figure 6-13b, insert). This hit 2 has a V144C mutation, which shrinks back the active site, allowing the nitro group to fit. Additionally, the F7H mutation in hit 2 is intriguing, as it both allows the active site to have more steric flexibility and may be involved in some unknown hydrogen bonding networks.

a)

EcWRS5.1	F7	S8	V43	H45	D135	I136	V144
1	V	A	F	H	D	I	V
2	H	A	F	H	D	I	C
3	L	S	F	H	D	I	V
4	T	A	F	H	D	I	V
5	S	A	V	H	D	I	V
6	L	S	F	H	D	I	V
7	L	A	F	H	D	I	C
8	V	A	F	H	D	I	C

b)

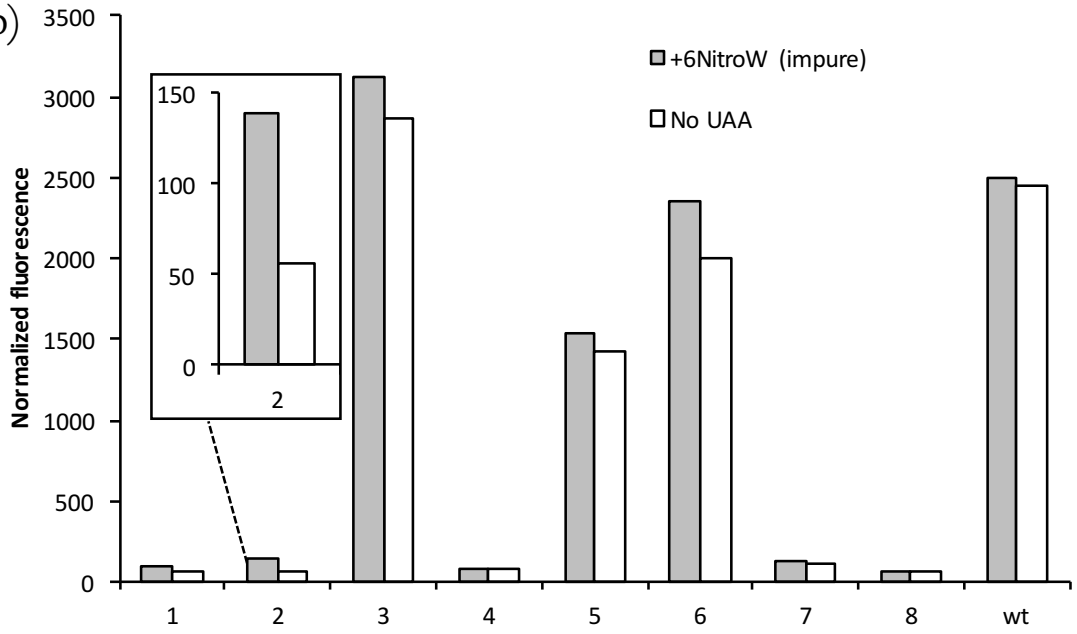


Figure 6-13. Selection hits from EcWRS5 selection for 6NTP. a) EcWRS5-6NTPmix selection hits. b) Normalized fluorescence of EcWRS5-6NTP synthetase hits expressed in ATMW1 pEvoltac-EcW-sfGFP151TGA with or without 1 mM 6NTPmix. Expression performed as previously described.[1]

Despite the apparent survival and fluorescence dependence EcWRS5-6NTP.h2 displayed, we were not certain of 6NTP incorporation until a mass was observed. To do this, we repeated the sfGFP-151TGA expression with either 2 mM 6NTPmix, 4 mM 6NTPmix, 2 mM 5NTP, or no UAA (Figure 6-14a). There was an increase in normalized fluorescence in the presence of all three UAA relative to the no UAA control, which indicates the incorporation of 5NTP rather than 6NTP. Surprisingly, LCMS analysis of these proteins indicated a pure nitrotryptohan mass (6NTP/5NTP have the same mass), while the protein expressed with pure 5NTP was a mixture of nitrotryptophan and tryptophan masses (Figure 6-14b). It is possible that this synthetase EcWRS5.h2 is polyspecific for 6- or 5- substituted nitrotryptophan analogs.

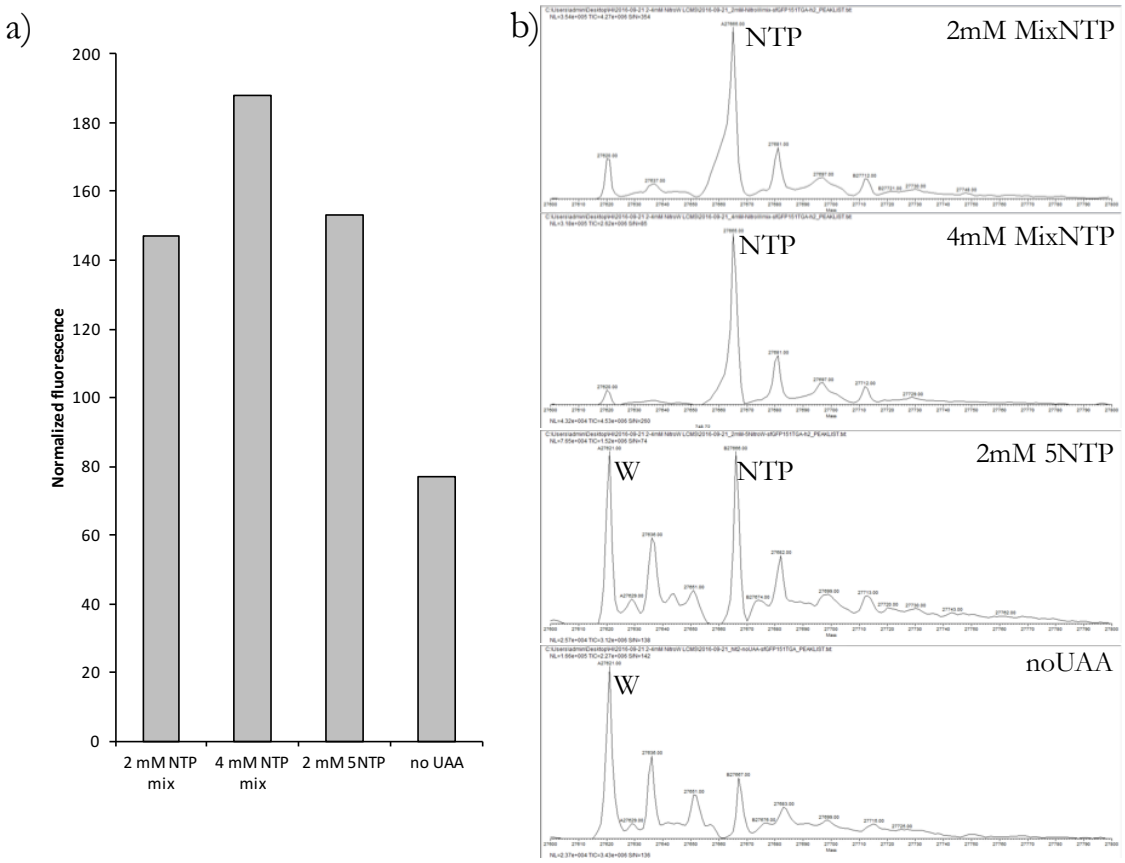


Figure 6-14. EcWRS5.h2 mass spectrometry analysis with 6NTPmix versus 5NTP. a) Fluorescence analysis with ATMW1 pEvoltac-EcW-sfGFP151TGA pBK-EcWRS5.h2 in the presence of 2 or 4 mM 6NTPmix, 2 mM 5NTP, or no UAA. 6NTPmix contains a mixture of nitrotryptophan analogs due to synthetic route. b) LCMS of sfGFP-NTP proteins. Masses are labeled corresponding to UAA incorporated.

EcWRS5.h2-6NTP was cloned into the tryptophanal pAcBac mammalian expression vector was next transfected into HEK293T with an EGFP-39* reporter.[1] A purer form of 6NTP was synthesized by Partha and used in these fluorescent studies, and 1 mM UAA was added to cells at the time of transfection (Figure 6-15). Very surprisingly, pure 6 and 5NTP resulted in very little fluorescence relative to no UAA, while the crudely prepared 6NTPmix resulted in a positive UAA dependence (Figure 6-15). This directly contradicts the bacterial expression data, as 5NTP resulted in an UAA dependence when expressed in ATMW1. The bacterial expression with the pure 6NTP also corroborated this surprising data from mammalian cells and did not show incorporation of nitrotryptophan (Figure 6-16).

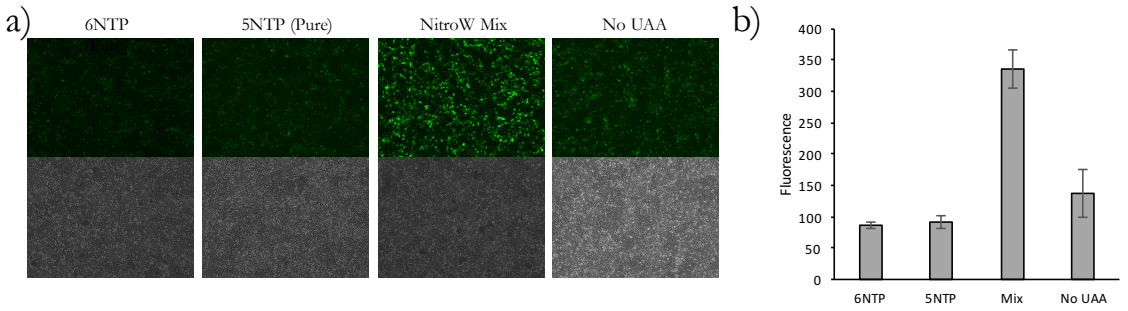


Figure 6-15. Mammalian expression of EcWRS5.h2-6NTP. HEK cells seeded in a 12 well plate were transfected with 500 ng pAcBac EGFP-U6-EcWtR-TAG [1:E1] and 500 ng pAcBac1 EcWRS5.1-6NTP-h2-EcWtR TAG [5:F7] via PEI Max transfection.[1] a) Transfection images in FITC channel, 50 ms exposure, 3 click. A more pure of 6NTP was synthesized by Partha, which should not contain 5NTP. b) Raw fluorescence of EGFP-39TAG transfected in Figure 6-15a measured from HEK cell lysate 48 hours post transfection. See Italia *et. al*[1] for HEK cell fluorescence protocol.

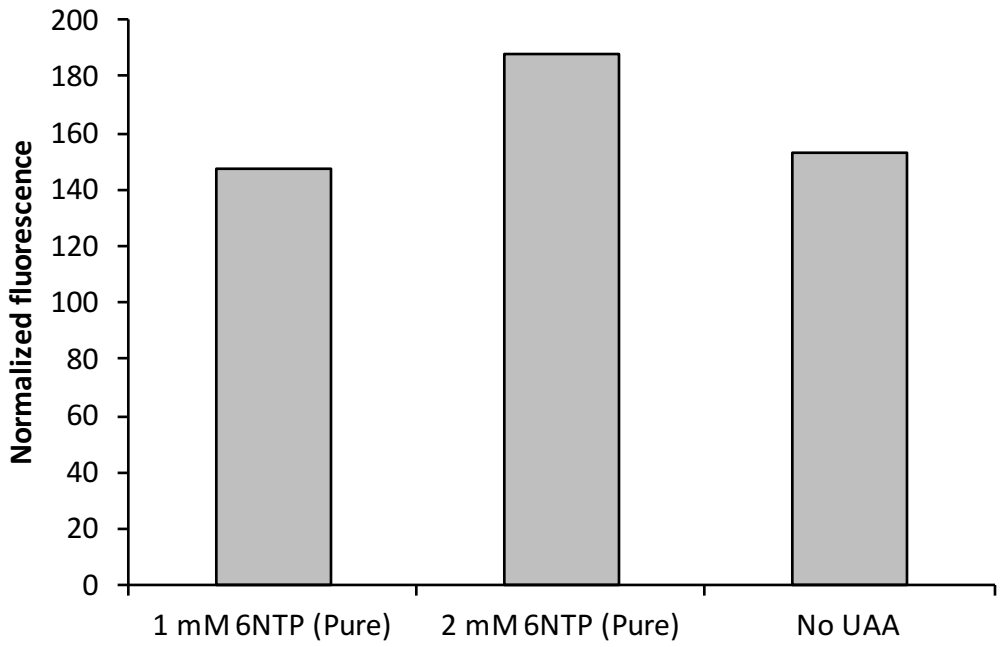


Figure 6-16. EcWRS5.h2 expression of sfGFP-151TGA in ATMW1 with pure 6NTP. Bacterial expression was performed (see Figure 6-13 and 6-14) as previously described[73] but instead with 2 mM pure 6NTP.

We hypothesized that something in the 6NTP mix was being charged, but it was not 6NTP or 5NTP. It is possible that there are trace amounts of 4NTP or 2NTP (Figure 6-9) in the 6NTPmix synthesized UAA due to the non-specificity of nitration.[27] We employed the use of the Gao lab prep HPLC to try to separate the incorporated UAA from the 6NTPmix (Thank you Kelly McCarthy). We attempted to separate the nitrotryptophan isomers, as elution profiles have been studied via mass spec[27] 2 mL (~10 mg) 6NTPmix was injected on the prep-HPLC (Figure 6-17). Three fractions corresponding to the major peaks were isolated for analysis by LCMS. Based on previous elution profiles,[27] we anticipated the first prep HPLC peak to contain the 2NTP isomer, while the second and third would contain the 6NTP or 4NTP isomer. Overall yields were low <20% recovery, likely due to inexperience with the

prep. However, only the first two prep-HPLC peaks contained NTP isomers (Figure 6-18). Our prep HPLC ocnditions were not sufficient to separate what appears to be the 6NTP and 4NTP isomers.

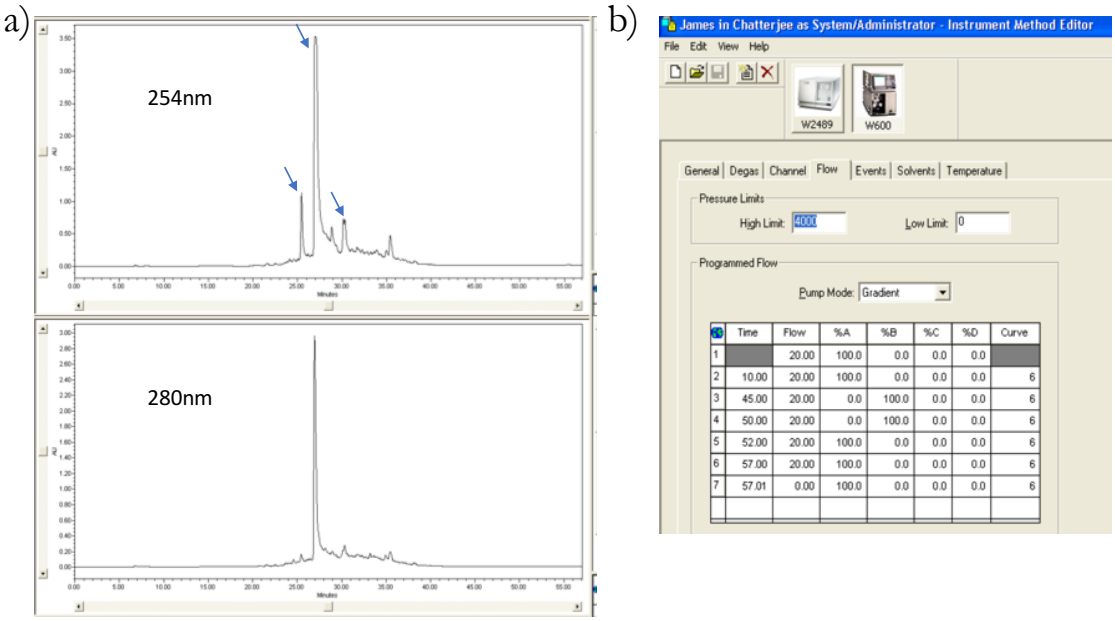


Figure 6-17. Prep HPLC elution profile. a) Elution profile of 6NTPmix on prep HPLC. Spectra was monitored at 254 and 280 nm. Fractions noted by arrows were collected for mass spec analysis. b) Prep-HPLC conditions. Insert prep HPLC buffers.

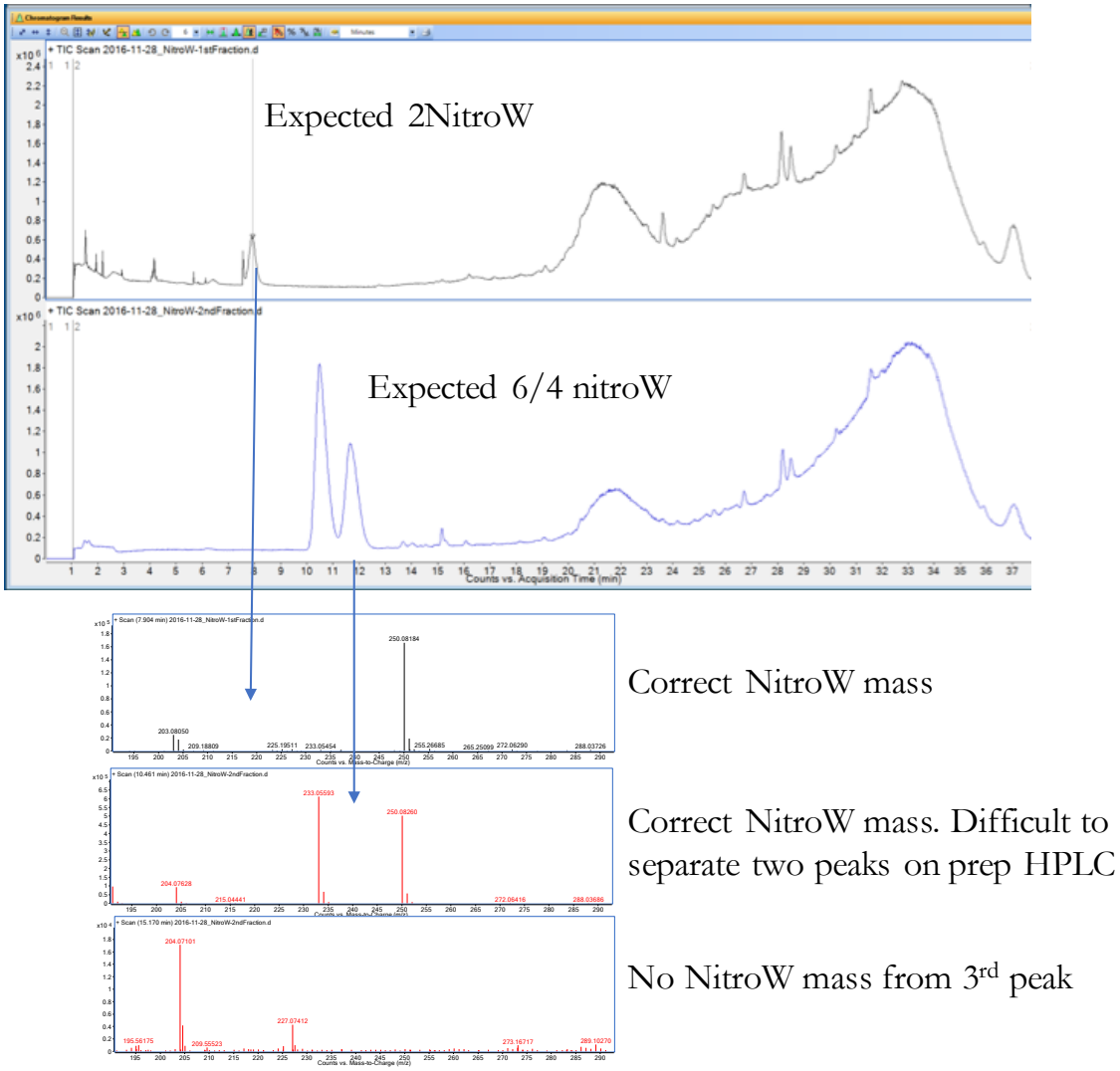


Figure 6-18. LCMS of purified peaks from prep HPLC. Figure depicts the TIC scan of first peak and second peak with their corresponding masses as noted via arrows. Both of the first two peaks have the correct NTP masses. The TIC for the third prep HPLC is not shown, but did not denote any NTP mass. Buffer A: 95% water, 5% acetonitrile, 0.1% TFA; Buffer B: 95% acetonitrile, 5% water, 0.1% TFA. (Data quality is poor)

Lastly, the prep HPLC purified NTP isomers were used in a transfection experiment to observe which portion of the 6NTPmix results in UAA incorporation. Unfortunately, this experiment resulted in more questions than answers, as only the crude NTP (6NTPmix) resulted in incorporation of some UAA (Figure 6-19), and the concentration of UAA from the prep HPLC purified peaks is unclear. Their low expression may simply be due to their low concentration and should be repeated with proper concentration controls. This was a surprising result and was a pause point in this exciting selection process.

Amazingly, we were unable to confirm the selective incorporation of 6NTP in any of the selections even after some positive expression results. The first selection with EcWRS1 resulted in toxicity, possibly due to 5NTP levels in the bacterial cells. The selection of crude 6NTP with EcWRS4 was unsuccessful due to selection of a highly conserved set of synthetases which appear to charge an UAA but only take tryptophan (unless there is some de-nitration). Lastly, the selection synthetase varaints with crude 6NTP found a relatively conserved but weak synthetase hit which actually charges an amino acid with an identical mass to nitrotryptophan. However, it is unclear which isomer of nitrotryptophan is accepted as substrate for EcWRS5.h2-NTP. While future work seems straightforward (determine which NTP isomer), the implications of the synthetase becomes less relevant since the UAA is not the physiologically relevant 6NTP. We know 4NTP and 2NTP have some minor physiological relevance, but the significance is much less than 6NTP. The challenge of evolving another tryptophan synthetase variant continues to be vexing, but the value of solving these tryptophan selection difficulties is monumental. Isolation of pure 6NTP and repeating the selection with the old/new libraries may result in a successful identification of 6NTP specific aaRS mutants. Also, it might be worth it to normalize the amount of NTP isolated from the prep HPLC, as we already likely have a hit but have not yet been able to confirm this finding.

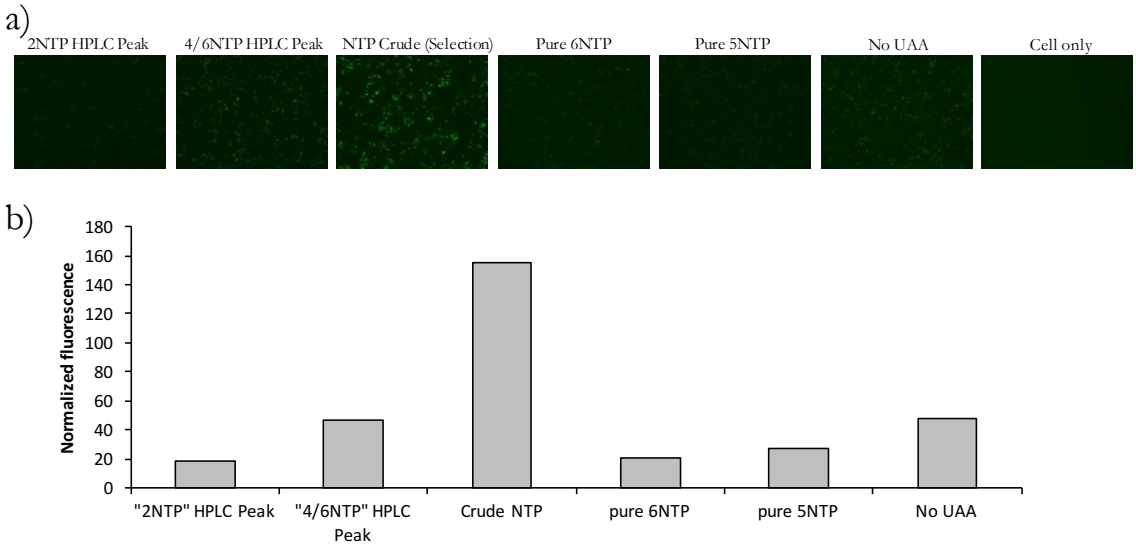


Figure 6-19. Mammalian expression of EcWRS5.h2-6NTP with prep HPLC peaks. a) Fluorescence images of HEK293T 48 hrs post transfection, 50 ms, 3 click. Transfection performed as previously described in 12 wells. b) Fluorescence measurement of HEK cell lysate, performed as previously described.

6.4 *In vitro* characterization of sulfotyrosine synthetase mutant specificity & stability

6.4.1 Brief Overview

Our lab has recently developed synthetase variants which selectively charge the eukaryotic PTM, sulfotyrosine, in bacteria and eukaryotic cells (Chapter 4). While both synthetase variants (VGL and VGM) have high selectivity and low background in ATMY *E. coli* (Figure 6-20a), the background is significantly increased in the absence of sTyr when overexpressed in HEK293T (Figure 6-20b). Our experience with previously evolved *E. coli* tryptophanyl/tyrosyl tRNA/aaRS or Pyrrolysine tRNA/aaRS pairs has indicated that expressions are proportionally selective between *E. coli* and eukaryotes,[28–30] so we sought to explore the cause of this variation. The bacterially purified sfGFP-151TGA expressed in the

absence of UAA (Figure 6-20a) results in a mass corresponding to the incorporation of glutamine, which could be due to much lower expression of the SulfoRS in bacteria (Figure 6-20d), as our mammalian transfections have relatively higher levels of expression of the tRNA/aaRS pair due to the nature of transient transfections. Additionally, we previously found that the EGFP-39* purified in the absence of UAA (Figure 6-20b) results in a mass corresponding to tyrosine (Figure 6-20e). We developed two hypotheses: 1) overexpression of SulfoRS leads to higher background charging of endogenous amino acids, such as tyrosine; 2) SulfoRS also charges an endogenously synthesized amino acid (with PTM) not present in *E. coli* but present at some level in eukaryotes. Hypothesis 1 is the most straightforward, previous work has shown that higher expression of aaRS can result in loss of specificity.[28] However, hypothesis 2 is more obtuse. If the SulfoRS does not charge an endogenous amino acid, it would hypothetically charge an amino acid with a similar structure to sulfotyrosine (sTyr). One such amino acid is phosphotyrosine (pTyr) (Figure 6-20c), which is present in proteins found in eukaryotes but not *E. coli*.[31,32] It is possible that SulfoRS charges pTyr in eukaryotes, but the phosphate is removed from the recombinant protein,[31] corresponding with the EGFP-39* no UAA mass (Figure 6-20e). We anticipated using an *in vitro* specificity test to determine which hypothesis may be correct.

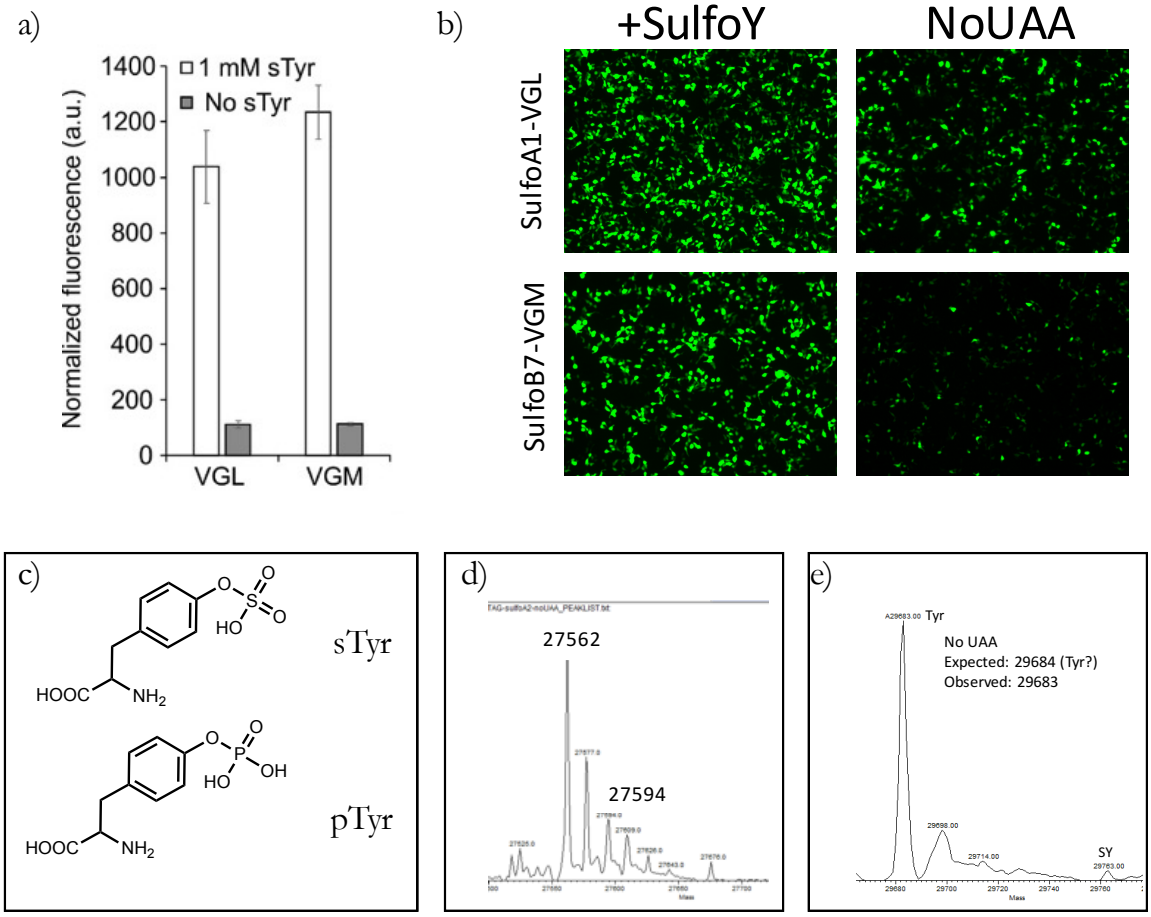


Figure 6-20. Background charging of SulfoRS in HEK293T. a) Normalized fluorescence of SulfoRS-VGL and –VGM in ATMY4, expressed as previously described.[30] b) Fluorescence of HEK293T cell lysate after transfection of EGFP-39TAG reporter with SulfoRS/$tRNA^{EcTyr}_{CUA}$ as described.[30] c) Structures of sulfotyrosine (sTyr) and phosphotyrosine (pTyr). d) LCMS analysis of sfGFP-151TAG expressed with ATMY4 pBK-SulfoRS-VGL with no UAA (no sTyr bars in Figure 6-20a, gray). e) LCMS analysis of EGFP-39* expressed in HEK293T. SulfoRS-VGL and –VGM were transfected with the EGFP reporter in the absence of UAA, as described in Figure 6-20b.

In vitro methods to determine aaRS specificity using a fluorescent reporter have been developed.[33,34] This assay takes advantage of malachite green, an organic reporter which forms a detectible color change proportional to the amount of inorganic phosphate in solution. We can use an abbreviated form of this assay without the tRNA, which can detect the adenylation of the amino acid charged by the synthetase of interest (Figure 6-21). A fantastic undergraduate, Christen Hillenbrand aka C3, optimized this assay for use in the Chatterjee lab as described in detail in the materials and methods section.

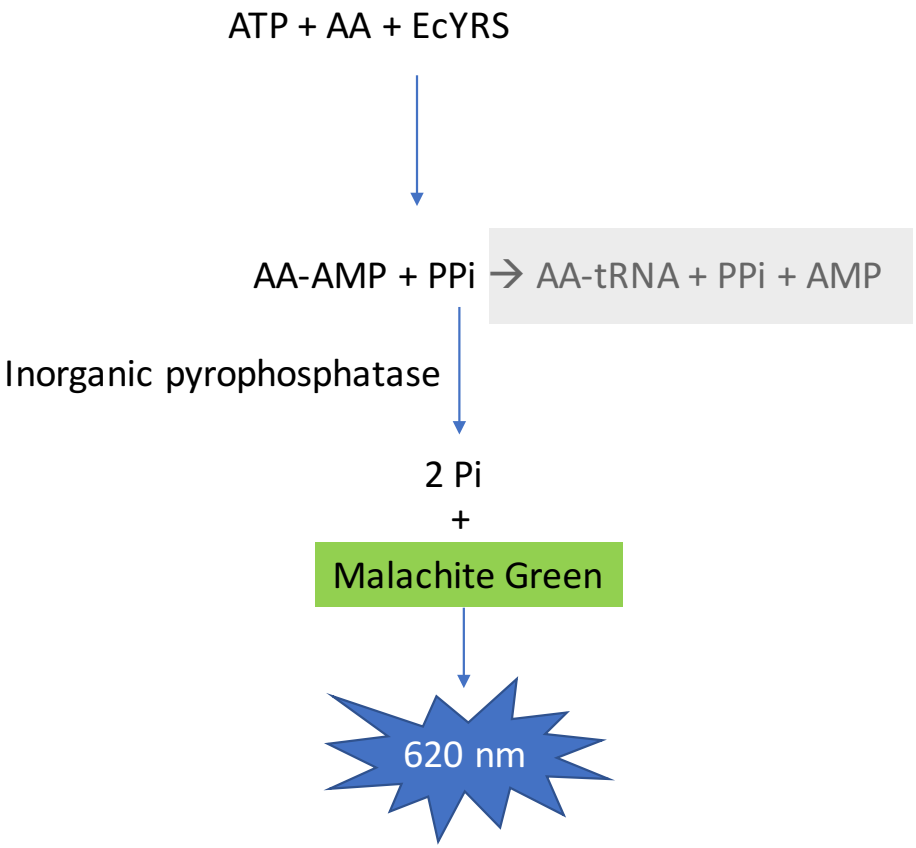


Figure 6-21. Malachite green assay scheme. The tyrosine synthetase adenylates the amino acid, which releases PPi. Inorganic pyrophosphatase cleaves PPi into 2Pi. Malachite green reacts with free phosphate which changes color and can be measured at 620 nm absorbance.

6.4.2 Results and Discussion

In order to purify the tyrosine synthetase variants, the synthetases were cloned by Christen Hillenbrand with an n-terminal 10xHisTag in pET22b (pBR322 ori) and under expression of a T5-lac promoter (Figure 6-22a). It should be noted that C-terminal tags were also effective, but N-terminal tags were chosed due to better cloning success. pET22b-EcYRS-HisN and other aaRS plasmids were individually transformed into DH10b. Starter cultures were used to inoculate expressions at a 1:100 ratio in 15 mL LB (in 50 mL falcon tube) and grown shaking at 250 rpm, 37°C to 0.5 OD_{600} in LB + Amp. Upon reaching 0.5 OD_{600}, cells were induced with 1 mM IPTG and grown for an additional 18 hrs at 37°C. After expression, cells were pelleted, resuspendend in Bper lysis buffer, and purified via Ni-NTA as previously described.[49] EcYRSwt was purified at 7.72 mg/L and analysed by SDS-PAGE (Figure 6-22b), but a mass was not obtained.

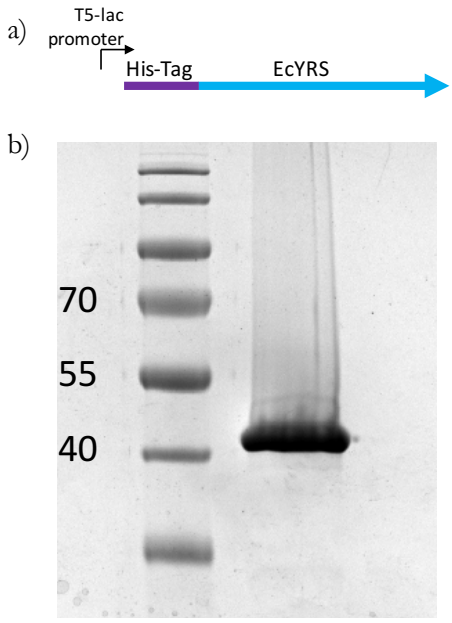


Figure 6-22. Purification of EcYRS. a) Scheme of pET22b-EcYRS-HisN expression system. An N-terminal His-Tag was cloned before the EcYRS, all under control of the T5-lac (IPTG inducible) promoter. b) SDS-PAGE analysis of purified EcYRSwt.

While purification of wild-type EcYRS was successful, yields of the SulfoRS mutants was negligible when purified using Bper lysis buffer (Figure 6-23). We discovered an instability in both the supernatant and purified mutant tyrosine synthetases when subjected to Bper lysis buffer as opposed purification by sonication (Figure 6-23). Additionally, we cloned and screened a series of tyrosine synthetase mutants (Figure 6-24a) and found every mutant has reduced stability in Bper buffer compared to the wild type (Figure 6-24b). It is currently unclear why Bper reduces the stability of the mutants but it does confirm the expected but the previously undemonstrated fact that selection of aaRS mutants can result in proteins with decreased stability. We do not necessarily see a correlation between mutation site and decreased stability, as there is generally two-modal phenotype: stability in Bper or not (Figure 6-24). It should be noted that the BzF2 mutant in this study has a point mutation, which should void the results seen here. Interestingly, pBoF-GSIE (sequence verified) is slightly less stable after sonication, as well as poor expression from OmeYRS-VSML in this experiment (possible point mutation?), but there is no consensus on a specific residue causing this stability difference. The stability decrease is likely due to an overall structural strain on the enzyme and less likely due to one specific residue. Regardless, the using sonication allows for isolation of EcYRS-Sulfo mutants with yields between 3-10 mg/L, which was sufficient for *in vitro* specificity studies.

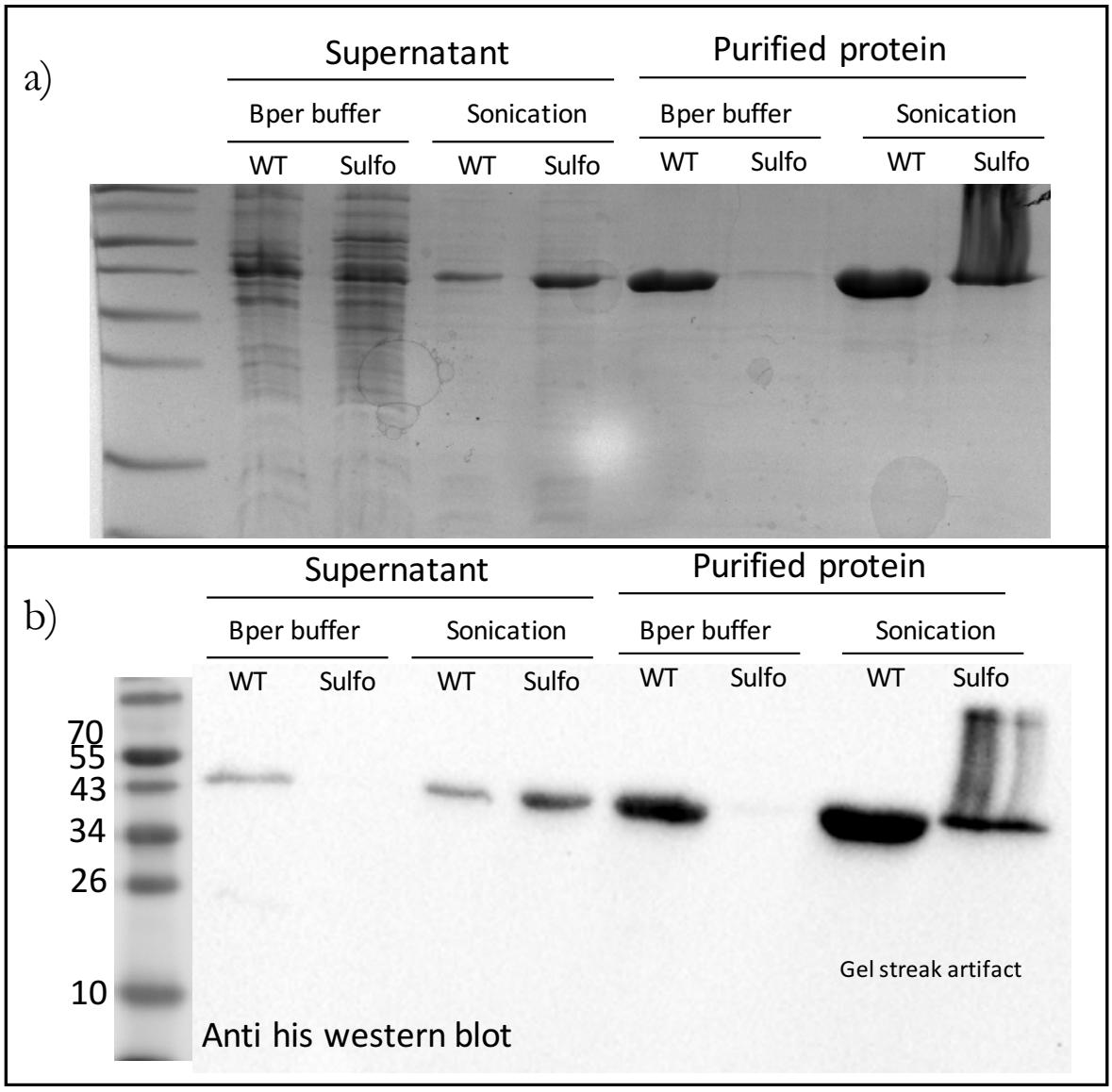


Figure 6-23. EcYRSwt and SulfoRS stability in Bper buffer versus sonication. Protein under the "supernatant" label is protein subjected to Bper lysis buffer[28,30] or sonication (10 mL pellet resuspended in 50 mM Hepes buffer, pH 7.5, 8x20 sec cycles, 40 Amp, in 4°C room with samples on ice, Weerapana tip) and subsequently centrifuged at 22,000 x g. Supernatant was directly analyzed by SDS-PAGE or western blot. For "purified protein," the supernatant containing EcYRS variants was purified using the standard Ni-NTA affinity purification protocol (using Hepes based buffers rather than phosphate).[28] a) SDS-PAGE analysis of

protein stability in the supernatant or after purification. b) Anti-His western blot, with identical protocol from Italia *et. al.*[30]

a)

#	Plasmid	Expected mutations (All contain D265R) Y37	L71	D165	D182	F183	L186	
1	pET22b-BzF1-GGYM-R265-HisN	G	L	D	G	Y	M	
2	pET22b-OmeY-D*-VSMA-D165G-R265-HisN	V	L	G	S	M	A	
3	pET22b-OmeYRS-HisN	V	L	D	S	M	L	
4	pET22b-BzF2-GGFA-R265-HisN	G	L	D	G	F	A	A304V point mutation
5	pET22b-EcOmeYRS-H3-VCMA-HisN	V	L	D	C	M	A	
6	pET22b-EcOmeYRS-D1-HisN	V	L	D	S	M	A	
7	pET22b-BoF-B5a-GSIE-R265-HisN	G	L	D	S	I	E	
8	pET22b-EcYRS-VGYI-D265-HisN	Y	V	D	G	Y	I	
9	pET22b-EcYRS-HisN	Y	L	D	D	F	L	
10	pET22b-EcYRS-SulfoVGL-HisN	Y	V	D	G	F	L	

b)

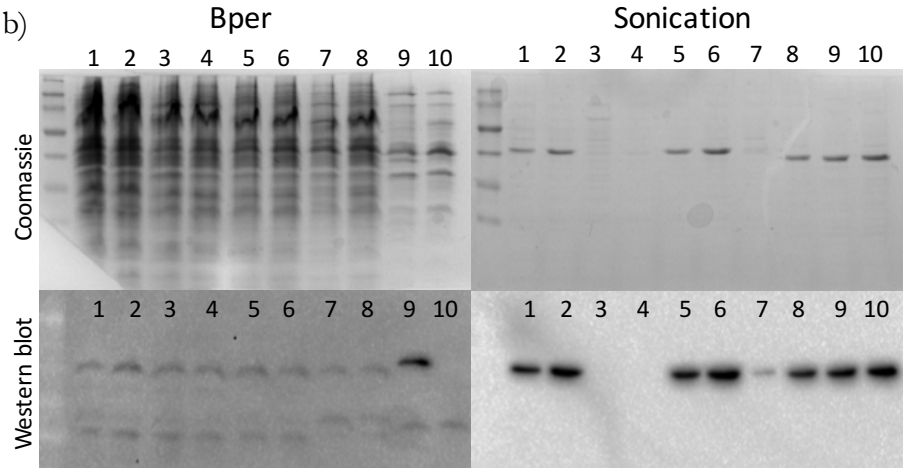


Figure 6-24. EcYRS mutant stability comparison in Bper or sonication supernatants. EcYRS mutants were expressed from pET22b-EcYRS-X-HisN (X=Mutant) in 20 mL cultures as previously described. a) A list of mutants. b) After overnight incubation, the 20 mL cultures were split into 2 x 10 mL cultures and processed via Bper lysis buffer or sonication as described in Figure 6-23. After the supernatant was clarified, 30 uL was analyzed by SDS-PAGE and anti-His western blots.

Despite the challenges associated with protein stability, the aaRS's were purified and increasing concentrations of aaRS were tested with the malachite green assay using 1 mM tyrosine, 1 mM glycine, 1 mM sulfotyrosine, or no UAA (Figure 6-25). Three SulfoRS mutants (VGYI-Katie Grasso hit, VGM and VGL; order of increasing background) were tested. Unsurprisingly, the EcYRSwt selectively charges tyrosine, with UAA measurements similar to no UAA controls (Figure 6-25a). SulfoRS-VGYI, which was shown in Katie's mammalian transfection (data not shown) to have minimal background with any amino acid other than 1 mM sTyr (Figure 6-25a). The charging of tyrosine increases with the VGM and VGL mutants, indicating their background charging *in vivo* is likely due to their overexpression and specificity toward endogenous tyrosine (Figure 6-25c, Figure 6-25d). Katie's less cross-reactive SulfoRS-VGYI is less active than the VGM and VGL mutants and therein lies the lower cross reactivity, as seen by the lower amount of phosphate formed in the VGYI assay (Figure 6-25b). A closer look into the background amino acids points out the respective charging of 1 mM tyrosine by the SulfoRS mutants, highlighting how the Tyr background charging by VGYI is below even the VGM-No UAA expression levels (Figure 6-25e). While the sTyr mutants do charge Tyr at a low level, it is also possible that pTyr is incorporated. However, our current optimization of the malachite green assay is insufficient for testing pTyr incorporation by aaRS variants (Figure 6-26) due to the low amount of inorganic phosphate present in the pTyr stock. It may be possible to eliminate the residual inorganic phosphate via HPLC, but these steps were not attempted at this time. Therefore, it appears that the background is actually due to the incorporation of Tyr rather than pTyr. Charging of pTyr cannot necessarily be ruled out, but the data is currently inconclusive.

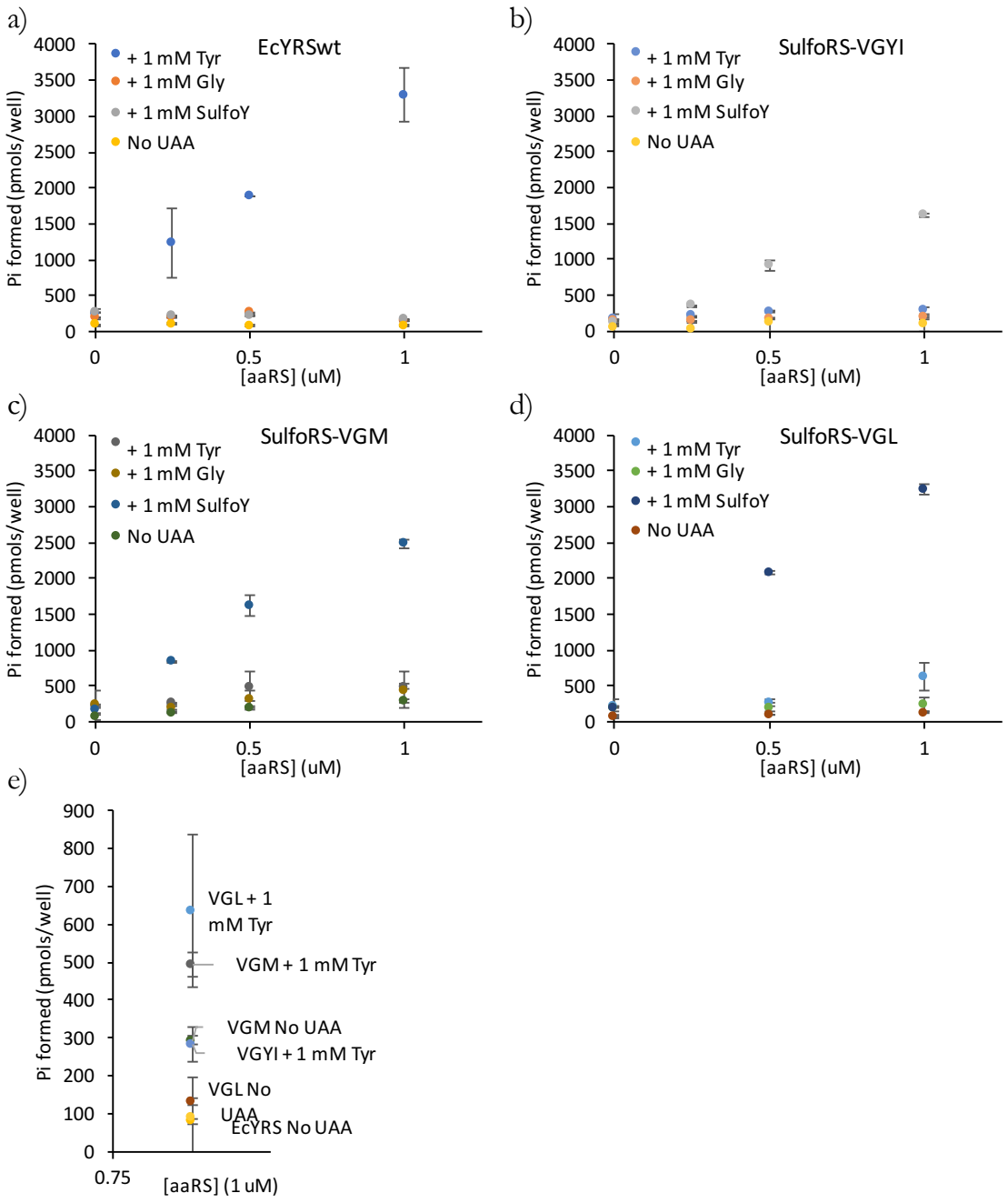


Figure 6-25. Malachite green assay of EcYRSwt and SulfoRS mutants. The malachite green protocol described in the Brief Overview was followed for these synthetase variants. Final concentrations of 1 mM Tyr, Gly, sTyr, or no UAA were used. The amount of Pi formed is proportional to the enzyme activity at a given enzyme concentration. a) EcYRSwt. b) SulfoRS-VGYI, a hit identified by Katie Grasso from her selection with her less cross reactive/active

tRNA, h1. c) SulfoRS-VGM. d) SulfoRS-VGL. e) Enhanced region of 1 uM aaRS (a-d) for the less active amino acids. Region is enhanced to show the hierarchy of activity of VGL/VGM relative to no UAA backgrounds.

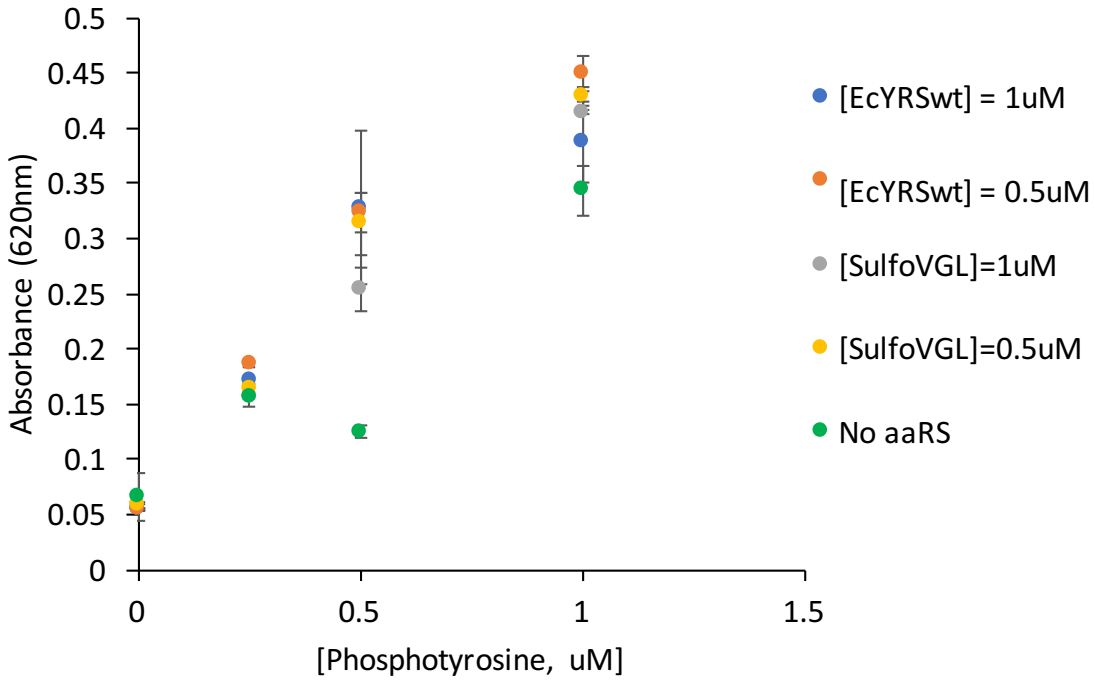


Figure 6-26. Malachite green assay with phosphotyrosine titration. Despite reducing the amount of pTyr, the signal of pTyr was too noisy, likely due to residual Pi in the pTyr stock.

We next tested the expression of the SulfoRS variants under stronger promoters in ATMY4. SulfoRS mutants under glns' (weaker) or T5-lac (stronger) promoters were transformed with an sfGFP-151TAG reporter and expressed as previously described (Figure 6-27a).[49] Transformations efficiency was poor for the SulfoRS mutants expressed from pET-22b under the T5-lac promoter (Figure 6-27b), implying the charging of Tyr to the entire proteome which might explain their toxicity even though they are under an inducible promoter. This toxicity is present in their fluorescence data, as the T5-lac expressed

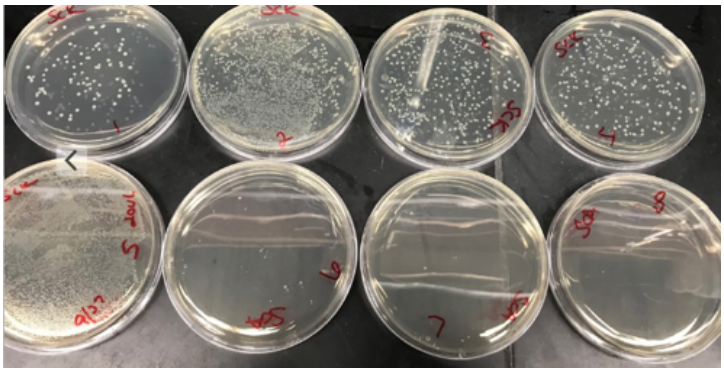
synthetases result in low sfGFP expression due to transformation artefacts (Figure 6-27c). Admittedly, these experiments are missing some controls but do corroborate the malachite green assay conclusions (Figure 6-25).

Overall, the pursuit of the malachite green assay has taught us two major points: 1) mutant synthetases are much less stable than their wild-type counterparts and 2) SulfoRS mutants charge Tyr in the absence of sTyr when overexpressed in mammalian cells. Point 1 could be an interesting branch point for new synthetase hits: would it be ideal to identify a thermostable synthetase for future selections? Is synthetase stability causing researchers to miss out on hits which can charge an UAA but are unstable in the cell? Point 2 is less clear, but is a great lesson as we proceed with synthetase hits in the future.

a)

#	Name	Plasmid 1	Plasmid 2
1	pBK-VGL	pEvolSPI-sfGFP-151TAG [5:G6]	pBK EcYRS1 sulfo Hit1-A2 [6:B2]
2	pBK-VGM	pEvolSPI-sfGFP-151TAG [5:G6]	pBK EcYRS1 sulfo Hit2-B7 [6:B3]
3	pBK-VGL-R265	pEvolSPI-sfGFP-151TAG [5:G6]	pBK-Sulfo-A1-VGL-R265 [7:B2]
4	pBK-VGM-R265	pEvolSPI-sfGFP-151TAG [5:G6]	pBK-Sulfo-B7-VGM-R265 [7:B3]
5	pBK-VGYI	pEvolSPI-sfGFP-151TAG [5:G6]	pBK-EcYRS-SulfoRS-VGYI [Katie]
6	pET22b-VGL	pEvolSPI-sfGFP-151TAG [5:G6]	pET22b-EcYRS-SulfoVGL-HisN [8:I4]
7	pET22b-VGM	pEvolSPI-sfGFP-151TAG [5:G6]	pET22b-EcYRS-SulfoVGM-HisN [8:I5]
8	pET22b-VGYI	pEvolSPI-sfGFP-151TAG [5:G6]	pET22b-EcYRSVGYI-D265-HisN [9:A8]

b)



c)

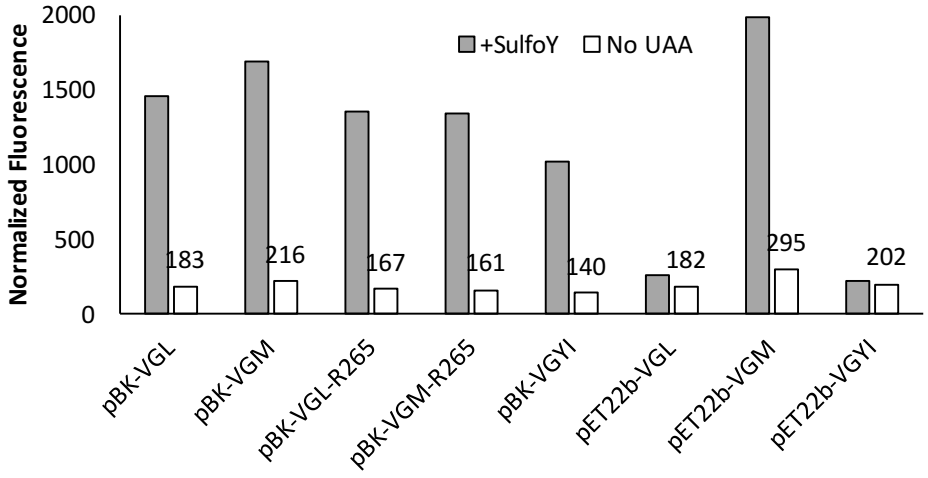


Figure 6-27. Analysis of bacterial overexpression of SulfoRS with sfGFP-151TAG reporter. a) List of plasmid combinations. b) Transformation plates 1-8 (read top left, to right) corresponding to the list. c) Normalized fluorescence of SulfoRS mutants under glnS' (pBK) or T5-lac (pET22b). The pET22b data is questionable, as the transformation efficiency was poor and therefore the lack of activity may be due to cheaters/non-legit colonies. We predict this is actually a toxicity concern, as the higher expression of the SulfoRS hits results in

proteome wide incorporation of tyrosine to TAG codons, killing the cell similarly to the EcYRSwt (which cannot be transformed even in pBK).

6.5 TPST HEK293T KO strain engineering via CRISPR

6.5.1 Brief Overview

We have recently evolved an *E. coli* TyrRS/$tRNA^{Tyr}_{CUA}$ pair which selectively charges sulfotyrosine (Chapter 4, Chapter 6.4). With this novel pair, we have studied the effects of sTyr on protein function, specifically with Heparin Cofactor II (HCII), a serine protease inhibitor which acts to regulate thrombin levels in the blood.[35] As mentioned, endogenous tyrosyl protein sulfotransferases transfer a sulfate group onto tyrosine residues in target proteins (Figure 6-28). For more detail on HCII or TPST, see Chapter 4. Our HCII studies in Chapter 4 are limited to some extent by the inconvenient fact that HCII's two Tyr sites of sulfation, Y79 (aka Y60 in Chapter 4) and Y92 (aka Y72 in Chapter 4), are both sulfated (~89, 95%) endogenously in HEK293T. While we can improve the sulfation to >95% at both sites using our sulfotyrosine tRNA/aaRS pair, the high level of endogenous sulfation in this specific protein in this cell line limits our ability to truly investigate each residue's role. While modification of Tyr to Phe (as done in Chapter 4) can be introduced, the Phe residue has much different structure and electronic properties to Tyr. Attempts to use anti-TPST shRNA or TPST inhibitors (sodium chlorate) were unsuccessful (Data not shown). We envisioned the development of a TPST KO cell line, which would allow us to make desulfated proteins which are normally significantly sulfated *in vivo*. For example, we could produce HCII-79Tyr-92sTyr, which is currently impossible and would allow for probing of each sTyr site's activity without perturbing the native amino acid sequence. To do this, we would employ CRISPR-Cas9[36–38] to sequentially inactivate TPST1 and TPST2 via frameshift mutations. Without TPSTs, there

should be no sulfation unless incorporated with our aaRS/tRNA pair. Many thanks to Tyler Bechtel (Weerapana Lab) for his help and guidance in getting the CRISPR system up and running in our lab.

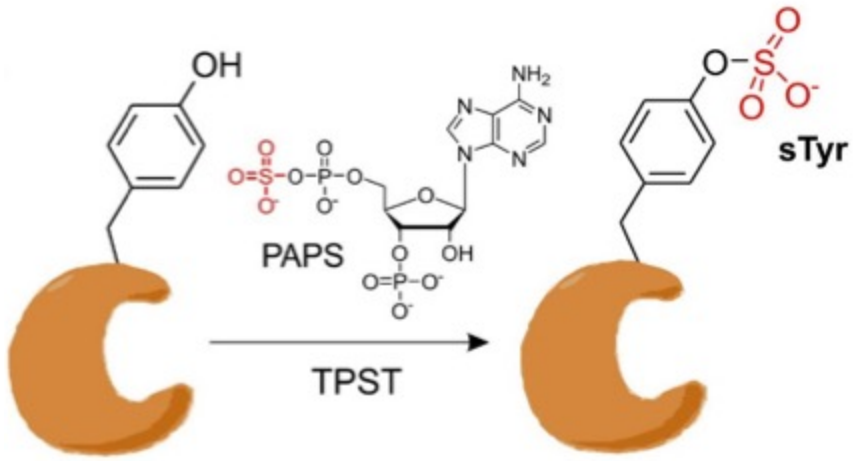
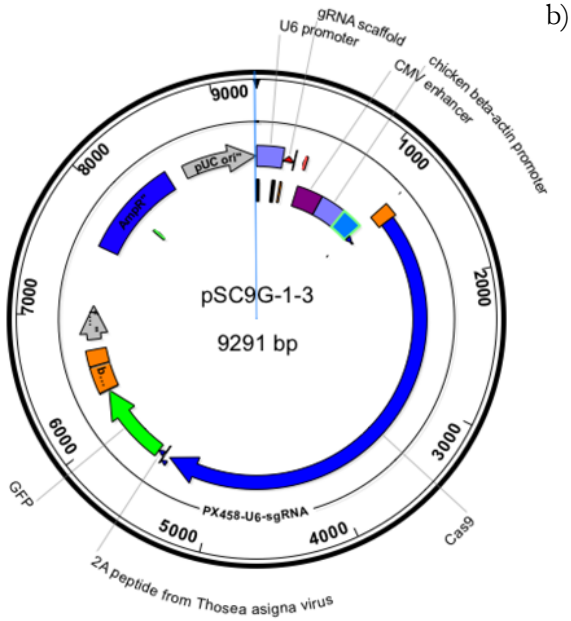


Figure 6-28. Tyrosylprotein sulfotransferase scheme.

6.5.2 Results and Discussion

pSpCas9(BB)-2A-GFP was purchased from Addgene (#48138). Target gRNA sequences (Figure 6-29b) were cloned into this plasmid with the Zhang Lab Target Sequence Cloning Protocol to create pSC9G-X (Figure 6-29a). 3000 ng pSC9G-X plasmids were transfected into HEK-293T under Chatterjee Lab transfection conditions[28] for SigmaPEI, MaxPEI, or MaxPEI + 2 mM sodium butyrate. MaxPEI maintained a high level of cell viability while achieving significant transfection efficiency and was therefore used for all future CRISPR transfections.

a)



b)

#	Plasmid	Target	sgRNA ID	gRNA sequence
8	pSC9G-1.3	TPST1	3	gcacatcctgacattcgctg
9	pSC9G-1.13	TPST1	13	ggaatgccatcaccggatag
1	pSC9G-2.2	TPST2	2	cgtggaataccgctatggca
2	pSC9G-2.3	TPST2	3	tgccatagcggtattccacg
3	pSC9G-2.14	TPST2	14	gcaggtgctagagtgccggg

Figure 6-29. CRISPR a) plasmid map and b) gRNA.

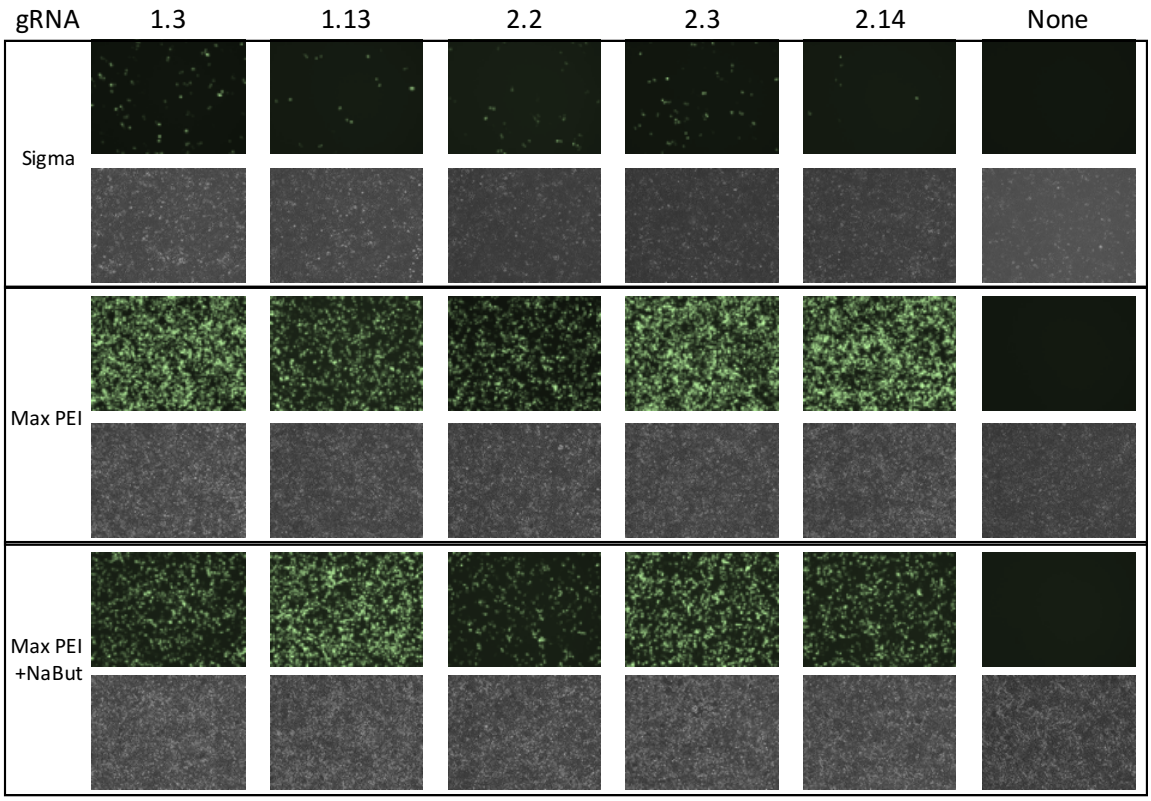
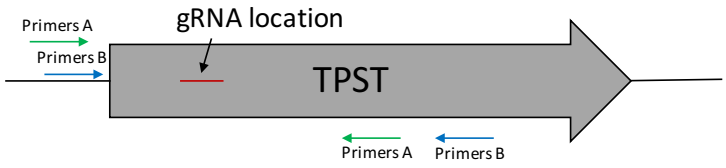


Figure 6-30. Transfection fluorescence and white light images of HEK293T transfected with pSC9G-X plasmids. gRNA notation at the top corresponds to Figure 6-29b.

We first tested the indel formation efficiency of the gRNAs in Cas9 transfected cells. Primers were designed flanking the gRNA location (Figure 6-31) as specified Ran *et. al.*[37] Generally, one set of primers is approximately 300 bp from the gRNA location, while the other is approximately 700 bp, resulting in a 1 kbp PCR product. Two sets of primers, A and B (TPST1) or C and D (TPST2), were designed due to prior knowledge regarding challenges with genomic PCR.[28] Cells were pelleted 48 post-transfection, the genome was extracted (see methods), and PCR products were amplified via Phusion Hot Start II polymerase.[28] Ideally, PCR reactions will result in a clean band which will make downstream image processing more accurate (reaction D, Figure 6-31c). While the DNA products look homogenous after PCR, there should be a proportion which have been edited by Cas9 and contain mutations varying from single base-pair insertions to larger deletions. These indel mutations would not appear by standard PCR but would via the surveyor assay.[37] Therefore, we first heteroduplexed the PCR products and characterized the indel formation by measuring band intensity on analytical gels (Figure 6-32). Successful indel formation should be visible by the appearance of the two bands ~300/700bp. gRNA 1.3 with Primer combo A and gRNA 2.14 with primer combo C were the best combinations, giving 31% and 15.9% indel formation, respectively (Figure 6-32).

a)



b)

Primer combo	Expected size (bp)
A (1.3, F/R)	1108
B (1.3, F2/R2)	1243
C (2.3, F/R)	1003
D (2.3, F2/R2)	908
Primers for Surveyor kit assay	
1.3-surv-F	agATGTTGGTTATCTTTCTGAAGTAGACTGTCC
1.3-surv-R	CCTCATCAATGAAGCTGTATCTGTTCCAGAAAAGG
1.3-surv-F2	GGTGATAACTGGTGACTGATTTGTAAGCAATCTCAATGTAATG
1.3-surv-R2	GAAAAAATGAAGTATAAATAAGTGAACATCTGTAATAAACTAGTAGAG
2.3-surv-F	AGATGTGAGCCTGGCGGGCTGCCCGCTAACCTGTC
2.3-surv-R	ctgattttcttgcctggcgcctctgactcc
2.3-surv-F2	CTAACCTGTCGCTGAAGCCCCAGAAGCG
2.3-surv-R2	CTTGGACAGGGAGACACCACCGGGCTTGCC

c)

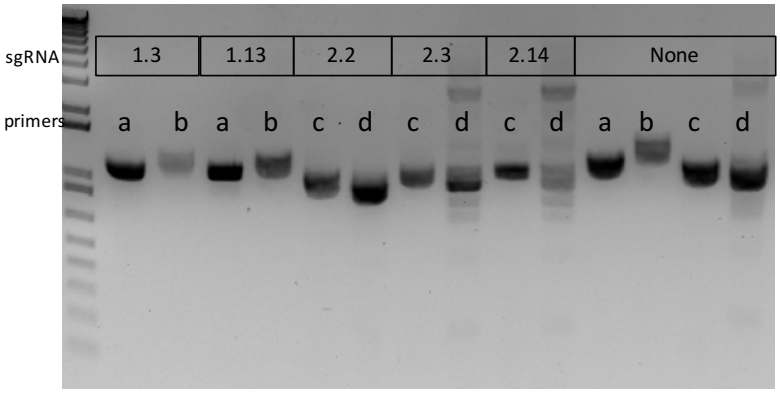
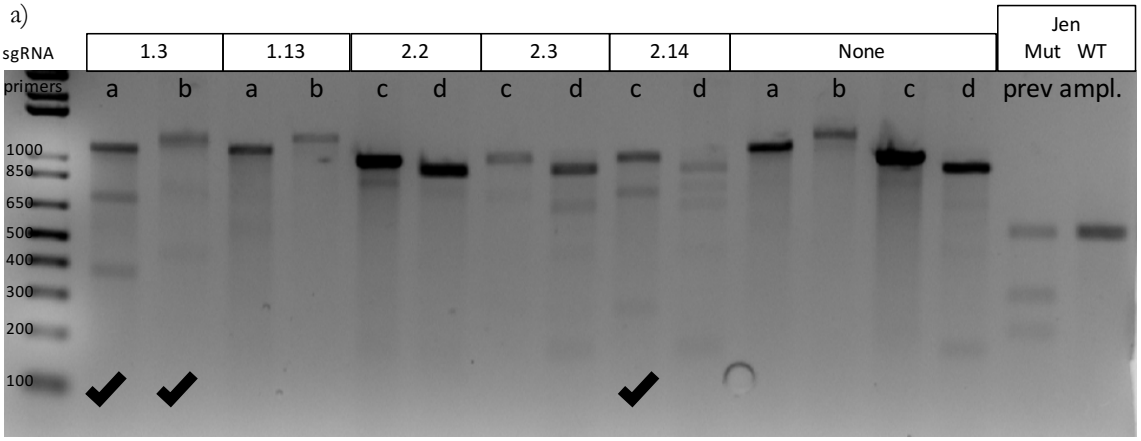


Figure 6-31. Amplification of genomic PCR products for surveyor assay. a) Scheme of TPST1 or 2 with gRNA locations. gRNA must sit 300 and 700 bp from the gRNA location. b) List of primer combos, expected sizes, and primers for the surveyor kit assay. c) Analytical gel of genomic PCR products. sgRNA is noted, as well as primer combo A or B for TPST1/2. Each TPST has two sets of sequencing primers used for testing. Some amplifications were cleaner than others. "None" is a no gRNA control.

a)



b)

Primer combo	Expected size (bp)	With cleavage of heterodimer				
		1.3	**1.13**	**2.2**	**2.3**	**2.14**
A	1108	452/735	209/919			
B	1243	465/801	278/985			
C	1003			187/836	286/737	286/737
D	908			162/766	261/667	261/667

c)

	indel %
1.3 - primer A	31
1.3 - primer B	35.2
2.14 - primer C	15.9
Jenny control	32.6

Figure 6-32. Indel formation analysis via surveyor assay. a) Analytical gel of surveyor assay. sgRNA labeling and primer combo labeling are the same as Figure 6-31. Check marks indicate indel formation. b) Expected sizes of bands. c) Indel formation percentages.

Initially, TPST1 was targeted for Cas9 mediated knockout with sgRNA 1.3 via standard PEIMax transfection with pSC9G-1.3 (Figure 6-33). Surveyor assay analysis of this transfection suggested 18% indel formation, down from the previous 31% (see Crispr methods for equation). This could be due to worse transfection efficiency. With low efficiencies, one could FACS sort the fluorescent population receiving the Cas9, but this route was not pursued. 48 hrs post transfection, cells were diluted to 0.5 cells/well in 96 well plates for single colony propogation. 20-30 clonal strains were tested via Tide analysis and one hit resulted in complete frameshift mutations in TPST1 (Figure 6-34). This strain was named 826 and contains a fully frameshifted TPST1.KO by Tide analysis. Western blot of cell pellets prepared with three different methods identified a knockdown of TPST1 (Figure 6-35).

Unfortunately, this anti-TPST1 antibody may have some cross-reactivity with TPST2. However, the blot does look promising, as the standard sample preparation indicates a lower amount of TPST1/2 in the 826 strain. Additionally, the modified protein preparation was performed to resolubilize any insoluble fractions which may be in the cell pellet. In the 826 lane of this prep (Figure 6-35, last lane), we see what could possible be the ~10 kDa truncated TPST1 which would be formed by a frameshifted stop-codon shortly after the gRNA site. While promising, better western blot data should be obtained to confirm TPST1/2 KO in the final 100% KO strain.

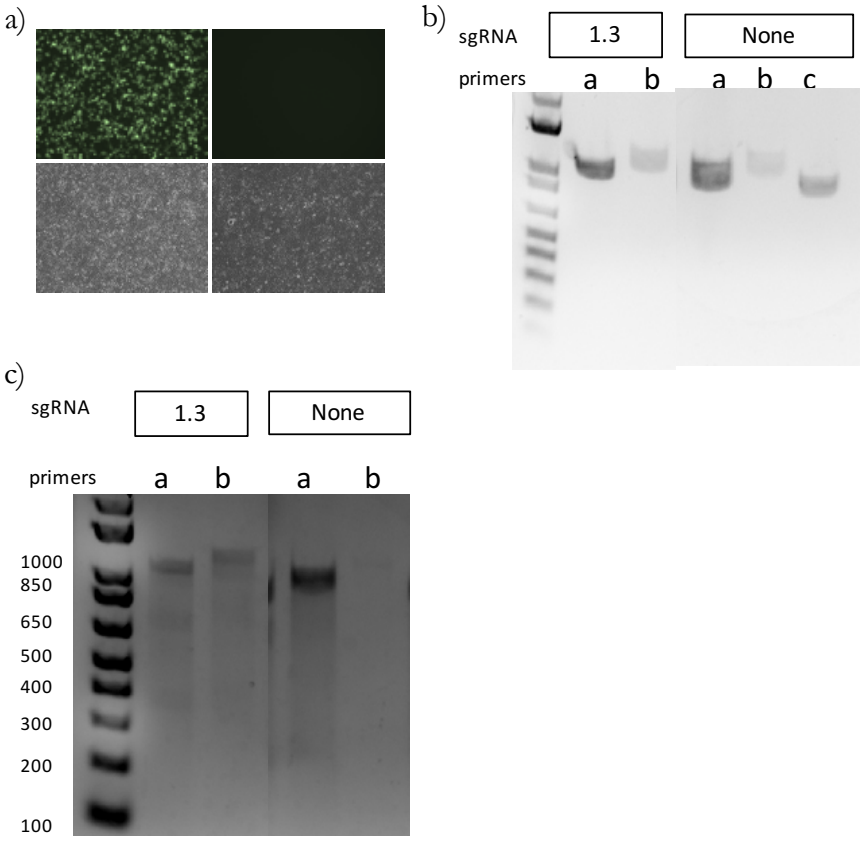


Figure 6-33. CRISPR targeting of TPST1 with gRNA 1.3. a) Fluorescence and white light images of pSC9G-1.3 transfection into HEK293T. b) Amplification of transfected and untransfected extracted genomic DNA. c) Surveyor assay, indicating 18% indel formation.

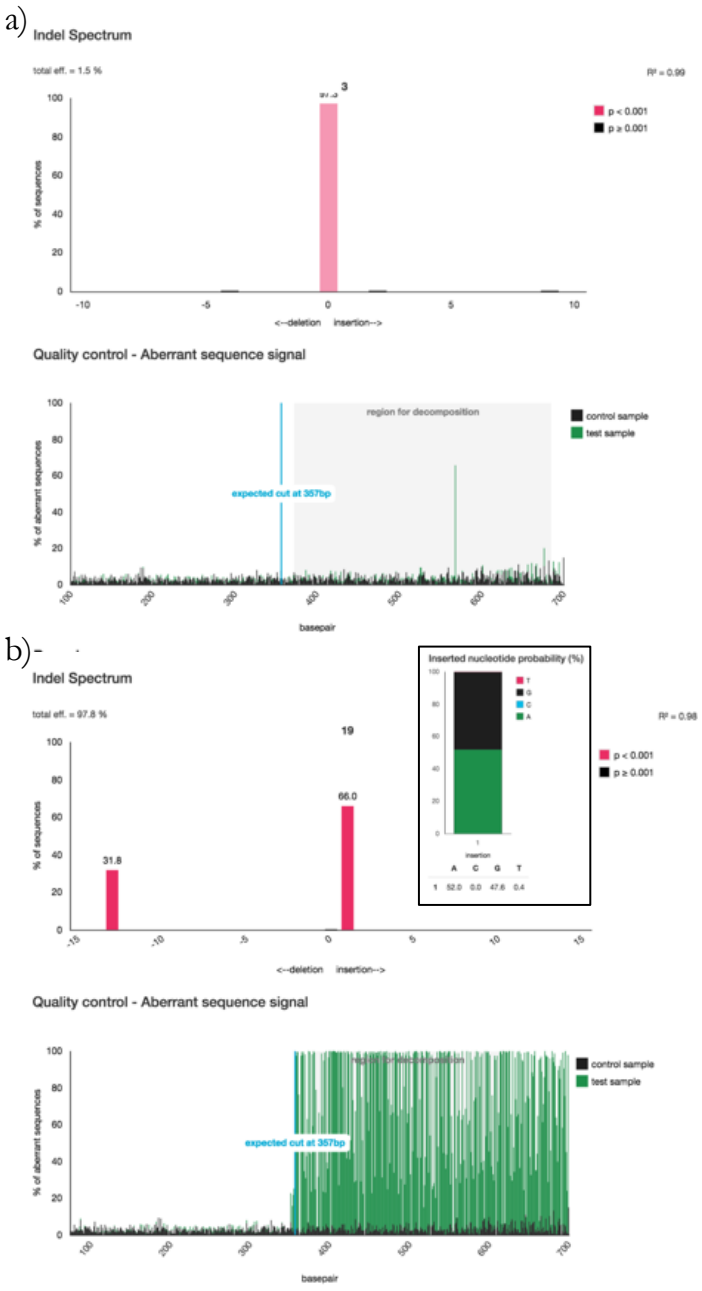


Figure 6-34. Tide Analysis of TPST1 KO strain. a) Representative UI of control sample and a non-CRISPR strain. The test sample (no CRISPR) and control sample sequencing traces overlap, indicating no mutations. b) Strain 826 Tide analysis, which shows significant deviation from the control sample. HEK293T are homotriploid. Strain 826 has two instances of single base-pair insertions (frameshift) as well as a -13 mutation which causes a frameshift.

Strain	WT	826	WT	826	WT	826
Cell Lytic M	x	x				
60 uL 1x SDS directly			x	x		
60 uL 1x SDS + Spin					x	x

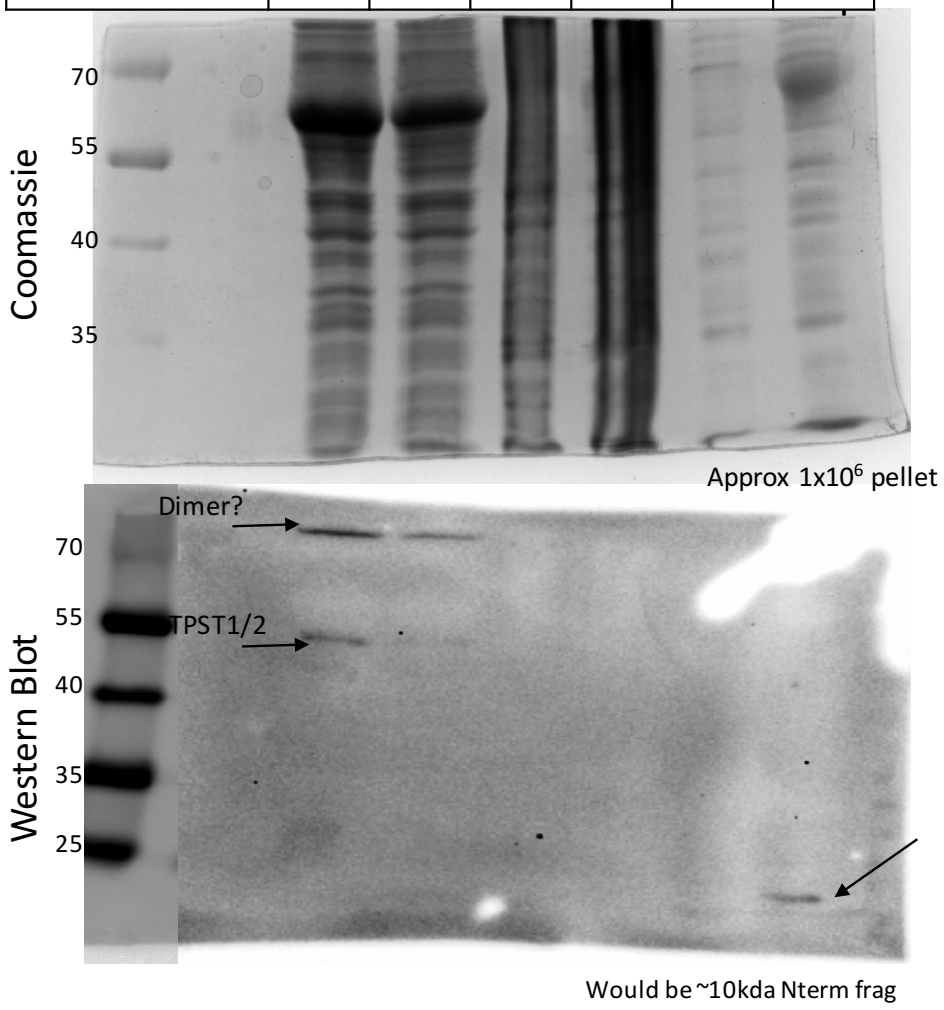
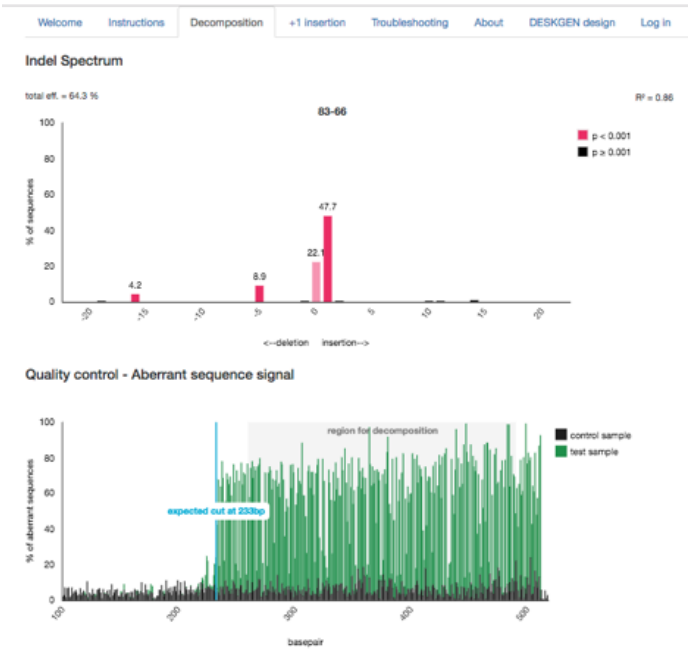


Figure 6-35. Western blot analysis of 826 Strain (TPST1.KO). Protocol in Methods section.

Next, we attempted to use strain 826 (TPST1.KO) to make a full TPST KO strain. gRNA 2.14 was used and all CRISPR protocols were followed to isolate 40 potential hits. Of the 40 hits, only 3 hits had mutations (poor efficiency) when analyzed by Tide/Synthego

analysis. However, one strain had 70% TPST2 KO (Figure 6-36). This strain was termed 83-66 and is HEK293T TPST1.KO TPST2.70%KO. Unfortunately, there is some residual TPST2 expression, which we are currently work on removing with a final round of CRISPR on strain 83-66. Note: a TPST1/2 KO cell line with >92% KO was recently isolated by JI and stocked (5/15/2019).

a)



b)

Figure 6-36. Tide and synthego analysis of strain 83-66. a) Tide analysis of strain 83-66 indicates 75% frameshift. b) Synthego analysis concurs and identifies sequence likelihood in framseshifts.

Lastly, we wanted to test the effect of TPST knockdown using HCII. We have previously shown how to isolated HCII (Chapter 4), which has shown sulfation of 79Y and 92Y of 85% and 95%, respectively, in HEK293T WT. We purified HCII-wt, HCII-79TAG + sTyr, and HCII-92TAG + sTyr from 5 x 10cm transfections of 83-66 and received poor yields due to poor cell line health (Figure 6-37a). Cell line health was remedied by recovery of a new stock. Tryptic/elastase spectrometry analysis indicated no detection of sTyr and only detected Tyr in both sites (Figure 6-37b). Mass spec results are shown below (Figure 6-38 and Figure 6-39). This result is indicative of significant knockdown of TPST1/2.

While the lack of Tyr in HCII is promising, it does not definitively show the elimination of any sulfation. It is possible the sulfation is below the detection limit. Regardless, better western blot analysis of TPST1/TPST2 will be performed, as well as tryptic mass spec controls which may allow for insight into the detection limits of sulfation.

a)

	F9	79TAG	92TAG
Abs	0.228	0.217	0.236
ng/ul	9.9	4.1	14.1
approx elution volume (ul)	500	500	500
Yield (ng)	4947	2053	7053
Yield per 10cm dish (ug)	0.8	0.3	0.9

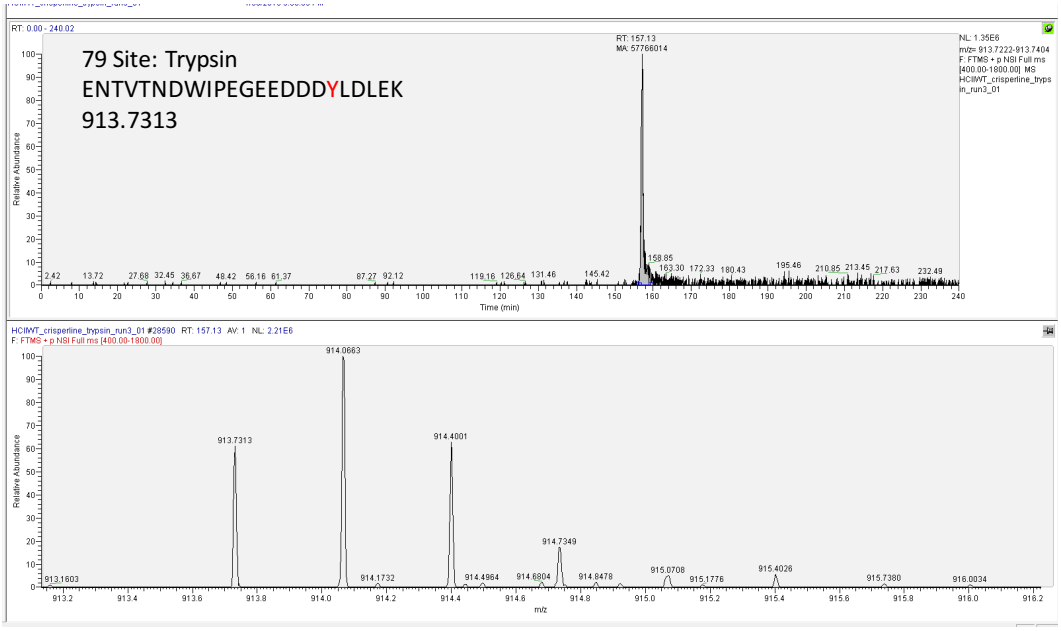
b)

79 site - Trypsin			AREA INTEGRATION	
ENTVTNDWIPEGEEDDDYLDLEK	Tyrosine	Sulfated	Tyr	Sulfo
M	2738.1719	2818.1288		
(M+H)+	2739.1792	2819.1360		
(M+2H)2+	1370.0933	1410.0717	11135026	
(M+3H)3+	913.7313	940.3836	57766014	
(M+4H)4+	685.5503	705.5395		
			68901040	0
			% sulfated	0.0000

92 site - Elastase			AREA INTEGRATION	
FSEDDDYIDIV	Tyrosine	Sulfated	Tyr	Sulfo
Exact mass	1329.5612	1409.5180		
(M+H)+	1330.5685	1410.5253	75778754	
(M+2H)2+	665.7879	705.7663	170059501	
(M+3H)3+	444.1943	470.8466		
			245838255	0
			% sulfated	0.0000

Figure 6-37. HCII tryptic mass spec overview. a) Yields of HCII. b) Ion abundance of HCII variants.

a)



b)

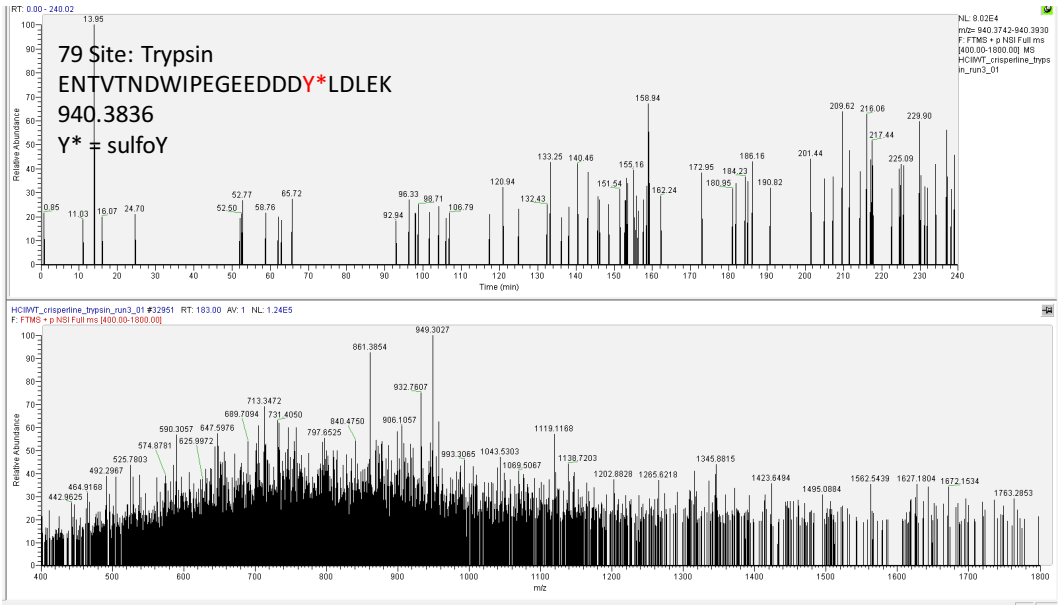
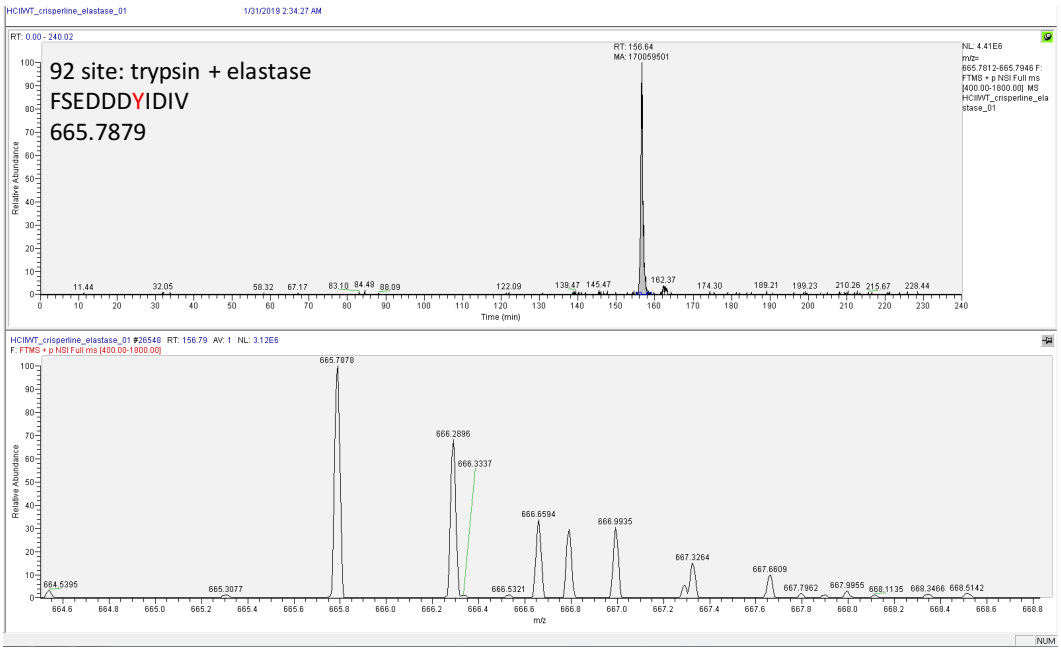


Figure 6-38. Tryptic mass spec analysis of site 79. Searching for a) tyrosine or b) sulfotyrosine in the target peptide.

a)



b)

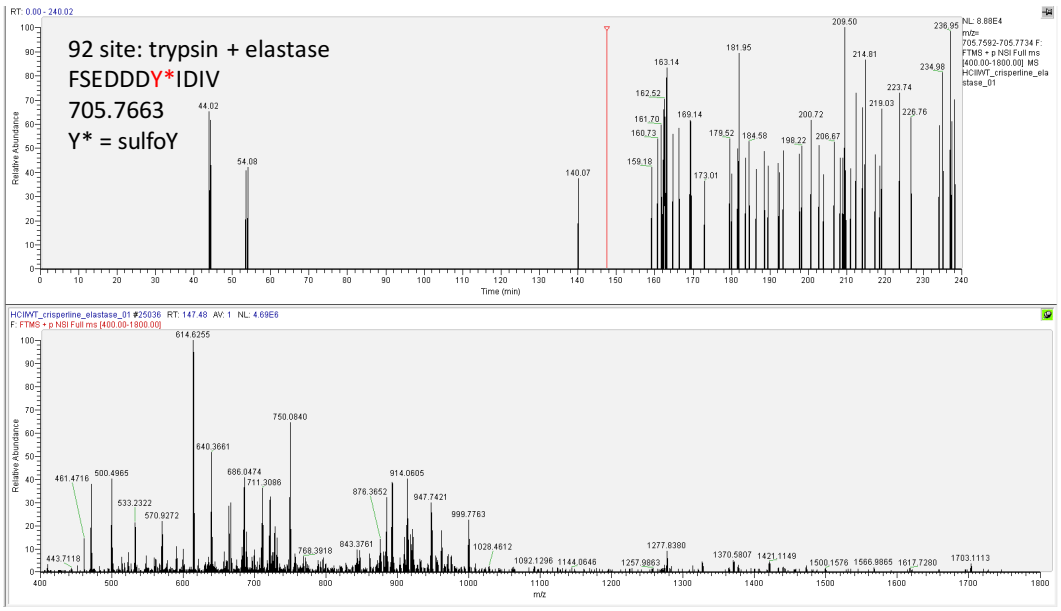


Figure 6-39. Tryptic mass spec analysis of site 92. Searching for a) tyrosine or b) sulfotyrosine in the target peptide.

6.6 Additional methodology and modifications

Additional methodology detail is described in this section. In some cases, protocols were provided in the context of the short story. This section is reserved for more extensive protocol detail. Commonly performed protocols such as transfection, growth curves, coomassie/anti-His western blots, and expression of sfGFP/EGFP as well as their purification have been extensively covered in previous thesis chapters and publications, with only the plasmid combinations subject to change (Specified in story context). These commonly performed protocols have been detailed in the story context.

6.6.1 pKD46 Recombination

First, the complementation plasmid (pUltra ScW40) and pKD46 were transformed into BL21. Due to temperature sensitivity, *E. coli* containing pKD46 must be grown at 30°C to maintain the plasmid. Overnight starter cultures were used to inoculate cultures at a 1:100 ratio in 2 mL LB + antibiotics. At 0.1 OD_{600}, 10 mM final L-arabinose was added to induce pKD46 λ-Red expression (include a no arabinose control). Allow the cultures to grow at 30°C to 0.4 OD_{600}. Aliquot 1 mL of each culture into two 1.5 mL centrifuge tubes and chill in ice-water bath for 10 min. After bath, centrifuge for 10 min at 4000 x g and 4°C. Aspirate supernatant and resuspend pellets in 1 mL ice-cold sterile ddH_2O. Repeat wash, centrifuging for 10 min at 4000 x g and 4°C. Resuspend pellet in 50 μL sterile ddH_2O. These 50 μL aliquots are now ready for electroporation with dsDNA or ssDNA oligo cassettes for λ-Red recombination. Add 5 pg to 0.5 μg DNA to cells and electroporate/recover via standard protocol with a minor modification to incubate for 120 min prior to plating on selection plates. If a subsequent round of recombination needs to be performed, recover and plate at 30°C. However, if this is the final round of recombination, recover and plate at 37°C to remove

pKD46. If pKD46 removal is proving tricky, one can try growth at up to 42°C for extended lengths of time. The absence of pKD46 can be confirmed by plating on ampicillin. Colony PCR was performed as described (Chapter 2).[28]

6.6.2 Malachite Green Assay

Aminoacylation assays were performed in aminoacylation buffer (50 mM Hepes-KOH pH 7.5, 20 mM NaCl, 10 mM MgCl_2, 1 mM DTT) incubated with 0.2 mM ATP, 2 units/mL PPiase, 1 mM amino acid (Tyr, sTyr, Gly, or no UAA) and [aaRS] at 1.0 μM, 0.5 μM, 0.25 μM for 30 min at 37°C. Total reaction volume was 25 μL. After 30 min, reactions were quenched with 100 μL Malachite Green (MG, Echelon Biosciences). After the addition of MG, reactions were allowed to develop at room temperature for 20 min. Absorbance was measured at 620 nm by SpectroMax Plate Reader.

Synthetase activity of the aaRS can be by quantified by the Pi present in the reaction. Absorbance was related to pmols of Pi per well through the equation, $y = 0.0001x+0.054$, where y = absorbance and x = pmol/well Pi. This equation was derived from a standard standard curve, repeated in triplicate of KH_2PO_4 serially diluted from 100 μM and quenched with malachite green as with the general assay. The reaction was processed as previously described.

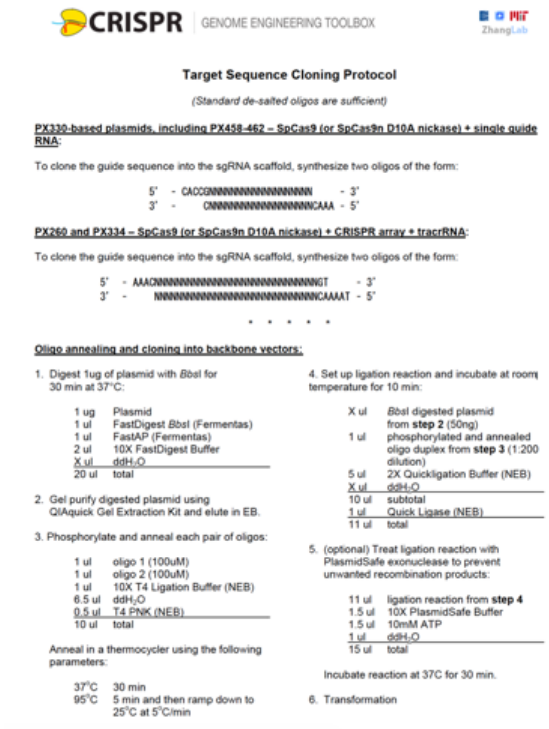
The phosphotyrosine gradient was performed as previously described with two protocol modifications. We attempted to optimize the assay by using a concentration gradient of phosphotyrosine (1.5 μM, 1 μM, 0.5 μM, 0 μM). Synthetase concentration was held constant at 1 μM, with otherwise identical conditions/incubation time.

6.6.3 Purification of EcYRS mutants

Starter cultures of ATMY5 pET22b-EcYRS-HisN (or mutants) were grown overnight in LB + Spec/Amp. Starter cultures were used to inoculate 100 mL cultures via 1:100 dilution and were grown at 37°C OD_{600} of 0.5 and induced with IPTG and expressed overnight (~12 hours) while shaking at 37°C, 250 rpm. AT 0.5 OD_{600}, cultures were induced with 1 mM IPTG and grown for ~18 hrs at 30°C. Cells were pelleted and processed for western blot analysis or for purification. Purification was performed as previously described,[28] except for use of 45 mM imidazole in wash buffers and using Hepes based buffers (50 mM Hepes-KOH pH 7.5, 20 mM NaCl, pH 7.5) rather than PBS.

6.6.4 CRISPR-Target Sequence Cloning Protocol

Target Sequence Cloning Protocol from Tyler was used to clone the gRNA sequences. It has been pasted below, as it seemed easier to display the enzyme/gRNA schematics directly.



CRISPR GENOME ENGINEERING TOOLBOX

ZhangLab

Target Sequence Cloning Protocol

(Standard de-salted oligos are sufficient)

PX330-based plasmids, including PX458-462 – SpCas9 (or SpCas9n D10A nickase) + single guide RNA:

To clone the guide sequence into the sgRNA scaffold, synthesize two oligos of the form:

5' - CACCGNNNNNNNNNNNNNNNNNNN - 3'
3' - CNNNNNNNNNNNNNNNNNNNCAAA - 5'

PX260 and PX334 – SpCas9 (or SpCas9n D10A nickase) + CRISPR array + tracrRNA:

To clone the guide sequence into the sgRNA scaffold, synthesize two oligos of the form:

5' - AAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT - 3'
3' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAAT - 5'

* * * * *

Oligo annealing and cloning into backbone vectors:

1. Digest 1ug of plasmid with *Bbs*I for 30 min at 37°C:

1 ug	Plasmid
1 ul	FastDigest *Bbs*I (Fermentas)
1 ul	FastAP (Fermentas)
2 ul	10X FastDigest Buffer
X ul	ddH_2O
20 ul	total

2. Gel purify digested plasmid using QIAquick Gel Extraction Kit and elute in EB.

3. Phosphorylate and anneal each pair of oligos:

1 ul	oligo 1 (100uM)
1 ul	oligo 2 (100uM)
1 ul	10X T4 Ligation Buffer (NEB)
6.5 ul	ddH_2O
0.5 ul	T4 PNK (NEB)
10 ul	total

Anneal in a thermocycler using the following parameters:

37°C	30 min
95°C	5 min and then ramp down to 25°C at 5°C/min

4. Set up ligation reaction and incubate at room temperature for 10 min:

X ul	*Bbs*I digested plasmid from **step 2** (50ng)
1 ul	phosphorylated and annealed oligo duplex from **step 3** (1:200 dilution)
5 ul	2X Quickligation Buffer (NEB)
X ul	ddH_2O
10 ul	subtotal
1 ul	Quick Ligase (NEB)
11 ul	total

5. (optional) Treat ligation reaction with PlasmidSafe exonuclease to prevent unwanted recombination products:

11 ul	ligation reaction from **step 4**
1.5 ul	10X PlasmidSafe Buffer
1.5 ul	10mM ATP
1 ul	ddH_2O
15 ul	total

Incubate reaction at 37C for 30 min.

6. Transformation

6.6.5 CRISPR-Genomic isolation and surveyor assay

All CRISPR protocols are extensively outlined in Ran *et. al.* [37] However, some modifications to this protocol were more efficient in our hands. When performing the genome extraction, this modified thermocycler protocol was used: 200 μL per 6 well, QuickExtract 65°C 15:00; 68°C 15:00, 98°C 10:00, 4°C 0:00. Step 79 Heteroduplex formation was performed with 10x GeneAmp PCR Buffer (Ref 4486219, Applied Biosciences Life Technologies). Reaction was annealed with the Heteroduplex protocol on the BioRad PCR: 95°C 10:00, 85°C 1:00 Ramp 2°C/s, 75°C 1:00 Ramp 0.3°C/s, 65°C 1:00 Ramp 0.3°C/s, 55°C 1:00 Ramp 0.3°C/s, 45°C 1:00 Ramp 0.3°C/s, 35°C 1:00 Ramp 0.3°C/s, 25°C 1:00 Ramp 0.3°C/s, 4°C 0:00 Ramp 0.3°C/s. Surveyor nuclease S digestion (Step 81) was performed as described with one modification: An additional 15 mM $MgCL_2$ was added to the reaction. PCR band intensity was measured with the ChemiDoc imager.

6.6.6 Isolation of single cell lines

After 48 hrs post transfection, cells were mechanically dissociated from the 6 well dish, serially diluted to 50 cells/mL, and 100 μL (~0.5 cells per well) was seeded in 96 well plates. Usually, 5-10x 96 well plates were seeded. Single wells were checked for growth 1.5 weeks after seeding. Wells with HEK293T growth radiating from a single source (rounded colonies radiating form a central point) should be noted for use, while wells that received multiple cells should be crossed out and not used. When wells reach ~60% confluence, add 100 μL DMEM (some DMEM has likely evaporated. It is a good idea to add more in week 1.5) mechanically pipet the cells derived from single cell sources and transfer half to a 24 well (Note: This is a bit of a "feel thing." If cells are growing slowly, it might be advantageous to seed more).

Technically, it is possible to perform PCR amplifications of the genomic DNA in this step for indel analysis, but waiting until the 24->6well/10cm dish is acceptable. After cells are >60% confluent in 24 well cultures, they can be transferred with standard trypsin protocols for propogation in 6 well or 10 cm dishes. Scale up is necessary to stock successful CRISPR knockouts.

6.6.7 CRISPR-Tide/Synthego Indel Analysis

Isolate leftover cell pellets from 96 to 24 or higher cell propogation and extract their respective genomes using the genome extraction protocol. PCR amplify the TPST1/2 region with surveyor primers, but heteroduplex/surveyor assay are not necessary. These PCR products can be directly sent to EtonBio/sequencing service for sequencing with a terminal forward primer. Make sure to include a control strain with no Cas9 expression. Upon receiving sequences from DNA sequencing services, use either Tide (tide.deskgen.com) or Synthego (synthego.com) to analyze the individual strain genotype. Strains with indels/mutations will result in disrupted spectrums that these computational softwares can identify. Their software is intuitive and I have found that I like the UI and detail of Synthego a bit more than Tide, but analysis with both is beneficial and can confirm their findings between each other.

6.6.8 CRISPR-Anti-TPST1 western blot

Antibodies used were anti-TPST1 Polyclonal antibody Rabbit IgG (PA5-61508, Thermo Fisher/Invitrogen, 1:1000) and Goat anti-Rabbit IgG H&L HRP (Abcam ab205718, 1:1000). Standard western blot protocols were followed (See Chapter 4). Samples were either HEK293T-wild type ($4x10^6$cells) or 826 TPST1 KO strain ($1.6x10^6$ cells). sSamples were prepared slightly differently than previous western blots. "Cell Lytic M" prepared samples

were standardly prepared supernatants. "60 µL 1x SDS directly" were samples were resuspended in 60 µL SDS directly (No CellLytic M) and 30 µL was loaded onto the gel without a clarification spin. "60 µL 1x SDS + Spin" is the same as the previous, but with a clarification spin.

CRISPR-Tryptic/Elastase mass spec analysis of HCII: Mass spec was performed identically to Chapter 4 analysis of HCII. Many thanks to Jenny and the Weerapana lab for analysis of these samples. Samples in this TPST KO study are as follows: HCII-wt, HCII-79TAG, HCII-92TAG, expressed and purified as previously described in Chapter 4. Yields were significantly worse, as expression was performed in the 83-66 strain, which contains the TPST1.KO and TPST2.75%KO. This 83-66 strain was relatively sick due to the stressful propogation during single strain isolation.

6.7 References

1. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat Chem Biol* **13,** 446–450 (2017).
2. Italia, J. S., Latour, C., Wrobel, C. J. J. & Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. *Cell Chem. Biol.* **25,** 1304–1312.e5 (2018).
3. Mandell, D. J. *et al.* Biocontainment of genetically modified organisms by synthetic protein design. *Nature* (2015). doi:10.1038/nature14121
4. Chan, C. T. Y., Lee, J. W., Cameron, D. E., Bashor, C. J. & Collins, J. J. 'Deadman' and 'Passcode' microbial kill switches for bacterial containment. *Nat. Chem. Biol.* **12,** 1–7 (2015).
5. Kleiner, P., Heydenreuter, W., Stahl, M., Korotkov, V. S. & Sieber, S. A. A Whole Proteome Inventory of Background Photocrosslinker Binding. *Angew. Chemie - Int. Ed.* **56,** 1396–1401 (2017).
6. Elliott, T. S. *et al.* Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat. Biotechnol.* **32,** 465–472 (2014).
7. Hoesl, M. G. *et al.* Chemical Evolution of a Bacterial Proteome. *Angew. Chemie Int. Ed.* **6520,** n/a-n/a (2015).
8. Budisa, N. *et al.* Global replacement of tryptophan with amino- tryptophans generates non-invasive protein- based optical pH sensors. *Angew. Chem. Int. Ed. Engl.* **41,** 4066–4069 (2002).

9. Rovner, A. J. *et al.* Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518,** 89–93 (2015).
10. DeVito, J. A. Recombineering with tolC as a Selectable/Counter-selectable Marker: remodeling the rRNA Operons of Escherichia coli. *Nucleic Acids Res.* **36,** e4–e4 (2007).
11. Addy, P. S., Erickson, S. B., Italia, J. S. & Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **139,** 11670–11673 (2017).
12. Mosberg, J. A., Lajoie, M. J. & Church, G. M. Lambda Red Recombineering in Escherichia coli Occurs Through a Fully Single-Stranded Intermediate. *Genetics* **186,** 791–799 (2010).
13. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5,** 1–17 (2014).
14. Fathi-Roudsari, M., Akhavian-Tehrani, A. & Maghsoudi, N. Comparison of three Escherichia coli strains in recombinant production of reteplase. *Avicenna J. Med. Biotechnol.* **8,** 16–22 (2016).
15. Kirill, A. D. & Barry, L. W. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97,** 6640–6645 (2000).
16. Nuriel, T., Hansler, A. & Gross, S. S. Protein nitrotryptophan: Formation, significance and identification. *J. Proteomics* **74,** 2300–2312 (2011).
17. Yamakura, F. & Ikeda, K. Modification of tryptophan and tryptophan residues in proteins by reactive nitrogen species. *Nitric Oxide - Biol. Chem.* **14,** 152–161 (2006).
18. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280,** 1–9 (1998).
19. Ikeda, K. *et al.* Detection of 6-nitrotryptophan in proteins by Western blot analysis and its application for peroxynitrite-treated PC12 cells. *Nitric Oxide - Biol. Chem.* **16,** 18–28 (2007).
20. Kawasaki, H. *et al.* Nitration of tryptophan in ribosomal proteins is a novel post-translational modification of differentiated and naïve PC12 cells. *Nitric Oxide - Biol. Chem.* **25,** 176–182 (2011).
21. Uda, M., Kawasaki, H., Shigenaga, A., Baba, T. & Yamakura, F. Proteomic analysis of endogenous nitrotryptophan-containing proteins in rat hippocampus and cerebellum. *Biosci. Rep.* **32,** 521–530 (2012).
22. Heart, R., Rebrin, I. & Bre, C. Nitration of Tryptophan 372 in Succinyl-CoA : 3-Ketoacid CoA Transferase during. 10130–10144 (2007).
23. Yamakura, F. *et al.* Modification of a single tryptophan residue in human Cu,Zn-superoxide dismutase by peroxynitrite in the presence of bicarbonate. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1548,** 38–46 (2001).
24. Rebrin, I., Brégère, C., Kamzalov, S., Gallaher, T. K. & Sohal, R. S. Nitration of tryptophan 372 in succinyl-CoA:3-ketoacid CoA transferase during aging in rat heart mitochondria. *Biochemistry* **46,** 10130–10144 (2007).
25. Park, H.-S. *et al.* Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science (80-.).* **333,** 1151–1154 (2011).
26. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24,** 1436–1440 (2006).
27. Ishii, Y. *et al.* Quantification of nitrated tryptophan in proteins and tissues by high-performance liquid chromatography with electrospray ionization tandem mass

spectrometry. *J. Pharm. Biomed. Anal.* **44,** 150–159 (2007).
28. Italia, J. S. *et al.* An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat. Chem. Biol.* 1–8 (2017). doi:10.1038/nchembio.2312
29. Zheng, Y., Lewis, T. L., Igo, P., Polleux, F. & Chatterjee, A. Virus-enabled optimization and delivery of the genetic machinery for efficient unnatural amino acid mutagenesis in mammalian cells and tissues. *ACS Synth. Biol.* **6,** 13–18 (2017).
30. Italia, J. S., Latour, C., Wrobel, C. J. J. & Chatterjee, A. Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. *Cell Chem. Biol.* 1–9 (2018). doi:10.1016/j.chembiol.2018.07.002
31. Luo, X. *et al.* Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria. *Nat. Chem. Biol.* 1–7 (2017). doi:10.1038/nchembio.2405
32. Chen, G., Zhang, Y., Trinidad, J. C. & Dann, C. Distinguishing Sulfotyrosine Containing Peptides from their Phosphotyrosine Counterparts Using Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **29,** 455–462 (2018).
33. Cestari, I. & Stuart, K. A spectrophotometric assay for quantitative measurement of aminoacyl-tRNA synthetase activity. *J. Biomol. Screen.* **18,** 490–497 (2013).
34. Sungwienwong, I. *et al.* Improving target amino acid selectivity in a permissive aminoacyl tRNA synthetase through counter-selection. *Org. Biomol. Chem.* **15,** 3603–3610 (2017).
35. Baglin, T. P., Carrell, R. W., Church, F. C., Esmon, C. T. & Huntington, J. A. Crystal structures of native and thrombin-complexed heparin cofactor II reveal a multistep allosteric mechanism. *Proc. Natl. Acad. Sci.* **99,** 11079–11084 (2002).
36. Lander, E. S. The Heroes of CRISPR. *Cell* **164,** 18–28 (2016).
37. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).
38. Cong, L. *et al.* Target Sequence Cloning Protocol. *Science (80-.).* **339,** 819–824 (2013).

Appendix I

Recombination cassettes

Chapter 2 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.

*trpS::*Zeo^R PCR cassette
EM-7 promoter (red); Zeo^R (green); CYC1 terminator (blue); TrpRS flanking homology (black)
ATCAGTCTATAAATGACCTTCTGCCCGCATTAGGGCTTCCGCATAGCGAAAATCAGGA
ATCGAAAAAGGTGTTGACAATTAATCATCGGCATAGTATATtGGCATAGTATAATACG
ACAAGGTGAGGAACTAAACCATGGCCAAGCTGACCAGTGCCGTTCCGGTGCTCACCG
CGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGG
GACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATC
AGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTGCG
CGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGG
ACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTC
GCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTG
ACACGTCCGACGGCGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTT
TGTCGATATCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCCG
CTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTT
TATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTA
CAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGA
CGCTCGAAGGCTTTAATTTGCAAGCTTTCACTATTGCTGGCAAATTGCGCTTTGTTCAT
GCCGGATGCGGCGTGAACGCCTTATCCGGCCTACA

trpT::Gent^R PCR cassette
Gent^R (green); trpT flanking homology (red)
CAGTCGGTTAGAATACCTGCCTGTCACGCAGGGGGTCGCGGGTTCGAGTCCCGTCCG
TTCCGCCACCCTAATTACGCACACCGTGGAAACGGATGAAGGCACGAACCCAGTTGAC
ATAAGCCTGTTCGGTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGG
TCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCT
TGTTATGACTGTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCGTT
ACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGCAAC
GATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTATGGGC
ATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTTG
ATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTC
CGATTACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTCGAC
CAAGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAGGTTTGAGCAGCCG
CGTAGTGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGC
ATTGCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATG
TGATCTACGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAAGTT
GGGCATACGGGAAGAAGTGATGCACTTTGATATCGACCCAAGTACCGCCACCTAACA
ATTCGTTCAAGCCGAGATCGGCTTCCCGGGAAATCATCCTTAGCGAAAGCTAAGGATT
TTTTTTATCTGAAATAACCCTCTCCGAAGTAAATCCTTCTACCG

λ-RED::*galK* PCR cassette
galK (green); galK promoter (blue); lambda deletion homology (red)
GCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAACTTGAGCGA
AACGGGAAGCCTGTTGACAATTAATCATCGGCAtagtatatcggcatagtataatacgacaaggtgaggaactaaa
cccaggaggcagatcatgagtctgaaagaaaaaacacaatctctgtttgccaacgcatttggctaccctgccactcacaccattcaggcgcctggccgc

gtgaatttgattggtgaacacaccgactacaacgacggtttcgttctgccctgcgcgattgattatcaaaccgtgatcagttgtgcaccacgcgatgacc
gtaaagttcgcgtgatggcagccgattatgaaaatcagctcgacgagttttccctcgatgcgcccattgtcgcacatgaaaactatcaatgggctaacta
cgttcgtggcgtggtgaaacatctgcaactgcgtaacaacagcttcggcggcgtggacatggtgatcagcggcaatgtgccgcagggtgccgggttaa
gttcttccgcttcactggaagtcgcggtcggaaccgtattgcagcagctttatcatctgccgctggacggcgcacaaatcgcgcttaacggtcaggaag
cagaaaaccagtttgtaggctgtaactgcgggatcatggatcagctaatttccgcgctcggcaagaaagatcatgccttgctgatcgattgccgctcact
ggggaccaaagcagtttccatgcccaaaggtgtggctgtcgtcatcatcaacagtaacttcaaacgtaccctggttggcagcgaatacaacacccgtcg
tgaacagtgcgaaaccggtgcgcgtttcttccagcagccagccctgcgtgatgtcaccattgaagagttcaacgctgttgcgcatgaactggacccgat
cgtggcaaaacgcgtgcgtcatatactgactgaaaacgcccgcaccgttgaagctgccagcgcgctggagcaaggcgacctgaaacgtatgggcga
gttgatggcggagtctcatgcctctatgcgcgatgatttcgaaatcaccgtgccgcaaattgacactctggtagaaatcgtcaaagctgtgattggcgac
aaaggtggcgtacgcatgaccggcggcggatttggcggctgtatcgtcgcgctgatcccggaagagctggtgcctgccgtacagcaagctgtcgctg
aacaatatgaagcaaaaacaggtattaaagagacttttacgtttgtaaaccatcacAAGGAGCAGGACAGTGCTGAATATT
ACAACGCGGCAGCATTATGAGCTGGCAGGAGAAAATCAACGCGGC

Chapter 3 Recombination Cassettes

PCR products: The following sequences are the dsDNA PCR products that were electroporated for recombineering.

tolC::tyrS PCR Cassette
tolC (blue); recombination homology (red)
TTTTTGTACCAGCGTGAAAATGATGCGTGGAAGATTGATCGTCTTGCACCCTG
AAAAGATGCAAAAATCTTGTTGAGGCACATTAACGCCCTATGGCACGTAACGC
CAACCTTTTGCGGTAGCGGCTTCTGCTAGAATCCGCAATAATTTTACAGTTTGA
TCGCGCTAAATACTGCTTCACCACAAGGAATGCAAATGAAGAAATTGCTCCCC
ATTCTTATCGGCCTGAGCCTTTCTGGGTTCAGTTCGTTGAGCCAGGCCGAGAA
CCTGATGCAAGTTTATCAGCAAGCACGCCTTAGTAACCCGGAATTGCGTAAGT
CTGCCGCCGATCGTGATGCTGCCTTTGAAAAAATTAATGAAGCGCGCAGTCCA
TTACTGCCACAGCTAGGTTTAGGTGCAGATTACACCTATAGCAACGGCTACCG
CGACGCGAACGGCATCAACTCTAACGCGACCAGTGCGTCCTTGCAGTTAACTC
AATCCATTTTTGATATGTCGAAATGGCGTGCGTTAACGCTGCAGGAAAAAGCA
GCAGGGATTCAGGACGTCACGTATCAGACCGATCAGCAAACCTTGATCCTCAA
CACCGCGACCGCTTATTTCAACGTGTTGAATGCTATTGACGTTCTTTCCTATAC
ACAGGCACAAAAAGAAGCGATCTACCGTCAATTAGATCAAACCACCCAACGTT
TTAACGTGGGCCTGGTAGCGATCACCGACGTGCAGAACGCCCGCGCACAGTA
CGATACCGTGCTGGCGAACGAAGTGACCGCACGTAATAACCTTGATAACGCG
GTAGAGCAGCTGCGCCAGATCACCGGTAACTACTATCCGGAACTGGCTGCGC
TGAATGTCGAAAACTTTAAAACCGACAAACCACAGCCGGTTAACGCGCTGCTG
AAAGAAGCCGAAAAACGCAACCTGTCGCTGTTACAGGCACGCTTGAGCCAGG
ACCTGGCGCGCGAGCAAATTCGCCAGGCGCAGGATGGTCACTTACCGACTCT
GGATTTAACGGCTTCTACCGGGATTTCTGACACCTCTTATAGCGGTTCGAAAA
CCCGTGGTGCCGCTGGTACCCAGTATGACGATAGCAATATGGGCCAGAACAA
AGTTGGCCTGAGCTTCTCGCTGCCGATTTATCAGGGCGGAATGGTTAACTCGC
AGGTGAAACAGGCACAGTACAACTTTGTCGGTGCCAGCGAGCAACTGGAAAG
TGCCCATCGTAGCGTCGTGCAGACCGTGCGTTCCTCCTTCAACAACATTAATGC
ATCTATCAGTAGCATTAACGCCTACAAACAAGCCGTAGTTTCCGCTCAAAGCTC
ATTAGACGCGATGGAAGCGGGCTACTCGGTCGGTACGCGTACCATTGTTGAT
GTGTTGGATGCGACCACCACGTTGTACAACGCCAAGCAAGAGCTGGCGAATG

CGCGTTATAACTACCTGATTAATCAGCTGAATATTAAGTCAGCTCTGGGTACG
TTGAACGAGCAGGATCTGCTGGCACTGAACAATGCGCTGAGCAAACCGGTTT
CCACTAATCCGGAAAACGTTGCACCGCAAACGCCGGAACAGAATGCTATTGCT
GATGGTTATGCGCCTGATAGCCCGGCACCAGTCGTTCAGCAAACATCCGCACG
CACTACCACCAGTAACGGTCATAACCCTTTCCGTAACTGAACAGGGAACATGA
TGAAAAATATTCTCGCTATCCAGTCTCACGTTGTTTATGGTCATGCGGGTAAC
A

tyrTV::*tolC* PCR Cassette
tolC (blue); recombination homology (red)
AAAATAACTGGTTACCTTTAATCCGTTACGGATGAAAATTACGCAACCAGTTG
AGGCACATTAACGCCCTATGGCACGTAACGCCAACCTTTTGCGGTAGCGGCTT
CTGCTAGAATCCGCAATAATTTTACAGTTTGATCGCGCTAAATACTGCTTCACC
ACAAGGAATGCAAATGAAGAAATTGCTCCCCATTCTTATCGGCCTGAGCCTTT
CTGGGTTCAGTTCGTTGAGCCAGGCCGAGAACCTGATGCAAGTTTATCAGCAA
GCACGCCTTAGTAACCCGGAATTGCGTAAGTCTGCCGCCGATCGTGATGCTGC
CTTTGAAAAAATTAATGAAGCGCGCAGTCCATTACTGCCACAGCTAGGTTTAG
GTGCAGATTACACCTATAGCAACGGCTACCGCGACGCGAACGGCATCAACTCT
AACGCGACCAGTGCGTCCTTGCAGTTAACTCAATCCATTTTTGATATGTCGAAA
TGGCGTGCGTTAACGCTGCAGGAAAAAGCAGCAGGGATTCAGGACGTCACGT
ATCAGACCGATCAGCAAACCTTGATCCTCAACACCGCGACCGCTTATTTCAACG
TGTTGAATGCTATTGACGTTCTTTCCTATACACAGGCACAAAAAGAAGCGATC
TACCGTCAATTAGATCAAACCACCCAACGTTTTAACGTGGGCCTGGTAGCGAT
CACCGACGTGCAGAACGCCCGCGCACAGTACGATACCGTGCTGGCGAACGAA
GTGACCGCACGTAATAACCTTGATAACGCGGTAGAGCAGCTGCGCCAGATCA
CCGGTAACTACTATCCGGAACTGGCTGCGCTGAATGTCGAAAACTTTAAAACC
GACAAACCACAGCCGGTTAACGCGCTGCTGAAAGAAGCCGAAAAACGCAACC
TGTCGCTGTTACAGGCACGCTTGAGCCAGGACCTGGCGCGCGAGCAAATTCG
CCAGGCGCAGGATGGTCACTTACCGACTCTGGATTTAACGGCTTCTACCGGGA
TTTCTGACACCTCTTATAGCGGTTCGAAAACCCGTGGTGCCGCTGGTACCCAG
TATGACGATAGCAATATGGGCCAGAACAAAGTTGGCCTGAGCTTCTCGCTGCC
GATTTATCAGGGCGGAATGGTTAACTCGCAGGTGAAACAGGCACAGTACAAC
TTTGTCGGTGCCAGCGAGCAACTGGAAAGTGCCCATCGTAGCGTCGTGCAGA
CCGTGCGTTCCTCCTTCAACAACATTAATGCATCTATCAGTAGCATTAACGCCT
ACAAACAAGCCGTAGTTTCCGCTCAAAGCTCATTAGACGCGATGGAAGCGGG
CTACTCGGTCGGTACGCGTACCATTGTTGATGTGTTGGATGCGACCACCACGT
TGTACAACGCCAAGCAAGAGCTGGCGAATGCGCGTTATAACTACCTGATTAAT
CAGCTGAATATTAAGTCAGCTCTGGGTACGTTGAACGAGCAGGATCTGCTGG
CACTGAACAATGCGCTGAGCAAACCGGTTTCCACTAATCCGGAAAACGTTGCA
CCGCAAACGCCGGAACAGAATGCTATTGCTGATGGTTATGCGCCTGATAGCCC
GGCACCAGTCGTTCAGCAAACATCCGCACGCACTACCACCAGTAACGGTCATA
ACCCTTTCCGTAACTGAAGTCCCTGAACTTCCCAACGAATCCGCAATTAAATAT
TCTGCCCATGCGG

tyrTV::gentR PCR Cassette
gent^{R} (green); recombination homology (red)
AAAATAACTGGTTACCTTTAATCCGTTACGGATGAAAATTACGCAACCAGACG
CACACCGTGGAAACGGATGAAGGCACGAACCCAGTTGACATAAGCCTGTTCG

GTTCGTAAACTGTAATGCAAGTAGCGTATGCGCTCACGCAACTGGTCCAGAAC
CTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTCATGGCTTGT
TATGACTGTTTTTTTTGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAGCGCG
TTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAG
CAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGC
TCAAGTATGGGCATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATC
CATGCGGGCTGCTCTTGATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCT
ACTCCCAACATCAGCCGGACTCCGATTACCTCGGGAACTTGCTCCGTAGTAAG
ACATTCATCGCGCTTGCTGCCTTCGACCAAGAAGCGGTTGTTGGCGCTCTCGC
GGCTTACGTTCTGCCCAGGTTTGAGCAGCCGCGTAGTGAGATCTATATCTATG
ATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGCATTGCCACCGCGCTCAT
CAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATGTGATCTACGTGC
AAGCAGATTACGGTGACGATCCCGCAGTGGCTCTCTATACAAAGTTGGGCATA
CGGGAAGAAGTGATGCACTTTGATATCGACCCAAGTACCGCCACCTAACAATT
CGTTCAAGCCGAGATCGGCTTCCCGGAGTCCCTGAACTTCCCAACGAATCCGC
AATTAAATATTCTGCCCATGCGG

tyrTV.inact-gblock.orig
tyrosyl tRNA (purple); recombination homology (red), anticodon underlined
AAAATAACTGGTTACCTTTAATCCGTTACGGATGAAAATTACGCAACCAGTTCA
TTTTTCTCAACGTAACACTTTACAGCGGCGCGTCATTTGATATGATGCGCCCCG
CTTCCCGATAAGGGAGCAGGCCAGTAATTCACCACAGCGATGTGGTAACTTCG
GTGAGATTGTGAAAGTTTGCACTCATCGGGGAAGGGTGAGAACCTTCGACTA
AGGTTCGATTCGAGCGAAAGCGAGAGAACGTTGCCACAGGCAACGGCCCGAA
GGGTGAGGAGCGAAGCGACGAATAATCCTTCCCCCACCACCATCTCAAGCATT
ACCTCAAAATTTGAATTACCCCTGGTGGGGGTTCCCGAGCGGCCAAAGGGAGC
AGACTAAATCTGCCGTCATCGACTTCGAAGGTTCGAATCCTTCCCCCACCACCA
TCACTTTCAAAAGTCCCTGAACTTCCCAACGAATCCGCAATTAAATATTCTGCC
CATGCGG

tyrTV.CUA-gblock.orig
tyrosyl tRNA (purple); recombination homology (red), anticodon underlined
AAAATAACTGGTTACCTTTAATCCGTTACGGATGAAAATTACGCAACCAGTTCA
TTTTTCTCAACGTAACACTTTACAGCGGCGCGTCATTTGATATGATGCGCCCCG
CTTCCCGATAAGGGAGCAGGCCAGTAATTCACCACAGCGATGTGGTAACTTCG
GTGAGATTGTGAAAGTTTGCACTCATCGGGGAAGGGTGAGAACCTTCGACTA
AGGTTCGATTCGAGCGAAAGCGAGAGAACGTTGCCACAGGCAACGGCCCGAA
GGGTGAGGAGCGAAGCGACGAATAATCCTTCCCCCACCACCATCTCAAGCATT
ACCTCAAAATTTGAATTACCCCTGGTGGGGGTTCCCGAGCGGCCAAAGGGAGC
AGACTCTAAATCTGCCGTCATCGACTTCGAAGGTTCGAATCCTTCCCCCACCAC
CATCACTTTCAAAAGTCCCTGAACTTCCCAACGAATCCGCAATTAAATATTCTG
CCCATGCGG

tyrU::gentR
gent^{R} (green); recombination homology (red)
GTAATCAGTAGGTCACCAGTTCGATTCCGGTAGTCGGCACCATCAAGTCCAAG
GAATGCAAATGTTACGCAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCCC
TAAAACAAAGTTAGGTGGCTCAAGTATGGGCATCATTCGCACATGTAGGCTCG

GCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTTGATCTTTTCGGTCGTGAG
TTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCCGATTACCTCGG
GAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTCGACCAAGAAG
CGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAGGTTTGAGCAGCCGCGT
AGTGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGG
GCATTGCCACCGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGT
GCTTATGTGATCTACGTGCAAGCAGATTACGGTGACGATCCCGCAGTGGCTCT
CTATACAAAGTTGGGCATACGGGAAGAAGTGATGCACTTTGATATCGACCCAA
GTACCGCCACCTAACAATTCGTTCAAGCCGAGATCGGCTTCCCGGGGCCACGC
GATGGCGTAGCCCGAGACGATAAGTTCGCTTACCGGCTCGAA

tyrU::*tolC*
tolC (blue); recombination homology (red)
GTAATCAGTAGGTCACCAGTTCGATTCCGGTAGTCGGCACCATCAAGTCCAAG
GAATGCAAatgAAGAAATTGCTCCCCATTCTTATCGGCCTGAGCCTTTCTGGGTT
CAGTTCGTTGAGCCAGGCCGAGAACCTGATGCAAGTTTATCAGCAAGCACGCC
TTAGTAACCCGGAATTGCGTAAGTCTGCCGCCGATCGTGATGCTGCCTTTGAA
AAAATTAATGAAGCGCGCAGTCCATTACTGCCACAGCTAGGTTTAGGTGCAGA
TTACACCTATAGCAACGGCTACCGCGACGCGAACGGCATCAACTCTAACGCGA
CCAGTGCGTCCTTGCAGTTAACTCAATCCATTTTTGATATGTCGAAATGGCGT
GCGTTAACGCTGCAGGAAAAAGCAGCAGGGATTCAGGACGTCACGTATCAGA
CCGATCAGCAAACCTTGATCCTCAACACCGCGACCGCTTATTTCAACGTGTTGA
ATGCTATTGACGTTCTTTCCTATACACAGGCACAAAAAGAAGCGATCTACCGTC
AATTAGATCAAACCACCCAACGTTTTAACGTGGGCCTGGTAGCGATCACCGAC
GTGCAGAACGCCCGCGCACAGTACGATACCGTGCTGGCGAACGAAGTGACCG
CACGTAATAACCTTGATAACGCGGTAGAGCAGCTGCGCCAGATCACCGGTAAC
TACTATCCGGAACTGGCTGCGCTGAATGTCGAAAACTTTAAAACCGACAAACC
ACAGCCGGTTAACGCGCTGCTGAAAGAAGCCGAAAAACGCAACCTGTCGCTG
TTACAGGCACGCTTGAGCCAGGACCTGGCGCGCGAGCAAATTCGCCAGGCGC
AGGATGGTCACTTACCGACTCTGGATTTAACGGCTTCTACCGGGATTTCTGAC
ACCTCTTATAGCGGTTCGAAAACCCGTGGTGCCGCTGGTACCCAGTATGACGA
TAGCAATATGGGCCAGAACAAAGTTGGCCTGAGCTTCTCGCTGCCGATTTATC
AGGGCGGAATGGTTAACTCGCAGGTGAAACAGGCACAGTACAACTTTGTCGG
TGCCAGCGAGCAACTGGAAAGTGCCCATCGTAGCGTCGTGCAGACCGTGCGT
TCCTCCTTCAACAACATTAATGCATCTATCAGTAGCATTAACGCCTACAAACAA
GCCGTAGTTTCCGCTCAAAGCTCATTAGACGCGATGGAAGCGGGCTACTCGG
TCGGTACGCGTACCATTGTTGATGTGTTGGATGCGACCACCACGTTGTACAAC
GCCAAGCAAGAGCTGGCGAATGCGCGTTATAACTACCTGATTAATCAGCTGAA
TATTAAGTCAGCTCTGGGTACGTTGAACGAGCAGGATCTGCTGGCACTGAAC
AATGCGCTGAGCAAACCGGTTTCCACTAATCCGGAAAACGTTGCACCGCAAAC
GCCGGAACAGAATGCTATTGCTGATGGTTATGCGCCTGATAGCCCGGCACCA
GTCGTTCAGCAAACATCCGCACGCACTACCACCAGTAACGGTCATAACCCTTTC
CGTAACtgAGGCCACGCGATGGCGTAGCCCGAGACGATAAGTTCGCTTACCGG
CTCGAA

tyrU.inact-gblock
tyrosyl tRNA (purple); recombination homology (red), anticodon underlined
GTAATCAGTAGGTCACCAGTTCGATTCCGGTAGTCGGCACCATCAAGTCCggtG
GGGTTCCCGAGCGGCCAAAGGGAGCAGACTAAATCTGCCGTCACAGACTTCG
AAGGTTCGAATCCTTCCCCCACCAccaATTTCGGCCACGCGATGGCGTAGCCCG
AGACGATAAGTTCGCTTACCGGCTCGAA

tyrU.CUA-gblock
tyrosyl tRNA (purple); recombination homology (red), anticodon underlined
GTAATCAGTAGGTCACCAGTTCGATTCCGGTAGTCGGCACCATCAAGTCCggtG
GGGTTCCCGAGCGGCCAAAGGGAGCAGACTCTAAATCTGCCGTCACAGACTT
CGAAGGTTCGAATCCTTCCCCCACCAccaATTTCGGCCACGCGATGGCGTAGCC
CGAGACGATAAGTTCGCTTACCGGCTCGAA

lambda::ZeoR
zeo^{R} (pink); recombination homology (red)
GCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAACTTGA
GCGAAACGGGAAGGGTGTTGACAATTAATCATCGGCATAGTATATtGGCATAG
TATAATACGACAAGGTGAGGAACTAAACCATGGCCAAGCTGACCAGTGCCGTT
CCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACC
GGCTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCG
GGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGAC
AACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGT
GGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGAC
CGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGC
CGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACACGTCCGACGG
CGGCCCACGGGTCCCAGGCCTCGGAGATCCGTCCCCCTTTTCCTTTGTCGATA
TCATGTAATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCT
AACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTT
TTATAGTTATGTTAGTATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTT
TCTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGA
AGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAGCTATATTACAACGCGGCA
GCATTATGAGCTGGCAGGAGAAAATCAACGCGGC

lambda::*tolC*
tolC (blue); recombination homology (red)
GCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAACTTGA
GCGAAACGGGAAGTTGAGGCACATTAACGCCCTATGGCACGTAACGCCAACC
TTTTGCGGTAGCGGCTTCTGCTAGAATCCGCAATAATTTTACAGTTTGATCGC
GCTAAATACTGCTTCACCACAAGGAATGCAAATGAAGAAATTGCTCCCCATTCT
TATCGGCCTGAGCCTTTCTGGGTTCAGTTCGTTGAGCCAGGCCGAGAACCTGA
TGCAAGTTTATCAGCAAGCACGCCTTAGTAACCCGGAATTGCGTAAGTCTGCC
GCCGATCGTGATGCTGCCTTTGAAAAAATTAATGAAGCGCGCAGTCCATTACT
GCCACAGCTAGGTTTAGGTGCAGATTACACCTATAGCAACGGCTACCGCGACG
CGAACGGCATCAACTCTAACGCGACCAGTGCGTCCTTGCAGTTAACTCAATCC
ATTTTTGATATGTCGAAATGGCGTGCGTTAACGCTGCAGGAAAAAGCAGCAG
GGATTCAGGACGTCACGTATCAGACCGATCAGCAAACCTTGATCCTCAACACC
GCGACCGCTTATTTCAACGTGTTGAATGCTATTGACGTTCTTTCCTATACACAG

GCACAAAAAGAAGCGATCTACCGTCAATTAGATCAAACCACCCAACGTTTTAAC
GTGGGCCTGGTAGCGATCACCGACGTGCAGAACGCCCGCGCACAGTACGATA
CCGTGCTGGCGAACGAAGTGACCGCACGTAATAACCTTGATAACGCGGTAGA
GCAGCTGCGCCAGATCACCGGTAACTACTATCCGGAACTGGCTGCGCTGAAT
GTCGAAAACTTTAAAACCGACAAACCACAGCCGGTTAACGCGCTGCTGAAAGA
AGCCGAAAAACGCAACCTGTCGCTGTTACAGGCACGCTTGAGCCAGGACCTG
GCGCGCGAGCAAATTCGCCAGGCGCAGGATGGTCACTTACCGACTCTGGATT
TAACGGCTTCTACCGGGATTTCTGACACCTCTTATAGCGGTTCGAAAACCCGT
GGTGCCGCTGGTACCCAGTATGACGATAGCAATATGGGCCAGAACAAAGTTG
GCCTGAGCTTCTCGCTGCCGATTTATCAGGGCGGAATGGTTAACTCGCAGGT
GAAACAGGCACAGTACAACTTTGTCGGTGCCAGCGAGCAACTGGAAAGTGCC
CATCGTAGCGTCGTGCAGACCGTGCGTTCCTCCTTCAACAACATTAATGCATCT
ATCAGTAGCATTAACGCCTACAAACAAGCCGTAGTTTCCGCTCAAAGCTCATTA
GACGCGATGGAAGCGGGCTACTCGGTCGGTACGCGTACCATTGTTGATGTGT
TGGATGCGACCACCACGTTGTACAACGCCAAGCAAGAGCTGGCGAATGCGCG
TTATAACTACCTGATTAATCAGCTGAATATTAAGTCAGCTCTGGGTACGTTGA
ACGAGCAGGATCTGCTGGCACTGAACAATGCGCTGAGCAAACCGGTTTCCACT
AATCCGGAAAACGTTGCACCGCAAACGCCGGAACAGAATGCTATTGCTGATG
GTTATGCGCCTGATAGCCCGGCACCAGTCGTTCAGCAAACATCCGCACGCACT
ACCACCAGTAACGGTCATAACCCTTTCCGTAACTGAATATTACAACGCGGCAG
CATTATGAGCTGGCAGGAGAAAATCAACGCGGC

Appendix II

Oligo list

Chapter 2 Oligos

TrpRS.Z.ab-F	ATCAGTCTATAAATGACCTTCTGCCCGCATTAGGGCTTCCGCATAGCGAAAATCAGGAAT CGAAAAAGGTGTTGACAATTAATCATCGGC
TrpRS.Z.ab-R	TGTAGGCCGGATAAGGCGTTCACGCCGCATCCGGCATGAACAAAGCGCAATTTGCCAGC AATAGTGAAAGCTTGCAAATTAAAGCCTTCG
TrpRS150F	GTCGGCGACTCACGCAATGATATTCAGGCGGC
TrpRS150R	AGCGAGATGTGGAAACGGCGAGGCACTTCAC
Zeo-iR	CTGGTCCTGGACCGCGCTGATGAAC
TrpRS_sqiR	ATCCTGGCGCACGGTGATCGCGTGTTG
trpTKO.Gent-F	CAGTCGGTTAGAATACCTGCCTGTCACGCAGGGGGTCGCGGGTTCGAGTCCCGTCCGTT CCGCCACCCTAATTACGCACACCGTGGAAAC
trpTKO.Gent-R	CGGTAGAAGGATTTACTTCGGAGAGGGTTATTTCAGATAAAAAAAATCCTTAGCTTTCGC TAAGGATGATTTCCCGGGAAGCCGATCTCG
trpT GsqF	GGGGTCTCCCCATGCGAGAGTAGGGAAC
trpT GsqR	CCGTTGTCGATAGCACAACACTTTCACGGCC
galK.90 del	CGCGCAGTCAGCGATATCCATTTTCGCGAATCCGGAGTGTAAGAACGCGCAGTCAGCGA TATCCATTTTCGCGAATCCGGAGTGTAAGAA
galK_KO_verf_F	TGGCAGAGACCCAGCG
galK_KO_verf_R	ACCCCAGTCCATCAGCG
dLambda.galK-F	GCTATGAAATAGAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAACTTGAGCGAAA CGGGAAGCCTGTTGACAATTAATCATCGGC
dlambda.galK dterm-R	GCCGCGTTGATTTTCTCCTGCCAGCTCATAATGCTGCCGCGTTGTAATATTCAGCACTGT CCTGCTCCTT
dlambda.sqF	GGTTTGATCAGAAGGACGTTGATCGGGCGG
dlambda.sqR	TTCAGATACTGGCGATCATCCGCCACCAG
dLambda.sqiR	AGCCCATTGATAGTTTTCATGTGCGACAATGGGCG
EcWRS_mut7_8-F	GAATCCCATATGATGACTAAGCCCATCGTTNBTNSTGGCGCACAGCCCTCAGGTGAATTG
libEcWRS-NdeI-F	TACGCTTTGAGGAATCCCATATGATGACTAAGCCCATCG
EcWRS1_mut-VPViR	CAGATTAGTTTGATACAGCAGGATGTCCGCTGCCATC
EcWRS1_mut144-6_F	GATGGCAGCGGACATCCTGCTGTATCAAACTAATCTGnnknnknnkGGTGAAGACCAGAAA CAGCACCTCGAACTGAGC
EcWRS_NcoI_PstI_termR	agcgtttgaaactgcagccatggtaccTTACGGCTTCGCCACAAAACCAATCGC
proK-F	GTTAGCCTGCAGGTAATTCCGCTTCGCAACATGTGAG
TrpH NcoI-R	GGCCGCCATGGCAAATTCGACCCTG
Trp40CCA-iR	GCAACCAGGCGCTTTGGAGGCGCCAGCTCTACCCTTGAG
Trp40CCA-iF	AGCTGGCGCCTCCAAAGCGCCTGGTTGCAGGTTC
SmR-R	CGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTG
pNP-SpeI-F	ATAATGGACTAGTGCGCTTGTTTCGGC
pNP-BAB-R	CTCCTTAGATCTTCCTAGGTGGATCCACCATTCC
pEvol CmR SpeI-F	AATAATACTAGTGTTGATACCGGGAAGCCCTGG
pEvol_CmR PstI-R	AATAATCTGCAGCGAATTTCTGCCATTCATCCGCTTATTATCAC
CmR-TGA-7	GCTAAGGAAGCTAAAATGGAGAAAAAAATCACTTGATATACCACCGTTGATATATCCCAA TGGC
CmR-TGA-84	GCAATGAAAGACGGTGAGCTGGTGTGATGGGATAGTGTTCACCCTTGTTACACC
CmR-TGAT-98	CCCTTGTTACACCGTTTTCCATGAGTGATCTGAAACGTTTTCATCGCTCTGGAG
pRep-KpnI-tR-F	AATAATaggtaccGTTCTGTTGCCCGTCTCACTGGTG
pRep-EcWtR-NdeI/AvrII-R.	AATAATAcatatgCCTAGGTGGCAGGGGCGGAGAGACTC
EcW-TGA-MSDM	GTTCAATTGGTAGAGCACCGGTCTTCAAAACCGGGTGTTGGGAGTTCGAG
T7F1	CAGGTTCGCAGCGTCAGCCGGAATGGTACCG
T7R3	GCGCCCGACAGCCTTCCAGTTCCTGTGAGAAATCAAGCCGGAAGCCGTAGCGTAC
T7F3	GTACGCTACGGCTTCCGGCTTGATTTCTCACAGGAACTGGAAGGCTGTCGGGCGC
T7R4b	CCATGACCATGATTACCGTGCACTGAAATACCATTAACATTGCTAAGAACG
T7F4	CGTTCTTAGCAATGTTAATGGTATTUCAGTGCACGGTAATCATGGTCATGG
T7R5	CGAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGG
T7_mut-SbfI-F	AATAATcctgcaggCTACTCAGGAGAGCGTTCACCGAC
T7_mut-NotI/SbfI-R	AATAATCCTGCAGGGCGGCCGCTACGGGAGGGCTTACCATCTGG
Barnase MSDM 3 TGA	TTTAACTTTAAGAAGGAGATATACATATGGCATgaGTTATCAACACGTTTGACGGGGTTG
Barnase MSDM 45 TGA	GTGGCATCAAAAGGGAACCTTGCATGAGTCGCTCCGGGGAAAAGCATC
EcWtR-PstI-F	TTAGCCTGCAGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTC
EcWtR-prok-oF	GCGCCCCGCATTTAGGGGCGTAGTTCAATTGGTAGAGCACCGGTC
EcWtR-proK-oR	ACTACGCCCCTAAATGCGGGGCGCATCTTACTGCGC
EcWtR-KpnI-R	ATATATGGTACCAAAAAATGGCAGGGGCGGAGAGACTCG
GFPflip-NotI-F +	TCGATCCCGCGAAATTAATACGACTCACTATAG
sfGFP-T7+lam-PstI-R.	ATATACTGCAGCGCCAAGCTAGCTTGGATTCTCACCAATAAAAAACGC

MjYtR del F	TGGCAGGGGCGGAGAGACTCGAACTCC
MjYtR del oR	CGAGTCTCTCCGCCCCTGCCA AATTCGAAAAGCCTGCTCAACGAGCAGG
EcWtR TGA MSDM	GTTCAATTGGTAGAGCACCGGTCTTCAAAACCGGGTGTTGGGAGTTCGAG
sfGFP (pEvol) TGA151	CTCGAGTACAACTTTAACTCACACAATGTATGAATCACGGCAGACAAACAAAAGAATGG
EcWRS1.FA.NotI-F	AATAATAgcggccgcATGACTAAGCCCATCGTTTTTGCTGGCGCAC
EcWRS-NotI-R	AATAATAgcggccgcTTACGGCTTCGCCACAAAACCAATCGC
pUltraII-tRsqR	GGTGCCCTTAAACGCCTGGTTGC
EcTrpRS-NdeI-F	AAtAAAcatatgATGACTAAGCCCATCGTTTTTAGTGGCGCAC
EcTrpRS-PstI-R	TTATTCTGCAGTTACGGCTTCGCCACAAAACCAATCGC
TrpRS-R EcoRI	ATTATTGAATTCTTACGGCTTCGCCACAA
TrpRS-F-NheI	AATAAATGCTAGCATGACTAAGCCCATC
U6-R tRNAtrp* AvrII	AATTATTGCTAGCAAAAAATGGCAGGGGCG
tRNAtrp* Nhe-R	AATTATTGCTAGCAAAAAATGGCAGGGGCG
EcWRS_mamNheI-F	aataataGCTAGCgccaccATGACTAAGCCCATCGTT
EcWRS_mamEcoRI-R	AATAATAgaattcTTACGGCTTCGCCACAAAACCAATCGC

Chapter3 Oligos

List of oligonucleotides (purchased from Integrated DNA Technologies)

Name	Sequence
EcYS.*tolC*-F2	TTTTTGTACCAGCGTGAAAATGATGCGTGGAAGATTGATCGTC TTGCACCCTGAAAAGATGCAAAAATCTTGTTGAGGCACATTAA CGCC
EcYS.*tolC*-R2	TGTTACCCGCATGACCATAAACAACGTGAGACTGGATAGCGA GAATATTTTTCATCATGTTCCCTGTTCAGTTACGGAAAGGGTT ATGAC
tyrS.90 del	TGGAAGATTGATCGTCTTGCACCCTGAAAAGATGCAAAAATCT TGACAGGGAACATGATGAAAAATATTCTCGCTATCCAGTCTCA CGTT
EcY_TV.*tolC*-F	AAAATAACTGGTTACCTTTAATCCGTTACGGATGAAAATTACG CAACCAGTTGAGGCACATTAACGCC
EcY_TV.*tolC*-R	CCGCATGGGCAGAATATTTAATTGCGGATTCGTTGGGAAGTT CAGGGACTTCAGTTACGGAAAGGGTTATGACCGTTAC
tyrTV.KOgent-F	AAAATAACTGGTTACCTTTAATCCGTTACGGATGAAAATTACG CAACCAGACGCACACCGTGGAAAC
tyrTV.KOgent-R	CCGCATGGGCAGAATATTTAATTGCGGATTCGTTGGGAAGTT CAGGGACTCCGGGAAGCCGATCTCG
tyrTV.90 del	AACTGGTTACCTTTAATCCGTTACGGATGAAAATTACGCAACC AGAGTCCCTGAACTTCCCAACGAATCCGCAATTAAATATTCTGC CCA
tyrTV.lhm-F	AAAATAACTGGTTACCTTTAATCCGTTACGG
tyrTV.rhm-R	CCGCATGGGCAGAATATTTAATTGCG
tyrU-*tolC*.F	GTAATCAGTAGGTCACCAGTTCGATTCCGGTAGTCGGCACCAT CAAGTCCAAGGAATGCAAATGAAGAAATTGCTCCCCATTC
tyrU-*tolC*.R	TTCGAGCCGGTAAGCGAACTTATCGTCTCGGGCTACGCCATCG CGTGGCCTCAGTTACGGAAAGGGTTATGACC
tyrU.lh-F	GTAATCAGTAGGTCACCAGTTCGATTC
tyrU.rh-R	TTCGAGCCGGTAAGCGAACTTATCGTCTC

tyrU.KO.gentR-RBS-F	GTAATCAGTAGGTCACCAGTTCGATTCCGGTAGTCGGCACCATCAAGTCCAAGGAATGCAAATGTTACGCAGCAGCAACGATGTTAC
tyrU.KO.gentR-R	TTCGAGCCGGTAAGCGAACTTATCGTCTCGGGCTACGCCATCGCGTGGCCCCGGGAAGCCGATCTCGGCTTGAACG
dlambda ZtattB-F	GCTATGAAATGAAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAACTTGAGCGAAACGGGAAGGGTGTTGACAATTAATCATCGGC
dlambda ZtattB-R	GCCGCGTTGATTTTCTCCTGCCAGCTCATAATGCTGCCGCGTTGTAATATGACTTGCAAATTAAAGCCTTCG
dlambda.*tolC*-F	GCTATGAAATGAAAAAATGAATCCGTTGAAGCCTGCTTTTTTATACTAACTTGAGCGAAACGGGAAGTTGAGGCACATTAACGCC
dlambda.*tolC*-R	GCCGCGTTGATTTTCTCCTGCCAGCTCATAATGCTGCCGCGTTGTAATATTCAGTTACGGAAAGGGTTATGACCGTTAC
tyrS.GsqF	CAGGGTGGTGAGCATCGCCTGC
tyrS.GsqR	CAATACGGCAAATGGACTGGCTGCGTGAT
tyrS.GsqR2	CACAGGTGTGTAATTTATCAATGGCGGCAATGCCTTGC
Ec tyrTV GsqF	TCGTGCACTATACAAAGTACTGGCACAGCGC
Ec tyrTV GsqR	TTGAGAGTTCAGGGACTTTTGAAAGTGATGGTGGTGG
Ec tyrTV GsqiR	TCGAACCTTAGTCGAAGGTTCTCACCCTTCCC
Ec tyrTV GsqiR2	CACAATCTCACCGAAGTTACCACATCG
Ec tyrTV GsqR2	GCATGGGCAGAATATTTGATTGCGGATTC
Ec tyrTV GsqR3	CGGATTCGCTTGAGAGTTCAGGGAC
tyrU Gsq-F	GTGCTAATCTGCCCTCCGTTCGGCTGTTTC
tyrU Gsq-R	GCCCGCATCCTGGAACTCGGCTACC
tyrU-GsqR2	GGGTGCGCTCTACCAACTGAGC
dlambda.sqF	GGTTTGATCAGAAGGACGTTGATCGGGCGG
dlambda.sqR	TTCAGATACTGGCGATCATCCGCCACCAG
gent-450sqiR	CCACCTAACTTTGTTTTAGGGCGACTGCCCTG
gentR-iF	GGCTTACGTTCTGCCCAGGTTTGAG
Zeo-iR	CTGGTCCTGGACCGCGCTGATGAAC
tolC sqiR	CTGTGGCAGTAATGGACTGCGCGCTTC

Chapter 4 Oligos

Primer Name	**Sequence**
pBK seqT-F	ATTACGCTGACTTGACGGGACGG
EcYRS-L71-oR	CGCAACCGGCTTGTGGCCCGCCTGC
EcYRS-L71-NBT-F	GCAGGCGGGCCACAAGCCGGTTGCGnbtGTAGGCGGCGCGACGGGTCTGATTG
EcYRS-N126-oR	GTTCGCCGCGATAGCAGAGTTTTC

EcYRS-N126x-F	GAAAACTCTGCTATCGCGGCGAACnnnTATGACTGGTTCGGCAATATGAATGTGCTGAC
pBK MCS JIsqR	GAGATCATGTAGGCCTGATAAGCGTAGC
EcYRS-NheI-F	GCTAGCGCCACCATGGCAAGCA
EcYRS-XhoI-R	aataatCTCGAGTTATTTCCAGCAAATCAGACAGTAATTCTTTTTACC
HCII-SfiI-F	TGGCAAAGAATTGGCCAAGGAGGCCACCATGAAACACTCATTAAACGCACTTC
10xHis-TGA-SfiI-R	TGGCGGCCGGCCAGGCCTCAATGATGGTGGTGATGATGATGGTGATGATG
HCII-79-Phe-R	GTCGTCGTCTTCACTGAATATCTTCTCCAGGTCCAGaaaGTCGTCGTCCTCCTCCCCC
HCII-79-TAG-R	GTCGTCGTCTTCACTGAATATCTTCTCCAGGTCCAGctaGTCGTCGTCCTCCTCCCCC
HCII-92-Phe-F	CTGGACCTGGAGAAGATATTCAGTGAAGACGACGACtttATCGACATCGTCGACAGTCTG
HCII-92-TAG-F	CTGGACCTGGAGAAGATATTCAGTGAAGACGACGACtagATCGACATCGTCGACAGTCTG
HCII-80-iF	CTGGACCTGGAGAAGATATTCAGTGAAGACGACGAC
HCII-80-iR	GTCGTCGTCTTCACTGAATATCTTCTCCAGGTCCAG

Chapter 5 Oligos

	Primer Name	Sequence
GTEV Prim	GFPflip-NotI-F	TCGATCCCGCGAAATTAATACGACTCACTATAG
	sfGFP-EcoRI-3TAG-F	ACACAGAATTCATTAAAGAGGAGAAATTACATATGGCATAGAGCAAAGGAGAAGAACTTTTC
	sfGFP-EcoRI-3TGA-F	ACACAGAATTCATTAAAGAGGAGAAATTACATATGGCATGAAGCAAAGGAGAAGAACTTTTC
	sfGFP-EcoRI-3TAA-F	ACACAGAATTCATTAAAGAGGAGAAATTACATATGGCATAAAGCAAAGGAGAAGAACTTTTC
	sfGFP-51T-oF	ACTGGAAAACTACCTGTTCCGTGGCCAAC
	sfGFP-51T-TAG-R	GTTGGCCACGGAACAGGTAGTTTTCCAGTCTAGCAAATAAATTTAAGGGTGAGTTTTCCGTTTG
	sfGFP-51T-TGA-R	GTTGGCCACGGAACAGGTAGTTTTCCAGTTCAGCAAATAAATTTAAGGGTGAGTTTTCCGTTTG
	sfGFP-51T-TAA-R	GTTGGCCACGGAACAGGTAGTTTTCCAGTTTAGCAAATAAATTTAAGGGTGAGTTTTCCGTTTG
	sfGFP-151Y-oR	TACATTGTGTGAGTTAAAGTTGTACTCGAG
	sfGFP-151Y-TAG-F	CTCGAGTACAACTTTAACTCACACAATGTATAGATCACGGCAGACAAACAAAAGAATGG
	sfGFP-151Y-TGA-F	CTCGAGTACAACTTTAACTCACACAATGTATGAATCACGGCAGACAAACAAAAGAATGG
	sfGFP-151Y-TAA-F	CTCGAGTACAACTTTAACTCACACAATGTATAAATCACGGCAGACAAACAAAAGAATGG
	sfGFP-clTEV-overlap-R	ACCTTGAAAATAAAGATTTTCATGATGATGATGGTGGTGGTGGTGGTGGG
	TEVp-overlap-1xGGS-F	CATCATCATCATGAAAATCTTTATTTTCAAGGTGGAGGAAGTGGAGAAAGTTTGTTTAAGGGGCCG
	TEV-HindIII-R	GGATCCAAGCTTttaTTATTAGCGACGGCGACGACGATTCATGAG
Other Consti	PylCas-F	AATAATAGCTagCGAGCTGTTGACAATTAATCATCGGCTCG
	PylCas-AvrII-R	TATTATTccTGAgACCCAAATTCGACCCTGAGCTGCTCG
	lacI-iF	gccagacgcagacgcgccgagac
	SmR-R	CGCGCGCAGATCAGTTGGAAGAATTTGTCCACTACGTG
	BKevol-oF2	GGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTGAGTTGAAGGATCCGCGGCCGC
	pBK-MCS-JIsqR	GAGATCATGTAGGCCTGATAAGCGTAGC
	pEvol-tacI-oR	ATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCCC

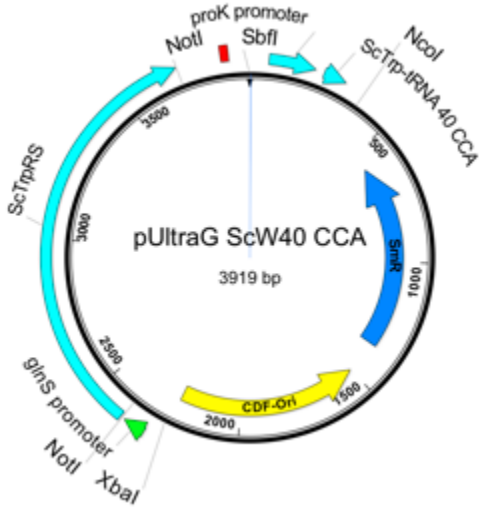
Appendix III

Plasmid sequences

Chapter 2 plasmid sequences

Plasmid map and full plasmid sequence are given below. For each system, GFP is highlighted in green, CAT/Barnase in orange, T7 RNA polymerase in purple, tRNA in red, and aaRS in blue unless otherwise specified.

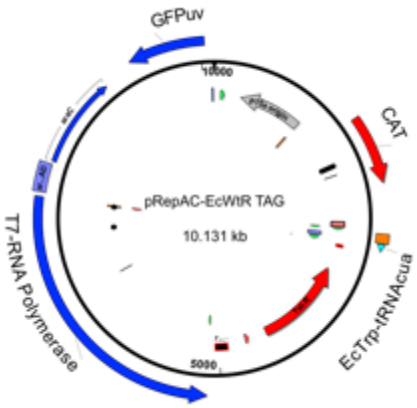
pUltra_ScW40$_{CCA}$



GGtaattccgcttcgcaacatgtgagcaccggtttattgactacAggaagcagtgtgaccgtgtgcttctcaaatgcctgaggccagtttgctcaggc
tctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaaggga
ttgacgagggcgtatctgcgcagtaagatgcgccccgcattGAAgcGGTGGCTCAAgGGTAGAGCTggcgcCTCcAA
AgcgcctGGtTGCAGGTTCAAgTCCTGcCCgtTTCACCAaattcgaaaagcctgctcaacgagcaggcttttttgcatgct
cgagcagctcagggtcgaatttgCCATGGcggccACCAGGTacCACCGGCGcctcaggcatttgagaagcacacggtcacact
gcttccggtagtcaataaaccggtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgacgaccgggtcatcgtggccggatc
ttgcggcccctcggcttgaacgaattgttagacattatttgccgactaccttggtgatctcgcctttcacgtagtggacaaattcttccaactgatctgcgc
gcgaggccaagcgatcttcttcttgtccaagataagcctgtctagcttcaagtatgacgggctgatactgggccggcaggcgctccattgcccagtcgg
cagcgacatccttcggcgcgattttgccggttactgcgctgtaccaaatgcgggacaacgtaagcactacatttcgctcatcgccagcccagtcgggcg
gcgagttccatagcgttaaggtttcatttagcgcctcaaatagatcctgttcaggaaccggatcaaagagttcctccgccgctggacctaccaaggcaac
gctatgttctcttgcttttgtcagcaagatagccagatcaatgtcgatcgtggctggctcgaagatacctgcaagaatgtcattgcgctgccattctccaaa
ttgcagttcgcgcttagctggataacgccacggaatgatgtcgtcgtgcacaacaatggtgacttctacagcgcggagaatctcgctctctccagggga
agccgaagtttccaaaaggtcgttgatcaaagctcgccgcgttgtttcatcaagccttacggtcaccgtaaccagcaaatcaatatcactgtgtggcttca
ggccgccatccactgcggagccgtacaaatgtacggccagcaacgtcggttcgagatggcgctcgatgacgccaactacctctgatagttgagtcgat
acttcggcgatcaccgcttccctcatactcttccttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttag
aaaaataaacaaatagctagctcactcggtcgctacgctccgggcgtgagactgcggcgggcgctgcggacacatacaaagttacccacagattccgt
ggataagcaggggactaacatgtgaggcaaaacagcagggccgcgccggtggcgtttttccataggctccgcccctcctgccagagttcacataaaca
gacgcttttccggtgcatctgtgggagccgtgaggctcaaccatgaatctgacagtacgggcgaaacccgacaggacttaaagatccccaccgtttcc
ggcgggtcgctccctcttgcgctctcctgttccgaccctgccgtttaccggatacctgttccgcctttctcccttacgggaagtgtggcgctttctcatag
ctcacacactggtatctcggctcggtgtaggtcgttcgctccaagctgggctgtaagcaagaactccccgttcagcccgactgctgcgccttatccggt
aactgttcacttgagtccaacccggaaaagcacggtaaaacgccactggcagcagccattggtaactgggagttcgcagaggatttgtttagctaaaca
cgcggttgctcttgaagtgtgcgccaaagtccggctacactggaaggacagatttggttgctgtgctctgcgaaagccagttaccacggttaagcagttc
cccaactgacttaaccttcgatcaaaccacctccccaggtggttttttcgtttacagggcaaaagattacgcgcagaaaaaaaggatctcaagaagatcc
tttgatcttttctactgaaccgctctagagtcatcaatcatcccataatccttgttagattatcaattttaaaaaactaacagttgtcagcctgtcccgcttta
atatcatacgccgttatacgttgtttacgctttAaggagGCGGCCGCATGAGCAACGACGAAAACTGTAGAGAAA
GTCACCCAACAAGTGTCGGAACTAAAAAGCACAGATGTTAAAGAGCAAGTAGTTACA
CCTTGGGATGTGGAAGGTGGGGTTGATGAACAAGGTAGAGCCCAAAATATTGATTAC
GACAAATTGATCAAACAATTCGGTACTAAGCCGGTCAACGAAGAAACCCTGAAGAGA

TTCAAGCAAGTGACGGGTCGTGAACCACATCATTTTTTGCGTAAGGGATTGTTTTTCA
GTGAGCGTGACTTCACTAAAATATTAGACCTTTACGAACAAGGCAAACCATTTTTCCTA
TACACTGGTAGAGGTCCTTCGAGCGATTCTATGCACTTGGGTCATATGATCCCTTTTG
TCTTCACCAAATGGTTACAGGAAGTGTTTGACGTACCATTAGTCATAGAGTTGACAGA
TGACGAAAAATTTTTATTCAAACACAAGTTGACCATCAATGACGTTAAGAATTTTGCCC
GTGAAAATGCCAAGGATATCATTGCTGTTGGCTTTGACCCAAAGAACACCTTTATCTT
TTCTGATTTGCAATACATGGGTGGTGCATTTTACGAAACTGTAGTAAGAGTTTCCAGA
CAAATTACAGGATCCACTGCAAAGGCTGTTTTCGGGTTTAATGACTCCGACTGTATTG
GTAAGTTCCATTTTGCCTCCATTCAAATTGCTACCGCATTCCCAAGCTCATTTCCTAATG
TGTTAGGCTTGCCTGATAAGACACCATGTTTGATTCCATGTGCAATTGACCAAGATCC
ATATTTCAGAGTTTGTAGGGATGTCGCGGATAAATTGAAGTACTCCAAACCTGCTTTG
CTTCATTCCAGATTCTTTCCAGCTTTGCAAGGTTCCACGACCAAAATGTCAGCCTCTGA
TGATACCACTGCCATTTTCATGACCGATACACCAAAGCAAATTCAAAAGAAAATTAACA
AGTACGCATTCAGCGGTGGTCAAGTGTCCGCCGACCTACATAGAGAATTAGGTGGTA
ATCCCGATGTCGATGTTGCATACCAATACTTGTCATTTTTCAAGGATGACGATGTTTTC
TTGAAGGAATGCTATGACAAATATAAGTCCGGTGAATTACTATCAGGTGAAATGAAG
AAACTGTGTATCGAGACTCTGCAAGAATTCGTTAAGGCGTTCCAGGAACGCAGAGCT
CAGGTGGACGAAGAGACCTTGGACAAATTCATGGTCCCACATAAGTTGGTTTGGGGC
GAAAAGGAAAGACTTGTCGCACCTAAGCCAAAAACTAAGCAAGAAAAGAAGTAAGCG
GCCGCtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatctctgcaatatattgagtttgcgtgcttttgtaggccggataag
gcgttcacgccgcatccggcaagaaacagcaaacaatccaaaacgccgcgttcagcggcgtttttctgcttttcttcgcgaattaattccgcttcgcaa
catgtgagcaccggtttattgactacctgca

pRepAC-EcWtR-TAG

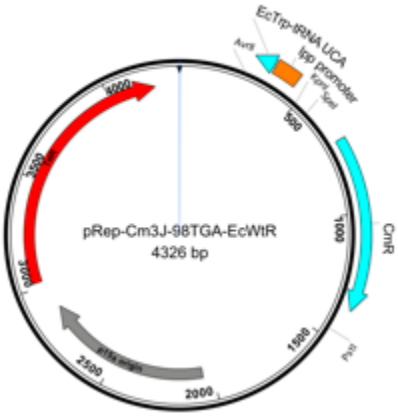


gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcc
cgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaa
ctgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaa
tcccttaacgtgagttttcgttccactgagcgtcagaccccttaataagatgatcttcttgagatcgttttggtctgcgcgtaatctcttgctctgaaaacga
aaaaaccgccttgcagggcggtttttcgaaggttctctgagctaccaactctttgaaccgaggtaactggcttggaggagcgcagtcaccaaaacttgtc
ctttcagtttagccttaaccggcgcatgacttcaagactaactcctctaaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttg
gactcaagacgatagttaccggataaggcgcagcggtcggactgaacgggggttcgtgcatacagtccagcttggagcgaactgcctacccggaac
tgagtgtcaggcgtggaatgagacaaacgcggccataacagcggaatgacaccggtaaaccgaaaggcaggaacaggagagcgcacgagggagc
cgccaggggggaaacgcctggtatctttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcggagcct
atggaaaaacggctttgccgcggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccatttccgctcg
ccgcagtcgaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttttct
cctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgcttttgccgtt
acgcaccaccccgtcagtagctgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacaccatcatacactaaatcagtaa

gttggcagcatcacccgacgcactttgcgccgaataaataccttgtgacggaagatcacttcgcagaataaataaatcctggtgtccctgttgataccggg
aagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttt
tgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaagaacatt
ttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagtttt
atccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcac
ccttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccactaggatttccggcagtttctacacatatattcgcaagatg
tggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgattta
aacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttc
atcatgccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaattttttttaaggcagt
tattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgacccggtcgtcggttc
agggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttattgactaccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggcc
agtttgctcaggctctccccgtggaggtaataattgacgatatgatcatttattctgcctcccagagcatgataaaaacggttagcgcttcgttaatacaga
tgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggACTAGTgcgcttgtttcggcgggactgttgtaactc
agaataagaaatgaggccgctcatggcgttctgttgcccgtctcactggtgaaaagaaaaacaaccctggcgccgcttctttgagcgaacgatcaaaaa
taagtggcgccccatcaaaaaaatattctcaacataaaaaactttgtgtaatacttgtaacgctGAATTCAGGGGCGTAGTTCAA
TTGGTAGAGCACCGGTCTctaAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGC
CACTGCAGATccttagcgaaagctaaggatttttttttaagcttggcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttac
ccaacttaatcgccttgcagcacatccccctttcgccagttgagcaccgccgccgcaaggaatggtGaattcaggaTCTAAGGAGcccgag
atgcgccgcgtgcggctgctggagatggcggacgcgatggatatgttctgccaagggttggtttgcgcattcacagttctccgcaagaattgattggct
ccaattcttggagtggtgaatccgttagcgaggtgccgccggcttccattcaggtcgaggtggcccggctccatgcaccgcgacgcaacgcggggag
gcagacaaggtatagggcggcgcctacaatccatgccaacccgttccatgtgctcgccgaggcggcataaatcgccgtgacgatcagcggtccaatg
atcgaagttaggctggtaagagccgcgagcgatccttgaagctgtccctgatggtcgtcatctacctgcctggacagcatggcctgcaacgcgggcat
cccgatgccgccggaagcgagaagaatcataatggggaaggccatccagcctcgcgtcgcgaacgccagcaagacgtagcccagcgcgtcggccg
ccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgc
aagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgata
aagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgacgc
tctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggc
gcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggt
gatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatccacaggacgggtgtggt
cgccatgatcgcgtagtcgatagtggctccaagtagcgaagcgagcaggactgggcggcggccaaagcggtcggacagtgctccgagaacgggtg
cgcatagaaattgcatcaacgcatatagcgctagcagcacgccatagtgactggcgatgctgtcggaatggacgatatcccgcaagaggcccggcag
taccggcataaccaagcctatgcctacagcatccagggtgacggtgccgaggatgacgatgagcgcattgttagatttcatacacggtgcctgactgcg
ttagcaatttaactgtgataaactaccgcattaaagcttatcgatgataagctgtcaaacatgagaattacaacttatatcgtatggggctgacttcaggtg
ctacattgctcaaagatgcaggggtaaaagctaaccgcatctttaccgacaaggcatccggcagttcaacagatcgggaagggctggatttgctgagg
atgaaggtggaggaaggtgatgtcattctggtgaagaagctcgaccgtcttggccgcgacaccgccgacatgatccaactgataaaagagtttgatgct
cagggtgtagcggttcggtttattgacgacgggatcagtaccgacggtgatatggggcaaatggtggtcaccatcctgtcggctgtggcacaggctga
acgccggaggatcctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaaacaacagat
aaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggcagttccctactctcgcatggggagaccccacactaccatcggcgct
acggcgtttcacttctgagttcggcatggggtcaggtgggaccaccgcgctactgccgccaggcaaattctgttttatcagaccgcttctgcgttctgatt
taatctgtatcaggctgaaaatcttctctcatccgacgtcttaggcgaaggcgaagtccgactctaagatgtcacggaggttcaagttacctttagccgga
agtgctggcattttgtccaattgagactcgtgcaactggtcagcgaactggtcgtagaaatcagccagtacatcacaagactcatatgtgtcaaccatagt
ttcgcgcactgctttgaacaggttcgcagcgtcagccggaatggtaccgaaggagtcgtgaatcagtgcaaaagattcgattccgtacttctcgtgtgcc
cacactacagtcttacgaaggtggctaccgtcttggctgtgtacaaagttaggagcgataccagactcctgtttgtgtgcatcaatctcgctatctttgttg
gtgttaatggtaggctgtaagcggaactgaccgaggaacatcaggttcaagcgcgtctgaataggcttcttgtattcctgccacacagggaaaccatca
ggagttacccaatgcacagcgcaacgcttgcgaagaatctctccagtcttcttatctttgacctcagcagccagcagcttagcagcagacttaagccagt
tcattgcttcaaccgcagctaccaccgtcacgctcacagattcccaaatcagcttagccatgtatccagcagcctgattcggctgagtgaacatcagacc
cttgccggaatcaatagctggctgaatggtatcttccagcacttgttgacggaagccgaactctttggacccgtaagccagcgtcatgactgaacgctta
gtcacactgcgagtaacaccgtaagccagccattgaccagccagtgccttagtgcccagcttgactttctcagagatttcaccagtgttctcatcggtca
cggtaactacttcgttatcggtcccattgattgcgtctgcttgtagaatctcgttgactttcttagcaacaatcccgtagatgtcctgaacggtttcactagg
aagcaagttaaccgcgcgaccacctacctcatctcggagcatcgcggagaagtgctggatgccagagcaagacccgtcaaacgccagcggaaggga
gcagttatagctcaggccgtggtgctgtaccccagcgtactcaaagcagaacgcaaggaagcagaacggagaatcttgctcagcccaccaagtgttct
ccagtggagacttagcgcaagccatgatgttctcgtggttttcctcaatgaacttgatgcgctcagggaacggaaccttatcgacacccgcacagtttgc

accgtggattttcagccagtagtaaccttccttaccgattggtttacctttcgccagcgtaagcagtcctttggtcatatcgttaccttgcgggttgaacatt
gacacagcgtaaacacgaccgcgccagtccatgttgtaagggaaccagatggccttatggttagcaaacttattggcttgctcaagcatgaactcaagg
ctgatacggcgagacttgcgagccttgtccttgcggtacacagcagcggcagcacgtttccacgcggtgagagcctcaggattcatgtcgatgtcttcc
ggtttcatcgggagttcttcacgctcaatcgcagggatgtcctcgaccggacaatgcttccacttggtgattacgttggcgaccgctaggactttcttgtt
gattttccatgcggtgttttgcgcaatgttaatcgctttgtacacctcaggcatgtaaacgtcttcgtagcgcatcagtgctttcttactgtgagtacgcacc
agcgccagaggacgacgaccgttagcccaatagccaccaccagtaatgccagtccacggcttaggaggaactacgcaaggttggaacatcggagag
atgccagccagcgcacctgcacgggttgcgatagcctcagcgtattcaggtgcgagttcgatagtctcagagtcttgacctactacgccagcattttgg
cggtgtaagctaaccattccggttgactcaatgagcatctcgatgcagcgtactcctacatgaatagagtcttccttatgccacgaagaccacgcctcgc
caccgagtagacccttagagagcatgtcagcctcgacaacttgcataaatgctttcttgtagacgtgccctacgcgcttgttgagttgttcctcaacgtttt
tcttgaagtgcttagcttcaaggtcacggatacgaccgaagcgagcctcgtcctcaatggcccgaccgattgcgcttgctacagcctgaacggttgtatt
gtcagcactggttaggcaagccagagtggtcttaatggtgatgtacgctacggcttccggcttgatttcctacaggaactggaaggctgtcgggcgcttg
ccgcgcttagctttcacttcctcaaaccagtcgttgatgcgtgcaatcatcttagggagtagggtagtgatgagaggcttggcggcagcgttatccgcaa
cctcaccagctttaagttgacgctcaaacatcttgcggaagcgtgcttcacccatctcgtaagactcatgctcaagggccaactgttcgcgagctaaacg
ctcaccgtaatggtcagccagagtgttgaacgggatagcagccagttcgatgtcagagaagtcgttcttagcaatgttaatggtattctagtgcacggta
atcatggtcatggttaattcctcctgttagcccaaaaaacgggtatggagaaacagtagagagttgcgataaaaagcgtcaggtaggatccgctaatctt
atggataaaaatgctatggcatagcaaagtgtgacgccgtgcaaataatcaatgtggacttttctgccgtgattatagacacttttgttacgcgtttttgtca
tggctttggtcccgctttgttacagaatgcttttaataagcggggttaccggtttggttagcgagaagagccagtaaaagacgcagtgacggcaatgtct
gatgcaatatggacaattggtttcttctctgaatggcgggagtatgaaaagtatggctgaagcgcaaaatgatcccctgctgccgggatactcgtttaatg
cccatctggtggcgggtttaacgccgattgaggccaacggttatctcgatttttttatcgaccgaccgctgggaatgaaaggttatattctcaatctcacca
ttcgcggtcaggggggtggtgaaaaatcagggacgagaatttgtttgccgaccgggtgatattttgctgttcccgccaggagagattcatcactacggtcg
tcatccggaggctcgcgaatggtatcaccagtgggtttactttcgtccgcgcgcctactggcatgaatggcttaactggccgtcaatatttgccaatacg
gggttctttcgcccggatgaagcgcaccagccgcatttcagcgacctgtttgggcaaatcattaacgccgggcaaggggaagggcgctattcggagct
gctggcgataaatctgcttgagcaattgttactgcggcgcatggaagcgattaacgagtcgctccatccaccgatggataatcgggtacgcgaggcttg
tcagtacatcagcgatcacctggcagacagcaattttgatatcgccagcgtcgcacagcatgtttgcttgtcgccgtcgcgtctgtcacatcttttccgcc
agcagttagggattagcgtcttaagctggcgcgaggaccaacgtatcagccaggcgaagctgcttttgagcaccacccggatgcctatcgccaccgtc
ggtcgcaatgttggttttgacgatcaactctatttctcgcgggtatttaaaaaatgcaccggggccagcccgagcgagttccgtgccggttgtgaagaaa
aagtgaatgatgtagccgtcaagttgtcataattggtaacgaatcagacaattgacggcttgacggagtagcatagggtttgcagaatccctgcttcgtcc
atttgacaggcacattatgcatgccgcttcgccttcgcgcgcgaattgatctgctgcctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgc
agctcccggagacggtcacagcctgcagcaaaaaaccccctcaagacccgtttagaggccccaaggggttatgctagttattgctcagcggtggcagc
agccaactcagcttcctttcgggctttgttatttgtagagctcatccatgccatgtgtaatcccagcagcagttacaaactcaagaaggaccatgtggtca
cgcttttcgttgggatctttcgaaagggcagattgtgtcgacaggtaatggttgtctggtaaaaggacagggccatcgccaattggagtattttgttgataa
tggtctgctagttgaacggatccatcttcaatgttgtggcgaattttgaagttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagt
tatagttgtactcgagtttgtgtccgagaatgtttccatcttctttaaaatcaataccttttaactcgatacgattaacaagggtatcaccttcaaacttgactt
cagcacgcgtcttgtagttcccgtcatctttgaaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaagtcatgccgtttcatat
gatccggataacgggaaaagcattgaacaccataagagaaagtagtgacaagtgttggccatggaacaggtagttttccagtagtgcaaataaatttaa
gggtaagttttccgtatgttgcatcaccttcacccctctccactgacagaaaatttgtgcccattaacatcaccatctaattcaacaagaattgggacaactc
cagtgaaaagttcttctcctttactcatatgtatatctccttcttaaagttaaacaaaattatttctagagggaaaccgttgtggtctccctatagtgagtcgta
ttaatttcgcgggatcg

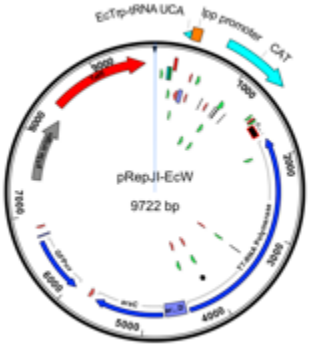
pRep-Cm3J-98TGA-EcWtR



cgcgtccgccatctccagcagccgcacgcggcgcatcttgggctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactac
tgggctgcttcctaatgcaggagtcgcataagggagagcgtctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcc
cagtcacgacgttgtaaaacgacggccagtgccaagcttaaaaaaaatccttagctttcgctaaggatcaTATGCCTAGGTGGCAGG
GGCGGAGAGACTCGAACTCCCAaCACCCGGTTTTGAAGACCGGTGCTCTACCAATTGA
ACTACGCCCCTGAATTCagcgttacaagtattacacaaagtttttatgttgagaatatttttttgatggggcgccacttatttttgatcgtt
cgctcaaagaagcggcgccagggttgtttttctttttcaccagtgagacgggcaacagaacGGTACCtctagacaattggtgcacttcaaaaatcg
atgagctgttgacaattaatcatcgaactagtgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttc
caactttcaccataatgaaataagatcactaccgggcgtatttttttgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactg
gatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggata
ttacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtat
ggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagTGATctgaaacgttttcatcgctctggagtg
aataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaata
tgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaata
ttatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtTtgtgatggcttccatgtcggcagaatgcttaatgaattacaa
cagtactgcgatgagtggcagggcggggcgtaattttttttaaggcagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagc
ggatgaatggcagaaatTCGCTGCAgcagcatcaaagttctggtgctggtagctgcgccagaaggtatcgctgcgctggaaaaagcgcaccc
ggacgtcgaactgtataccgcatcgattgatcagggactgaacgagcacggatacattattccgggcctcggcgatgccggtgacaaaatctttggtac
gaaataaagaattcgaagcttgggcccgaacaaaaactcatctcagaagaggatctgaatagcgccgtcgaccatcaccatcatcattgagtttaaa
cgacgtccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcct
ggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtgaggcctcccatgcgagag
tagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgatatctgcttttcttcg
gatccctcgagagatctccatgggctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaa
aaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgagc
ggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcggcaaagccgtttttccata
ggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggc
ggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgg
gtaggcagttcgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgga
aagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaa
gttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttt
tcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaaggggtctgacgcacatgtaatgtagcacctgaagtcagccccca
tacgatataagttgtaattctcatgtttgacagcttatcatcgataagctttaatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgtgtat
gaaatctaacaatgcgctcatcgtcatcctcggcaccgtcaccctggatgctgtaggcataggcttggttatgccggtactgccgggcctcttgcgggat
atcgtccattccgacagcatcgccagtcactatggcgtgctgctagcgctatatgcgttgatgcaatttctatgcgcacccgttctcggagcactgtccga
ccgctttggccgccgcccagtcctgctcgcttcgctacttggagccactatcgactacgcgatcatggcgaccacacccgtcctgtggattctctacgc
cggacgcatcgtggccggcatcaccggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttc
gggctcatgagcgcttgtttcggcgtgggtatggtggcaggccccgtggccgggggactgttgggcgccatctccttgcatgcaccattccttgcggc
ggcggtgctcaacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtcgaccgatgcccttgagagccttcaacc
cagtcagctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgactgtcttctttatcatgcaactcgtaggacaggtgccggcagcgct

ctgggtcattttcggcgaggaccgctttcgctggagcgcgacgatgatcggcctgtcgcttgcggtattcggaatcttgcacgccctcgctcaagccttc
gtcactggtcccgccaccaaacgtttcggcgagaagcaggccattatcgccggcatggcggccgacgcgctgggctacgtcttgctggcgttcgcga
cgcgaggctggatggccttccccattatgattcttctcgcttccggcggcatcgggatgcccgcgttgcaggccatgctgtccaggcaggtagatgacg
accatcagggacagcttcaaggatcgctcgcggctcttaccagcctaacttcgatcattggaccgctgatcgtcacggcgatttatgccgcctcggcga
gcacatggaacgggttggcatggattgtaggcgccgccctataccttgtctgcctccccgcgttgcgtcgcggtgcatggagccgggccacctcgacc
tgaatggaagccggcggcacctcgctaacggattcaccactccaagaattggagccaatcaattcttgcggagaactgtgaatgcgcaaaccaaccct
tggcagaacatatccat

pRepJI-EcW

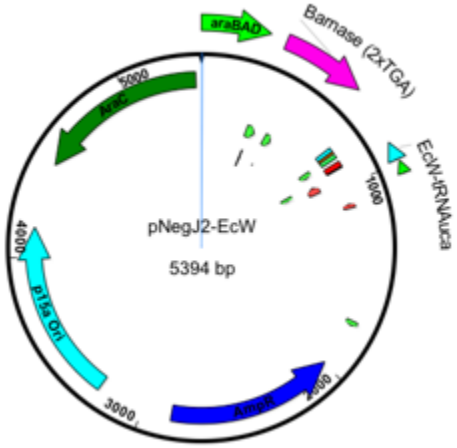


cgcgtccgccatctccagcagccgcacgcggcgcatcttgggctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactac
tgggctgcttcctaatgcaggagtcgcataagggagagcgtctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcc
cagtcacgacgttgtaaaacgacggccagtgccaagcttaaaaaaaatccttagctttcgctaaggatcaTATGCCTAGGTGGCAGG
GGCGGAGAGACTCGAACTCCCAaCACCCGGTTTTGAAGACCGGTGCTCTACCAATTGA
ACTACGCCCCTGAATTCagcgttacaagtattacacaaagtttttatgttgagaatatttttttgatggggcgccacttatttttgatcgtt
cgctcaaagaagcggcgccagggttgtttttctttttcaccagtgagacgggcaacagaacGGTACCctctagacaattggtgcacttcaaaaatcga
tgagctgttgacaattaatcatcgaactagtgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttcc
aactttcaccataatgaaataagatcactaccgggcgtatttttttgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactgg
atataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttcagctggatatt
acggccttttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaattccgtatg
gcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagTGATctgaaacgttttcatcgctctggagtga
ataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatttccctaaagggtttattgagaatat
gtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaatat
tatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtTtgtgatggcttccatgtcggcagaatgcttaatgaattacaac
agtactgcgatgagtggcagggcggggcgtaattttttttaaggcagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcg
gatgaatggcagaaatTCGCTGCAggctactcaggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcc
tttcgttttatttgatgcctggcagttccctactctcgcatggggagaccccacactaccatcggcgctacggcgtttcacttctgagttcggcatggggtc
aggtgggaccaccgcgctactgccgccaggcaaattctgttttatcagaccgcttctgcgttctgatttaatctgtatcaggctgaaaatcttctctcatcc
gacgtcttaggcgaaggcgaagtccgactctaagatgtcacggaggttcaagttacctttagccggaagtgctggcattttgtccaattgagactcgtgc
aactggtcagcgaactggtcgtagaaatcagccagtacatcacaagactcatatgtgtcaaccatagtttcgcgcactgctttgaacaggttcgcagcgt
cagccggaatggtaccgaaggagtcgtgaatcagtgcaaaagattcgattccgtacttctcgtgtgcccacactacagtcttacgaaggtggctaccgt
cttggctgtgtacaaagttaggagcgataccagactcctgtttgtgtgcatcaatctcgctatctttgttggtgttaatggtaggctgtaagcggaactgac
cgaggaacatcaggttcaagcgcgtctgaataggcttcttgtattcctgccacacagggaaaccatcaggagttacccaatgcacagcgcaacgcttgc
gaagaatctctccagtcttcttatctttgacctcagcagccagcagcttagcagcagacttaagccagttcattgcttcaaccgcagctaccaccgtcacg
ctcacagattcccaaatcagcttagccatgtatccagcagcctgattcggctgagtgaacatcagacccttgccggaatcaatagctggctgaatggtat
cttccagcacttgttgacggaagccgaactctttggacccgtaagccagcgtcatgactgaacgcttagtcacactgcgagtaacaccgtaagccagcc
attgaccagccagtgccttagtgcccagcttgactttctcagagatttcaccagtgttctcatcggtcacggtaactacttcgttatcggtcccattgattgc
gtctgcttgtagaatctcgttgactttcttagcaacaatcccgtagatgtcctgaacggtttcactaggaagcaagttaaccgcgcgaccacctacctcat
ctcggagcatcgcggagaagtgctggatgccagagcaagacccgtcaaacgccagcggaagggagcagttatagctcaggccgtggtgctgtaccc
cagcgtactcaaagcagaacgcaaggaagcagaacggagaatcttgctcagcccaccaagtgttctccagtggagacttagcgcaagccatgatgttc

tcgtggttttcctcaatgaacttgatgcgctcagggaacggaaccttatcgacacccgcacagtttgcaccgtggattttcagccagtagtaaccttcctt
accgattggtttacctttcgccagcgtaagcagtcctttggtcatatcgttaccttgcgggttgaacattgacacagcgtaaacacgaccgcgccagtcc
atgttgtaagggaaccagatggccttatggttagcaaacttattggcttgctcaagcatgaactcaaggctgatacggcgagacttgcgagccttgtcctt
gcggtacacagcagcggcagcacgtttccacgcggtgagagcctcaggattcatgtcgatgtcttccggtttcatcgggagttcttcacgctcaatcgc
agggatgtcctcgaccggacaatgcttccacttggtgattacgttggcgaccgcAaggactttcttgttgattttccatgcggtgttttgcgcaatgttaat
cgctttgtacacctcaggcatgtaaacgtcttcgtagcgcatcagtgctttcttactgtgagtacgcaccagcgccagaggacgacgaccgttagccca
atagccaccaccagtaatgccagtccacggcttaggaggaactacgcaaggttggaacatcggagagatgccagccagcgcacctgcacgggttgc
gatagcctcagcgtattcaggtgcgagttcgatagtctcagagtcttgacctactacgccagcattttggcggtgtaagctaaccattccggttgactcaa
tgagcatctcgatgcagcgtactcctacatgaatagagtcttccttatgccacgaagaccacgcctcgccaccgagtagacccttagagagcatgtcag
cctcgacaacttgcataaatgctttcttgtagacgtgccctacgcgcttgttgagttgttcctcaacgttttcttgaagtgcttagcttcaaggtcacggat
acgaccgaagcgagcctcgtcctcaatggcccgaccgattgcgcttgctacagcctgaacggttgtattgtcagcactggttaggcaagccagagtgg
tcttaatggtgatgtacgctacggcttccggcttgatttctcacaggaactggaaggctgtcgggcgcttgccgcgcttagctttcacttcctcaaaccag
tcgttgatgcgtgcaatcatcttagggagtagggtagtgatgagaggcttggcggcagcgttatccgcaacctcaccagctttaagttgacgctcaaaca
tcttgcggaagcgtgcttcacccatctcgtaagactcatgctcaagggccaactgttcgcgagctaaacgctcaccgtaatggtcagccagagtgttga
acgggatagcagccagttcgatgtcagagaagtcgttcttagcaatgttaatggtattTCAgtgcacggtaatcatggtcatggttaattcctcctgtta
gcccaaaaaacgggtatggagaaacagtagagagttgcgataaaaagcgtcaggtaggatccgctaatcttatggataaaaatgctatggcatagcaaa
gtgtgacgccgtgcaaataatcaatgtggacttttctgccgtgattatagacacttttgttacgcgtttttgtcatggctttggtcccgctttgttacagaatg
cttttaataagcggggttaccggtttggttagcgagaagagccagtaaaagacgcagtgacggcaatgtctgatgcaatatggacaattggtttcttctct
gaatggcgggagtatgaaaagtatggctgaagcgcaaaatgatcccctgctgccgggatactcgtttaatgcccatctggtggcgggtttaacgccgat
tgaggccaacggttatctcgatttttttatcgaccgaccgctgggaatgaaaggttatattctcaatctcaccattcgcggtcaggggtggtgaaaaatc
agggacgagaatttgtttgccgaccgggtgatattttgctgttcccgccaggagagattcatcactacggtcgtcatccggaggctcgcgaatggtatca
ccagtgggtttactttcgtccgcgcgcctactggcatgaatggcttaactggccgtcaatatttgccaatacggggttctttcgcccggatgaagcgcac
cagccgcatttcagcgacctgtttgggcaaatcattaacgccgggcaaggggaagggcgctattcggagctgctggcgataaatctgcttgagcaattg
ttactgcggcgcatggaagcgattaacgagtcgctccatccaccgatggataatcgggtacgcgaggcttgtcagtacatcagcgatcacctggcaga
cagcaattttgatatcgccagcgtcgcacagcatgtttgcttgtcgccgtcgcgtctgtcacatcttttccgccagcagttagggattagcgtcttaagctg
gcgcgaggaccaacgtatcagccaggcgaagctgcttttgagcaccacccggatgcctatcgccaccgtcggtcgcaatgttggttttgacgatcaact
ctatttctcgcgggtatttaaaaaatgcaccggggccagcccgagcgagttccgtgccggttgtgaagaaaaagtgaatgatgtagccgtcaagttgtc
ataattggtaacgaatcagacaattgacggcttgacggagtagcatagggtttgcagaatccctgcttcgtccatttgacaggcacattatgcatgccgct
tcgccttcgcgcgcgaattgatctgctgcctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcctgc
agcaaaaaaccccctcaagacccgtttagaggccccaaggggttatgctagttattgctcagcggtggcagcagccaactcagcttcctttcgggctttgt
tatttgtagagctcatccatgccatgtgtaatcccagcagcagttacaaactcaagaaggaccatgtggtcacgcttttcgttgggatctttcgaaagggc
agattgtgtcgacaggtaatggttgtctggtaaaaggacagggccatcgccaattggagtattttgttgataatggtctgctagttgaacggatccatcttc
aatgttgtggcgaattttgaagttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagttatagttgtactcgagtttgtgtccgaga
atgtttccatcttctttaaaatcaatacctttttaactcgatacgattaacaagggtatcaccttcaaacttgacttcagcacgcgtcttgtagttcccgtcatct
ttgaaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaagtcatgccgtttcatatgatccggataacgggaaaagcattgaac
accataagagaaagtagtgacaagtgttggccatggaacaggtagttttccagtagtgcaaataaatttaagggtaagttttccgtatgttgcatcaccttc
accctctccactgacagaaaatttgtgcccattaacatcaccatctaattcaacaagaattgggacaactccagtgaaaagttcttctcctttactcatatgt
atatctccttcttaaagttaaacaaaattatttctagagggaaaccgttgtggtctccctatagtgagtcgtattaatttcgcgggatcggcccttccggctg
gctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtagcggccgcc
ctgcagcagcatcaaagttctggtgctggtagctgcgccagaaggtatcgctgcgctggaaaaagcgcacccggacgtcgaactgtataccgcatcga
ttgatcagggactgaacgagcacggatacattattccgggcctcggcgatgccggtgacaaaatctttggtacgaaataaagaattcgaagcttgggcc
cgaacaaaaactcatctcagaagaggatctgaatagcgccgtcgaccatcaccatcatcatcattgagtttaaacgacgtccagcttggctgttttggcg
gatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtccca
cctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtgaggcctcccatgcgagagtagggaactgccaggcatcaaataa
aacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgatatctgcttttcttcggatccctcgagagatctccatgggct
agcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcag
aatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcgga
gatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcggcaaagccgtttttccataggctccgcccccctgacaagcatcac
gaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcct
gcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggact
gtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggca
gcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagc

cagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggttttttcgttttcagagcaagagattacgcgcagac
caaaacgatctcaagaagatcatcttattaaggggtctgacgcacatgtaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgttt
gacagcttatcatcgataagctttaatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgtgtatgaaatctaacaatgcgctcatcgtca
tcctcggcaccgtcaccctggatgctgtaggcataggcttggttatgccggtactgccgggcctcttgcgggatatcgtccattccgacagcatcgcca
gtcactatggcgtgctgctagcgctatatgcgttgatgcaatttctatgcgcacccgttctcggagcactgtccgaccgctttggccgccgcccagtcct
gctcgcttcgctacttggagccactatcgactacgcgatcatggcgaccacacccgtcctgtggattctctacgccggacgcatcgtggccggcatcac
cggcgccacaggtgcggttgctggcgcctatatcgccgacatcaccgatggggaagatcgggctcgccacttcgggctcatgagcgcttgtttcggcg
tgggtatggtggcaggccccgtggccgggggactgttgggcgccatctccttgcatgcaccattccttgcggcggcggtgctcaacggcctcaaccta
ctactgggctgcttcctaatgcaggagtcgcataagggagagcgtcgaccgatgcccttgagagccttcaacccagtcagctccttccggtgggcgcg
gggcatgactatcgtcgccgcacttatgactgtcttctttatcatgcaactcgtaggacaggtgccggcagcgctctgggtcattttcggcgaggaccgc
tttcgctggagcgcgacgatgatcggcctgtcgcttgcggtattcggaatcttgcacgccctcgctcaagccttcgtcactggtcccgccaccaaacgtt
tcggcgagaagcaggccattatcgccggcatggcggccgacgcgctgggctacgtcttgctggcgttcgcgacgcgaggctggatggccttccccat
tatgattcttctcgcttccggcggcatcgggatgcccgcgttgcaggccatgctgtccaggcaggtagatgacgaccatcagggacagcttcaaggatc
gctcgcggctcttaccagcctaacttcgatcattggaccgctgatcgtcacggcgatttatgccgcctcggcgagcacatggaacgggttggcatggat
tgtaggcgccgccctataccttgtctgcctccccgcgttgcgtcgcggtgcatggagccgggccacctcgacctgaatggaagccggcggcacctcg
ctaacggattcaccactccaagaattggagccaatcaattcttgcggagaactgtgaatgcgcaaaccaaccccttggcagaacatatccat

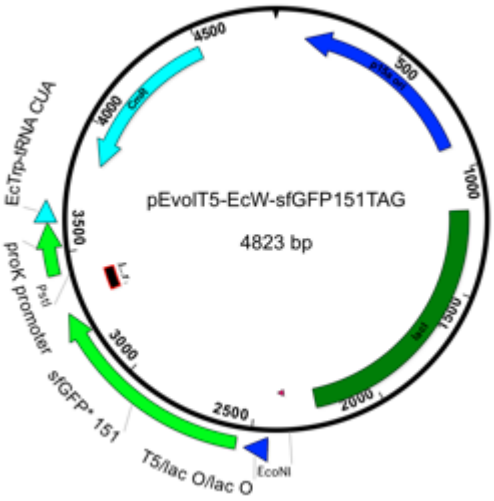
pNegJ2-EcW



aaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactggctcttctcgctaaccaaaccggtaaccccgcttattaaaagcattctgta
acaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagtgtctataatcacggcagaaaagtccacattgattatttgcacggcgtcacactttg
ctatgccatagcatttttatccataagattagcggatcctacctgacgctttttatcgcaactctctactgtttctccatacccgtttttttgggctagaaataa
ttttgtttaactttaagaaggagatatacatatggcatGAgttatcaacacgtttgacggggttgcggattatcttcagacatatcataagctacctgataa
ttacattacaaaatcagaagcacaagccctcggctgggtggcatcaaaagggaaccttgcatGAgtcgctccggggaaaagcatcggcggagacat
cttctcaaacagggaaggcaaactcccgggcaaaagcggacgaacatggcgtgaagcggatattaactatacatcaggcttcagaaattcagaccgga
ttctttactcaagcgactggctgatttacaaaacaacggaccattatcagacctttacaaaaatcagataagcatgcaccattccttgcggcggcggtgct
caacggcctcaacctactactgggctgcttcctaatgcaggagtcgcataagggagagcgtctggcgaaaggggatgtgctgcaaggcgattaagtt
gggtaacgccagggttGaattcttcccagtcacgacgttgtaaaacgacggccagtgccaagcttaaaaaaaatccttagctttcgctaaggatcaTA
TGCCTAGGTGGCAGGGGCGGAGAGACTCGAACTCCCAaCACCCGGTTTTGAAGACCG
GTGCTCTACCAATTGAACTACGCCCCTGaattcagcgttacaagtattacacaaagttttttatgttgagaatatttttttgat
ggggcgccacttatttttgatcgttcgctcaaagaagcggcgccagggttgtttttcttttgttctgttgcccgtctcactggtggccatgagcggcctcat
ttcttattctgagttacaacagtccgcaccgctgccggtagctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgactgtcttctttatc
atgcaactcgtaggacaggtgccggcagcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgccctgcaccatt
atgttccggatctgcatcgcaggatgctgctggctaccctgtggaacacctacatctgtattaacgaagcgctaaccgtttttatcatgctctgggaggca
gaataaatgatcatatcgtcaattattacctccacggggagagcctgagcaaactggcctcaggcatttgagaagcacacggtcacactgcttccggta
gtcaataaaccggtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgacgaccgggtcgaatttgctttcgaatttctgccatt
catccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgcag
tactgttgtaattcattaagcattctgccgacatggaagccatcacaaacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgcg

tataatatttgcccatggggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaag
ttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcca
tagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccg
gctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgcc
gggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctgcaggcatcgtggtgtcacgctcgtcgtttggtatggcttcat
tcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaa
gttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaacc
aagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaacacgggataataccgcgccacatagcagaactttaaaagtgctc
atcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagc
atcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcat
actcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcg
cacatttccccgtacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcga
agtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtg
gcttctgtttctatcagctgtccctcctgttcagctactgacggggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatact
ggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacagg
atatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagat
gccaggaagatacttaacagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctc
aaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttacc
ggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaaccc
cccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaa
ttgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttc
aaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaa
gaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaatt
ctcatgtttgacagcattatcatcgataagctttaatgcggtagtttatcacagttaaattgctaacgcagtcaggcaccgtgtatgaaatctaacaatgcg
ctcatcgtcatcctcggcaccgtcaccctggatgctgtaggcataggcttggttatgccggtactgccgggcctcttgcgggatggccacgatgcgtcc
ggcgtagaggatctgctcatgtttgacagcttatcatcgatgcataatgtgcctgtcaaatggacgaagcagggattctgcaaaccctatgctactccgtc
aagccgtcaattgtctgattcgttaccaattatgacaacttgacggctacatcattcacttttttcttcacaaccggcacggaactcgctcgggctggcccc
ggtgcatttttttaaataccgcgagaaatagagttgatcgtcaaaaccaacattgcgaccgacggtggcgataggcatccgggtggtgctcaaaagcag
cttcgcctggctgatacgttggtcctcgcgccagcttaagacgctaatccctaactgctggcggaaaagatgtgacagacgcgacggcgacaagcaaa
catgctgtgcgacgctggcgatatcaaaattgctgtctgccaggtgatcgctgatgtactgacaagcctcgcgtacccgattatccatcggtggatggag
cgactcgttaatcgcttccatgcgccgcagtaacaattgctcaagcagatttatcgccagcagctccgaatagcgcccttcccccttgcccggcgttaatg
atttgcccaaacaggtcgctgaaatgcggctggtgcgcttcatccgggcgaaagaaccccgtattggcaaatattgacggccagttaagccattcatgc
cagtaggcgcgcggacgaaagtaaacccactggtgataccattcgcgagcctccggatgacgaccgtagtgatgaatctctcctggcgggaacagca
aaatatcacccggtcggcaaacaaattctcgtccctgatttttcaccaccccctgaccgcgaatggtgagattgagaatataacctttcattcccagcggt
cggtcgataaaaaaatcgagataaccgttggcctcaatcggcgttaaacccgccaccagatgggcattaaacgagtatcccggcagcaggggatcatt
ttgcgcttcagccatacttttcatactcccgccattcagagaag

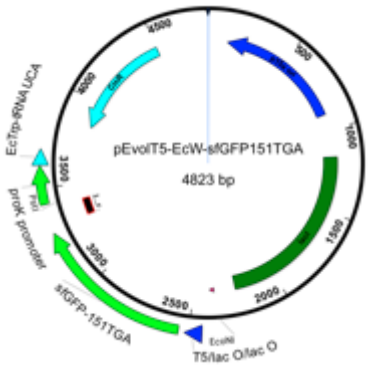
pEvolT5-EcW-sfGFP151TAG



gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgc
ttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtc
gttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcgg
caaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagatac
caggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacg
cctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcg
tcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaag
gctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccct
gcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagt
gcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctct
tcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaagg
ctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtc
gtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaaca
gctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacg
gcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcg
cccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccag
tcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggccc
gctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtct
ggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagccca
ctgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcg
gcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagtt
gttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttccgcgttttcgcagaaacgtggctggcctggttcaccacgcg
ggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatc
atgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggctcactatagggaattgtg
agcggataacaattcccctctagagtttgacagcattatcatcgatctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattatataatagat
tcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatATGAGCAAAGGAGAAGAACTTTTCA
CTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTC
TGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTAT
TTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTATG
GTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAG
TGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACC
TACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGT
TAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAA
CTTTAACTCACACAATGTATAGATCACGGCAGACAAACAAAAGAATGGAATCAAAGCT
AACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCA
ACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGA

CACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGA
GTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCCAC
CACCACCACCACCACTAAaagcttaattagctgagcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgt
tcagaacgctcggttgccgccgggcgtttttttattggtgagaatccaagctagcttggcgCTGCAgtgtgcttctcaaatgcctgaggccagtttgc
tcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctatt
aagggattgacgagggcgTATCTgcgcagtaagatgcgccccgcatttAGGGGGCGTAGTTCAATTGGTAGAGCAC
CGGTCTCtAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAAATTCGAAAA
GCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCTT
TCGAatttctgccattcatccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgc
cctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatca
gcaccttgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcac
ccagggattggctgagacgaaaaacatattctcaataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgta
gaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatat
caccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtgcttatttt
tctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgcc
attgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgat
cttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacacc
aggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtg
tttttgaggtgctccagtggcttctgtttctatcagctgtccctcctgttcagctactgacg

pEvolT5-EcW-sfGFP151TGA



gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgc
ttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtc
gttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcgg
caaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagatac
caggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacg
cctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcg
tcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaag
gctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccct
gcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagt
gcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctct
tcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaagg
ctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtc
gtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaaca
gctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacg
gcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcg
cccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccag
tcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggccc
gctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtct

ggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagccca
ctgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcg
gcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagtt
gttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcg
ggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatc
atgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggctcactataggggaattgtg
agcggataacaattcccctctagagtttgacagcattGtcatcgatctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataataga
ttcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatATGAGCAAAGGAGAAGAACTTTTC
ACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTT
CTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTA
TTTGCACTACTGGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCTatG
GTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAG
TGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACC
TACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGT
TAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAA
CTTTAACTCACACAATGTATgaATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTA
ACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAA
CAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGAC
ACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAG
TTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAAGGATCCCACC
ACCACCACCACCACTAAaagcttaattagctgagcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttc
agaacgctcggttgccgccgggcgtttttttattggtgagaatccaagctagcttggcgCTGCAgtgtgcttctcaaatgcctgaggccagtttgctc
aggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaa
gggattgacgagggcgTATCTgcgcagtaagatgcgccccgcatttAGGGGCGTAGTTCAATTGGTAGAGCACC
GGTCTTCAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAAATTCGAAAA
GCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAGGGTCGAATTTGCTT
TCGAatttctgccattcatccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgc
cctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatca
gcaccttgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcac
ccagggattggctgagacgaaaaacatattctcaataaacccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgta
gaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatat
caccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtgcttatttt
tctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgcc
attgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgat
cttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacacc
aggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtg
tttttgaggtgctccagtggcttctgtttctatcagctgtccctcctgttcagctactgacg

pEvoltac-EcW-TGA-h14

gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgc
ttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtc
gttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcgg
caaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagatac
caggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccacg
cctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcg
tcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaag
gctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccct
gcaaggcggttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagt
gcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctct
tcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaagg
ctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtc
gtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttctttttcaccagtgagacgggcaaca

gctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacg
gcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcg
cccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccag
tcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggccc
gctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtct
ggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagccca
ctgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcg
gcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagtt
gttgtgccacgcggttgggaatgtaattcagctcccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc
gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctat
catgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggGAGCTGTTTGA
CAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACAaA
GGAggtGCggccgcatgactaagcccatcgtttttGCTggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtc
agtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaagcga
cgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttaggctgg
gcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaacgctggtctgtttg
actatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgGGTCCTTGTggtgaagaccagaaacagcacctcgaactgagc
cgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctgg
agccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcg
gtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggcca
gagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgca
ggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgcta
aaagcggtgtacgaagcgattggttttgtggcgaagccgtaagcGGCCGCgtttaaacggtctccagcttggctgttttggcggatgagagaaga
ttttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgc
cgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctc
agtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgttagCCTGCAgtgtgcttctcaaatgcctgaggccag
tttgctcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttg
ctattaagggattgacgagggcgTATCTgcgcagtaagatgcgccccgcatttAGGGGCGTAGTTCAATTGGTAGAG
CACCGGTCTTCAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAAATTCG
AAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGctcgagcagctcagggtcgaatttgCtttcgaatttctgcca
ttcatccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgca
gtactgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttg
cgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctgag
acgaaaaacatattctcaataaacccttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatc
gtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctt
tcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaa
aggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacg
gtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtga
aagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgc
gaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagt
ggcttctgtttctatcagctgtccctcctgttcagctactgacg

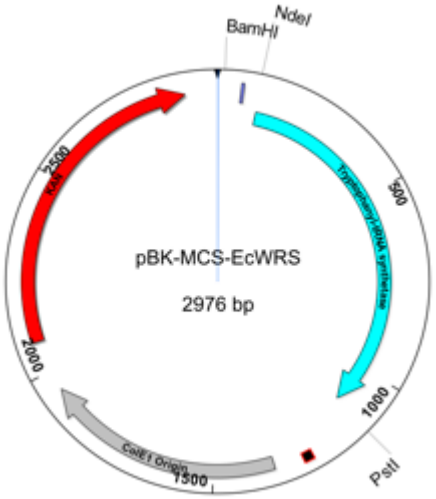
pEvoltac-EcW-TGA-h9

The sequence is identical to pEvoltac-EcW-TGA-h14 except for the V144-V146 region. The -h9 aaRS is listed below, with mutations in blue:

ATGACTAAGCCCATCGTTTTTgctGGCGCACAGCCCTCAGGTGAATTGACCATTGGTAA
CTACATGGGTGCGCTGCGTCAGTGGGTAAACATGCAGGATGACTACCATTGCATTTA
CTGTATCGTTGACCAACACGCGATCACCGTGCGCCAGGATGCACAGAAGCTGCGTAA
AGCGACGCTGGATACGCTGGCCTTGTATCTGGCTTGTGGTATCGATCCTGAGAAAAG
CACCATTTTTGTTCAGTCCCACGTGCCGGAACATGCACAGTTAGGCTGGGCACTGAAC
TGCTATACCTACTTCGGCGAACTGAGTCGCATGACGCAGTTTAAAGATAAATCTGCGC
GTTATGCCGAGAACATCAACGCTGGTCTGTTTGACTATCCGGTGCTGATGGCAGCGG

ACATCCTGCTGTATCAAACTAATCTGagtcctgctGGTGAAGACCAGAAACAGCACCTCGAA
CTGAGCCGCGATATTGCCCAGCGTTTCAACGCGCTGTATGGCGAGATCTTTAAGGTG
CCGGAGCCGTTTATTCCGAAATCTGGCGCGCGCGTAATGTCGCTGCTGGAGCCGACC
AAGAAGATGTCCAAGTCTGACGATAATCGCAATAACGTTATCGGCCTGCTGGAAGAT
CCGAAATCGGTAGTGAAGAAAATCAAACGTGCGGTCACTGACTCCGACGAGCCGCCG
GTAGTTCGCTACGATGTGCAGAACAAAGCGGGCGTTTCCAACCTGTTGGATATCCTTT
CAGCGGTAACGGGCCAGAGCATCCCAGAACTGGAAAAACAGTTCGAAGGCAAGATGT
ATGGTCATCTGAAAGGTGAAGTGGCTGATGCCGTTTCCGGTATGCTGACTGAATTGC
AGGAACGCTATCACCGTTTCCGCAACGATGAAGCCTTCCTGCAACAGGTGATGAAAG
ATGGCGCGGAAAAAGCCAGCGCGCACGCTTCCCGTACGCTAAAAGCGGTGTACGAAG
CGATTGGTTTTGTGGCGAAGCCGTAA

pBK-EcWRS



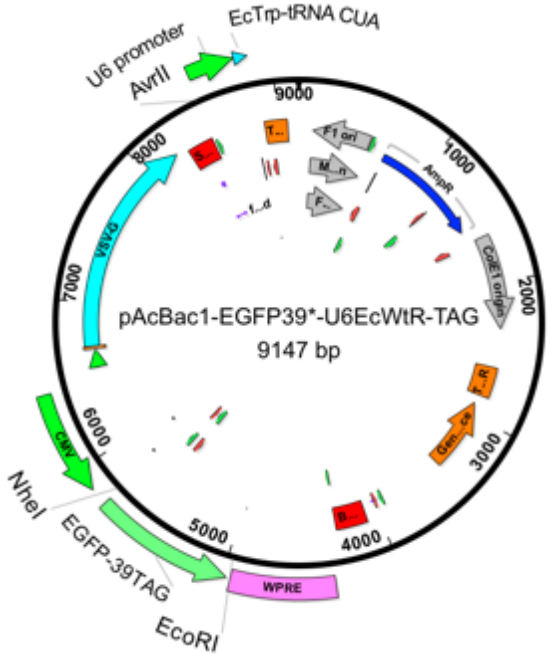
cttttgctgagttgaaggatccGCGGCCGCtcgggttgtcagcctgtcccgcttataagatcatacgccgttatacgttgtttacgctttgaggaa
tcccaTATGatgactaagcccatcgtttttagtggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaa
acatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaagcgacgctggata
cgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaa
ctgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccgg
tgctgatggcagcggacatcctgctgtatcaaactaatctggtaccggtgggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccag
cgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaag
atgtccaagtctgacgataatcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgac
gagccgccggtagttcgctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaact
ggaaaaacagttcgaaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccg
tttccgcaacgatgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaa
gcgattggttttgtggcgaagccgtaaCTGCAgtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatctctgcaatatattg
agtttgcgtgcttttgtaggccggataaggcgttcacgccgcatccggcaagaaacagcaaacaatccaaaacgcccgcgttcagcggcgttttttctgct
tttcttcgcgaattaattccgcttcgcacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggct
ccgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagc
tccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtag
gtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtctt
gagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttctt
gaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttga
tccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttgatcttttc
tacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgtt
atgagccatattcaacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgg

gcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagat
gagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatc
cccgggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcc
tgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgag
cgtaatggctggcctgttgaacaagtctggaaagaaatgcataagcttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataac
cttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcgg
tgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaa
tcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggacggcggctttgttgaataaatcgaa

pBK-EcWRS-h14
Same pBK sequence as above, with EcWRS-h14, below with mutations in blue:
Atgactaagcccatcgttttgctggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggat
gactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttg
tatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctataccta
cttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggca
gcggacatcctgctgtatcaaactaatctgggtccttgtggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcg
ctgtatggcgagatctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtct
gacgataatcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggt
agttcgctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagtt
cgaaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgat
gaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgt
ggcgaagccgtaa

pBK-EcWRS-h9
Same pBK sequence as above, with EcWRS-h9, below with mutations in blue:
ATGACTAAGCCCATCGTTTTTgctGGCGCACAGCCCTCAGGTGAATTGACCATTGGTAA
CTACATGGGTGCGCTGCGTCAGTGGGTAAACATGCAGGATGACTACCATTGCATTTA
CTGTATCGTTGACCAACACGCGATCACCGTGCGCCAGGATGCACAGAAGCTGCGTAA
AGCGACGCTGGATACGCTGGCCTTGTATCTGGCTTGTGGTATCGATCCTGAGAAAAG
CACCATTTTTGTTCAGTCCCACGTGCCGGAACATGCACAGTTAGGCTGGGCACTGAAC
TGCTATACCTACTTCGGCGAACTGAGTCGCATGACGCAGTTTAAAGATAAATCTGCGC
GTTATGCCGAGAACATCAACGCTGGTCTGTTTGACTATCCGGTGCTGATGGCAGCGG
ACATCCTGCTGTATCAAACTAATCTGagtcctgctGGTGAAGACCAGAAACAGCACCTCGAA
CTGAGCCGCGATATTGCCCAGCGTTTCAACGCGCTGTATGGCGAGATCTTTAAGGTG
CCGGAGCCGTTTATTCCGAAATCTGGCGCGCGCGTAATGTCGCTGCTGGAGCCGACC
AAGAAGATGTCCAAGTCTGACGATAATCGCAATAACGTTATCGGCCTGCTGGAAGAT
CCGAAATCGGTAGTGAAGAAAATCAAACGTGCGGTCACTGACTCCGACGAGCCGCCG
GTAGTTCGCTACGATGTGCAGAACAAAGCGGGCGTTTCCAACCTGTTGGATATCCTTT
CAGCGGTAACGGGCCAGAGCATCCCAGAACTGGAAAAACAGTTCGAAGGCAAGATGT
ATGGTCATCTGAAAGGTGAAGTGGCTGATGCCGTTTCCGGTATGCTGACTGAATTGC
AGGAACGCTATCACCGTTTCCGCAACGATGAAGCCTTCCTGCAACAGGTGATGAAAG
ATGGCGCGGAAAAAGCCAGCGCGCACGCTTCCCGTACGCTAAAAGCGGTGTACGAAG
CGATTGGTTTTGTGGCGAAGCCGTAA

pAcBac1-EGFP39*-U6-EcWtR-TAG

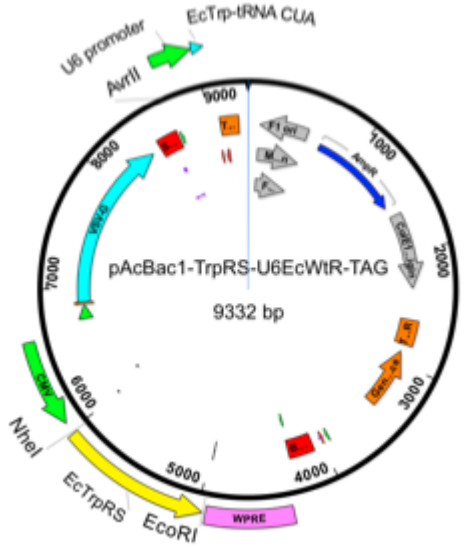


ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaatgggacgc
gccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttctt
cccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgacccca
aaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactct
tgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaa
caaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttctaaat
acattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatggcgtagagtattcaacatttccgtgt
cgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagt
gggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtgg
cgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaa
gcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggagg
accgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacga
gcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactgg
atggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgc
ggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagaca
gatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaag
gatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggat
cttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactc
tttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccg
cctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataa
ggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattg
agaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccaggg
ggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaa
cgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccg
cctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttct
ccttacgcatctgtgcggtatttcacaccgcagaccagccgcgtaacctggcaaaatcggttacggttgagtaataaatggatgccctgcgtaagcggg
tgtgggcggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactagacagaatagttgtaaactgaaatcagtccag
ttatgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcattttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaa
gggcatggtaaagactatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcgg
tacttgggtcgatatcaaagtgcatcacttcttcccgtatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcacgtagatc
acataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtggcaatgccctgcctccggtgctcgccggagactgcgagatcata

gatatagatctcactacgcggctgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatg
aatgtcttactacggagcaagttcccgaggtaatcggagtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaaga
gcagcccgcatggatttgacttggtcagggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgccctgct
gcgtaacatcgttgctgctgcgtaacatcgttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcata
gactgtacaaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttctggaccagttgcgtgagcg
catacgctacttgcattacagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccacggtgtgcgtcacccggcaaccttgg
gcagcagcgaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtctccacgcatcgtcaggcattggcggccttgctgttcttct
acggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggccgtcgcggcgcttgccggtggtgctgaccccggatgaagt
ggttcgcatcctcggttttctggaaggcgagcatcgtttgttcgcccaggactctagctatagttctagtggttggctacgtacccgtagtggctatggca
gggcttgcgcttaatgcgccgctacagggcgcgtggggataccccctagagccccagctggttctttccgcctcagaagccatagagcccaccgcatc
cccagcatgcctgctattgtcttcccaatcctcccccttgctgtcctgccccaccccaccccccagaatagaatgacacctactcagacaatgcgatgca
atttcctcattttattaggaaaggacagtgggagtggcaccttccagggtcaaggaaggcacgggggagggcaaacaacagatggctggcaactag
aaggcacagtcgaggctgatcagcgggtttaaacgggccctctagactcgagttaaagtcgacgcggggaggcggcccaaagggagatccgactcg
tctgagggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaagggacgt
agcagaaggacgtcccgcgcagaatccaggtggcaacacaggcgagcagccaaggaaaggacgatgatttccccgacaacaccacggaattgtca
gtgcccaacagccgagcccctgtccagcagcgggcaaggcaggcggcgatgagttccgccgtggcaatagggaggggaaagcgaaagtcccgg
aaaggagctgacaggtggtggcaatgccccaaccagtgggggttgcgtcagcaaacacagtgcacaccacgccacgttgcctgacaacgggccaca
actcctcataaagagacagcaaccaggatttatacaaggaggagaaaatgaaagccatacgggaagcaatagcatgatacaaaggcattaaagcagc
gtatccacatagcgtaaaaggagcaacatagttaagaataccagtcaatctttcacaaattttgtaatccagaggttgattgtcgacttaacgcgttgaatt
cTCAATGGTGATGGTGATGATGACCGGTATGCATATTCAGATCCTCTTCTGAGATGAG
TTTTTGTTCGAAGGGCCCCTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGC
GGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAG
GGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGA
TGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTCGATGTTGT
GGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATATAGAC
GTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTG
AAGTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCACCT
CGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGT
AGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAG
CGGCTGAAGCACTGCACGCCGTAgGTCAGgGTGGTCACGAGGGTGGGCCAGGGCACG
GGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCCctaAGTGGCATCGC
CCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCCGTCCAGCTCGAC
CAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGCTCACCATggtggcggcGctagcc
agcttgggtctccctatagtgagtcgtattaatttcgataagccagtaagccagtaagcagtgggttctctagttagccagagagctctgcttatatagacc
tcccaccgtacacgcctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccgttgattttggtgccaaaacaaactcccatt
gacgtcaatggggtggagacttggaaatccccgtgagtcaaaccgctatccacgcccattgatgtactgccaaaaccgcatcaccatggtaatagcgat
gactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgccaggcgggccatttaccgtcattgacgtcaatagg
gggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaaatagtccacccattgacgtcaatggaaagtccctattggcgtta
ctatgggaacatacgtcattattgacgtcaatgggcgggggtcgttgggcggtcagccaggcgggccatttaccgtaagttatgtaacgcggaactcca
tatatgggctatgaactaatgaccccgtaattgattactattaataactagtcaataatcaatgtcaacgcgtatatctggcccgtacatcgcgaagcagcg
caaaacGGATCCtgcaggtatttGCGGCCGCggtccgtatactccggaatatttaatagatcatggagataattaaaatgataaccatctcg
caaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccaccatcgggcgcgAACT
CCTAAAAAACCGCCACCatgaagtgccttttgtacttagccttttattcattggggtgaattgcaagttcaccatagtttttccacacaac
caaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgacttaataggcacagccttacaagtc
aaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgtgatttccgctggtatggaccgaagt
atataacacattccatccgatccttcactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttccc
tcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgg
gttgattcacagttcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaacctggcattctgactataaggtcaaagggctatg
tgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactacttt
gcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggat
ctctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttgagag
gatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcctaaaaacccag

gaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgctccaatcctctcaagaatggtc
ggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaattggacccaatggagttctgaggacca
gttcaggatataagtttcctttatacatgattggacatggtatgttggactccgatcttcatcttagctcaaaggctcaggtgttcgaacatcctcacattca
agacgctgcttcgcaacttcctgatgatgagagtttatttttggtgatactgggctatccaaaaatccaatcgagcttgtagaaggttggttcagtagttgg
aaaagctctattgcctctttttctttatcatagggttaatcattggactattcttggttctccgagttggtatccatctttgcattaaattaaagcacaccaaga
aaagacagatttatacagacatagagatgaaccgacttggaaagtgataaggccaggccggccaagcttgtcgagaagtactagaggatcataatcag
ccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgttta
ttgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaat
gtatcttatcatgtctggatctgatcactgcttgagcCTAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT
CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGA
CTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGG
GTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTT
GAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCAGGGG
CGTAGTTCAATTGGTAGAGCACCGGTCTCtAAAACCGGGTGtTGGGAGTTCGAGTCTC
TCCGCCCCTGCCATTTTTTgctaggctCAAgcagtgatctccgaaccagataagtgaaatctagttccaaactattttgtcattttta
attttcgtattagcttacgacgctacacccagttcccatctattttgtcactcttccctaaataatccttaaaaactccatttccacccctcccagttcccaac
tattttgtccgcccacagcggggcatttttcttcctgttatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttcggctacttt

pAcBac1-TrpRS-U6EcWtR-TAG



ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaatgggacgc
gccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttctt
cccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgacccca
aaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactct

tgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaa
caaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaat
acattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttt
attcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttac
atcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggt
attatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatctt
acggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaa
ggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtga
caccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggag
gcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatca
ttgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgct
gagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctag
gtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttga
gatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttccg
aaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacat
acctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgca
gcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaag
cgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaac
gcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccag
caacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttga
gtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacg
catctgtgcggtatttcacaccgcagaccagccgcgtaacctggcaaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgtgggc
ggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactagacagaatagttgtaaactgaaatcagtccagttatgctg
tgaaaaagcatactggacttttgttatggctaaagcaaactcttcattttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaagggcatg
gtaaagactatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtacttgg
gtcgatatcaaagtgcatcacttcttcccgtatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcacgtagatcacataag
caccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtggcaatgccctgcctccggtgctcgccggagactgcgagatcatagatatag
atctcactacgcggctgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtct
tactacggagcaagttcccgaggtaatcggagtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagcagcc
cgcatggatttgacttggtcagggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgccctgctgcgtaa
catcgttgctgctgcgtaacatcgttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatagactgta
caaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttctggaccagttgcgtgagcgcatacgc
tacttgcattacagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccacggtgtgcgtcacccggcaaccttgggcagcag
cgaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtctccacgcatcgtcaggcattggcggccttgctgttcttctacggcaa
ggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggccgtcgcggcgcttgccggtggtgctgaccccggatgaagtggttcgca
tcctcggttttctggaaggcgagcatcgtttgttcgcccaggactctagctatagttctagtggttggctacgtacccgtagtggctatggcagggcttgc
gcttaatgcgccgctacagggcgcgtggggatacccccctagagccccagctggttctttccgcctcagaagccatagagcccaccgcatccccagca
tgcctgctattgtcttcccaatcctcccccttgctgtcctgccccaccccaccccccagaatagaatgacacctactcagacaatgcgatgcaatttcctc
attttattaggaaaggacagtgggagtggcaccttccagggtcaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaaggcac
agtcgaggctgatcagcgggtttaaacgggccctctagactcgagttaaagtcgacgcggggaggcggcccaaagggagatccgactcgtctgagg
gcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaagggacgtagcagaa
ggacgtcccgcgcagaatccaggtggcaacacaggcgagcagccaaggaaaggacgatgatttccccgacaacaccacggaattgtcagtgcccaa
cagccgagcccctgtccagcagcgggcaaggcaggcggcgatgagttccgccgtggcaatagggaggggaaagcgaaagtcccggaaaggagc
tgacaggtggtggcaatgccccaaccagtgggggttgcgtcagcaaacacagtgcacaccacgccacgttgcctgacaacgggccacaactcctcat
aaagagacagcaaccaggatttatacaaggaggagaaaatgaaagccatacgggaagcaatagcatgatacaaaggcattaaagcagcgtatccaca
tagcgtaaaaggagcaacatagttaagaataccagtcaatctttcacaaattttgtaatccagaggttgattgtcgacttaacgcgttGaattCTTAC
GGCTTCGCCACAAAACCAATCGCTTCGTACACCGCTTTTAGCGTACGGGAAGCGTGCG
CGCTGGCTTTTTCCGCGCCATCTTTCATCACCTGTTGCAGGAAGGCTTCATCGTTGCG
GAAACGGTGATAGCGTTCCTGCAATTCAGTCAGCATACCGGAAACGGCATCAGCCACT
TCACCTTTCAGATGACCATACATCTTGCCTTCGAACTGTTTTTCCAGTTCTGGGATGCT
CTGGCCCGTTACCGCTGAAAGGATATCCAACAGGTTGGAAACGCCCGCTTTGTTCTGC
ACATCGTAGCGAACTACCGGCGGCTCGTCGGAGTCAGTGACCGCACGTTTGATTTTCT

TCACTACCGATTTCGGATCTTCCAGCAGGCCGATAACGTTATTGCGATTATCGTCAGA
CTTGGACATCTTCTTGGTCGGCTCCAGCAGCGACATTACGCGCGCGCCAGATTTCGGA
ATAAACGGCTCCGGCACCTTAAAGATCTCGCCATACAGCGCGTTGAAACGCTGGGCAA
TATCGCGGCTCAGTTCGAGGTGCTGTTTCTGGTCTTCACCCACCGGTACCAGATTAGT
TTGATACAGCAGGATGTCCGCTGCCATCAGCACCGGATAGTCAAACAGACCAGCGTT
GATGTTCTCGGCATAACGCGCAGATTTATCTTTAAACTGCGTCATGCGACTCAGTTCG
CCGAAGTAGGTATAGCAGTTCAGTGCCCAGCCTAACTGTGCATGTTCCGGCACGTGG
GACTGAACAAAAATGGTGCTTTTCTCAGGATCGATACCACAAGCCAGATACAAGGCCA
GCGTATCCAGCGTCGCTTTACGCAGCTTCTGTGCATCCTGGCGCACGGTGATCGCGT
GTTGGTCAACGATACAGTAAATGCAATGGTAGTCATCCTGCATGTTTACCCACTGACG
CAGCGCACCCATGTAGTTACCAATGGTCAATTCACCTGAGGGCTGTGCGCCACTAAAA
ACGATGGGCTTAGTCATgctagccagcttgggtctccctatagtgagtcgtattaatttcgataagccagtaagcagtgggttctctag
ttagccagagagctctgcttatatagacctcccaccgtacacgcctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccg
ttgattttggtgccaaaacaaactcccattgacgtcaatggggtggagacttggaaatccccgtgagtcaaaccgctatccacgcccattgatgtactgc
caaaaccgcatcaccatggtaatagcgatgactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgccaggcg
ggccatttaccgtcattgacgtcaataggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaaatagtccacccattg
acgtcaatggaaagtccctattggcgttactatgggaacatacgtcattattgacgtcaatgggcgggggtcgttgggcggtcagccaggcgggccatt
taccgtaagttatgtaacgcggaactccatatatgggctatgaactaatgaccccgtaattgattactattaataactagtcaataatcaatgtcaacgcgt
atatctggcccgtacatcgcgaagcagcgcaaaacGGATCCtgcaggtatttGCGGCCGCggtccgtatactccggaatattaatagat
catggagataattaaaatgataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattca
taccgtcccaccatcgggcgcgAACTCCTAAAAAACCGCCACCatgaagtgcctttgtacttagccttttattcattggggtgaat
tgcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaattggcata
atgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactac
ttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaac
aaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtccaggtgactcctcaccatgt
gctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaacctgg
cattctgactataaggtcaaagggctatgtgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctgggaaagga
gggcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcag
gtgtctggttcgagatggctgataaggatctctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtg
gatgtaagtctaattcaggacgttgagaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtgg
atctcagctatcttgctcctaaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgat
attgctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtgga
aattggacccaatggagttctgaggaccagttcaggatataagtttccttatacatgattggacatggtatgttggactccgatcttcatcttagctcaaag
gctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttggtgatactgggctatccaaaaatccaatc
gagcttgtagaaggttggttcagtagttggaaaagctctattgcctctttttctttatcatagggttaatcattggactattcttggttctccgagttggtatc
catctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttggaaagtgataaggccaggccggccaagct
tgtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaacctgaaacata
aaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgc
attctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagcCTAGGTCGGGCAGGAAGAG
GGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGAT
AATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGA
AAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCA
TATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAG
GACGAAACACCAGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTCtAAAACCGGGTGt
TGGGAGTTCGAGTCTCTCCGCCCCTGCCATTTTTTTgctagggctaggagatccgaaccagataagtgaaatctag
ttccaaactattttgtcatttttaattttcgtattagcttacgacgctacacccagttcccatctattttgtcactcttccctaaataatccttaaaaactccattt
ccacccctcccagttcccaactattttgtccgcccacagcggggcatttttcttcctgttatgtttttaatcaaacatcctgccaactccatgtgacaaaccg
tcatcttcggctacttt

pAcBac1-EcWRS-h14-U6EcWtR-TAG
Same plasmid map as pAcBac1-TrpRS-U6EcWtR-TAG replaced with EcTrpRS–h14 sequence, shown above

pAcBac1-EcWRS-h9-U6EcWtR-TAG
Same plasmid map as pAcBac1-TrpRS-U6EcWtR-TAG replaced with EcTrpRS–h9 sequence, shown

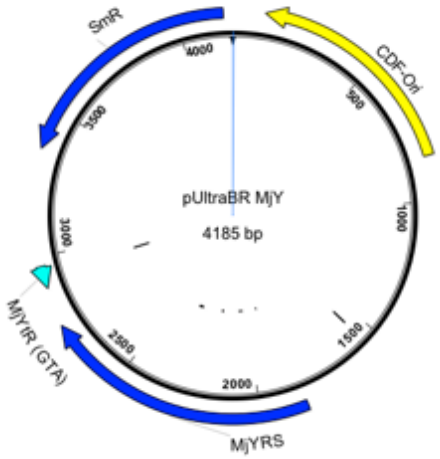
above

*****

Chapter 3 plasmid sequences

Final plasmids: Plasmid map and full plasmid sequence are given below. For each plasmid text, GFP is highlighted in green, CAT/Amp^R/Barnase in orange, T7 RNA polymerase in purple, tRNA in red, and aaRS in blue unless otherwise specified. Images are not color coded.

pUltraBR MjY



atcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagctagctcactcggtcgctacgctccgggcg
tgagactgcggcgggcgctgcggacacatacaaagttacccacagattccgtggataagcaggggactaacatgtgaggcaaaacagc
agggccgcgccggtggcgtttttccataggctccgcccctcctgccagagttcacataaacagacgcttttccggtgcatctgtgggagccg
tgaggctcaaccatgaatctgacagtacgggcgaaacccgacaggacttaaagatccccaccgtttccggcgggtcgctccctcttgcgc
tctcctgttccgaccctgccgtttaccggatacctgttccgcctttctcccttacgggaagtgtggcgctttctcatagctcacacactggtat
ctcggctcggtgtaggtcgttcgctccaagctgggctgtaagcaagaactccccgttcagcccgactgctgcgccttatccggtaactgtt
cacttgagtccaacccggaaaagcacggtaaaacgccactggcagcagccattggtaactgggagttcgcagaggatttgtttagctaaa
cacgcggttgctcttgaagtgtgcgccaaagtccggctacactggaaggacagatttggttgctgtgctctgcgaaagccagttaccacgg
ttaagcagttccccaactgacttaaccttcgatcaaaccacctccccaggtggttttttcgtttacagggcaaaagattacgcgcagaaaaa
aaggatctcaagaagatcctttgatcttttctactgaaccgctctagaCGTCCATTCCGACAGCATCGTCACTA
TGGCGTGCTGCTAGCATCGACGAGTCTGGCTTTGACATTCGACTAGAAGTGG
ACGGTGGCGTGAAGGTGAACAACATTGGCGAAATCGCTGCGGCGGGCGCGG
ATATGTTCGTCGCCGGTTCGGCAATCTTCGACCAGCCAGACTACAAAAAAGTC
ATTGATGAAATGCGCAGTGAACTGGCAAAGGTAAGTCATGAATAAGTTTGAA
GATATTCGCGGCGTCGCTTTTGATCTTGATGGTACGCTGGTCGACAGTGCTCC

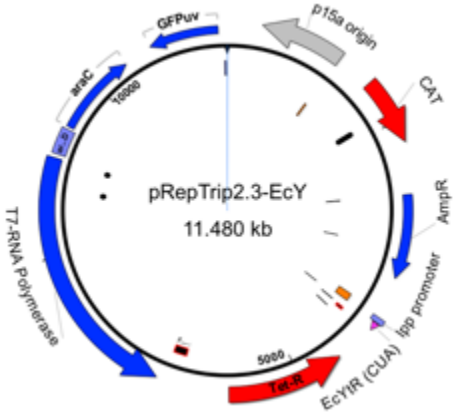
TGGTCTTGCTGCTGCGGTAGATATGGCGCTGTATGCGCTGGAGTTGCCCGTC
GCAGGTGAAGAACGCGTTATTACCTGGATTGGTAACGGCGCAGATGTTCTGA
TGGAGCGCGCATTGACCTGGGCGCGTCAGGAACGTGCGACTCAGCGTAAAAC
AATGGGTAAACCGCCCGTTGATGACGACATTCCGGCAGAAGAACAGGTACGT
ATTCTGCGTAAACTGTTCGATCGCTACTATGGCGAGGTTGCCGAAGAGGGGA
CGTTTTTGTTCCCGCACGTTGCCGATACGTTGGGCGCGTTGCAGGCTAAAGGC
CTGCCGCTAGGCCTGGTCACCAACAAACCGACGCCGTTCGTCGCGCCGCTGCT
CGAAGCCTTAGATATCGCCAAATACTTCAGCGCGGTGATTGGTGGTGATGAT
GTGCAAAACAAAAAACCGCATCCGGACCCGCTGTTACTGGTGGCTGAGCGGA
TGGGAATTGCCCCACAACAGATGCTGTTTGTCGGCGACTCACGCAATGATATT
CAGGCGGCAAAAGCGGCAGGTTGCCCATCAGTTGGCTTAACCTACGGATATA
ACTACGGCGAGGCTATCGATCTCAGCCAGCCTGATGTAATTTATCAGTCTATA
AATGACCTTCTGCCCGCATTAGGGCTTCCGCATAGCGAAAATCAGGAATCGAC
ATatggacgaatttgaaatgataaagagaaacacatctgaaattatcagcgaggaagagttaagagaggttttaaaaaaagatgaaaaat
ctgcttacataggttttgaaccaagtggtaaaatacatttagggcattatctccaaataaaaaagatgattgatttacaaaatgctggatttgat
ataattatattgttggctgatttacacgcctatttaaaccagaaaggagagttggatgagattagaaaaataggagattataacaaaaaagttt
ttgaagcaatggggttaaaggcaaaatatgtttatggaagtgaattccagcttgataaggattatacactgaatgtctatagattggctttaaa
aactaccttaaaaagagcaagaaggagtatggaacttatagcaagagaggatgaaaatccaaaggttgctgaagttatctatccaataatg
caggttaatgatattcattatttaggcgttgatgttgcagttggagggatggagcagagaaaaatacacatgttagcaagggagcttttacca
aaaaaggttgtttgtattcacaaccctgtcttaacgggtttggatggagaaggaaagatgagttcttcaaaagggaattttatagctgttgatg
actctccagaagagattagggctaagataaagaaagcatactgcccagctggagttgttgaaggaaatccaataatggagatagctaaata
cttccttgaatatcctttaaccataaaaaggccagaaaaatttggtggagatttgacagttaatagctatgaggagttagagagtttatttaaa
aataaggaattgcatccaatggatttaaaaaatgctgtagctgaagaacttataaagattttagagccaattagaaagagattataaagaTC
TGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCGGATGAAC
AGGCAGACATCTGTGAATCGCTTCACGACCACGCATCAAAAAAAATCCTTAGC
TTTCGCTAAGGATTTTATTTTGGTCCGGCGGGCCGGATTTGAACCAGCGACAT
GCGGATCTACAGTCCGCCGTTCTACCAGGCTGAACTACCGCCGGAGGAAAGT
TACAAGTATTACACAAAGTTTTTTATGTTGAGAATATTTTTTTGATGGGGCGCC
ACTTATTTTTGATCGTTCGCTCAAAGAAGCCTCGAGTCACTTTCGGCCGACGC
GCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCC
ATTATGgcatgctcgagcagctcagggtcgaatttgCCATGGcggccACCAGGTacCACCGGCGcctcag
gcatttgagaagcacacggtcacactgcttccggtagtcaataaaccggtaaaccagcaatagacataagcggctatttaacgaccctgcc
ctgaaccgacgaccgggtcatcgtggccggatcttgcggcccctcggcttgaacgaattgttagacattatttgccgactaccttggtgatc
tcgcctttcacgtagtggacaaattcttccaactgatctgcgcgcgaggccaagcgatcttcttcttgtccaagataagcctgtctagcttca
agtatgacgggctgatactgggccggcaggcgctccattgcccagtcggcagcgacatccttcggcgcgattttgccggttactgcgctg
taccaaatgcgggacaacgtaagcactacatttcgctcatcgccagcccagtcgggcggcgagttccatagcgttaaggtttcatttagcg
cctcaaatagatcctgttcaggaaccggatcaaagagttcctccgccgctggacctaccaaggcaacgctatgttctcttgcttttgtcagc
aagatagccagatcaatgtcgatcgtggctggctcgaagatacctgcaagaatgtcattgcgctgccattctccaaattgcagttcgcgctt
agctggataacgccacggaatgatgtcgtcgtgcacaacaatggtgacttctacagcgcggagaatctcgctctctccaggggaagccga
agtttccaaaaggtcgttgatcaaagctcgccgcgttgtttcatcaagccttacggtcaccgtaaccagcaaatcaatatcactgtgtggctt
caggccgccatccactgcggagccgtacaaatgtacggccagcaacgtcggttcgagatggcgctcgatgacgccaactacctctgata
gttgagtcgatacttcggcgatcaccgcttccctcatactcttccttttcaatattattgaagcattt
pRepTrip2.3 TAG
gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcac
tgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagat
cctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagacccctaataagatgatcttcttgagatcgttt
tggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggcggtttttcgaaggttctctgagctaccaactctttgaaccg

aggtaactggcttggaggagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaactcctctaaat
caattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttggactcaagacgatagttaccggataaggcgcagcggtcgg
actgaacgggggttcgtgcatacagtccagcttggagcgaactgcctacccggaactgagtgtcaggcgtggaatgagacaaacgcgg
ccataacagcggaatgacaccggtaaaccgaaaggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctggtatc
tttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttg
ccgcggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccatttccgctcgccgcagtc
gaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttttct
cctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgct
ttgccgttacgcaccaccccgtcagtagctgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacaccatc
atacactaaatcagtaagttggcagcatcacccgacgcactttgcgccgaataaatacctgtgacggaagatcacttcgcagaataaataa
atcctggtgtccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttca
ccataatgaaataagatcactaccgggcgtatttttttgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactg
gatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttca
gctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgct
catccggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactga
aacgttttcatcgctctggagtgaataccactaggatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacct
ggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaata
tggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcat
gccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaattttttttaagg
cagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgaccc
ggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttattgactaccggaagcagtgtgaccgtgtg
cttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattgacgatatgatcatttattctgcctcccagagcatgat
aaaaacggttagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggAc
tagcggtggcactttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatggcgtagagtattcaacatttccgtgtcgcccttattcccttttttgcggcatt
ttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactgg
atctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatt
atcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaag
catcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgat
cggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagc
cataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagct
tcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctg
ataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacac
gacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagacca
agtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcc
cttaacgtgagttttCACTAGCTaGGAaGATCTaaggagcccgagatgcgccgcgtgcggctgctggagatggcggac
gcgatggatatgttctgccaagggttggtttgcgcattcacagttctccgcaagaattgattggctccaattcttggagtggtgaatccgtta
gcgaggtgccgccggcttccattcaggtcgaggtggcccggctccatgcaccgcgacgcaacgcggggaggcagacaaggtataggg
cggcgcctacaatccatgccaacccgttccatgtgctcgccgaggcggcataaatcgccgtgacgatcagcggtccaatgatcgaagtta
ggctggtaagagccgcgagcgatccttgaagctgtccctgatggtcgtcatctacctgcctggacagcatggcctgcaacgcgggcatcc
cgatgccgccggaagcgagaagaatcataatggggaaggccatccagcctcgcgtcgcgaacgccagcaagacgtagcccagcgcgt
cggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaa
gattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggc
acctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgg
gttgaaggctctcaagggcatcggtcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagc
accgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaa
caagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgcc
ggtgatgccggccacgatgcgtccggcgtagaggatccacaggacgggtgtggtcgccatgatcgcgtagtcgatagtggctccaagta

gcgaagcgagcaggactgggcggcggccaaagcggtcggacagtgctccgagaacgggtgcgcatagaaattgcatcaacgcatata
gcgctagcagcacgccatagtgactggcgatgctgtcggaatggacgatatcccgcaagaggcccggcagtaccggcataaccaagcc
tatgcctacagcatccagggtgacggtgccgaggatgacgatgagcgcattgttagatttcatacacggtgcctgactgcgttagcaattta
actgtgataaactaccgcattaaagcttatcgatgataagctgtcaaacatgagaattacaacttatatcgtatggggctgacttcaggtgct
acattgctcaaagatgcaggggtaaaagctaaccgcatctttaccgacaaggcatccggcagttcaacagatcgggaagggctggatttg
ctgaggatgaaggtggaggaaggtgatgtcattctggtgaagaagctcgaccgtcttggccgcgacaccgccgacatgatccaactgata
aaagagtttgatgctcagggtgtagcggttcggtttattgacgacgggatcagtaccgacggtgatatggggcaaatggtggtcaccatcc
tgtcggctgtggcacaggctgaacgccggaggatcctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactc
aggagagcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggcagttcccta
ctctcgcatggggagaccccacactaccatcggcgctacggcgtttcacttctgagttcggcatggggtcaggtgggaccaccgcgctac
tgccgccaggcaaattctgttttatcagaccgcttctgcgttctgatttaatctgtatcaggctgaaaatcttctctcatccgacgtcttaggcg
aaggcgaagtccgactctaagatgtcacggaggttcaagttacctttagccggaagtgctggcattttgtccaattgagactcgtgcaactg
gtcagcgaactggtcgtagaaatcagccagtacatcacaagactcatatgtgtcaaccatagtttcgcgcactgctttgaacaggttcgcag
cgtcagccggaatggtaccgaaggagtcgtgaatcagtgcaaaagattcgattccgtacttctcgtgtgcccacactacagtcttacgaag
gtggctaccgtcttggctgtgtacaaagttaggagcgataccagactcctgtttgtgtgcatcaatctcgctatctttgttggtgttaatggta
ggctgtaagcggaactgaccgaggaacatcaggttcaagcgcgtctgaataggcttcttgtattcctgccacacagggaaaccatcagga
gttacccaatgcacagcgcaacgcttgcgaagaatctctccagtcttcttatctttgacctcagcagccagcagcttagcagcagacttaag
ccagttcattgcttcaaccgcagctaccaccgtcacgctcacagattcccaaatcagcttagccatgtatccagcagcctgattcggctgag
tgaacatcagacccttgccggaatcaatagctggctgaatggtatcttccagcacttgttgacggaagccgaactctttggacccgtaagc
cagcgtcatgactgaacgcttagtcacactgcgagtaacaccgtaagccagccattgaccagccagtgccttagtgcccagcttgactttc
tcagagatttcaccagtgttctcatcggtcacggtaactacttcgttatcggtcccattgattgcgtctgcttgtagaatctcgttgactttctta
gcaacaatcccgtagatgtcctgaacggtttcactaggaagcaagttaaccgcgcgaccacctacctcatctcggagcatcgcggagaag
tgctggatgccagagcaagacccgtcaaacgccagcggaagggagcagttatagctcaggccgtggtgctgtaccccagcgtactcaaa
gcagaacgcaaggaagcagaacggagaatcttgctcagcccaccaagtgttctccagtggagacttagcgcaagccatgatgttctcgtg
gttttcctcaatgaacttgatgcgctcagggaacggaaccttatcgacacccgcacagtttgcaccgtggattttcagccagtagtaacctt
ccttaccgattggtttacctttcgccagcgtaagcagtcctttggtcatatcgttaccttgcgggttgaacattgacacagcgtaaacacgac
cgcgccagtccatgttgtaagggaaccagatggccttatggttagcaaacttattggcttgctcaagcatgaactcaaggctgatacggcg
agacttgcgagccttgtccttgcggtacacagcagcggcagcacgtttccacgcggtgagagcctcaggattcatgtcgatgtcttccggt
ttcatcgggagttcttcacgctcaatcgcagggatgtcctcgaccggacaatgcttccacttggtgattacgttggcgaccgcAaggactt
tcttgttgattttccatgcggtgttttgcgcaatgttaatcgctttgtacacctcaggcatgtaaacgtcttcgtagcgcatcagtgctttcttac
tgtgagtacgcaccagcgccagaggacgacgaccgttagcccaatagccaccaccagtaatgccagtccacggcttaggaggaactac
gcaaggttggaacatcggagagatgccagccagcgcacctgcacgggttgcgatagcctcagcgtattcaggtgcgagttcgatagtctc
agagtcttgacctactacgccagcattttggcggtgtaagctaaccattccggttgactcaatgagcatctcgatgcagcgtactcctacat
gaatagagtcttccttatgccacgaagaccacgcctcgccaccgagtagacccttagagagcatgtcagcctcgacaacttgcataaatgc
tttcttgtagacgtgccctacgcgcttgttgagttgttcctcaacgtttttcttgaagtgcttagcttcaaggtcacggatacgaccgaagcga
gcctcgtcctcaatggcccgaccgattgcgcttgctacagcctgaacggttgtattgtcagcactggttaggcaagccagagtggtcttaat
ggtgatgtacgctacggcttccggcttgatttcCTAcaggaactggaaggctgtcgggcgcttgccgcgcttagctttcacttcctcaaa
ccagtcgttgatgcgtgcaatcatcttagggagtagggtagtgatgagaggcttggcggcagcgttatccgcaacctcaccagctttaagtt
gacgctcaaacatcttgcggaagcgtgcttcacccatctcgtaagactcatgctcaagggccaactgttcgcgagctaaacgctcaccgta
atggtcagccagagtgttgaacgggatagcagccagttcgatgtcagagaagtcgttcttagcaatgttaatggtattCTAgtgcacggt
aatcatggtcatggttaattcctcctgttagcccaaaaaacgggtatggagaaacagtagagagttgcgataaaaagcgtcaggtaggatc
cgctaatcttatggataaaaatgctatggcatagcaaagtgtgacgccgtgcaaataatcaatgtggacttttctgccgtgattatagacactt
ttgttacgcgtttttgtcatggctttggtcccgctttgttacagaatgcttttaataagcggggttaccggtttggttagcgagaagagccagta
aaagacgcagtgacggcaatgtctgatgcaatatggacaattggtttcttctctgaatggcgggagtatgaaaagtatggctgaagcgcaa
aatgatcccctgctgccgggatactcgtttaatgcccatctggtggcgggtttaacgccgattgaggccaacggttatctcgatttttttatcg
accgaccgctgggaatgaaaggttatattctcaatctcaccattcgcggtcaggggtggtgaaaaatcagggacgagaatttgtttgccg
accgggtgatattttgctgttcccgccaggagagattcatcactacggtcgtcatccggaggctcgcgaatggtatcaccagtgggtttact
ttcgtccgcgcgcctactggcatgaatggcttaactggccgtcaatatttgccaatacggggttctttcgcccggatgaagcgcaccagcc

gcatttcagcgacctgtttgggcaaatcattaacgccgggcaaggggaagggcgctattcggagctgctggcgataaatctgcttgagca
attgttactgcggcgcatggaagcgattaacgagtcgctccatccaccgatggataatcgggtacgcgaggcttgtcagtacatcagcgat
cacctggcagacagcaattttgatatcgccagcgtcgcacagcatgtttgcttgtcgccgtcgcgtctgtcacatcttttccgccagcagtta
gggattagcgtcttaagctggcgcgaggaccaacgtatcagccaggcgaagctgcttttgagcaccacccggatgcctatcgccaccgtc
ggtcgcaatgttggttttgacgatcaactctatttctcgcgggtatttaaaaaatgcaccggggccagcccgagcgagttccgtgccggttg
tgaagaaaaagtgaatgatgtagccgtcaagttgtcataattggtaacgaatcagacaattgacggcttgacggagtagcatagggtttgca
gaatccctgcttcgtccatttgacaggcacattatgcatgccgcttcgccttcgcgcgcgaattgatctgctgcctcgcgcgtttcggtgatg
acggtgaaaacctctgacacatgcagctcccggagacggtcacagcctgcagcaaaaaacccctcaagacccgtttagaggccccaag
gggttatgctagttattgctcagcggtggcagcagccaactcagcttcctttcgggctttgttatttgtagagctcatccatgccatgtgtaat
cccagcagcagttacaaactcaagaaggaccatgtggtcacgcttttcgttgggatctttcgaaagggcagattgtgtcgacaggtaatgg
ttgtctggtaaaaggacagggccatcgccaattggagtattttgttgataatggtctgctagttgaacggatccatcttcaatgttgtggcgaa
ttttgaagttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagttatagttgtactcgagtttgtgtccgagaatgttt
ccatcttctttaaaatcaataccttttaactcgatacgattaacaagggtatcaccttcaaacttgacttcagcacgcgtcttgtagttcccgtc
atctttgaaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaagtcatgccgtttcatatgatccggataacggga
aaagcattgaacaccataagagaaagtagtgacaagtgttggccatggaacaggtagttttccagtagtgcaaataaatttaagggtaagtt
ttccgtatgttgcatcaccttcaccctctccactgacagaaaatttgtgcccattaacatcaccatctaattcaacaagaattgggacaactcc
agtgaaaagttcttctcctttactcatatgtatatctccttcttaaagttaaacaaaattatttctagagggaaaccgttgtggtctccctatagt
gagtcgtattaatttcgcgggatcg

pRepTrip2.3-EcY-TAG



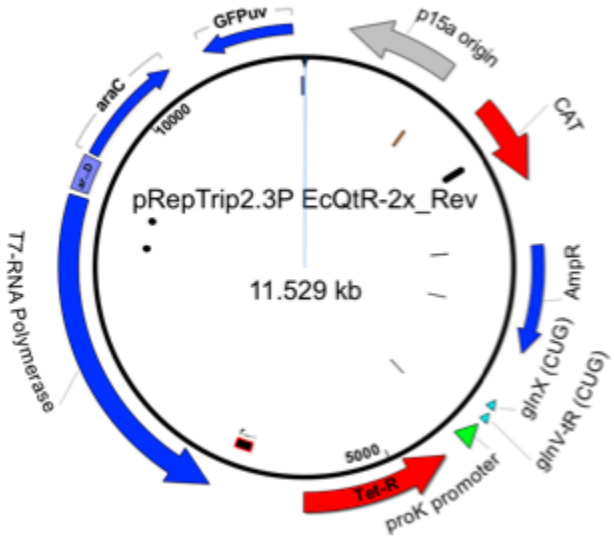
gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcac
tgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagat
cctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagacccccttaataagatgatcttcttgagatcgttt
tggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggcggtttttcgaaggttctctgagctaccaactctttgaaccg
aggtaactggcttggaggagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaactcctctaaat
caattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttggactcaagacgatagttaccggataaggcgcagcggtcgg
actgaacgggggttcgtgcatacagtccagcttggagcgaactgcctacccggaactgagtgtcaggcgtggaatgagacaaacgcgg
ccataacagcggaatgacaccggtaaaccgaaaggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctggtatc
tttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttg
ccgcggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccatttccgctcgccgcagtc
gaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttttct
cctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgct
tttgccgttacgcaccaccccgtcagtagctgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacaccatc

atacactaaatcagtaagttggcagcatcacccgacgcactttgcgccgaataaatacctgtgacggaagatcacttcgcagaataaataa
atcctggtgtccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttca
ccataatgaaataagatcactaccgggcgtatttttttgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactg
gatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttca
gctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgct
catccggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactga
aacgttttcatcgctctggagtgaataccactaggatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacct
ggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaata
tggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcat
gccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaattttttttaagg
cagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgaccc
ggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttattgactaccggaagcagtgtgaccgtgtg
cttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattgacgatatgatcatttattctgcctcccagagcatgat
aaaaacggttagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggAc
tagcggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatggcgtagagtattcaacatttccgtgtcgcccttattcccttttttgcggcatt
ttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactgg
atctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatt
atcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaag
catcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgat
cggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagc
cataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagct
tcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctg
ataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacac
gacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagacca
agtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcc
cttaacgtgagttttcaCTAGTgcgcttgtttcggcgtgggtatggtggcaggccccgtggccgggggactgttgggcgctgccggc
acctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctaccggc
agcggtgcggactgttgtaactcagaataagaaatgaggccgctcatggcgttctgttgcccgtctcactggtgaaaagaaaaacaaccct
ggcgccgcttctttgagcgaacgatcaaaaataagtggcgccccatcaaaaaaatattctcaacataaaaaactttgtgtaatacttgtaac
gctgaattcGGTGGGGTTCCCGAGCGGCCAAAGGGAGCAGACTCTAAATCTGCCGT
CATCGACTTCGAAGGTTCGAATCCTTCCCCCACCACCActgcagatccttagcgaaagctaagga
ttttttttaagcttggcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacGAA
TTCatcccccttttcgccagacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgcc
gccgcaaggaatggtCCTAGGaAGATCTaaggagcccgagatgcgccgcgtgcggctgctggagatggcggacgcgatg
gatatgttctgccaagggttggtttgcgcattcacagttctccgcaagaattgattggctccaattcttggagtggtgaatccgttagcgagg
tgccgccggcttccattcaggtcgaggtggcccggctccatgcaccgcgacgcaacgcggggaggcagacaaggtatagggcggcgc
ctacaatccatgccaacccgttccatgtgctcgccgaggcggcataaatcgccgtgacgatcagcggtccaatgatcgaagttaggctgg
taagagccgcgagcgatccttgaagctgtccctgatggtcgtcatctacctgcctggacagcatggcctgcaacgcgggcatcccgatgc
cgccggaagcgagaagaatcataatggggaaggccatccagcctcgcgtcgcgaacgccagcaagacgtagcccagcgcgtcggccg
ccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccg
aataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtc
ctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgcc
gccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcg
ctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgat
gccggccacgatgcgtccggcgtagaggatccacaggacgggtgtggtcgccatgatcgcgtagtcgatagtggctccaagtagcgaa
gcgagcaggactgggcggcggccaaagcggtcggacagtgctccgagaacgggtgcgcatagaaattgcatcaacgcatatagcgcta

gcagcacgccatagtgactggcgatgctgtcggaatggacgatatcccgcaagaggcccggcagtaccggcataaccaagcctatgcct
acagcatccagggtgacggtgccgaggatgacgatgagcgcattgttagatttcatacacggtgcctgactgcgttagcaatttaactgtg
ataaactaccgcattaaagcttatcgatgataagctgtcaaacatgagaattacaacttatatcgtatggggctgacttcaggtgctacattg
ctcaaagatgcaggggtaaaagctaaccgcatctttaccgacaaggcatccggcagttcaacagatcgggaagggctggatttgctgagg
atgaaggtggaggaaggtgatgtcattctggtgaagaagctcgaccgtcttggccgcgacaccgccgacatgatccaactgataaaagag
tttgatgctcagggtgtagcggttcggtttattgacgacgggatcagtaccgacggtgatatggggcaaatggtggtcaccatcctgtcgg
ctgtggcacaggctgaacgccggaggatcctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactcaggag
agcgttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggcagttccctactctcg
catggggagaccccacactaccatcggcgctacggcgtttcacttctgagttcggcatggggtcaggtgggaccaccgcgctactgccg
ccaggcaaattctgttttatcagaccgcttctgcgttctgatttaatctgtatcaggctgaaaatcttctctcatccgacgtcttaggcgaagg
cgaagtccgactctaagatgtcacggaggttcaagttacctttagccggaagtgctggcattttgtccaattgagactcgtgcaactggtca
gcgaactggtcgtagaaatcagccagtacatcacaagactcatatgtgtcaaccatagtttcgcgcactgctttgaacaggttcgcagcgtc
agccggaatggtaccgaaggagtcgtgaatcagtgcaaaagattcgattccgtacttctcgtgtgcccacactacagtcttacgaaggtgg
ctaccgtcttggctgtgtacaaagttaggagcgataccagactcctgtttgtgtgcatcaatctcgctatctttgttggtgttaatggtaggct
gtaagcggaactgaccgaggaacatcaggttcaagcgcgtctgaataggcttcttgtattcctgccacacagggaaaccatcaggagtta
cccaatgcacagcgcaacgcttgcgaagaatctctccagtcttcttatctttgacctcagcagccagcagcttagcagcagacttaagcca
gttcattgcttcaaccgcagctaccaccgtcacgctcacagattcccaaatcagcttagccatgtatccagcagcctgattcggctgagtga
acatcagacccttgccggaatcaatagctggctgaatggtatcttccagcacttgttgacggaagccgaactctttggacccgtaagccag
cgtcatgactgaacgcttagtcacactgcgagtaacaccgtaagccagccattgaccagccagtgccttagtgcccagcttgactttctca
gagatttcaccagtgttctcatcggtcacggtaactacttcgttatcggtcccattgattgcgtctgcttgtagaatctcgttgactttcttagc
aacaatcccgtagatgtcctgaacggtttcactaggaagcaagttaaccgcgcgaccacctacctcatctcggagcatcgcggagaagtg
ctggatgccagagcaagacccgtcaaacgccagcggaagggagcagttatagctcaggccgtggtgctgtaccccagcgtactcaaag
cagaacgcaaggaagcagaacggagaatcttgctcagcccaccaagtgttctccagtggagacttagcgcaagccatgatgttctcgtgg
ttttcctcaatgaacttgatgcgctcagggaacggaaccttatcgacacccgcacagtttgcaccgtggattttcagccagtagtaaccttcc
ttaccgattggtttacctttcgccagcgtaagcagtcctttggtcatatcgttaccttgcgggttgaacattgacacagcgtaaacacgaccg
cgccagtccatgttgtaagggaaccagatggccttatggttagcaaacttattggcttgctcaagcatgaactcaaggctgatacggcgag
acttgcgagccttgtccttgcggtacacagcagcggcagcacgtttccacgcggtgagagcctcaggattcatgtcgatgtcttccggtttc
atcgggagttcttcacgctcaatcgcagggatgtcctcgaccggacaatgcttccacttggtgattacgttggcgaccgcAaggacttttct
tgttgattttccatgcggtgttttgcgcaatgttaatcgctttgtacacctcaggcatgtaaacgtcttcgtagcgcatcagtgctttcttactgt
gagtacgcaccagcgccagaggacgacgaccgttagcccaatagccaccaccagtaatgccagtccacggcttaggaggaactacgca
aggttggaacatcggagagatgccagccagcgcacctgcacgggttgcgatagcctcagcgtattcaggtgcgagttcgatagtctcaga
gtcttgacctactacgccagcattttggcggtgtaagctaaccattccggttgactcaatgagcatctcgatgcagcgtactcctacatgaat
agagtcttccttatgccacgaagaccacgcctcgccaccgagtagacccttagagagcatgtcagcctcgacaacttgcataaatgctttc
ttgtagacgtgccctacgcgcttgttgagttgttcctcaacgtttttcttgaagtgcttagcttcaaggtcacggatacgaccgaagcgagcc
tcgtcctcaatggcccgaccgattgcgcttgctacagcctgaacggttgtattgtcagcactggttaggcaagccagagtggtcttaatggt
gatgtacgctacggcttccggcttgatttcCTAcaggaactggaaggctgtcgggcgcttgccgcgcttagctttcacttcctcaaacca
gtcgttgatgcgtgcaatcatcttagggagtagggtagtgatgagaggcttggcggcagcgttatccgcaacctcaccagctttaagttga
cgctcaaacatcttgcggaagcgtgcttcacccatctcgtaagactcatgctcaagggccaactgttcgcgagctaaacgctcaccgtaat
ggtcagccagagtgttgaacgggatagcagccagttcgatgtcagagaagtcgttcttagcaatgttaatggtattCTAgtgcacggtaa
tcatggtcatggttaattcctcctgttagcccaaaaaacgggtatggagaaacagtagagagttgcgataaaaagcgtcaggtaggatccg
ctaatcttatggataaaaatgctatggcatagcaaagtgtgacgccgtgcaaataatcaatgtggacttttctgccgtgattatagacacttttt
gttacgcgtttttgtcatggctttggtcccgctttgttacagaatgcttttaataagcggggttaccggtttggttagcgagaagagccagtaa
aagacgcagtgacggcaatgtctgatgcaatatggacaattggtttcttctctgaatggcgggagtatgaaaagtatggctgaagcgcaaa
atgatcccctgctgccgggatactcgtttaatgcccatctggtggcgggtttaacgccgattgaggccaacggttatctcgatttttttatcga
ccgaccgctgggaatgaaaggttatattctcaatctcaccattcgcggtcaggggggtggtgaaaaatcagggacgagaatttgtttgccga
ccgggtgatattttgctgttcccgccaggagagattcatcactacggtcgtcatccggaggctcgcgaatggtatcaccagtgggtttacttt
cgtccgcgcgcctactggcatgaatggcttaactggccgtcaatatttgccaatacggggttctttcgcccggatgaagcgcaccagccg
catttcagcgacctgtttgggcaaatcattaacgccgggcaaggggaagggcgctattcggagctgctggcgataaatctgcttgagcaat

tgttactgcggcgcatggaagcgattaacgagtcgctccatccaccgatggataatcgggtacgcgaggcttgtcagtacatcagcgatca
cctggcagacagcaattttgatatcgccagcgtcgcacagcatgtttgcttgtcgccgtcgcgtctgtcacatcttttccgccagcagttag
ggattagcgtcttaagctggcgcgaggaccaacgtatcagccaggcgaagctgcttttgagcaccacccggatgcctatcgccaccgtcg
gtcgcaatgttggttttgacgatcaactctatttctcgcgggtatttaaaaaatgcaccggggccagcccgagcgagttccgtgccggttgt
gaagaaaaagtgaatgatgtagccgtcaagttgtcataattggtaacgaatcagacaattgacggcttgacggagtagcatagggtttgca
gaatccctgcttcgtccatttgacaggcacattatgcatgccgcttcgccttcgcgcgcgaattgatctgctgcctcgcgcgtttcggtgatg
acggtgaaaacctctgacacatgcagctcccggagacggtcacagcctgcagcaaaaaaccccctcaagacccgtttagaggccccaag
gggttatgctagttattgctcagcggtggcagcagccaactcagcttcctttcgggctttgttatttgtagagctcatccatgccatgtgtaat
cccagcagcagttacaaactcaagaaggaccatgtggtcacgcttttcgttgggatctttcgaaagggcagattgtgtcgacaggtaatgg
ttgtctggtaaaaggacagggccatcgccaattggagtattttgttgataatggtctgctagttgaacggatccatcttcaatgttgtggcgaa
ttttgaagttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagttatagttgtactcgagtttgtgtccgagaatgttt
ccatcttctttaaaatcaataccttttaactcgatacgattaacaagggtatcaccttcaaacttgacttcagcacgcgtcttgtagttcccgtc
atctttgaaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaagtcatgccgtttcatatgatccggataacggga
aaagcattgaacaccataagagaaagtagtgacaagtgttggccatggaacaggtagttttccagtagtgcaaataaatttaagggtaagtt
ttccgtatgttgcatcaccttcaccctctccactgacagaaaatttgtgcccattaacatcaccatctaattcaacaagaattgggacaactcc
agtgaaaagttcttctcctttactcatatgtatatctccttcttaaagttaaacaaaattatttctagagggaaaccgttgtggtctccctatagt
gagtcgtattaatttcgcgggatcg

pRepTrip2.3P-EcQtR-2x TAG



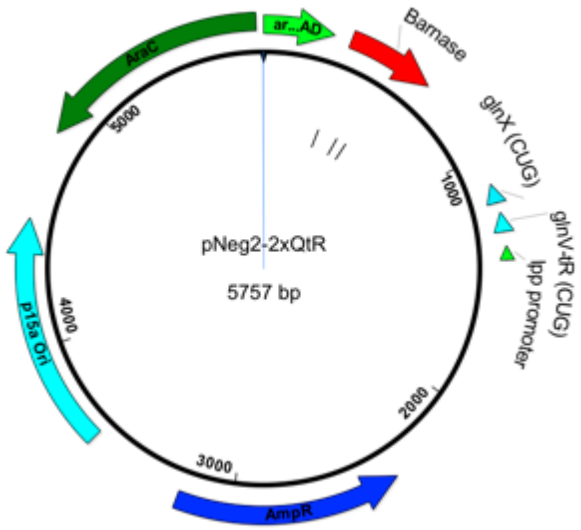
gcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggta
agccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcac
tgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagat
cctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagacccccttaataagatgatcttcttgagatcgttt
tggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagggcggtttttcgaaggttctctgagctaccaactctttgaaccg
aggtaactggcttggaggagcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaactcctctaaat
caattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttggactcaagacgatagttaccggataaggcgcagcggtcgg
actgaacgggggttcgtgcatacagtccagcttggagcgaactgcctacccggaactgagtgtcaggcgtggaatgagacaaacgcgg
ccataacagcggaatgacaccggtaaaccgaaaggcaggaacaggagagcgcacgagggagccgccaggggaaacgcctggtatc
tttatagtcctgtcgggtttcgccaccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcggagcctatggaaaaacggctttg
ccgcggccctctcacttccctgttaagtatcttcctggcatcttccaggaaatctccgccccgttcgtaagccatttccgctcgccgcagtc
gaacgaccgagcgtagcgagtcagtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttttct

cctgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgctgatgtccggcggtgct
tttgccgttacgcaccaccccgtcagtagctgaacaggagggacagctgatagaaacagaagccactggagcacctcaaaaacaccatc
atacactaaatcagtaagttggcagcatcacccgacgcactttgcgccgaataaatacctgtgacggaagatcacttcgcagaataaataa
atcctggtgtccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggcacgtaagaggttccaactttca
ccataatgaaataagatcactaccgggcgtatttttttgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactg
gatataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgttca
gctggatattacggccttttaaagaccgtaaagaaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgct
catccggaattccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttgttacaccgttttccatgagcaaactga
aacgttttcatcgctctggagtgaataccactaggatttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacct
ggcctatttccctaaagggtttattgagaatatgtttttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtggccaata
tggacaacttcttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcat
gccgtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcagggcggggcgtaattttttttaagg
cagttattggtgcccttaaacgcctggttgctacgcctgaataagtgataataagcggatgaatggcagaaattcgaaagcaaattcgaccc
ggtcgtcggttcagggcagggtcgttaaatagccgcttatgtctattgctggtttaccggtttattgactaccggaagcagtgtgaccgtgtg
cttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattgacgatatgatcatttattctgcctcccagagcatgat
aaaaacggttagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggAc
tagcggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataa
ccctgataaatgcttcaataatattgaaaaaggaagagtatggcgtagagtattcaacatttccgtgtcgcccttattcccttttttgcggcatt
ttgccttcctgttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactgg
atctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatt
atcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaag
catcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgat
cggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagc
cataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagct
tcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctg
ataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacac
gacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagacca
agtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcc
cttaacgtgagttttcaCTAGgGGTATGGATCTGTTTTATGCCGGATTCAGCAATAATCTG
CCACCACTGCGTTTTATGCGTAATCTCGGGTTAATGGCGGCGGAGCGTGCTG
GCGTGTTGAAACGTCAGGCGCTGAAATATGCGTTAGGGTTGTAGCCTTACAA
CATTGCCGGATGCGTGCCAACCGTAGGTCGGATAAGACGCGGCAGCGTCGCA
TCCGACATTGAAGGATAAGACGTGTCAACATCGCATTCGACATTGAATGAACG
CAGAAAAGCAAAAAGCTCGCCGAAGCGAGCTTTTTTAATGtggCTGGGGTACGA
GGATTCGAACCTCGGAATGCCGGAATCAGAATCCGGTGCCTTACCGCTTGGC
GATACCccaACTGGGTGCACTTACAAGGTAAGCGTCTTGAATAAATTGGCTGGG
GTACGAGGATTCGAACCTCGGAATGCCGGAATCAGAATCCGGTGCCTTACCG
CTTGGCGATACCCCAACAAATTGGTTTTGAATTTGCCGAACATATTCGATACAT
TCAGAATTaatgcggggcgcatcttactgcgcAGATAcgccctcgtcaatcccttaatagcaaaatgccttttgatcggcga
gaaagtcagcaggccgcttagttagccgtgcactgatcatatcgtcaattattacctccacggggagagcctgagcaaactggcctcagg
catttgagaagcacacACTaggaaGATCTaaggagcccgagatgcgccgcgtgcggctgctggagatggcggacgcgatggat
atgttctgccaagggttggtttgcgcattcacagttctccgcaagaattgattggctccaattcttggagtggtgaatccgttagcgaggtgc
cgccggcttccattcaggtcgaggtggcccggctccatgcaccgcgacgcaacgcggggaggcagacaaggtatagggcggcgccta
caatccatgccaacccgttccatgtgctcgccgaggcggcataaatcgccgtgacgatcagcggtccaatgatcgaagttaggctggtaa
gagccgcgagcgatccttgaagctgtccctgatggtcgtcatctacctgcctggacagcatggcctgcaacgcgggcatcccgatgccg
ccggaagcgagaagaatcataatggggaaggccatccagcctcgcgtcgcgaacgccagcaagacgtagcccagcgcgtcggccgcc
atgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaa
taccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcct

acgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggc
tctcaagggcatcggtcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgc
cgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgct
catgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgc
cggccacgatgcgtccggcgtagaggatccacaggacgggtgtggtcgccatgatcgcgtagtcgatagtggctccaagtagcgaagc
gagcaggactgggcggcggccaaagcggtcggacagtgctccgagaacgggtgcgcatagaaattgcatcaacgcatatagcgctagc
agcacgccatagtgactggcgatgctgtcggaatggacgatatcccgcaagaggcccggcagtaccggcataaccaagcctatgcctac
agcatccagggtgacggtgccgaggatgacgatgagcgcattgttagatttcatacacggtgcctgactgcgttagcaatttaactgtgata
aactaccgcattaaagcttatcgatgataagctgtcaaacatgagaattacaacttatatcgtatggggctgacttcaggtgctacattgctc
aaagatgcaggggtaaaagctaaccgcatctttaccgacaaggcatccggcagttcaacagatcgggaagggctggatttgctgaggatg
aaggtggaggaaggtgatgtcattctggtgaagaagctcgaccgtcttggccgcgacaccgccgacatgatccaactgataaaagagttt
gatgctcagggtgtagcggttcggtttattgacgacgggatcagtaccgacggtgatatggggcaaatggtggtcaccatcctgtcggctg
tggcacaggctgaacgccggaggatcctccgggccgttgcttcgcaacgttcaaatccgctcccggcggatttgtcctactcaggagagc
gttcaccgacaaacaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggcagttccctactctcgcat
ggggagaccccacactaccatcggcgctacggcgtttcacttctgagttcggcatggggtcaggtgggaccaccgcgctactgccgcca
ggcaaattctgttttatcagaccgcttctgcgttctgatttaatctgtatcaggctgaaaatcttctctcatccgacgtcttaggcgaaggcga
agtccgactctaagatgtcacggaggttcaagttacctttagccggaagtgctggcattttgtccaattgagactcgtgcaactggtcagcg
aactggtcgtagaaatcagccagtacatcacaagactcatatgtgtcaaccatagtttcgcgcactgctttgaacaggttcgcagcgtcagc
cggaatggtaccgaaggagtcgtgaatcagtgcaaaagattcgattccgtacttctcgtgtgcccacactacagtcttacgaaggtggcta
ccgtcttggctgtgtacaaagttaggagcgataccagactcctgtttgtgtgcatcaatctcgctatctttgttggtgttaatggtaggctgta
agcggaactgaccgaggaacatcaggttcaagcgcgtctgaataggcttcttgtattcctgccacacagggaaaccatcaggagttaccc
aatgcacagcgcaacgcttgcgaagaatctctccagtcttcttatctttgacctcagcagccagcagcttagcagcagacttaagccagttc
attgcttcaaccgcagctaccaccgtcacgctcacagattcccaaatcagcttagccatgtatccagcagcctgattcggctgagtgaacat
cagacccttgccggaatcaatagctggctgaatggtatcttccagcacttgttgacggaagccgaactctttggacccgtaagccagcgtc
atgactgaacgcttagtcacactgcgagtaacaccgtaagccagccattgaccagccagtgccttagtgcccagcttgactttctcagaga
tttcaccagtgttctcatcggtcacggtaactacttcgttatcggtcccattgattgcgtctgcttgtagaatctcgttgactttcttagcaacaa
tcccgtagatgtcctgaacggtttcactaggaagcaagttaaccgcgcgaccacctacctcatctcggagcatcgcggagaagtgctgga
tgccagagcaagacccgtcaaacgccagcggaagggagcagttatagctcaggccgtggtgctgtaccccagcgtactcaaagcagaa
cgcaaggaagcagaacggagaatcttgctcagcccaccaagtgttctccagtggagacttagcgcaagccatgatgttctcgtggttttcc
tcaatgaacttgatgcgctcagggaacggaaccttatcgacacccgcacagtttgcaccgtggattttcagccagtagtaaccttccttacc
gattggtttacctttcgccagcgtaagcagtcctttggtcatatcgttaccttgcgggttgaacattgacacagcgtaaacacgaccgcgcc
agtccatgttgtaagggaaccagatggccttatggttagcaaacttattggcttgctcaagcatgaactcaaggctgatacggcgagacttg
cgagccttgtccttgcggtacacagcagcggcagcacgtttccacgcggtgagagcctcaggattcatgtcgatgtcttccggtttcatcg
ggagttcttcacgctcaatcgcagggatgtcctcgaccggacaatgcttccacttggtgattacgttggcgaccgcAaggactttcttgttg
attttccatgcggtgttttgcgcaatgttaatcgctttgtacacctcaggcatgtaaacgtcttcgtagcgcatcagtgctttcttactgtgagt
acgcaccagcgccagaggacgacgaccgttagcccaatagccaccaccagtaatgccagtccacggcttaggaggaactacgcaaggt
tggaacatcggagagatgccagccagcgcacctgcacgggttgcgatagcctcagcgtattcaggtgcgagttcgatagtctcagagtct
tgacctactacgccagcattttggcggtgtaagctaaccattccggttgactcaatgagcatctcgatgcagcgtactcctacatgaataga
gtcttccttatgccacgaagaccacgcctcgccaccgagtagacccttagagagcatgtcagcctcgacaacttgcataaatgctttcttgt
agacgtgccctacgcgcttgttgagttgttcctcaacgtttttcttgaagtgcttagcttcaaggtcacggatacgaccgaagcgagcctcgt
cctcaatggcccgaccgattgcgcttgctacagcctgaacggttgtattgtcagcactggttaggcaagccagagtggtcttaatggtgatg
tacgctacggcttccggcttgatttcCTAcaggaactggaaggctgtcgggcgcttgccgcgcttagctttcacttcctcaaaccagtcg
ttgatgcgtgcaatcatcttagggagtagggtagtgatgagaggcttggcggcagcgttatccgcaacctcaccagctttaagttgacgctc
aaacatcttgcggaagcgtgcttcacccatctcgtaagactcatgctcaagggccaactgttcgcgagctaaacgctcaccgtaatggtca
gccagagtgttgaacgggatagcagccagttcgatgtcagagaagtcgttcttagcaatgttaatggtattCTAgtgcacggtaatcatg
gtcatggttaattcctcctgttagcccaaaaaacgggtatggagaaacagtagagagttgcgataaaaagcgtcaggtaggatccgctaat
cttatggataaaaatgctatggcatagcaaagtgtgacgccgtgcaaataatcaatgtggacttttctgccgtgattatagacacttttgttac
gcgtttttgtcatggctttggtcccgctttgttacagaatgcttttaataagcggggttaccggtttggttagcgagaagagccagtaaaaga

cgcagtgacggcaatgtctgatgcaatatggacaattggtttcttctctgaatggcgggagtatgaaaagtatggctgaagcgcaaaatgat
cccctgctgccgggatactcgtttaatgcccatctggtggcgggtttaacgccgattgaggccaacggttatctcgatttttttatcgaccga
ccgctgggaatgaaaggttatattctcaatctcaccattcgcggtcaggggggtggtgaaaaatcagggacgagaatttgtttgccgaccgg
gtgatattttgctgttcccgccaggagagattcatcactacggtcgtcatccggaggctcgcgaatggtatcaccagtgggtttactttcgtc
cgcgcgcctactggcatgaatggcttaactggccgtcaatatttgccaatacggggttctttcgcccggatgaagcgcaccagccgcattt
cagcgacctgtttgggcaaatcattaacgccgggcaaggggaagggcgctattcggagctgctggcgataaatctgcttgagcaattgtta
ctgcggcgcatggaagcgattaacgagtcgctccatccaccgatggataatcgggtacgcgaggcttgtcagtacatcagcgatcacctg
gcagacagcaattttgatatcgccagcgtcgcacagcatgtttgcttgtcgccgtcgcgtctgtcacatcttttccgccagcagttagggatt
agcgtcttaagctggcgcgaggaccaacgtatcagccaggcgaagctgcttttgagcaccacccggatgcctatcgccaccgtcggtcg
caatgttggttttgacgatcaactctatttctcgcgggtatttaaaaaatgcaccggggccagcccgagcgagttccgtgccggttgtgaag
aaaaagtgaatgatgtagccgtcaagttgtcataattggtaacgaatcagacaattgacggcttgacggagtagcatagggtttgcagaatc
cctgcttcgtccatttgacaggcacattatgcatgccgcttcgccttcgcgcgcgaattgatctgctgcctcgcgcgtttcggtgatgacgg
tgaaaacctctgacacatgcagctcccggagacggtcacagcctgcagcaaaaaaccccctcaagacccgtttagaggccccaaggggtt
atgctagttattgctcagcggtggcagcagccaactcagcttcctttcgggctttgttatttgtagagctcatccatgccatgtgtaatcccag
cagcagttacaaactcaagaaggaccatgtggtcacgcttttcgttgggatctttcgaaagggcagattgtgtcgacaggtaatggttgtct
ggtaaaaggacagggccatcgccaattggagtattttgttgataatggtctgctagttgaacggatccatcttcaatgttgtggcgaattttga
agttagctttgattccattcttttgtttgtctgccgtgatgtatacattgtgtgagttatagttgtactcgagtttgtgtccgagaatgtttccatct
tctttaaaatcaataccttttaactcgatacgattaacaagggtatcaccttcaaacttgacttcagcacgcgtcttgtagttcccgtcatctttg
aaagatatagtgcgttcctgtacataaccttcgggcatggcactcttgaaaaagtcatgccgtttcatatgatccggataacgggaaaagca
ttgaacaccataagagaaagtagtgacaagtgttggccatggaacaggtagttttccagtagtgcaaataaatttaagggtaagttttccgta
tgttgcatcaccttcaccctctccactgacagaaaatttgtgcccattaacatcaccatctaattcaacaagaattgggacaactccagtgaa
aagttcttctcctttactcatatgtatatctccttcttaaagttaaacaaaattatttctagagggaaaccgttgtggtctccctatagtgagtcgt
attaatttcgcgggatcg

pNeg2-2xQtR

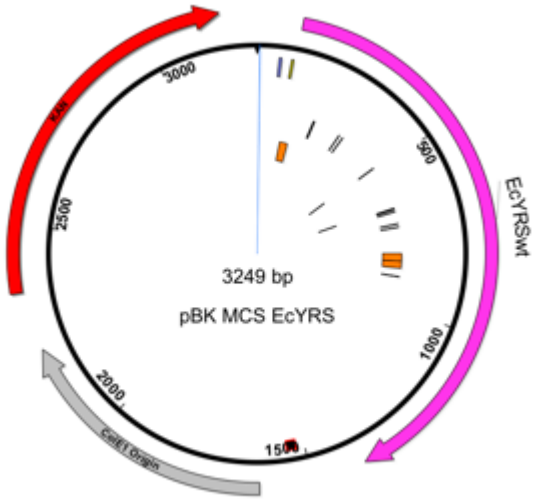


aaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactggctcttctcgctaaccaaaccggtaaccccgcttattaaaag
cattctgtaacaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagtgtctataatcacggcagaaaagtccacattgattatttg
cacggcgtcacactttgctatgccatagcatttttatccataagattagcggatcctacctgacgctttttatcgcaactctctactgtttctcc
atacccgtttttttgggctagaaataattttgtttaactttaagaaggagatatacatatggcataggttatcaacacgtttgacggggttgcgg
attatcttcagacatatcataagctacctgataattacattacaaaatcagaagcacaagccctcggctgggtggcatcaaaagggaaccttt
gcataggtcgctccggggaaaagcatcggcggagacatcttctcaaacagggaaggcaaactcccgggcaaaagcggacgaacatgg

cgtgaagcggatattaactatacatcaggcttcagaaattcagaccggattctttactcaagcgactggctgatttacaaaacaacggacca
ttatcagacctttacaaaaatcagataagcatgcaccattccttgcggcggcggtgctcaacggcctcaacctactactgggctgcttccta
atgcaggagtcgcataagggagagcgtctggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttGaattcg
GGTATGGATCTGTTTTATGCCGGATTCAGCAATAATCTGCCACCACTGCGTTT
TATGCGTAATCTCGGGTTAATGGCGGCGGAGCGTGCTGGCGTGTTGAAACGT
CAGGCGCTGAAATATGCGTTAGGGTTGTAGCCTTACAACATTGCCGGATGCG
TGCCAACCGTAGGTCGGATAAGACGCGGCAGCGTCGCATCCGACATTGAAGG
ATAAGACGTGTCAACATCGCATTCGACATTGAATGAACGCAGAAAAGCAAAAA
GCTCGCCGAAGCGAGCTTTTTTAATGtggCTGGGGGTACGAGGATTCGAACCTCG
GAATGCCGGAATCAGAATCCGGTGCCTTACCGCTTGGCGATACCccaACTGGGT
GCACTTACAAGGTAAGCGTCTTGAATAAATTGGCTGGGGGTACGAGGATTCGA
ACCTCGGAATGCCGGAATCAGAATCCGGTGCCTTACCGCTTGGCGATACCCCA
ACAAATTGGTTTTGAATTTGCCGAACATATTCGATACATTCAGAATtgaattcagcgtta
caagtattacacaaagtttttatgttgagaatatttttttgatggggcgccacttatttttgatcgttcgctcaaagaagcggcgccagggttg
ttttttcttttcaccagtgagacgggcaacagaacgccatgagcggcctcatttcttattctgagttacaacagtccgcaccgctgccggtag
ctccttccggtgggcgcggggcatgactatcgtcgccgcacttatgactgtcttctttatcatgcaactcgtaggacaggtgccggcagcg
cccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgccctgcaccattatgttccggatctgcatcgcagg
atgctgctggctaccctgtggaacacctacatctgtattaacgaagcgctaaccgtttttatcatgctctgggaggcagaataaatgatcata
tcgtcaattattacctccacggggagagcctgagcaaactggcctcaggcatttgagaagcacacggtcacactgcttccggtagtcaata
aaccggtaaaccagcaatagacataagcggctatttaacgaccctgccctgaaccgacgaccgggtcgaatttgctttcgaatttctgcca
ttcatccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgcca
ctcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacaaacggcatgatgaacctgaatcgccagcggcatc
agcaccttgtcgccttgcgtataatatttgcccatggggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttc
acctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtga
ggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatc
tggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgag
cgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttg
cgcaacgttgttgccattgctgcaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcg
agttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcact
catggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctga
gaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaacacgggataataccgcgccacatagcagaactttaaaagtgctcatca
ttggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatctt
cagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaat
gttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaat
aaacaaataggggttccgcgcacatttccccgtacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaa
cagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttac
tgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagctgtccctcctgttcagctactgacggggtggtgcgtaa
cggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatg
tggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcg
gtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgag
agggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaaccc
gacaggactataaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgct
gttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaacccccgtt
cagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggt
aattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagccagt
tacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggttttttcgttttcagagcaagagattacgcgc
agaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagt
cagccccatacgatataagttgtaattctcatgtttgacagcattatcatcgataagctttaatgcggtagtttatcacagttaaattgctaacg

cagtcaggcaccgtgtatgaaatctaacaatgcgctcatcgtcatcctcggcaccgtcaccctggatgctgtaggcataggcttggttatgc
cggtactgccgggcctcttgcgggatggccacgatgcgtccggcgtagaggatctgctcatgtttgacagcttatcatcgatgcataatgt
gcctgtcaaatggacgaagcagggattctgcaaaccctatgctactccgtcaagccgtcaattgtctgattcgttaccaattatgacaacttg
acggctacatcattcacttttttcttcacaaccggcacggaactcgctcgggctggccccggtgcatttttttaaataccgcgagaaatagag
ttgatcgtcaaaaccaacattgcgaccgacggtggcgataggcatccgggtggtgctcaaaagcagcttcgcctggctgatacgttggtc
ctcgcgccagcttaagacgctaatccctaactgctggcggaaaagatgtgacagacgcgacggcgacaagcaaacatgctgtgcgacgc
tggcgatatcaaaattgctgtctgccaggtgatcgctgatgtactgacaagcctcgcgtacccgattatccatcggtggatggagcgactc
gttaatcgcttccatgcgccgcagtaacaattgctcaagcagatttatcgccagcagctccgaatagcgcccttcccccttgcccggcgttaa
tgatttgcccaaacaggtcgctgaaatgcggctggtgcgcttcatccgggcgaaagaaccccgtattggcaaatattgacggccagttaag
ccattcatgccagtaggcgcgcggacgaaagtaaacccactggtgataccattcgcgagcctccggatgacgaccgtagtgatgaatctc
tcctggcgggaacagcaaaatatcacccggtcggcaaacaaattctcgtccctgatttttcaccacccccctgaccgcgaatggtgagattg
agaatataacctttcattcccagcggtcggtcgataaaaaaatcgagataaccgttggcctcaatcggcgttaaacccgccaccagatggg
cattaaacgagtatcccggcagcaggggatcattttgcgcttcagccatactttcatactcccgccattcagagaag

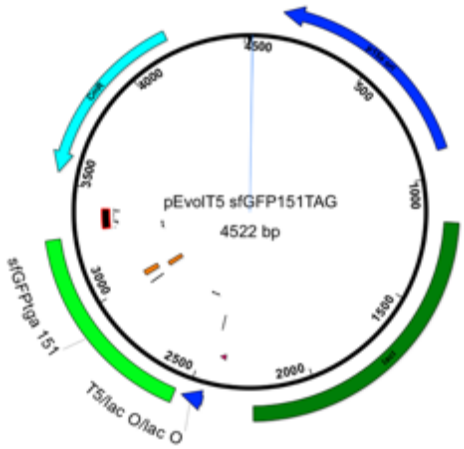
pBK EcYRS (mutated residues underlined)



cttttgctgagttgaaggatccGCGGCCGCtcgggttgtcagcctgtcccgcttataagatcatacgccgttatacgttgtttacgct
ttgaggaatcccatatgGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTG
GTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGC
CCGATCGCGCTC**TAT**TGCGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGG
GCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAAGC
CGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAA
AGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGAC
AAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTC
TGCTATCGCGGCGAACAACTATGACTGGTTCGGCAATATGAATGTGCTGACCT
TCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAA
GCGGTTAAGCAGCGTCTCAACCGTGAA**GAT**CAGGGGATTTCGTTCACTGAGT
TTTCCTACAACCTGTTGCAGGGTTAT**GACTTC**GCaTGT**CTG**AACAAACAGTAC
GGTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTG
GTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTT
CCGCTGATCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCG
CAGTCTGGTTaGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAGTTCTGGA
TCAACACTGCG**GAT**GCCGACGTTTACCGCTTCCTGAAGTTCTTCACCTTTATGA

GCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAACAGCGGTAAAGC
ACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGACTCGTCTGGTTCACGGT
GAAGAAGGTTTACAGGCGGCAAAACGTATTACCGAATGCCTGTTCAGCGGTT
CTTTGAGTGCGCTGAGTGAAGCGGACTTCGAACAGCTGGCGCAGGACGGCGT
ACCGATGGTTGAGATGGAAAAGGGCGCAGACCTGATGCAGGCACTGGTCGA
TTCTGAACTGCAACCTTCCCGTGGTCAGGCACGTAAAACTATCGCCTCCAATGC
CATCACCATTAACGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAAG
AAGATCGTCTGTTTGGTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTAC
TGTCTGATTTGCTGGAAAtaaccatggctgcagtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctaca
tgatctctgcaatatattgagtttgcgtgcttttgtaggccggataaggcgttcacgccgcatccggcaagaaacagcaaacaatccaaaac
gccgcgttcagcggcgtttttttctgctttcttcgcgaattaattccgcttcgcacatgtgagcaaaaggccagcaaaaggccaggaaccgt
aaaaaggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaac
ccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgt
ccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggct
gtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcc
actggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggcta
cactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaacca
ccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggt
ctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgt
tatgagccatattcaacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcg
ataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgt
tgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgat
gatgcatggttactcaccactgcgatccccgggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgc
gctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatca
cgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataagct
tttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgt
tggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggctttt
tcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaa
cactggcagagcattacgctgacttgacgggacggcggctttgttgaataaatcgaa

pEvolT5 sfGFP151TAG

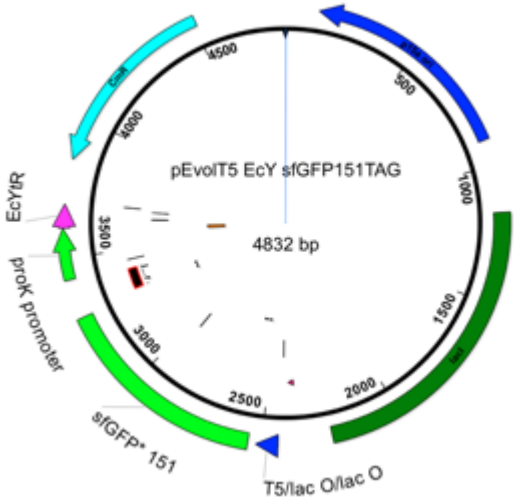


gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcag
tgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactg
actcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatactta

acagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccccctgacaagcatcacgaaatctgacgctcaaatcagt
ggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac
cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg
cacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactg
gcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg
ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagag
caagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaa
atgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatg
tagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg
aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggc
gaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacg
cgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattt
atgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccaga
tgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa
cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatc
gccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc
cacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc
gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttcc
gggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggc
tcactataggggaattgtgagcggataacaattcccctctagagtttgacagcattGtcatcgatctcgagaaatcataaaaaatttatttgc
tttgtgagcggataacaattataatagattcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatATGA
GCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGAT
GGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATG
CTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTG
TTCCGTGGCCAACACTTGTCACTACTCTGACCTatGGTGTTCAATGCTTTTCCCG
TTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAG
GTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACG
CGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAA
GGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACA
ACTTTAACTCACACAATGTATagATCACGGCAGACAAACAAAAGAATGGAATCA
AAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCA
GACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGAC
AACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGT
GACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCAT
GGATGAGCTCTACAAAGGATCCCACCACCACCACCACCACTAAaagcttaattagctgagct
tggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgag
aatccaagctagcttggcgcCTCGAGCAGCTCAGGGTCGAATTTGCTTTCGAatttctgccattcatcc
gcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcg
cagtactgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcacct
tgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactc
acccagggattggctgagacgaaaaacatattctcaataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcg
aatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaa
gggtgaacactatcccatatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaat
aaaggccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagca

actgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttcctta
gctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacg
tctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggt
atttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttcta
tcagctgtccctcctgttcagctactgacg

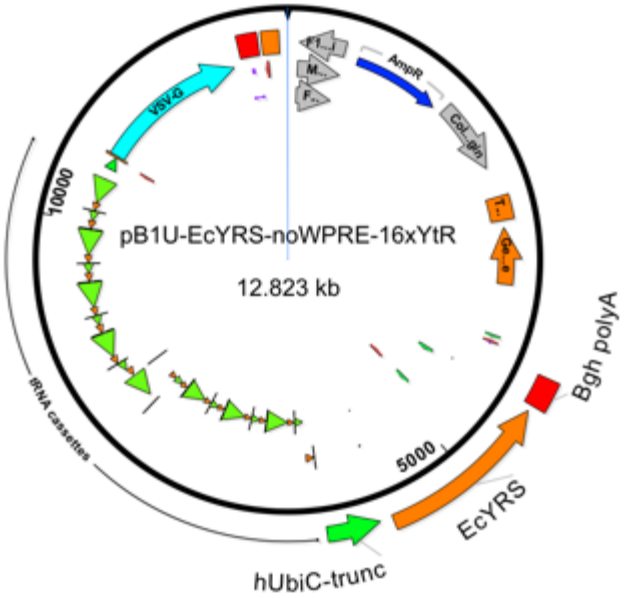
pEvolT5 EcY sfGFP151TAG



gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcag
tgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactg
actcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatactta
acagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagt
ggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac
cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg
cacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactg
gcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg
ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagag
caagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaa
atgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatg
tagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg
aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttcttttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggc
gaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacg
cgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattt
atgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccaga
tgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa
cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatc
gccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc
cacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc
gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttcc
gggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggc

tcactataggggaattgtgagcggataacaattcccctctagagtttgacagcattatcatcgatctcgagaaatcataaaaaatttatttgct
ttgtgagcggataacaattatataatagattcaattgtgagcggataacaatttcacacagaattcattaaagaggagaaattacatATGA
GCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGAT
GGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATG
CTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTG
TTCCGTGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCC
GTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAA
GGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGAC
GCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAA
AGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTAC
AACTTTAACTCACACAATGTATAGATCACGGCAGACAAACAAAAGAATGGAAT
CAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAG
CAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAG
ACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGC
GTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGC
ATGGATGAGCTCTACAAAGGATCCCACCACCACCACCACCACTAAaagcttaattagctg
agcttggactcctgttgatagatccagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttattgg
tgagaatccaagctagcttggcgCTGCAgtgtgcttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataatt
gacgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagggcg
TATCTgcgcagtaagatgcgccccgcatttGGTGGGGGTTCCCGAGCGGCCAAAGGGAGCAGA
CTCTAAATCTGCCGTCATCGACTTCGAAGGTTCGAATCCTTCCCCCACCACCAA
ATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCAGCTCAG
GGTCGAATTTGCTTTCGAatttctgccattcatccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggc
accaataactgccttaaaaaaattacgccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagcca
tcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgcgtataatatttgcccatggtgaaaacgggggcg
aagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaacatattctcaataaaccc
tttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactcca
gagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcattgcca
tacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaa
aggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgccattgggata
tatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgat
cttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaaca
gggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactg
atttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagctgtccctcctgttcagctactgacg

pAcBac-EcYRS-NoWPRE-16xtRY-TAG



ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaa
tgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgc
ccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattt
agtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttt
gacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataaggga
ttttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtg
gcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataa
atgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttg
ctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggta
agatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgc
cgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggc
atgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaag
gagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacg
agcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaatt
aatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagcc
ggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcag
gcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatat
actttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttt
tcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaa
aaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac
caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttac
cagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaa
cgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgccac
gcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaa
acgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatgga
aaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggat
aaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaag
agcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcagaccagccgcgtaacctggcaaaatcggttacggttga

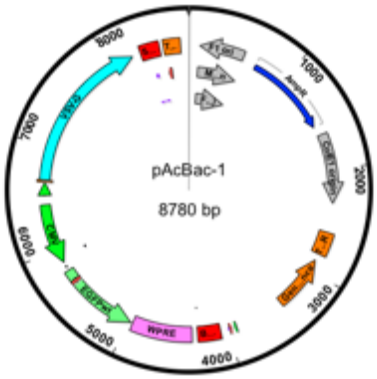
gtaataaatggatgccctgcgtaagcgggtgtgggcggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtctt
aaactagacagaatagttgtaaactgaaatcagtccagttatgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcatt
ttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaagggcatggtaaagactatattcgcggcgttgtgacaatttaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttcttcccg
tatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccaagcgcgttggcctcat
gcttgaggagattgatgagcgcggtggcaatgccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcg
gctgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtcttacta
cggagcaagttcccgaggtaatcggagtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagc
agcccgcatggatttgacttggtcagggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgc
cctgctgcgtaacatcgttgctgctgcgtaacatcgttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttgga
tgcccgaggcatagactgtacaaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggt
tctggaccagttgcgtgagcgcatacgctacttgcattacagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgttt
ccacggtgtgcgtcacccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtct
ccacgcatcgtcaggcattggcggccttgctgttcttctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacc
tcggccgtcgcggcgcttgccggtggtgctgaccccggatgaagtggttcgcatcctcggttttctggaaggcgagcatcgtttgttcgcc
caggactctagctatagttctagtggttggctacgtacccgtagtggctatggcagggcttgcgcttaatgcgccgctacagggcgcgtgg
ggataccccctagagccccagctggttctttccgcctcagaagccatagagcccaccgcatccccagcatgcctgctattgtcttcccaat
cctccccccttgctgtcctgccccaccccaccccccagaatagaatgacacctactcagacaatgcgatgcaatttcctcattttattaggaa
aggacagtgggagtggcaccttccagggtcaaggaaggcacgggggagggcaaacaacagatggctggcaactagaaggcacagtc
gaggctgatcagcgggtttaaacgggccctctagactcgagttaTTTCCAGCAAATCAGACAGTAATTCTTT
TTACCGCGACGCAGTAAGGTAAAACGACCAAACAGACGATCTTCTTCTTTAAA
GAAGTATTCAGGATCGGACTGTTTTTCACCGTTAATGGTGATGGCATTGGAG
GCGATAGTTTTACGTGCCTGACCACGGGAAGGTTGCAGTTCAGAATCGACCA
GTGCCTGCATCAGGTCTGCGCCCTTTTCCATCTCAACCATCGGTACGCCGTCCT
GCGCCAGCTGTTCGAAGTCCGCTTCACTCAGCGCACTCAAAGAACCGCTGAAC
AGGCATTCGGTAATACGTTTTGCCGCCTGTAAACCTTCTTCACCGTGAACCAG
ACGAGTCACCTGCTCCGCCAGTACATACTGGGCGCGCGGTGCTTTACCGCTGT
TTTTATCTTCTTCTTCCAGGGCGTTGATCTCTTCAATGCTCATAAAGGTGAAGA
ACTTCAGGAAGCGGTAAACGTCGGC**ATC**CGCAGTGTTGATCCAGAACTGGTA
GAATTTGTACGGGCTGGTTTTCTTCGGATCtAACCAGACTGCGCCGCCTTCAGT
TTTACCAAATTTGGTGCCATCTGCTTTAGTGATCAGCGGAACGGTCAGGCCAA
ACACCTGATTCTGATGCAGACGACGGGTCAGGTCGATACCAGAAGTGATGTT
ACCCCACTGGTCAGAACCACCAATTTGCAGCACCACACCGTACTGTTTGTT**CAG**
ACAtGC**GAAGTC**ATAACCCTGCAACAGGTTGTAGGAAAACTCAGTGAACGAAA
TCCCCTG**ATC**TTCACGGTTGAGACGCTGCTTAACCGCTTCTTTGTTGATCATCT
GGTTAACGGAGAAGTGTTTGCCAATATCGCGCAGGAAGGTCAGCACATTCAT
ATTGCCGAACCAGTCATAGTTGTTCGCCGCGATAGCAGAGTTTTCTCCACAGT
CGAAATCGAGGAACGGGGCAACCTGCTTACGGATTTTGTCCACCCACTCCTGA
ACAGTTTCTTCGGTGTTCAGCTTACGCTCGGCAGCTTTGAAGCTCGGGTCGCC
AATCAGACCCGTCGCGCCGCCTACCAGCGCAACCGGCTTGTGGCCCGCCTGCT
GGAAGCGTTTCAGGCATAACAATGGAACAAGATGCCCCAAATGCAAGCTGTC
AGCGGTAGGATCGAAGCCGCA**ATA**GAGCGCGATCGGGGCCTTGCGCCAGTCGC
TCTGCTAACGCTTCCTCGTCCGTCACCTGGGCTACCAGCCCCCGCTCTTGCAAT
TGTTTAATCAAGTTACTGCTTGCcatGGTGGCgctagccagcttgggtctccctatagtgagtcgtattaat
ttcgataagccagtaagcagtgggttctctagttagccagagagctCtagaccaagtgacgatcacagcgatccacaaacaagaaccgc
gacccaaatcccggctgcgacggaactagctgtgccacacccggcgcgtccttatataatcatcggcgttcaccgccccacggagatcc
ctccgcagaatcgccgagaagggactacttttcctcgcctgttccgctctctggaaagaaaaccagtgccctagagtcacccaagtcccgt
cctaaaatgtccttctgctgatactggggttctaaggccgagtcttatgagcagcgggccgctgtcctgagcgtccgggcggaaggatca

ggacgctcgctgcgcccttcgtctgacgtggcagcgctcgccgtgaggaggggggcgcccgcgggaggcgccaaaacccggcgcgg
aggccttcgaacggCCACTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCGAACC
CAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCG
GTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATTTC
ACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAATG
GTGGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTG
CTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATAT
ATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCC
ATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTAC
GTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTAT
CTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATA
GGCCCTCTTCCTGCCCGAcCTAGCAAAAAATGGAGGGGGACGGATTCGAACC
GCCGAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCT
ACCCCTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGAT
AGGGATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATAT
TAAAAAATGGTGGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATT
TAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCC
ACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATA
TGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTAT
TACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAA
TTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCA
TGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATGGAGGGGGACGGG
ATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGC
CACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCT
ATCACTGATAGGGATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACA
TGCAAATATTAAAAAATGGTGGGGGGAAGGATTCGAACCTTCGAAGTCTGTGA
CGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTT
CGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACG
GTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCA
AGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACA
GTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATAT
GAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAcCTAGCAAAAAATGGA
GGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAGAGTCCGC
CGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAGGGAACTT
ATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGGTGATTTCCCAGAACAC
ATAGCGACATGCAAATATTAAAAAATGGTGGGGGGAAGGATTCGAACCTTCGA
AGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCA
CCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTT
CAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCA
AACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTG
TGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATG
CAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAA
AAATGGAGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAG
AGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAG
GGAACTTATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGGTGATTTCCC
AGAACACATAGCGACATGCAAATATTAAAAAATGGTGGGGGGAAGGATTCGAA
CCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGG
AACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAA

ATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAA
AACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAA
TATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTAT
CGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAc
CTAGCAAAAAATGGAGGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGC
GGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATC
ACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGG
TGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAATGGTGGGGGAA
GGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGG
CCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCA
AGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAAC
ATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATT
TTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACA
GCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTC
CTGCCCGACCTAGCAAAAAATGGAGGGGGGACGGATTCGAACCGCCGAACCCA
AAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGT
GTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATTTCAC
GTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAATGGT
GGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCT
CCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATAT
AAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCAT
TTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGT
CACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCT
CTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGG
CCCTCTTCCTGCCCGAcCTAGCAAAAAATGGAGGGGGGACGGATTCGAACCGCC
GAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCC
CTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGG
GATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAA
AAATGGTGGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGA
GTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAA
GATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGAT
AGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTT
TCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTA
ATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGG
AAATAGGCCCTCTTCCTGCCCGACctagtcaataatcaatgtcaacgcgtatatctggcccgtacatcgcgaagc
agcgcaaaacGGATCCtgcaggtatttGCGGCCGCggtccgtatactccggaatattaatagatcatggagataattaaaat
gataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtccc
accatcgggcgcgAACTCCTAAAAAACCGCCACCatgaagtgccttttgtacttagccttttttattcattggggtgaatt
gcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaa
ttggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgctt
ccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgca
aggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccg
aagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgc
agcaattacatatgccccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccatgg
acatcaccttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaactgg
aggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctctttt
gctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttga
gaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctc

ctaaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgct
ccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaa
attggacccaatggagttctgaggaccagttcaggatataagtttcctttatacatgattggacatggtatgttggactccgatcttcatctta
gctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttttggtgatactgggct
atccaaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctctttttctttatcatagggttaatcattggact
attcttggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgactt
ggaaagtgataaggccaggccggccaagcttgtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctt
taaaaaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaa
taaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtct
ggatctgatcactgcttgagcctaggagatccgaaccagataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttac
gacgctacacccagttcccatctattttgtcactcttccctaaataatccttaaaaactccatttccacccctcccagttcccaactattttgtc
cgcccacagcggggcatttttcttcctgttatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttcggctacttt

pAcBac-EGFP39* TAG



ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaa
tgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgc
ccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattt
agtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttt
gacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataaggga
ttttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtg
gcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataa
atgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgttttg
ctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggta
agatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgc
cgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggc
atgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaag
gagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacg
agcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaatt
aatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagcc
ggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcag
gcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatat
actttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttt
tcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaa
aaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac
caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttac
cagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaa

cgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgccac
gcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaa
acgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatgga
aaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggat
aaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaag
agcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcagaccagccgcgtaacctggcaaaatcggttacggttga
gtaataaatggatgccctgcgtaagcgggtgtgggcggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtctt
aaactagacagaatagttgtaaactgaaatcagtccagttatgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcatt
ttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaagggcatggtaaagactatattcgcggcgttgtgacaatttaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttcttcccg
tatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccaagcgcgttggcctcat
gcttgaggagattgatgagcgcggtggcaatgccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcg
gctgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtcttacta
cggagcaagttcccgaggtaatcggagtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagc
agcccgcatggatttgacttggtcagggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgc
cctgctgcgtaacatcgttgctgctgcgtaacatcgttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttgga
tgcccgaggcatagactgtacaaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggt
tctggaccagttgcgtgagcgcatacgctacttgcattacagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgttt
ccacggtgtgcgtcacccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtct
ccacgcatcgtcaggcattggcggccttgctgttcttctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacc
tcggccgtcgcggcgcttgccggtggtgctgaccccggatgaagtggttcgcatcctcggttttctggaaggcgagcatcgtttgttcgcc
caggactctagctatagttctagtggttggctacgtacccgtagtggctatggcagggcttgcgcttaatgcgccgctacagggcgcgtgg
ggataccccctagagccccagctggttctttccgcctcagaagccatagagcccaccgcatccccagcatgcctgctattgtcttcccaat
cctccccccttgctgtcctgccccacccccacccccccagaatagaatgacacctactcagacaatgcgatgcaatttcctcattttattaggaa
aggacagtgggagtggcaccttccagggtcaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaaggcacagtc
gaggctgatcagcgggtttaaacgggccctctagactcgagttaaagtcgacgcggggaggcggcccaaagggagatccgactcgtctg
agggcgaaggcgaagacgcggaagaggccgcagagccggcagcaggccgcgggaaggaaggtccgctggattgagggccgaaggg
acgtagcagaaggacgtcccgcgcagaatccaggtggcaacacaggcgagcagccaaggaaaggacgatgatttccccgacaacacc
acggaattgtcagtgcccaacagccgagcccctgtccagcagcgggcaaggcaggcggcgatgagttccgccgtggcaataggggagg
gggaaagcgaaagtcccggaaaggagctgacaggtggtggcaatgccccaaccagtgggggttgcgtcagcaaacacagtgcacacc
acgccacgttgcctgacaacgggccacaactcctcataaagagacagcaaccaggatttatacaaggaggagaaaatgaaagccatacg
ggaagcaatagcatgatacaaaggcattaaagcagcgtatccacatagcgtaaaaggagcaacatagttaagaataccagtcaatctttca
caaattttgtaatccagaggttgattgtcgacttaacgcgttgaattcTCAATGGTGATGGTGATGATGACCG
GTATGCATATTCAGATCCTCTTCTGAGATGAGTTTTTGTTCGAAGGGCCCCTT
GTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGC
AGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGG
TGCTCAGGTAGTGGTTGTCGGGCAGCAGCACGGGGCCGTCGCCGATGGGGGG
TGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTCGATGTTGTG
GCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATAT
AGACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCG
TCCTCCTTGAAGTCGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTGTCGCC
CTCGAACTTCACCTCGGCGCGGGTCTTGTAGTTGCCGTCGTCCTTGAAGAAGA
TGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCGTG
CTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCAGG
GTGGTCACGAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATG
AACTTCAGGGTCAGCTTGCC**cTA**AGTGGCATCGCCCTCGCCCTCGCCGGACAC
GCTGAACTTGTGGCCGTTTACGTCGCCGTCCAGCTCGACCAGGATGGGCACCA
CCCCGGTGAACAGCTCCTCGCCCTTGCTCACCATGgTGGCGGCgctagccagcttgggtct

ccctatagtgagtcgtattaatttcgataagccagtaagcagtgggttctctagttagccagagagctctgcttatatagacctcccaccgta
cacgcctaccgcccatttgcgtcaatggggcggagttgttacgacattttggaaagtcccgttgattttggtgccaaaacaaactcccattg
acgtcaatggggtggagacttggaaatccccgtgagtcaaaccgctatccacgcccattgatgtactgccaaaaccgcatcaccatggta
atagcgatgactaatacgtagatgtactgccaagtaggaaagtcccataaggtcatgtactgggcataatgccaggcgggccatttaccgt
cattgacgtcaataggggggcgtacttggcatatgatacacttgatgtactgccaagtgggcagtttaccgtaaatagtccacccattgacgt
caatggaaagtccctattggcgttactatgggaacatacgtcattattgacgtcaatgggcgggggtcgttgggcggtcagccaggcggg
ccatttaccgtaagttatgtaacgcggaactccatatatgggctatgaactaatgaccccgtaattgattactattaataactagtcaataatc
aatgtcaacgcgtatatctggcccgtacatcgcgaagcagcgcaaaacGGATCCtgcaggtatttGCGGCCGCggtccgt
atactccggaatattaatagatcatggagataattaaaatgataccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaat
aaaaaaacctataaatattccggattattcataccgtcccaccatcgggcgcgAACTCCTAAAAAACCGCCACCatg
aagtgccttttgtacttagccttttattcattggggtgaattgcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgt
tccttctaattaccattattgcccgtcaagctcagatttaaattggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagt
cacaaggctattcaagcagacggttggatgtgtcatgcttccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataaca
cattccatccgatccttcactccatctgtagaacaatgcaaggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttcc
ctcctcaaagttgtggatatgcaactgtgacggatgccgaagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacaca
ggagaatgggttgattcacagttcatcaacggaaaatgcagcaattacatatgccccactgtccataactctacaacctggcattctgactat
aaggtcaaagggctatgtgattctaacctcatttccatggacatcaccttcttctcagaggacggagagctatcatccctgggaaaggagg
gcacagggttcagaagtaactactttgcttatgaaactggaggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactccc
atcaggtgtctggttcgagatggctgataaggatctctttgctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccat
ctcagacctcagtggatgtaagtctaattcaggacgttgagaggatcttggattattccctctgccaagaaacctggagcaaaatcagagc
gggtcttccaatctctccagtggatctcagctatcttgctcctaaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaa
atactttgagaccagatacatcagagtcgatattgctgctccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagg
gaactgtgggatgactgggcaccatatgaagacgtggaaattggacccaatggagttctgaggaccagttcaggatataagtttcctttata
catgattggacatggtatgttggactccgatcttcatcttagctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgca
acttcctgatgatgagagtttatttttggtgatactgggctatccaaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagc
tctattgcctctttttctttatcatagggttaatcattggactattcttggttctccgagttggtatccatctttgcattaaattaaagcacaccaa
gaaaagacagatttatacagacatagagatgaaccgacttggaaagtgataaggccaggccggccaagcttgtcgagaagtactagagg
atcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgca
attgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcatt
ctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttgagcctagggctaggagatccgaaccagataa
gtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttacgacgctacacccagttcccatctattttgtcactcttccctaaat
aatccttaaaaactccatttccaccctctcccagttcccaactattttgtccgcccacagcggggcatttttcttcctgttatgttttaatcaaac
atcctgccaactccatgtgacaaaccgtcatcttcggctacttt

pAcBac-EGFP39*151* TAG

Same as above, but EGFP sequence contains a second TAG site at 151* as shown below.

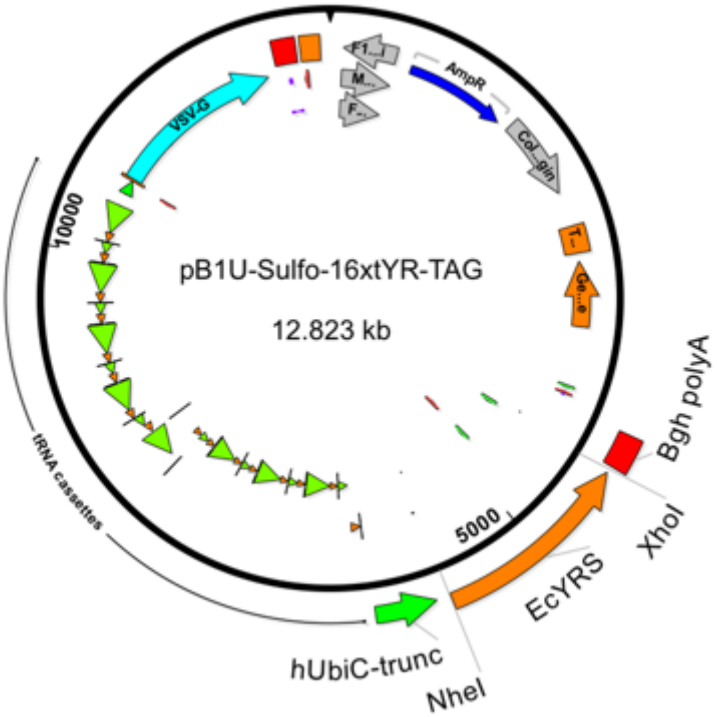
TCAATGGTGATGGTGATGATGACCGGTATGCATATTCAGATCCTCTTCTGAGA
TGAGTTTTTGTTCGAAGGGCCCCTTGTACAGCTCGTCCATGCCGAGAGTGATC
CCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGG
GGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCA
GCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCT
GCACGCTGCCGTCCTCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCG
TTCTTCTGCTTGTCGGCCATGAT**ATA**GACGTTGTGGCTGTTGTAGTTGTACTC
CAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGTCGATGCCCTTCAGCT
CGATGCGGTTCACCAGGGTGTCGCCCTCGAACTTCACCTCGGCGCGGGTCTTG
TAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGG
GCATGGCGGACTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCT
GAAGCACTGCACGCCGTAGGTCAGGGTGGTCACGAGGGTGGGCCAGGGCAC

GGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCC**CTA**AGTG
GCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTCGCC
GTCCAGCTCGACCAGGATGGGCACCACCCCGGTGAACAGCTCCTCGCCCTTGC
TCACCAT

*****

Chapter 4 plasmid sequences

pB1U-Sulfo-16xtYR-TAG: EcTyrRS VGL and VGM variants were amplified from pBK with oligonucleotides EcYRS-NheI-F and ExYRS-XhoI-R and subcloned into pB1U-OMeYRS-16xtYR-TAG between NheI and XhoI.



ttctctgtcacagaatgaaaatttttctgtcatctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaa
tgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgc
ccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgattt
agtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttt
gacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataaggga
ttttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtg
gcacttttcggggaaatgtgcgcggaaccccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataa
atgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttg
ctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggta
agatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgc
cgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggc
atgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaag
gagctaaccgctttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacg

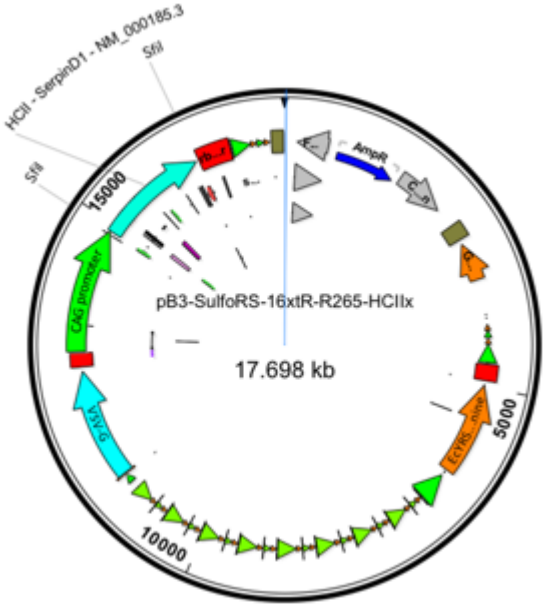
agcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaatt
aatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagcc
ggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcag
gcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatat
actttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttt
tcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaa
aaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac
caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttac
cagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaa
cgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgccac
gcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaa
acgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatgga
aaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggat
aaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaag
agcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcagaccagccgcgtaacctggcaaaatcggttacggttga
gtaataaatggatgcctgcgtaagcgggtgtgggcggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtctt
aaactagacagaatagttgtaaactgaaatcagtccagttatgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcatt
ttctgaagtgcaaattgcccgtcgtattaaagaggggcgtggccaagggcatggtaaagactatattcgcggcgttgtgacaatttaccga
acaactccgcggccgggaagccgatctcggcttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttcttcccg
tatgcccaactttgtatagagagccactgcgggatcgtcaccgtaatctgcttgcacgtagatcacataagcaccaagcgcgttggcctcat
gcttgaggagattgatgagcgcggtggcaatgccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcg
gctgctcaaacctgggcagaacgtaagccgcgagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtcttacta
cggagcaagttcccgaggtaatcggagtccggctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagc
agcccgcatggatttgacttggtcagggccgagcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgc
cctgctgcgtaacatcgttgctgctgcgtaacatcgttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttgga
tgcccgaggcatagactgtacaaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggt
tctggaccagttgcgtgagcgcatacgctacttgcattacagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgttt
ccacggtgtgcgtcacccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtct
ccacgcatcgtcaggcattggcggccttgctgttcttctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacc
tcggccgtcgcggcgcttgccggtggtgctgaccccggatgaagtggttcgcatcctcggttttctggaaggcgagcatcgtttgttcgcc
caggactctagctatagttctagtggttggctacgtacccgtagtggctatggcagggcttgcgcttaatgcgccgctacagggcgcgtgg
ggataccccctagagccccagctggttctttccgcctcagaagccatagagcccaccgcatccccagcatgcctgctattgtcttcccaat
cctcccccttgctgtcctgccccaccccaccccccagaatagaatgacacctactcagacaatgcgatgcaatttcctcattttattaggaa
aggacagtgggagtggcaccttccagggtcaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaaggcacagtc
gaggctgatcagcgggtttaaacgggccctctagactcgagttatttccagcaaatcagacagtaattctttttaccgcgacgcagtaaggt
aaaacgaccaaacagacgatcttcttctttaaagaagtattcaggatcggactgtttttcaccgttaatggtgatggcattggaggcgatagt
tttacgtgcctgaccacgggaaggttgcagttcagaatcgaccagtgcctgcatcaggtctgcgccctttccatctcaaccatcggtacgc
cgtcctgcgccagctgttcgaagtccgcttcactcagcgcactcaaagaaccgctgaacaggcattcggtaatacgttttgccgcctgtaa
accttcttcaccgtgaaccagacgagtcacctgctccgccagtacatactgggcgcgcggtgctttaccgctgtttttatcttcttcttccag
ggcgttgatctcttcaatgctcataaaggtgaagaacttcaggaagcggtaaacgtcggcatccgcagtgttgatccagaactggtagaatt
tgtacgggctggttttcttcggatccaaccagactgcgccgccttcagttttaccaaatttggtgccatctgctttagtgatcagcggaacgg
tcaggccaaacacctgattctgatgcagacgacgggtcaggtcgataccagaagtgatgttaccccactggtcagaaccaccaatttgca
gcaccacaccgtactgtttgttCAAacaggccatacTataaccctgCagcaggttgtaggaaaactcagtgaacgaaatcccctgaT
cttcacggttgagacgctgcttaaccgcttctttgttgatcatctggttaacggagaagtgtttgccaatatcgcgcaggaaggtcagcacat
tcatattgccgaaccagtcataattattggccgcgatagcagagttttctccacagtcgaaatcgaggaacggggcaacctgcttacggatt
ttgtccacccactcctgaacagtttcttcggtgttcagcttacgctcggcagctttgaagctcgggtcgccaatcagacccgtcgcgccgcc
taccagcgcaaccggcttgtggcccgcctgctggaagcgtttcaggcataacaatggaacaagatgccccaaatgcaagctgtcagcgg
taggatcgaagccacaCACgagtgcgatcgggccttgcgccagtcgctctgctaacgcttcctcgtccgtcacctgggctaccagccc

ccgctcttgcaattgtttaatcaagttactgcttgccatGGTGGCgctagccagcttgggtctccctatagtgagtcgtattaatttcga
taagccagtaagcagtgggttctctagttagccagagagctCtagaccaagtgacgatcacagcgatccacaaacaagaaccgcgaccc
aaatcccggctgcgacggaactagctgtgccacacccggcgcgtccttatataatcatcggcgttcaccgccccacggagatccctccgc
agaatcgccgagaagggactacttttcctcgcctgttccgctctctggaaagaaaaccagtgccctagagtcacccaagtcccgtcctaaa
atgtccttctgctgatactggggttctaaggccgagtcttatgagcagcgggccgctgtcctgagcgtccgggcggaaggatcaggacgc
tcgctgcgcccttcgtctgacgtggcagcgctcgccgtgaggaggggggcgcccgcgggaggcgccaaaacccggcgcggaggccttt
cgaacggCCACTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCGAACCCAAA
GGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGT
CTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATTTCACGT
TTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAATGGTGG
GGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCC
CTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAA
AGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTT
TAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTC
ACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTC
TCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGC
CCTCTTCCTGCCCGAcCTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCG
AACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCC
TCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGG
ATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAA
AATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAG
TCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAG
ATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATA
GTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTT
CTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAA
TTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGA
AATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATGGAGGGGGACGGATTCG
AACCGCCGAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTC
GCTACCCCTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACT
GATAGGGATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAA
TATTAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAG
ATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTT
TCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGC
ATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAAT
TATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAAT
TAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAA
TCATGGGAAATAGGCCCTCTTCCTGCCCGAcCTAGCAAAAAATGGAGGGGGAC
GGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTA
GCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCT
CTATCACTGATAGGGATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGA
CATGCAAATATTAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGT
GACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGT
TTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTA
CGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACC
CAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTAC
AGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATAT
GAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATGG
AGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAGAGTCCG

CCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAGGGAACT
TATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGGTGATTTCCCAGAACA
CATAGCGACATGCAAATATTAAAAAATGGTGGGGGAAGGATTCGAACCTTCG
AAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCC
ACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTT
TCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGC
AAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTT
GTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATAT
GCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAcCTAGCAA
AAAATGGAGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTA
GAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATA
GGGAACTTATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGGTGATTTC
CCAGAACACATAGCGACATGCAAATATTAAAAAATGGTGGGGGAAGGATTCG
AACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCG
GGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATC
GAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTT
TAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTAC
TAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTG
TATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCC
GACCTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCGAACCCAAAGGG
AGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTC
TATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATTTCACGTTTA
TGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAATGGTGGGG
GAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTT
TGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAG
CCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAA
AACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACG
TATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCT
AACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCT
CTTCCTGCCCGAcCTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCGAAC
CCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCC
GGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATT
TCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAAT
GGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCT
GCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATA
TATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTC
CATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTA
CGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTA
TCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAAT
AGGCCCTCTTCCTGCCCGACctagtcaataatcaatgtcaacgcgtatatctggcccgtacatcgcgaagcagcgca
aaacGGATCCtgcaggtatttGCGGCCGCggtccgtatactccggaatattaatagatcatggagataattaaaatgataac
catctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtcccaccatc
gggcgcgAACTCCTAAAAAACCGCCACCatgaagtgccttttgtacttagccttttattcattggggtgaattgcaag
ttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaattggc
ataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgcttccaaa
tgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgcaaggaa
agcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccgaagca
gtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgcagcaa

ttacatatgccccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccatggacatc
accttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaactggaggca
aggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctctttgctgc
agccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttgagagg
atcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctcctaaa
aacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgctccaat
cctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaaattgg
acccaatggagttctgaggaccagttcaggatataagtttcctttatacatgattggacatggtatgttggactccgatcttcatcttagctca
aaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttttggtgatactgggctatcca
aaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctctttttctttatcatagggttaatcattggactattctt
ggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgacttggaa
agtgataaggccaggccggccaagcttgtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctttaaa
aaacctcccacacctcccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaa
gcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgga
tctgatcactgcttgagcctaggagatccgaaccagataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttacgac
gctacacccagttcccatctattttgtcactcttccctaaataatccttaaaaactccatttccaccctcccagttcccaactattttgtccgc
ccacagcggggcatttttcttcctgttatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttcggctacttt

pB3-SulfoRS-16xYtR-TAG-HCII: pAcBac3 OMeYRS was used as a starting vector to construct this plasmid.[2] pB3 (abbreviated pAcBac3) is identical to pB1u except it contains a CAG promoter upstream from an SfiI site as well as 4 additional tRNA cassette copies. OMeYRS was replaced with SulfoRS via NheI/XhoI as previously described in pB1U cloning description. The SfiI site was used to insert HCII. HCII-SfiI-F and 10xHis-TGA-SfiI-R were used to amplify HCII from pCMV-SerpinD1 (Origine, SC120039). Mutations were introduced via overlap extension (see primer list for 79, 92, and 80 overlap primers – 79 and 92 correspond to 60 and 73 sites, respectively). HCII insert and vector were digested with SfiI and ligated via traditional RE cloning.



cctgttatgtttttaatcaaacatcctgccaactccatgtgacaaaccgtcatcttcggctacttttttctctgtcacagaatgaaaatttttctgt
catctcttcgttattaatgtttgtaattgactgaatatcaacgcttatttgcagcctgaatggcgaatgggacgcgccctgtagcggcgcatta
agcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttct

cgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaa
aacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtgg
actcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaa
tgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaatgtgcgcggaacc
cctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagt
atgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaa
gatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaac
gttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcata
cactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgc
cataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggg
gatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatg
gcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttg
caggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgca
gcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatc
gctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaattta
aaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaa
agatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgcc
ggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttag
gccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgt
cttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttgga
gcgaacgacctacaccgaactgagatacctacagcgtgagcattgagaaagcgccacgcttcccgaagggagaaaggcggacaggtat
ccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcg
ccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttc
ctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgatacc
gctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatct
gtgcggtatttcacaccgcagaccagccgcgtaacctggcaaaatcggttacggttgagtaataaatggatgccctgcgtaagcgggtgt
gggcggacaataaagtcttaaactgaacaaaatagatctaaactatgacaataaagtcttaaactagacagaatagttgtaaactgaaatca
gtccagttatgctgtgaaaaagcatactggacttttgttatggctaaagcaaactcttcattttctgaagtgcaaattgcccgtcgtattaaaga
ggggcgtggccaagggcatggtaaagactatattcgcggcgttgtgacaatttaccgaacaactccgcggccgggaagccgatctcggc
ttgaacgaattgttaggtggcggtacttgggtcgatatcaaagtgcatcacttcttcccgtatgcccaactttgtatagagagccactgcggg
atcgtcaccgtaatctgcttgcacgtagatcacataagcaccaagcgcgttggcctcatgcttgaggagattgatgagcgcggtggcaatg
ccctgcctccggtgctcgccggagactgcgagatcatagatatagatctcactacgcggctgctcaaacctgggcagaacgtaagccgc
gagagcgccaacaaccgcttcttggtcgaaggcagcaagcgcgatgaatgtcttactacggagcaagttcccgaggtaatcggagtccg
gctgatgttgggagtaggtggctacgtctccgaactcacgaccgaaaagatcaagagcagcccgcatggatttgacttggtcagggccga
gcctacatgtgcgaatgatgcccatacttgagccacctaactttgttttagggcgactgccctgctgcgtaacatcgttgctgctgcgtaaca
tcgttgctgctccataacatcaaacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatagactgtacaaaaaaacagt
cataacaagccatgaaaaccgccactgcgccgttaccaccgctgcgttcggtcaaggttctggaccagttgcgtgagcgcatacgctactt
gcattacagtttacgaaccgaacaggcttatgtcaactgggttcgtgccttcatccgtttccacggtgtgcgtcacccggcaaccttgggca
gcagcgaagtcgaggcatttctgtcctggctggcgaacgagcgcaaggtttcggtctccacgcatcgtcaggcattggcggccttgctgtt
cttctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggtagacctcggccgtcgcggcgcttgccggtggtgctgac
cccggatgaagtggttcgcatcctcggttttctggaaggcgagcatcgtttgttcgcccaggactctagctatagttctagtggttggctacg
tacccgtagtggctatggcagggcttgcgcttaatgcgccgctacagggcgcgtggggataccccctagagccccagctgAAAAA
ATGGAGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAGAG
TCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAGGG
AACTTATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGGTGATTTCCCAG
AACACATAGCGACATGCAAATATTAAAAAATGGTGGGGGAAGGATTCGAACC
TTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAA
CCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAAT

ACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAA
CTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTTGTACTAATAT
CTTTGTGTTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGT
ATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAGATC
TGgttctttccgcctcagaagccatagagcccaccgcatccccagcatgcctgctattgtcttcccaatcctccccccttgctgtcctgccc
caccccaccccccagaatagaatgacacctactcagacaatgcgatgcaatttcctcattttattaggaaaggacagtgggagtggcacct
tccagggtcaaggaaggcacgggggaggggcaaacaacagatggctggcaactagaaggcacagtcgaggctgatcagcgggtttaa
acgggccctctagactcgagttaaagtcgacttaacgcgttgaattcttaaacgggccctttccagcaaatcagacagtaattctttttaccg
cgacgcagtaaggtaaaacgaccaaacagacgatcttcttctttaaagaagtattcaggatcggactgtttttcaccgttaatggtgatggca
ttggaggcgatagttttacgtgcctgaccacgggaaggttgcagttcagaatcgaccagtgcctgcatcaggtctgcgccctttttccatctc
aaccatcggtacgccgtcctgcgccagctgttcgaagtccgcttcactcagcgcactcaaagaaccgctgaacaggcattcggtaatacg
ttttgccgcctgtaaaccttcttcaccgtgaaccagacgagtcacctgctccgccagtacatactgggcgcgcggtgctttaccgctgttttt
atcttcttcttccagggcgttgatctcttcaatgctcataaaggtgaagaacttcaggaagcggtaaacgtcggcatccgcagtgttgatcca
gaactggtagaatttgtacgggctggttttcttcgggtccaaccagactgcgccgccttcagtttaccaaatttggtgccatctgctttagtg
atcagcggaacggtcaggccaaacacctgattctgatgcagacgacgggtcaggtcgataccagaagtgatgttaccccattggtcagaa
ccaccaatttgcagcaccacaccgtactgtttgttagcacaggccataccataaccctggagcaggttgtaggaaaactcagtgaacgaaa
tcccctgaccttcacggttgagacgctgcttaaccgcttctttgttgatcatctggttaacggagaagtgtttgccaatatcgcgcaggaagg
tcagcacattcatattgccgaaccagtcataattattggccgcgatagcagagttttctccacagtcgaaatcgaggaacggggcaacctg
cttacggattttgtccacccactcctgaacagtttcttcggtgttcagcttacgctcggcagctttgaagctcgggtcgccaatcagacccgt
cgcgccgcctaccagcgcaaccggcttgtggcccgcctgctggaagcgtttcaggcataacaatggaacaagatgccccaaatgcaagc
tgtcagcggtaggatcgaagccacaaatgagtgcgatcgggccttgcgccagtcgctctgctaacgcttcctcgtccgtcacctgggcta
ccagcccccgctcttgcaattgtttaatcaagttactgcttgccatGGTGGCgctagccagcttgggtctccctatagtgagtcgtatt
aatttcgataagccagtaagcagtgggttctctagttagccagagagctCtagaccaagtgacgatcacagcgatccacaaacaagaacc
gcgacccaaatcccggctgcgacggaactagctgtgccacacccggcgcgtccttatataatcatcggcgttcaccgccccacggagat
ccctccgcagaatcgccgagaagggactacttttcctcgcctgttccgctctctggaaagaaaaccagtgccctagagtcacccaagtcc
cgtcctaaaatgtccttctgctgatactggggttctaaggccgagtcttatgagcagcgggccgctgtcctgagcgtccgggcggaaggat
caggacgctcgctgcgcccttcgtctgacgtggcagcgctcgccgtgaggaggggggcgcccgcgggaggcgccaaaacccggcgc
ggaggccttcgaacggCCACTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCGAA
CCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTC
CGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGAT
TTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAA
TGGTGGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTC
TGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGAT
ATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGT
CCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCT
ACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATT
ATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAA
TAGGCCCTCTTCCTGCCCGAcCTAGCAAAAAATGGAGGGGGACGGATTCGAAC
CGCCGAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCT
ACCCCTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGAT
AGGGATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATAT
TAAAAAATGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATT
TAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCC
ACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATA
TGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTAT
TACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAA
TTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCA
TGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAAAAATGGAGGGGGACGG

ATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGC
CACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCT
ATCACTGATAGGGATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACA
TGCAAATATTAAAAAATGGTGGGGGGAAGGATTCGAACCTTCGAAGTCTGTGA
CGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTT
CGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACG
GTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCAAACTACCCA
AGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACA
GTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATAT
GAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAcCTAGCAAAAAATGGA
GGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAGAGTCCGC
CGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAGGGAACTT
ATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGGTGATTTCCCAGAACAC
ATAGCGACATGCAAATATTAAAAAATGGTGGGGGGAAGGATTCGAACCTTCGA
AGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGGAACCCCA
CCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTT
CAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAAAACTGCA
AACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAATATCTTTG
TGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTATCGTATATG
CAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGACCTAGCAAA
AAATGGAGGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGCGGATTTAG
AGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATCACTGATAG
GGAACTTATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGGTGATTTCCC
AGAACACATAGCGACATGCAAATATTAAAAAATGGTGGGGGGAAGGATTCGAA
CCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGGCCGCTCGGG
AACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAA
ATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAACATAATTTTAA
AACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATTTTGTACTAA
TATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACAGCCTTGTAT
CGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTCCTGCCCGAc
CTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCGAACCCAAAGGGAGC
GGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGTGTCTCTATC
ACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATTTCACGTTTATGG
TGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAATGGTGGGGGAA
GGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCTCCCTTTGG
CCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCA
AGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCATTTTAAAAC
ATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGTCACGTATT
TTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCTCTCTAACA
GCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGGCCCTCTTC
CTGCCCGACCTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCCGAACCCA
AAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGGT
GTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGGGATTTCAC
GTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAAAAATGGT
GGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGAGTCTGCT
CCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAAGATATAT
AAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCCAT
TTTAAAACATAATTTTAAAACTGCAAACTACCCAAGAAATTATTACTTTCTACGT

CACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTAATTATCT
CTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGGAAATAGG
CCCTCTTCCTGCCCGAcCTAGCAAAAAATGGAGGGGGACGGATTCGAACCGCC
GAACCCAAAGGGAGCGGATTTAGAGTCCGCCGCGTTTAGCCACTTCGCTACCC
CTCCGGTGTCTCTATCACTGATAGGGAACTTATAAGTCTCTATCACTGATAGG
GATTTCACGTTTATGGTGATTTCCCAGAACACATAGCGACATGCAAATATTAAA
AAATGGTGGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTAGA
GTCTGCTCCCTTTGGCCGCTCGGGAACCCCACCGGTGTTTCGTCCTTTCCACAA
GATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGAT
AGTCCATTTTTAAAACATAATTTTTAAAACTGCAAACTACCCAAGAAATTATTACTT
TCTACGTCACGTATTTTGTACTAATATCTTTGTGTTTACAGTCAAATTAATTCTA
ATTATCTCTCTAACAGCCTTGTATCGTATATGCAAATATGAAGGAATCATGGG
AAATAGGCCCTCTTCCTGCCCGACctagtcaataatcaatgtcaacgcgtatatctggcccgtacatcgcgaagc
agcgcaaaacGGATCCtgcaggtatttGCGGCCGCggtccgtatactccggaatattaatagatcatggagataattaaaat
gataaccatctcgcaaataaataagtattttactgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgtccc
accatcgggcgcgAACTCCTAAAAAACCGCCACCatgaagtgccttttgtacttagccttttttattcattggggtgaatt
gcaagttcaccatagtttttccacacaaccaaaaaggaaactggaaaaatgttccttctaattaccattattgcccgtcaagctcagatttaaa
ttggcataatgacttaataggcacagccttacaagtcaaaatgcccaagagtcacaaggctattcaagcagacggttggatgtgtcatgctt
ccaaatgggtcactacttgtgatttccgctggtatggaccgaagtatataacacattccatccgatccttcactccatctgtagaacaatgca
aggaaagcattgaacaaacgaaacaaggaacttggctgaatccaggcttccctcctcaaagttgtggatatgcaactgtgacggatgccg
aagcagtgattgtccaggtgactcctcaccatgtgctggttgatgaatacacaggagaatgggttgattcacagttcatcaacggaaaatgc
agcaattacatatgccccactgtccataactctacaacctggcattctgactataaggtcaaagggctatgtgattctaacctcatttccatgg
acatcaccttcttctcagaggacggagagctatcatccctgggaaaggagggcacagggttcagaagtaactactttgcttatgaaactgg
aggcaaggcctgcaaaatgcaatactgcaagcattggggagtcagactcccatcaggtgtctggttcgagatggctgataaggatctctttt
gctgcagccagattccctgaatgcccagaagggtcaagtatctctgctccatctcagacctcagtggatgtaagtctaattcaggacgttga
gaggatcttggattattccctctgccaagaaacctggagcaaaatcagagcgggtcttccaatctctccagtggatctcagctatcttgctc
ctaaaaacccaggaaccggtcctgctttcaccataatcaatggtaccctaaaatactttgagaccagatacatcagagtcgatattgctgct
ccaatcctctcaagaatggtcggaatgatcagtggaactaccacagaaagggaactgtgggatgactgggcaccatatgaagacgtggaa
attggacccaatggagttctgaggaccagttcaggatataagtttcctttatacatgattggacatggtatgttggactccgatcttcatctta
gctcaaaggctcaggtgttcgaacatcctcacattcaagacgctgcttcgcaacttcctgatgatgagagtttatttttggtgatactgggct
atccaaaaatccaatcgagcttgtagaaggttggttcagtagttggaaaagctctattgcctctttttctttatcatagggttaatcattggact
attcttggttctccgagttggtatccatctttgcattaaattaaagcacaccaagaaaagacagatttatacagacatagagatgaaccgactt
ggaaagtgataagtcgagaagtactagaggatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctc
cccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaa
atttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatctgatcactgcttga
gcCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATAT
GGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCA
ACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAA
TAGGGACTTTCCATTGACGTCAATGGGTGGAgTATTTACGGTAAACTGCCCAC
TTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAAT
GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGT
GAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATT
TTGTATTTATTTATTTTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGG
GGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGG
GCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAG
TTTCCTTTTATGGCGAGGCGGCGGCGGCGGCGGCCCTATAAAAAGCGAAGCG
CGCGGCGGGCGGGAGTCGCTGCGTTGCCTTCGCCCCGTGCCCCGCTCCGCGC

CGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGA
GCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTAATG
ACGGCTCGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTAAAGGGCTCCGGGAG
GGCCCTTTGTGCGGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGT
GCGTGGGGAGCGCCGCGTGCGGCCCGCGCTGCCCGGCGGCTGTGAGCGCTG
CGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGCGAGGGGAGCGC
GGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGCTGCGAGGGGAACAAAGG
CTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCG
GCGGTCGGGCTGTAACCCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACG
GCCCGGCTTCGGGTGCGGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGCCG
TGCCGGGCGGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCC
GCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCGGAGCGC
CGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCG
TGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGGCGGAGCCGAAATC
TGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGCGAAGCGGTGCGGC
GCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGC
CGTCCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGCAGGGGGACGGCTGCCTT
CGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGG
CTCTAGAGCCTCTGCTAACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTG
GGCAACGTGCTGGTTaTTGTGCTGTCTCATCATTTTGGCAAAGAATTGGCCAA
GGAGGCCACCatgaaacactcattaaacgcacttctcattttcctcatcataacatctgcgtggggtgggagcaaaggcccgct
ggatcagctagagaaaggaggggaaactgctcagtctgcagatccccagtgggagcagttaaataacaaaaacctgagcatgcctcttct
ccctgccgacttccacaaggaaaacaccgtcaccaacgactggattccagaggggggaggaggacgacgactatctggacctggagaag
atattcagtgaagacgacgactacatcgacatcgtcgacagtctgtcagtttccccgacagactctgatgtgagtgctgggaacatcctcca
gctttttcatggcaagagccggatccagcgtcttaacatcctcaacgccaagttcgctttcaacctctaccgagtgctgaaagaccaggtca
acactttcgataacatcttcatagcacccgttggcatttctactgcgatgggtatgatttccttaggtctgaagggagagacccatgaacaag
tgcactcgattttgcattttaaagactttgttaatgccagcagcaagtatgaaatcacgaccattcataatctcttccgtaagctgactcatcg
cctcttcaggaggaattttgggtacacactgcggtcagtcaatgacctttatatccagaagcagtttccaatcctgcttgacttcaaaactaa
agtaagagagtattactttgctgaggcccagatagctgacttctcagaccctgccttcatatcaaaaaccaacaaccacatcatgaagctca
ccaagggcctcataaaagatgctctggagaatatagaccctgctacccagatgatgattctcaactgcatctacttcaaaggatcctgggtg
aataaattcccagtggaaatgacacacaaccacaacttccggctgaatgagagagaggtagttaaggtttccatgatgcagaccaagggg
aacttcctcgcagcaaatgaccaggagctggactgcgacatcctccagctggaatacgtgggggggcatcagcatgctaattgtggtccca
cacaagatgtctgggatgaagaccctcgaagcgcaactgacaccccgggtggtggagagatggcaaaaaagcatgacaaacagaactc
gagaagtgcttctgccgaaattcaagctggagaagaactacaatctagtggagtccctgaagttgatggggatcaggatgctgtttgacaa
aaatggcaacatggcaggcatctcagaccaaaggatcgccatcgacctgttcaagcaccaaggcacgatcacagtgaacgaggaaggc
acccaagccaccactgtgaccacggtggggttcatgccgctgtccacccaagtccgcttcactgtcgaccgccccttttcttttcctcatcta
cgagcaCcgcaccagctgcctgctcttcatgggaagagtggccaaccccagcaggtccCATCATCACCATCATCATC
ACCACCATCATtGAggcctggccggccgccagcacagtggtcgatcgaCCAATGCCCTGGCTCACAA
ATACCACTGAGATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCATGAAGCC
CCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAATAGT
GTGTTGGAATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATC
ATTTAAAACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCcATA
TGCTGGCTGCCATGAACAAAGGTtGGCTATAAAGAGGTCATCAGTATATGAAA
CAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGAGGTTAG
ATTTTTTTTTATATTTTGTTTTGTGTTATTTTTTTTCTTTAACATCCCTAAAATTTTC
CTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTGACTACTCCCAGTCATA
GCTGTCCCTCTTCTCTTATGgAGATCCCTCGACCTGCcctaggTCGGGCAGGAAGA
GGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAG

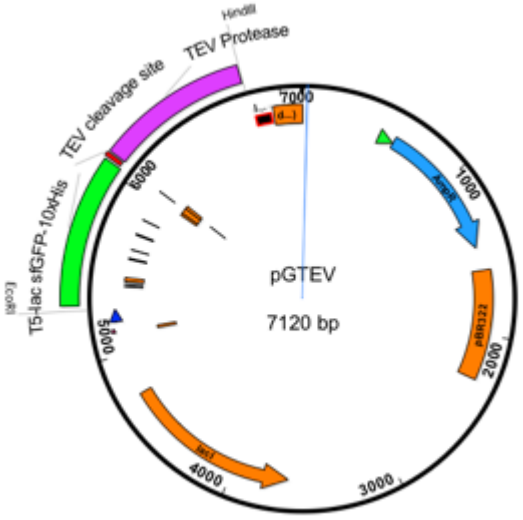
AGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATA
CGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGT
TTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTG
GCTTTATATATCTTGTGGAAAGGACGAAACACCGGTGGGGTTCCCGAGCGGC
CAAAGGGAGCAGACTCTAAATCTGCCGTCACAGACTTCGAAGGTTCGAATCCT
TCCCCCACCATTTTTTAATATTTGCATGTCGCTATGTGTTCTGGGAAATCACCA
TAAACGTGAAATCCCTATCAGTGATAGAGACTTATAAGTTCCCTATCAGTGATA
GAGACACCGGAGGGGTAGCGAAGTGGCTAAACGCGGCGGACTCTAAATCCG
CTCCCTTTGGGTTCGGCGGTTCGAATCCGTCCCCcTCCATTTTTTTgctaggagatccgaac
cagataagtgaaatctagttccaaactattttgtcatttttaattttcgtattagcttacgacgctacacccagttcccatctattttgtcactctt
ccctaaataatccttaaaaactccatttccacccctcccagttcccaactattttgtccgcccacagcggggcatttttct

****

Chapter 5 plasmid sequences

Plasmid map and full plasmid sequence are given below. *E. coli* tryptophanyl tRNA/synthetase pairs are in blue, *M. jannaschii* tyrosine tRNA/synthetase pairs are in gold, *Pyrrolysyl* tRNA/synthetase pairs are in purple, GFP-TEV reporters are in green, promoters are in red.

pGTEV



tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcc
ctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggt
tccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttc
gcccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttat
aagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataacc
ctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcc
tgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaaca
gcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtat
tgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacg
gatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggac
cgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggca
acaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctg
gagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggga
gtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactca
tatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtg
agttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaac
aaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcag
ataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctg
ttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggc
tgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcg
ccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccaggg
ggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagccta
tggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgt
ggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcg
gaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgc
cgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgcc
ctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatca
ccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcg
ttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgta
agggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggtta
ctggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaat
acagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttcc
agactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgc
gtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtg
gggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaa
gattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggc
acctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgg
gttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttcc
agtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttt
cttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgcccc
agcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgc
accaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatg
ccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtg
agatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgc
gaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataac
gccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcg

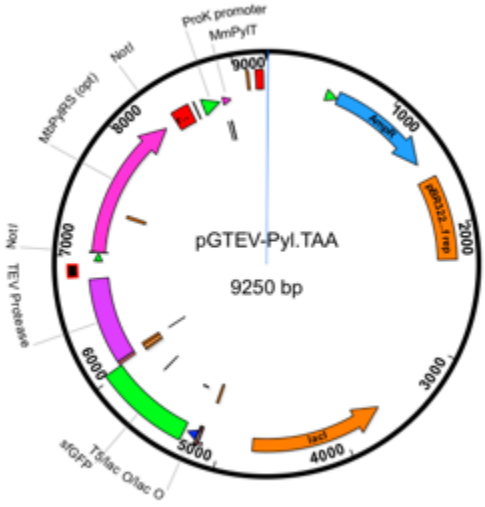
agaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcga
gatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccag
ttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggtt
caccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattga
ctctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcct
gcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacag
tcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtg
atgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgat
cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagagtttgacagcattatcatcgatctcgagaa
atcataaaaaatttatttgctttgtgagcggataacaattataatagattcaattgtgagcggataacaatttcacacagaattcattaaagagg
agaaattacatATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGT
TGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGT
GAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGG
AAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCtatGGTGTTCAA
TGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCC
ATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAC
CTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTA
TCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAA
CTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAA
GAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCG
TTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCC
TTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCA
ACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATT
ACACATGGCATGGATGAGCTCTACAAAGGATCCCACCACCACCACCACCACcatc
atcatcatGAAAATCTTATTTTCAAggtggaggaagtggagaaagTttgtttaaggggccgcgtgattacaacccga
tatcgagcaccatttgtcatttgacgaatgaatctgatgggcacacaacatcgttgtatggtattggatttggtcccttcatcattacaaacaa
gcacttgtttagaagaaataatggaacactgttggtccaatcactacatggtgtattcaaggtcaagaacaccacgactttgcaacaacacc
tcattgatgggagggacatgataattattcgcatgcctaaggatttcccaccatttcctcaaaagctgaaatttagagagccacaaagggaa
gagcgcatatgtcttgtgacaaccaacttccaaactaagagcatgtctagcatggtgtcagacactagttgcacattcccttcatctgatggc
atattctggaagcattggattcaaaccaaggatgggcagtgtggcagtccattagtatcaactagagatgggttcattgttggtatacactca
gcatcgaatttcaccaacacaaacaattatttcacaagcgtgccgaaaaacttcatggaattgttgacaaatcaggaggcgcagcagtggg
ttagtggttggcgattaaatgctgactcagtattgtgggggggccataaagttttcatggtgaaacctgaagagccttttcagccagttaagg
aagcgactcaactcatgaatcgtcgtcgccgtcgctaaTAATAAaagcttaattagctgagcttggactcctgttgatagatccagta
atgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaatccaagctagcttggcggcgg
ccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctga
gcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat

pGTEV-STOP Variants: Throughout this work, there are multiple variations of the GTEV reporter construct with different stop codons. Mutated codons are underlined, bolded, and numerically labeled in **red** font in the GFP-TEV fusion sequence below. Variants contain either a TAG, TAA, or TGA stop codon as described in each experiment.

ATG**gcatat(3)**AGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGT
TGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGT
GAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGC**ACT(51)**ACT
GGAAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCtatGGTGTTC
AATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGT
GCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGG
GACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATC
GTATCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACAC

AAACTCGAGTACAACTTTAACTCACACAATGTA**TAC(151)**ATCACGGCAGACAAA
CAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGG
TTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCC
TGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGA
TCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTG
GGATTACACATGGCATGGATGAGCTCTACAAAGGATCCCACCACCACCACCAC
CACcatcatcatcatGAAAATCTTTATTTTCAAggtggaggaagtggagaaagTttgtttaaggggccgcgtgatta
caacccgatatcgagcaccatttgtcatttgacgaatgaatctgatgggcacacaacatcgttgtatggtattggatttggtcccttcatcatt
acaaacaagcacttgtttagaagaaataatggaacactgttggtccaatcactacatggtgtattcaaggtcaagaacaccacgactttgca
acaacacctcattgatgggagggacatgataattattcgcatgcctaaggatttcccaccatttcctcaaaagctgaaatttagagagccac
aaagggaagagcgcatatgtcttgtgacaaccaacttccaaactaagagcatgtctagcatggtgtcagacactagttgcacattcccttca
tctgatggcatattctggaagcattggattcaaaccaaggatgggcagtgtggcagtccattagtatcaactagagatgggttcattgttggt
atacactcagcatcgaatttcaccaacacaaacaattatttcacaagcgtgccgaaaaacttcatggaattgttgacaaatcaggaggcgca
gcagtgggttagtggttggcgattaaatgctgactcagtattgtgggggggccataaagttttcatggtgaaacctgaagagccttttcagc
cagttaaggaagcgactcaactcatgaatcgtcgtcgccgtcgctaaTAATAA

pGTEV-Variant-Pyl.TAA



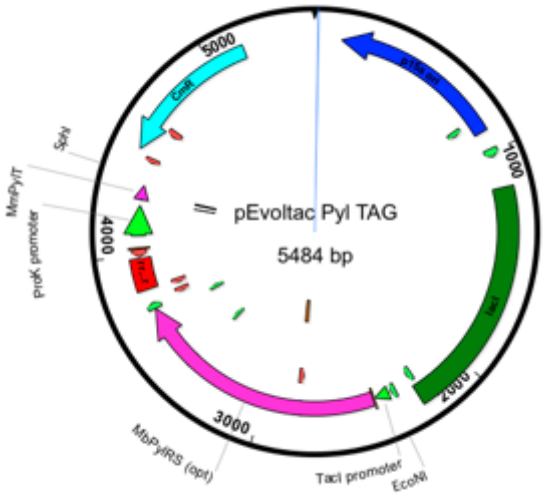
tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgcc
ctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggt
tccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttc
gccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttat
aagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattt
caggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataacc
ctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcc
tgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaaca
gcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtat
tgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacg
gatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggac
cgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaa
cgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggca
acaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctg
gagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacgggga

gtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactca
tatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttttgataatctcatgaccaaaatcccttaacgtg
agttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaac
aaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcag
ataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctg
ttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggc
tgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcg
ccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccaggg
ggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagccta
tggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgt
ggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcg
gaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatatggtgcactctcagtacaatctgctctgatgc
cgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgcc
ctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatca
ccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcg
ttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgta
agggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggtta
ctggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgcttcgttaat
acagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttccgcgtttcc
agactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgc
gtatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtg
gggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaa
gattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggc
acctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgg
gttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttcc
agtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttt
cttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgcccc
agcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgc
accaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatg
ccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtg
agatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgc
gaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataac
gccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcg
agaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcga
gatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccag
ttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggtt
caccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattga
ctctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcct
gcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacag
tccccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtg
atgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgat
cccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagagtttgacagcattatcatcgatctcgagaa
atcataaaaaatttatttgctttgtgagcggataacaattatataatagattcaattgtgagcggataacaatttcacacagaattcattaaagagg
agaaattacatATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGT
TGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGT
GAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGG
AAAACTACCTGTTCCGTGGCCAACACTTGTCACTACTCTGACCtatGGTGTTCAA
TGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCC

ATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGAC
CTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTA
TCGAGTTAAAGGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAAA
CTCGAGTACAACTTTAACTCACACAATGTATATATCACGGCAGACAAACAAAAG
AATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGT
TCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCT
TTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAA
CGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTA
CACATGGCATGGATGAGCTCTACAAAGGATCCCACCACCACCACCACCACcatcat
catcatGAAAATCTTTATTTTCAAggtggaggaagtggagaaagTttgtttaaggggccgcgtgattacaacccgata
tcgagcaccatttgtcatttgacgaatgaatctgatgggcacacaacatcgttgtatggtattggatttggtcccttcatcattacaaacaagc
acttgtttagaagaaataatggaacactgttggtccaatcactacatggtgtattcaaggtcaagaacaccacgactttgcaacaacacctc
attgatgggagggacatgataattattcgcatgcctaaggatttcccaccatttcctcaaaagctgaaatttagagagccacaaagggaaga
gcgcatatgtcttgtgacaaccaacttccaaactaagagcatgtctagcatggtgtcagacactagttgcacattcccttcatctgatggcat
attctggaagcattggattcaaaccaaggatgggcagtgtggcagtccattagtatcaactagagatgggttcattgttggtatacactcagc
atcgaatttcaccaacacaaacaattatttcacaagcgtgccgaaaaacttcatggaattgttgacaaatcaggaggcgcagcagtgggtta
gtggttggcgattaaatgctgactcagtattgtgggggggccataaagttttcatggtgaaacctgaagagccttttcagccagttaaggaa
gcgactcaactcatgaatcgtcgtcgccgtcgctaaTAATAAaagcttaattagctgagcttggactcctgttgatagatccagtaat
gacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaatccaagctagcttggcggcGG
CCCATTGAGGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATT
GTGAGCGGATAACAATTTCACAaAGGAggtGCGGCCGCATGGATAAAAAACCGC
TGGACGTTCTGATCTCCGCTACGGGTCTGTGGATGAGCCGCACGGGTACGCT
GCATAAAATTAAACACCACGAAGTGTCACGTTCGAAAATCTATATCGAAATGG
CGTGCGGTGATCATCTGGTGGTTAACAATAGCCGTTCTTGTCGCACCGCGCGT
GCCTTTCGCCATCACAAATACCGCAAAACGTGCAAACGTTGTCGCGTGTCAGA
TGAAGACATTAACAATTTCCTGACCCGTAGTACGGAATCCAAAAACTCAGTGA
AAGTTCGCGTCGTGAGTGCTCCGAAAGTTAAAAAAGCGATGCCGAAAAGTGT
CTCCCGTGCCCCGAAACCGCTGGAAAACTCAGTGTCGGCAAAAGCTTCCACCA
ATACGAGCCGCTCTGTTCCGTCGCCGGCAAAAAGCACCCCGAACAGCTCTGTC
CCGGCAAGCGCACCGGCACCGTCTCTGACGCGTAGTCAGCTGGATCGCGTGG
AAGCCCTGCTGTCCCCGGAAGACAAAATCTCACTGAATATGGCAAAACCGTTT
CGTGAACTGGAACCGGAACTGGTTACCCGTCGCAAAAACGATTTCCAACGTCT
GTATACGAATGATCGCGAAGACTACCTGGGTAAACTGGAACGTGATATCACCA
AATTTTTCGTGGACCGCGGCTTTCTGGAAATCAAATCTCCGATTCTGATCCCG
GCTGAATATGTTGAACGCATGGGTATTAACAATGATACCGAACTGAGTAAACA
GATTTTTCGTGTGGATAAAAACCTGTGCCTGCGGCCGATGCTGGCACCGACGC
TGTATAATTACCTGCGTAAACTGGATCGCATTCTGCCGGGTCCGATTAAAATC
TTTGAAGTGGGCCCGTGTTATCGTAAAGAATCGGATGGCAAAGAACACCTGG
AAGAATTTACCATGGTTAACTTCTGCCAAATGGGCAGCGGTTGTACGCGCGAA
AATCTGGAAGCGCTGATCAAAGAATTCCTGGATTACCTGGAAATCGACTTCGA
AATCGTCGGTGATTCTTGCATGGTGTTTGGCGATACCCTGGACATCATGCATG
GTGACCTGGAACTGAGTTCCGCTGTTGTCGGTCCGGTCAGCCTGGATCGTGA
ATGGGGCATTGACAAACCGTGGATCGGCGCGGGTTTTGGCCTGGAACGCCTG
CTGAAAGTTATGCACGGCTTCAAAAACATCAAACGTGCGTCTCGCTCGGAATC
GTATTACAACGGCATCTCAACCAATCTGTAATAAGCGGCCGCgtttaaacggtctccagctt
ggctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcg
gcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcg
agagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccc

cataatccttgttagCCTGCAGGtaattccgcttcgcaacatgtgagcaccggtttattgactaccggaaGCAGTGTGA
CCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAG
GTAATAATTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTT
TCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTG
CGCAGTAAGATGCGCCCCGCATTGGAAACCTGATCATGTAGATCGAACGGACT
tTAAATCcGTTCAGCCGGGTTAGATTCCCGGGGGTTTCCGCCAAATTCGAAAAGC
CTGCTCAACGAGCAGGCTTTTTTGCATGctcgagcagctcagggtcgaatttgctttcgaatttctgccattc
atccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcACCAATAAGggccgcactcgagcaccaccaccacc
accactgagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataaccccttggg
gcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat

pEvoltac Pyl TAG

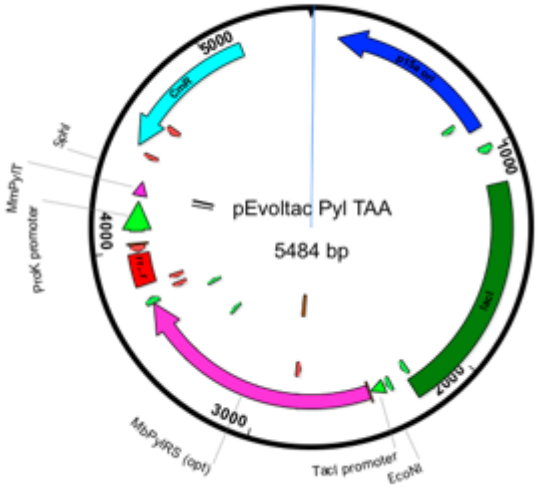


gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcag
tgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactg
actcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatactta
acagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagt
ggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac
cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg
cacgaacccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactg
gcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg
ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagag
caagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaa
atgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatg
tagcacctgaagtcagccccatacgatataagttgttaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg
aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttcttttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggc
gaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacg
cgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattt
atgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccaga
tgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa

cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatc
gccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc
cacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc
gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttcc
gggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagg
GAGCTGTTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGG
ATAACAATTTCACAaAGGAggtGCGGCCGCATGGATAAAAAACCGCTGGACGTT
CTGATCTCCGCTACGGGTCTGTGGATGAGCCGCACGGGTACGCTGCATAAAA
TTAAACACCACGAAGTGTCACGTTCGAAAATCTATATCGAAATGGCGTGCGGT
GATCATCTGGTGGTTAACAATAGCCGTTCTTGTCGCACCGCGCGTGCCTTTCG
CCATCACAAATACCGCAAAACGTGCAAACGTTGTCGCGTGTCAGATGAAGACA
TTAACAATTTCCTGACCCGTAGTACGGAATCCAAAAACTCAGTGAAAGTTCGC
GTCGTGAGTGCTCCGAAAGTTAAAAAAGCGATGCCGAAAAGTGTCTCCCGTG
CCCCGAAACCGCTGGAAAACTCAGTGTCGGCAAAAGCTTCCACCAATACGAGC
CGCTCTGTTCCGTCGCCGGCAAAAAGCACCCCGAACAGCTCTGTCCCGGCAAG
CGCACCGGCACCGTCTCTGACGCGTAGTCAGCTGGATCGCGTGGAAGCCCTG
CTGTCCCCGGAAGACAAAATCTCACTGAATATGGCAAAACCGTTTCGTGAACT
GGAACCGGAACTGGTTACCCGTCGCAAAAACGATTTCCAACGTCTGTATACGA
ATGATCGCGAAGACTACCTGGGTAAACTGGAACGTGATATCACCAAATTTTTC
GTGGACCGCGGCTTTCTGGAAATCAAATCTCCGATTCTGATCCCGGCTGAATA
TGTTGAACGCATGGGTATTAACAATGATACCGAACTGAGTAAACAGATTTTTC
GTGTGGATAAAAACCTGTGCCTGCGGCCGATGCTGGCACCGACGCTGTATAA
TTACCTGCGTAAACTGGATCGCATTCTGCCGGGTCCGATTAAAATCTTTGAAG
TGGGCCCGTGTTATCGTAAAGAATCGGATGGCAAAGAACACCTGGAAGAATT
TACCATGGTTAACTTCTGCCAAATGGGCAGCGGTTGTACGCGCGAAAATCTGG
AAGCGCTGATCAAAGAATTCCTGGATTACCTGGAAATCGACTTCGAAATCGTC
GGTGATTCTTGCATGGTGTTTGGCGATACCCTGGACATCATGCATGGTGACCT
GGAACTGAGTTCCGCTGTTGTCGGTCCGGTCAGCCTGGATCGTGAATGGGGC
ATTGACAAACCGTGGATCGGCGCGGGTTTTGGCCTGGAACGCCTGCTGAAAG
TTATGCACGGCTTCAAAAACATCAAACGTGCGTCTCGCTCGGAATCGTATTAC
AACGGCATCTCAACCAATCTGTAATAAGCGGCCGCgtttaaacggtctccagcttggctgttttggc
ggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcg
gtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaa
ctgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgtt
agCCTGCAGGtaattccgcttcgcaacatgtgagcaccggtttattgactaccggaaGCAGTGTGACCGTGTG
CTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAA
TTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCC
GATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGT
AAGATGCGCCCCGCATTGGAAACCTGATCATGTAGATCGAACGGACTcTAAAT
CcGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGCCAAATTCGAAAAGCCTGCT
CAACGAGCAGGCTTTTTTGCATGctcgagcagctcagggtcgaatttgctttcgaatttctgccattcatccgctta
ttatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgcagta
ctgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcg
ccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcaccca
gggattggctgagacgaaaaacatattctcaataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatata
tgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtga
acactatcccatatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaagg

ccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactga
ctgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcct
gaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcatt
ttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttatt
cggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagct
gtccctcctgttcagctactgacg

pEvoltac Pyl TAA

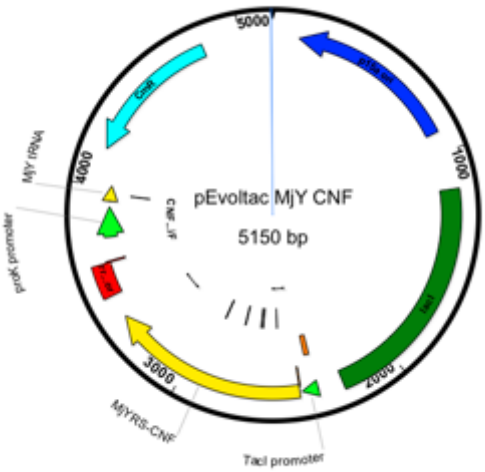


gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcag
tgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactg
actcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatactta
acagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagt
ggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac
cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg
cacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactg
gcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg
ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttttcagag
caagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaa
atgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatg
tagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg
aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggc
gaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacg
cgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattt
atgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccaga
tgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa
cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatc
gccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc
cacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc

gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttcc
gggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagg
GAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGG
ATAACAATTTCACAaAGGAggtGCGGCCGCATGGATAAAAAACCGCTGGACGTT
CTGATCTCCGCTACGGGTCTGTGGATGAGCCGCACGGGTACGCTGCATAAAA
TTAAACACCACGAAGTGTCACGTTCGAAAATCTATATCGAAATGGCGTGCGGT
GATCATCTGGTGGTTAACAATAGCCGTTCTTGTCGCACCGCGCGTGCCTTTCG
CCATCACAAATACCGCAAAACGTGCAAACGTTGTCGCGTGTCAGATGAAGACA
TTAACAATTTCCTGACCCGTAGTACGGAATCCAAAAACTCAGTGAAAGTTCGC
GTCGTGAGTGCTCCGAAAGTTAAAAAAGCGATGCCGAAAAGTGTCTCCCGTG
CCCCGAAACCGCTGGAAAACTCAGTGTCGGCAAAAGCTTCCACCAATACGAGC
CGCTCTGTTCCGTCGCCGGCAAAAAGCACCCCGAACAGCTCTGTCCCGGCAAG
CGCACCGGCACCGTCTCTGACGCGTAGTCAGCTGGATCGCGTGGAAGCCCTG
CTGTCCCCGGAAGACAAAATCTCACTGAATATGGCAAAACCGTTTCGTGAACT
GGAACCGGAACTGGTTACCCGTCGCAAAAACGATTTCCAACGTCTGTATACGA
ATGATCGCGAAGACTACCTGGGTAAACTGGAACGTGATATCACCAAATTTTTC
GTGGACCGCGGCTTTCTGGAAATCAAATCTCCGATTCTGATCCCGGCTGAATA
TGTTGAACGCATGGGTATTAACAATGATACCGAACTGAGTAAACAGATTTTTC
GTGTGGATAAAAACCTGTGCCTGCGGCCGATGCTGGCACCGACGCTGTATAA
TTACCTGCGTAAACTGGATCGCATTCTGCCGGGTCCGATTAAAATCTTTGAAG
TGGGCCCGTGTTATCGTAAAGAATCGGATGGCAAAGAACACCTGGAAGAATT
TACCATGGTTAACTTCTGCCAAATGGGCAGCGGTTGTACGCGCGAAAATCTGG
AAGCGCTGATCAAAGAATTCCTGGATTACCTGGAAATCGACTTCGAAATCGTC
GGTGATTCTTGCATGGTGTTTGGCGATACCCTGGACATCATGCATGGTGACCT
GGAACTGAGTTCCGCTGTTGTCGGTCCGGTCAGCCTGGATCGTGAATGGGGC
ATTGACAAACCGTGGATCGGCGCGGGTTTTGGCCTGGAACGCCTGCTGAAAG
TTATGCACGGCTTCAAAAACATCAAACGTGCGTCTCGCTCGGAATCGTATTAC
AACGGCATCTCAACCAATCTGTAATAAGCGGCCGCgtttaaacggtctccagcttggctgttttggc
ggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcg
gtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaa
ctgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgtt
agCCTGCAGGtaattccgcttcgcaacatgtgagcaccggtttattgactaccggaaGCAGTGTGACCGTGTG
CTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCGTGGAGGTAATAA
TTGACGATATGATCAGTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCC
GATCAAAAGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCAGT
AAGATGCGCCCCGCATTGGAAACCTGATCATGTAGATCGAACGGACTtTAAAT
CcGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGCCAAATTCGAAAAGCCTGCT
CAACGAGCAGGCTTTTTTGCATGctcgagcagctcagggtcgaatttgctttcgaatttctgccattcatccgctta
ttatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgcagta
ctgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcg
ccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcaccca
gggattggctgagacgaaaaacatattctcaataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatata
tgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtga
acactatcccatatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaagg
ccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactga
ctgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcct
gaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcatt
ttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttatt

cggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagct
gtccctcctgttcagctactgacg

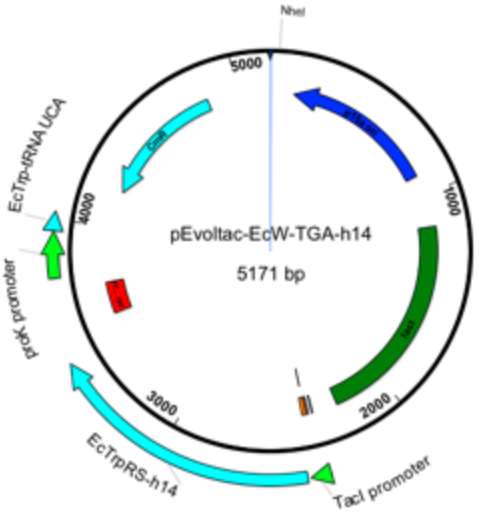
pEvoltac MjY CNF



gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcag
tgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactg
actcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatactta
acagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagt
ggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac
cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg
cacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactg
gcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg
ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttttcagag
caagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaa
atgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatg
tagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg
aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggc
gaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacg
cgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattt
atgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccaga
tgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa
cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatc
gccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc
cacgcggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc
gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttcc
gggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagg
GAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGG
ATAACAATTTCACAaAGGAggtGCGGCCGCatggacgaatttgaaatgataaagagaaacacatctgaaatta

tcagcgaggaagagttaagagaggttttaaaaaaagatgaaaaatctgctCTgataggttttgaaccaagtggtaaaatacatttagggca
ttatctccaaataaaaaagatgattgatttacaaaatgctggatttgatataattataGTTttggctgatttaCATgcctatttaaaccaga
aaggagagttggatgagattagaaaaataggagattataacaaaaaagtttttgaagcaatggggttaaaggcaaaatatgtttatggaagt
gaaTGGATGcttgataaggattatacactgaatgtctatagattggctttaaaaactaccttaaaaagagcaagaaggagtatggaac
ttatagcaagagaggatgaaaatccaaaggttgctgaagttatctatccaataatgCAGgttaatGGTGCTcattatCTTggcg
ttgatgttgcagttggGgggatggagcagagaaaaatacacatgttagcaagggagcttttaccaaaaaaggttgtttgtattcacaaccc
tgtcttaacgggtttggatggagaaggaaagatgagttcttcaaaagggaattttatagctgttgatgactctccagaagagattagggctaa
gataaagaaagcatactgcccagctggagttgttgaaggaaatccaataatggagatagctaaatacttccttgaatatccttaaccataaa
aaggccagaaaaatttggtggagatttgacagttaatagctatgaggagttagagagtttatttaaaaataaggaattgcatccaatggattt
aaaaaatgctgtagctgaagaacttataaagattttagagccaattagaaagagattataaGCGGCCGCgtttaaacggtctccag
cttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctgg
cggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatg
cgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatc
cccataatccttgttagCCTGCAGGATCCGGCCATGGCGGCCACCAGGTACCcaccggtttattgac
taccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattgacgatatgatca
gtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagggcgtatctgcgcagta
agatgcgccccgcattccggcggtagttcagcagggcagaacggcggactctaaatccgcatggcaggggttcaaatcccctccgccg
gaccaaattcgaaaagcctgctcaacgagcaggcttttttgcatgctcgagcagctcagggtcgaatttgctttcgaatttctgccattcatc
cgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatc
gcagtactgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcac
cttgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaact
cacccagggattggctgagacgaaaaacatattctcaataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgc
gaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaaca
agggtgaacactatcccatatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtga
ataaaggccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgag
caactgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttcct
tagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaa
cgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacag
gtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttc
tatcagctgtccctcctgttcagctactgacg

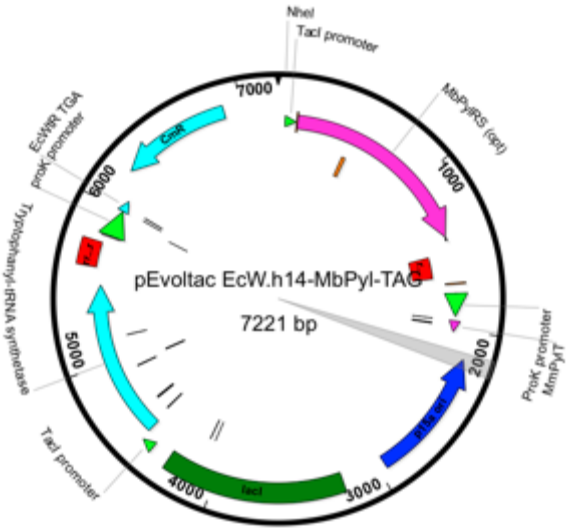
pEvoltac EcW.h14



gggtggtgcgtaacggcaaaagcaccgccggacatcagcgcctagcggagtgtatactggcttactatgttggcactgatgagggtgtca
gtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcact
gactcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatactt
aacagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcag
tggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac
cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg
cacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactg
gcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg
ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagag
caagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaa
atgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatg
tagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg
aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttctttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggc
gaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacg
cgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattt
atgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccaga
tgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa
cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatc
gccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc
cacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc
gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttcc
gggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagg
GAGCTGTTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGG
ATAACAATTTCACAaAGGAggtGCggccgcatgactaagcccatcgtttttagtggcgcacagccctcaggtgaatt
gaccattggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcga
tcaccgtgcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaa
gcaccatttttgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatg
acgcagtttaaagataaatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgct
gtatcaaactaatctggtaccggtgggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtat
ggcgagatctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaag
tctgacgataatcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacg
agccgccggtagttcgctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccagagcatcc
cagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgca
ggaacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttccc
gtacgctaaaagcggtgtacgaagcgattggttttgtggcgaagccgtaagcGGCCGCgtttaaacggtctccagcttggctgtttt
ggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagc
gcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagg
gaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatcct
tgttagCCTGCAgtgtgcttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattgacgatatgatcagtgc
acggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagggcgTATCTgcgcagtaa
gatgcgccccgcatttAGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTTCAAAACCGG
GTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAAATTCGAAAAGCCTGCTCAAC
GAGCAGGCTTTTTTGCATGctcgagcagctcagggtcgaatttgCtttcgaatttctgccattcatccgcttattatca
cttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgcagtactgttg

taattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttg
cgtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggat
tggctgagacgaaaaacatattctcaataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgta
gaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacact
atcccatatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccgga
taaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaa
tgcctcaaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaat
ctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgcc
aaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgc
aaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagctgtccctc
ctgttcagctactgacg

pEvoltac EcW.h14-MbPyl-TAG

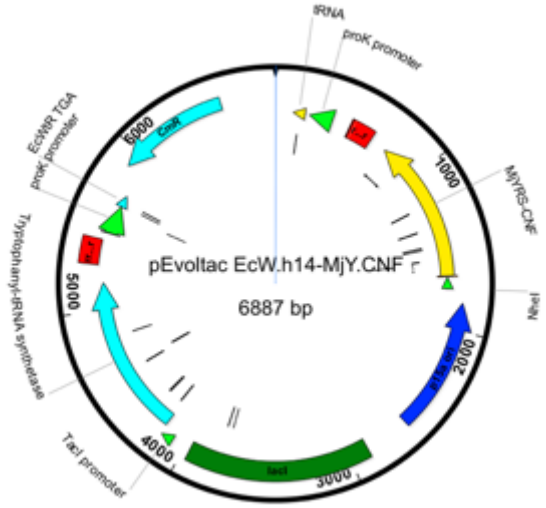


gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcGAGCTGTTGACAATTAATCATCGGCT
CGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACAaAGGAggtGCGGCC
GCATGGATAAAAAACCGCTGGACGTTCTGATCTCCGCTACGGGTCTGTGGAT
GAGCCGCACGGGTACGCTGCATAAAATTAAACACCACGAAGTGTCACGTTCGA
AAATCTATATCGAAATGGCGTGCGGTGATCATCTGGTGGTTAACAATAGCCGT
TCTTGTCGCACCGCGCGTGCCTTTCGCCATCACAAATACCGCAAAACGTGCAA
ACGTTGTCGCGTGTCAGATGAAGACATTAACAATTTCCTGACCCGTAGTACGG
AATCCAAAAACTCAGTGAAAGTTCGCGTCGTGAGTGCTCCGAAAGTTAAAAAA
GCGATGCCGAAAAGTGTCTCCCGTGCCCCGAAACCGCTGGAAAACTCAGTGTC
GGCAAAAGCTTCCACCAATACGAGCCGCTCTGTTCCGTCGCCGGCAAAAAGCA
CCCCGAACAGCTCTGTCCCGGCAAGCGCACCGGCACCGTCTCTGACGCGTAGT
CAGCTGGATCGCGTGGAAGCCCTGCTGTCCCCGGAAGACAAAATCTCACTGAA
TATGGCAAAACCGTTTCGTGAACTGGAACCGGAACTGGTTACCCGTCGCAAAA
ACGATTTCCAACGTCTGTATACGAATGATCGCGAAGACTACCTGGGTAAACTG
GAACGTGATATCACCAAATTTTTCGTGGACCGCGGCTTTCTGGAAATCAAATC
TCCGATTCTGATCCCGGCTGAATATGTTGAACGCATGGGTATTAACAATGATA
CCGAACTGAGTAAACAGATTTTTCGTGTGGATAAAAACCTGTGCCTGCGGCCG
ATGCTGGCACCGACGCTGTATAATTACCTGCGTAAACTGGATCGCATTCTGCC

GGGTCCGATTAAAATCTTTGAAGTGGGCCCGTGTTATCGTAAAGAATCGGAT
GGCAAAGAACACCTGGAAGAATTTACCATGGTTAACTTCTGCCAAATGGGCAG
CGGTTGTACGCGCGAAAATCTGGAAGCGCTGATCAAAGAATTCCTGGATTACC
TGGAAATCGACTTCGAAATCGTCGGTGATTCTTGCATGGTGTTTGGCGATACC
CTGGACATCATGCATGGTGACCTGGAACTGAGTTCCGCTGTTGTCGGTCCGG
TCAGCCTGGATCGTGAATGGGGCATTGACAAACCGTGGATCGGCGCGGGTTT
TGGCCTGGAACGCCTGCTGAAAGTTATGCACGGCTTCAAAAACATCAAACGTG
CGTCTCGCTCGGAATCGTATTACAACGGCATCTCAACCAATCTGTAATAAGCG
GCCGCgtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagc
ggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgc
cgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttg
tttgtgagctcccggtcatcaatcatccccataatccttgttagCCTGCAGGtaattccgcttcgcaacatgtgagcaccggtttatt
gactaccggaaGCAGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCA
GGCTCTCCCCGTGGAGGTAATAATTGACGATATGATCAGTGCACGGCTAACTA
AGCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATTAAGGGATT
GACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTGGAAACCTGATCA
TGTAGATCGAACGGACTtTAAATCcGTTCAGCCGGGTTAGATTCCCGGGGTTTC
CGCCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGCTCGAGCA
GCTCAGGGTCGAATTTGGGTcctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagt
gcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgc
tacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagg
gaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtgg
cgaaacccgacaggactataaagataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgt
cattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaa
ccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagca
gccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctc
caagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagagcaaga
gattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtag
cacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatgtagcac
ctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctc
aagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacct
gtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttcttttcaccagtga
gacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaat
cctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcag
cccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcat
ttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgcca
gccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctcc
acgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattag
tgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgca
ccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccg
cgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacg
cggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggga
aacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggc
gctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggGAG
CTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAA
CAATTTCACAaAGGAggtGCggccgcatgactaagcccatcgtttttagtggcgcacagccctcaggtgaattgaccatt
ggtaactacatgggtgcgctgcgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgt
gcgccaggatgcacagaagctgcgtaaagcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccat

ttttgttcagtcccacgtgccggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagt
ttaaagataaatctgcgcgttatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaa
actaatctggtaccggtgggtgaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgag
atctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacg
ataatcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgc
cggtagttcgctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaact
ggaaaaacagttcgaaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgc
tatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgcta
aaagcggtgtacgaagcgattggttttgtggcgaagccgtaagcGGCCGCgtttaaacggtctccagcttggctgttttggcggatg
agagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggt
cccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgcc
aggcatcaaataaaacgaaaggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgttagCC
TGCAgtgtgcttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaa
ctaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagggcgTATCTgcgcagtaagatgcgcc
ccgcatttAGGGGGCGTAGTTCAATTGGTAGAGCACCGGTCTTCAAAACCGGGTGtT
GGGAGTTCGAGTCTCTCCGCCCCTGCCAAATTCGAAAAGCCTGCTCAACGAGC
AGGCTTTTTTGCATGctcgagcagctcagggtcgaatttgCtttcgaatttctgccattcatccgcttattatcacttattca
ggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgcagtactgttgtaattcat
taagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgcgtataat
atttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctga
gacgaaaaacatattctcaataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaact
gccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatccca
tatcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaac
ttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctc
aaaatgttctttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgat
aactcaaaaaatacgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagt
tggcccagggcttcccggtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagt
gcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagctgtccctcctgtt
cagctactgacg

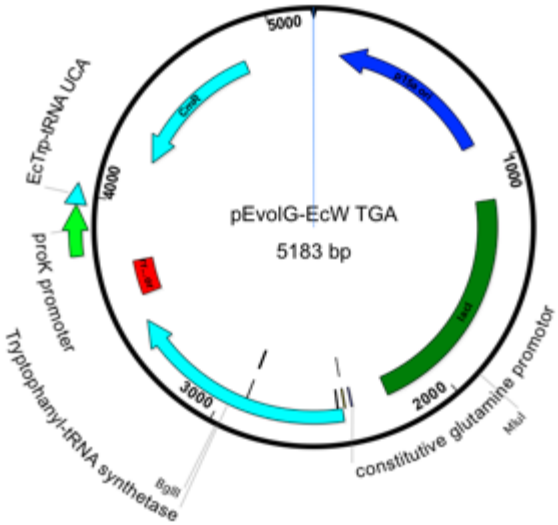
pEvoltac EcW.h14-MjY.CNF



gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctaggACCCAAATTCGACCCTGAGCTGCTCG
agcatgcaaaaaagcctgctcgttgagcaggcttttcgaatttggtccggcggaggggatttgaaccccctgccatgcggatttagagtccg
ccgttctgccctgctgaactaccgccggaatgcggggcgcatcttactgcgcagatacgccctcgtcaatcccttaatagcaaaatgccttt
tgatcggcgagaaagtcagcaggccgcttagttagccgtgcactgatcatatcgtcaattattacctccacggggagagcctgagcaaact
ggcctcaggcatttgagaagcacacggtcacactgcttccggtagtcaataaaccggtgGGTACCTGGTGGCCGCCA
TGGCCGGATCCTGCAGGctaacaaggattatggggatgattgatgaccgggagctcacaaacaaggcccagtctttcg
actgagcctttcgttttatttgatgcctggcagttccctactctcgcatggggagaccccacactaccatcggcgctacggcgtttcacttct
gagttcggcatggggtcaggtgggaccaccgcgctactgccgccaggcaaattctgttttatcagaccgcttctgcgttctgatttaatctg
tatcaggctgaaaatcttctctcatccgccaaaacagccaagctggagaccgtttaaacGCGGCCGCttataatctctttctaattg
gctctaaaatctttataagttcttcagctacagcatttttaaatccattggatgcaattccttatttttaaataaactctctaactcctcatagcta
ttaactgtcaaatctccaccaaatttttctggccttttatggttaaaggatattcaaggaagtatttagctatctccattattggatttccttcaa
caactccagctgggcagtatgctttctttatcttagccctaatctcttctggagagtcatcaacagctataaaattccctttgaagaactcatc
tttccttctccatccaaacccgttaagacagggttgtgaatacaaacaacctttttggtaaaagctcccttgctaacatgtgtatttttctctgc
tccatcccCccaactgcaacatcaacgccAAGataatgAGCACCattaacCTGcattattggatagataacttcagcaaccttt
ggattttcatcctctcttgctataagttccatactccttcttgctctttttaaggtagttttaaagccaatctatagacattcagtgtataatcctt
atcaagCATCCAttcacttccataaacatattttgcctttaaccccattgcttcaaaaactttttttgttataatctcctatttttctaatctcat
ccaactctccttttctggtttaaataggcATGtaaatcagccaaAACtataattatatcaaatccagcattttgtaaatcaatcatctttttt
atttggagataatgccctaaatgtattttaccacttggttcaaaacctatcAGagcagatttttcatctttttttaaaacctctcttaactcttcc
tcgctgataatttcagatgtgtttctctttatcatttcaaattcgtccatGCGGCCGCacCTCCTtTGTGAAATTGTT
ATCCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTGTCAACAGCTC
Gctagcggagtgtatactggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccg
gtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaa
tggcttacgaacggggcggagatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcggcaaagccgtttttcc
ataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcg
tttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgttatggccgcgtttgtctcattccac
gcctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaacccccccgttcagtccgaccgctgcgccttatccg
gtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttagaggagttagtcttga
agtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcggttcaaagagttggtagct
cagagaaccttcgaaaaaccgccctgcaaggcggtttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatc
atcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaat
tctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattct
catgttagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagt
gagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcg
cggggagaggcggtttgcgtattgggcgccagggtggtttttctttcaccagtgagacgggcaacagctgattgcccttcaccgcctggc
cctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatg
agctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcg
ccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactcca
gtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaactt
aatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaata
atactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcat
ccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctac
catcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactgg
aggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttc
cactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgaca
tcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcg
atggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggGAGCTGTTGACAATTAATCATCGGC
TCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACAaAGGAggtGCggccg
catgactaagcccatcgtttttagtggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtcagtgggtaaaca

tgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaagcgacgct
ggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttag
gctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaa
cgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctggtaccggtgggtgaagaccagaaacag
cacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgtttattccgaaatctg
gcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctggaagatc
cgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggc
gtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatc
tgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacgatgaagccttcctgcaa
caggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcgattggttttgtggcgaa
gccgtaagcGGCCGCgtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcaga
acgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaac
gccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaaga
ctgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgttagCCTGCAgtgtgcttctcaaatgcctgaggccagttt
gctcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaagg
cattttgctattaagggattgacgagggcgTATCTgcgcagtaagatgcgccccgcatttAGGGGGCGTAGTTCAAT
TGGTAGAGCACCGGTCTTCAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCC
CCTGCCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGctcgagcagct
cagggtcgaatttgCtttcgaatttctgccattcatccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataact
gccttaaaaaaattacgccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacaAacg
gcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgcgtataatatttgcccatggtgaaaacgggggcgaagaagttgt
ccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaacatattctcaataaaccctttagggaaa
taggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatga
aaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctttcattgccatacggaattc
cggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaat
atccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacggt
ggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcatta
tggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccag
gatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatg
atggtgtttttgaggtgctccagtggcttctgtttctatcagctgtccctcctgttcagctactgacg

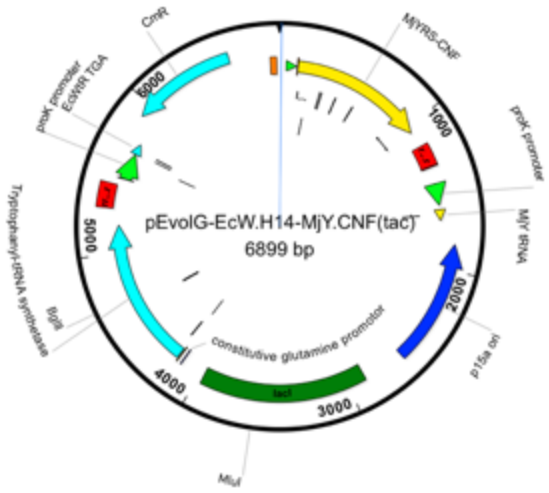
pEvolG EcW.h14



gggtggtgcgtaacggcaaaagcaccgccggacatcagcgctagcggagtgtatactggcttactatgttggcactgatgagggtgtcag
tgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcgctcactg
actcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaagatgccaggaagatactta
acagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacgctcaaatcagt
ggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggtttac
cggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatg
cacgaaccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactg
gcagcagccactggtaattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcg
ctcctccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttttcagag
caagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaa
atgtagcacctgaagtcagccccatacgatataagttgtaattctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatg
tagcacctgaagtcagccccatacgatataagttgtaattctcatgttagtcatgccccgcgcccaccggaaggagctgactgggttgaag
gctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcggg
aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggttttttctttcac
cagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggc
gaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacg
cgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcatt
cagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatattt
atgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccaga
tgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaa
cattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagatt
gtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatc
gccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgc
cacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgc
gggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttcc
gggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattgagt
tgaaggatccGCGGCCGCtcgggttgtcagcctgtcccgcttataagatcatacgccgttatacgttgtttacgctttgaggaatcc
caTATGatgactaagcccatcgtttttgctggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctgcgtcagtg
ggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaa
gcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgc
acagttaggctgggcactgaactgctatacctacttcggcgaagtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgcc
ggaacatgcacagttaggctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgt
tatgccgagaacatcaacgctggtctgtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgggtccttgtggt
gaagaccagaaacagcacctcgaactgagccgcgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagcc
gtttattccgaaatctggcgcgcgcgtaatgtcgctgctggagccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatc
ggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggtcactgactccgacgagccgccggtagttcgctacgatgtg
cagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccagagcatcccagaactggaaaaacagttcgaaggc
aagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcaggaacgctatcaccgtttccgcaacga
tgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaagcggtgtacgaagcg
attggttttgtggcgaagccgtaagcGGCCGCgtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagcctg
atacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccg
aactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaa
aggctcagtcgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgttagCCTGCAgtgtgcttctcaa
atgcctgaggccagtttgctcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgacttt
ctcgccgatcaaaaggcattttgctattaagggattgacgagggcgTATCTgcgcagtaagatgcgccccgcatttAGGGGC
GTAGTTCAATTGGTAGAGCACCGGTCTTCAAAACCGGGTGtTGGGAGTTCGA
GTCTCTCCGCCCCTGCCAAATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTTG
CATGctcgagcagctcagggtcgaatttgCtttcgaatttctgccattcatccgcttattatcacttattcaggcgtagcaaccaggcgt

ttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatg
gaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgcgtataatatttgcccatggtgaaaac
gggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctgagacgaaaaacatattctca
ataaaccctttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggta
ttcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctt
tcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtgcttatttttctttacgg
tctttaaaaaggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgcc
attgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgccc
ggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccg
gtatcaacagggacaccaggatttatttattctgcgaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgcc
aacttactgatttagtgtatgatggtgtttttgaggtgctccagtggcttctgtttctatcagctgtccctcctgttcagctactgacg

pEvolG EcW.h14-MjY.CNF(tac)

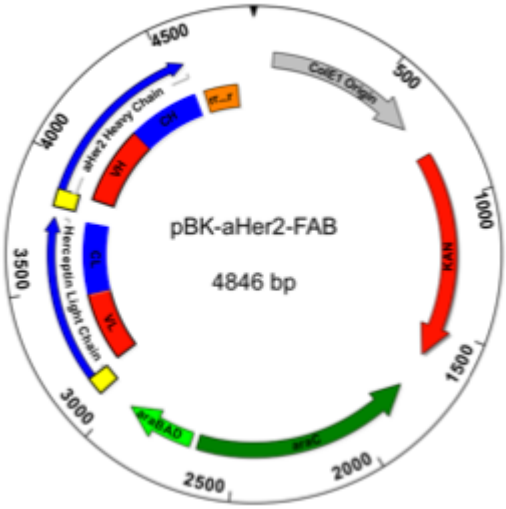


gggtggtgcgtaacggcaaaagcaccgccggacatcagcgCTAGCGAGCTGTTGACAATTAATCATCGGCTCG
TATAATGTGTGGAATTGTGAGCGGATAACAATTTCACAaAGGAggtGCGGCCGCatggacga
atttgaaatgataaagagaaacacatctgaaattatcagcgaggaagagttaagagaggttttaaaaaaagatgaaaaatctgctCTgataggttttgaa
ccaagtggtaaaatacatttagggcattatctccaaataaaaaagatgattgatttacaaaatgctggatttgatataattataGTTttggctgatttaC
ATgcctatttaaaccagaaaggagagttggatgagattagaaaaataggagattataacaaaaaagtttttgaagcaatggggttaaaggcaaaatatg
tttatggaagtgaaTGGATGcttgataaggattatacactgaatgtctatagattggctttaaaaactaccttaaaaagagcaagaaggagtatgga
acttatagcaagagaggatgaaaatccaaaggttgctgaagttatctatccaataatgCAGgttaatGGTGCTcattatCTTggcgttgatgt
tgcagttggGgggatggagcagagaaaaatacacatgttagcaagggagcttttaccaaaaaaggttgtttgtattcacaaccctgtcttaacgggtttg
gatggagaaggaaagatgagttcttcaaaagggaattttatagctgttgatgactctccagaagagattagggctaagataaagaaagcatactgcccag
ctggagttgttgaaggaaatccaataatggagatagctaaatacttccttgaatatcctttaaccataaaaaggccagaaaaatttggtggagatttgaca
gttaatagctatgaggagttagagagtttatttaaaaataaggaattgcatccaatggatttaaaaaatgctgtagctgaagaacttataaagattttagagc
caattagaaagagattataaGCGGCCGCgtttaaacggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaa
atcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgc
cgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggccttg
tttgtgagctcccggtcatcaatcatccccataatccttgttagCCTGCAGGATCCGGCCATGGCGGCCACCAGGTA
CCcaccggtttattgactaccggaagcagtgtgaccgtgtgcttctcaaatgcctgaggccagtttgctcaggctctccccgtggaggtaataattgac
gatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgctattaagggattgacgagggcgtatctgcgcagt
aagatgcgccccgcattccggcggtagttcagcagggcagaacggcggactctaaatccgcatggcaggggttcaaatcccctccgccggaccaaat
tcgaaaagcctgctcaacgagcaggcttttttgcatgctCGAGCAGCTCAGGGTCGAATTTGGGTcctagcggagtgtat

actggcttactatgttggcactgatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgcaccggtgcgtcagcagaatatgtgatac
aggatatattccgcttcctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcggagatttcctggaa
gatgccaggaagatacttaacagggaagtgagagggccgcggcaaagccgtttttccataggctccgccccctgacaagcatcacgaaatctgacg
ctcaaatcagtggtggcgaaacccgacaggactataaagataccaggcgtttccccctggcggctccctcgtgcgctctcctgttcctgcctttcggttt
accggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcagttcgctccaagctggactgtatgcacgaa
ccccccgttcagtccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactgg
taattgatttagaggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcctccaagccagttacctcgg
ttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcggtttttcgtttcagagcaagagattacgcgcagaccaaaacgatctc
aagaagatcatcttattaatcagataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaa
ttctcatgtttgacagcttatcatcgataagcttgcaatttatctcttcaaatgtagcacctgaagtcagccccatacgatataagttgtaattctcatgttagt
catgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaa
ttgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattg
ggcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgct
ggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtcc
gcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcat
tcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagc
cagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtc
gcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacag
caatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgc
ttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactg
gaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccacttttt
cccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactg
gtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctc
tcccttatgcgactcctgcattgagttgaaggatccGCGGCCGCtcgggttgtcagcctgtcccgcttataagatcatacgccgttatacgttgtt
tacgctttgaggaatcccaTATGatgactaagcccatcgtttttgctggcgcacagccctcaggtgaattgaccattggtaactacatgggtgcgctg
cgtcagtgggtaaacatgcaggatgactaccattgcatttactgtatcgttgaccaacacgcgatcaccgtgcgccaggatgcacagaagctgcgtaaa
gcgacgctggatacgctggccttgtatctggcttgtggtatcgatcctgagaaaagcaccatttttgttcagtcccacgtgccggaacatgcacagttag
gctgggcactgaactgctatacctacttcggcgaactgagtcgcatgacgcagtttaaagataaatctgcgcgttatgccgagaacatcaacgctggtct
gtttgactatccggtgctgatggcagcggacatcctgctgtatcaaactaatctgggtccttgtggtgaagaccagaaacagcacctcgaactgagccg
cgatattgcccagcgtttcaacgcgctgtatggcgagatctttaaggtgccggagccgtttattccgaaatctggcgcgcgcgtaatgtcgctgctggag
ccgaccaagaagatgtccaagtctgacgataatcgcaataacgttatcggcctgctggaagatccgaaatcggtagtgaagaaaatcaaacgtgcggt
cactgactccgacgagccgccggtagttcgctacgatgtgcagaacaaagcgggcgtttccaacctgttggatatcctttcagcggtaacgggccaga
gcatcccagaactggaaaaacagttcgaaggcaagatgtatggtcatctgaaaggtgaagtggctgatgccgtttccggtatgctgactgaattgcagg
aacgctatcaccgtttccgcaacgatgaagccttcctgcaacaggtgatgaaagatggcgcggaaaaagccagcgcgcacgcttcccgtacgctaaaa
gcggtgtacgaagcgattggttttgtggcgaagccgtaagcGGCCGCgtttaaacggtctccagcttggctgttttggcggatgagagaagatttt
cagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccga
actcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagt
cgaaagactgggccttgtttgtgagctcccggtcatcaatcatccccataatccttgttagCCTGCAgtgtgcttctcaaatgcctgaggccagttt
gctcaggctctccccgtggaggtaataattgacgatatgatcagtgcacggctaactaagcggcctgctgactttctcgccgatcaaaaggcattttgct
attaagggattgacgagggcgTATCTgcgcagtaagatgcgccccgcatttAGGGGGCGTAGTTCAATTGGTAGAGC
ACCGGTCTTCAAAACCGGGTGtTGGGAGTTCGAGTCTCTCCGCCCCTGCCAAATTCGA
AAAGCCTGCTCAACGAGCAGGCTTTTTTGCATGctcgagcagctcagggtcgaatttgCtttcgaatttctgccatt
catccgcttattatcacttattcaggcgtagcaaccaggcgtttaagggcaccaataactgccttaaaaaaattacgccccgccctgccactcatcgcag
tactgttgtaattcattaagcattctgccgacatggaagccatcacaAacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgc
gtataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcacccagggattggctgaga
cgaaaaacatattctcaataaacccttttagggaaataggccaggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatc
gtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccatatcaccagctcaccgtctt
tcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaa
aggccgtaatatccagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgccattgggatatatcaacg
gtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaatacgcccggtagtgatcttatttcattatggtga
aagttggaacctcttacgtgccgatcaacgtctcattttcgccaaaagttggcccagggcttcccggtatcaacagggacaccaggatttatttattctgc

gaagtgatcttccgtcacaggtatttattcggcgcaaagtgcgtcgggtgatgctgccaacttactgatttagtgtatgatggtgtttttgaggtgctccagt
ggcttctgtttctatcagctgtccctcctgttcagctactgacg

pBK-aHer2-FAB: Color scheme based on plasmid map image.



GAATTCctgcttttcttcgcgaattaattccgcttcgcacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccg
cgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggac
tataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctc
ccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacga
acccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagc
agccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaag
gacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggta
gcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctc
agtggaacgaaaactcacgttaagggattttggtcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagcca
tattcaacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcg
ggcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatga
tgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatg
gttactcaccactgcgatccccgggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcag
tgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaat
aacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataagcttttgccattc
tcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagt
cggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaaata
tggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaattggttgtaacactggca
gagcattacgctgacttgacgggacggcggctttgttgaataaatcgCCTAGGcgatataagttgtaattctcatgtttgacagcttat
catcgataagcttggtacccaattattgacaacttgacggctacatcattcacttttttcttcacaaccggcacggaactcgctcgggctggcc
ccggtgcatttttaaatacccgcgagaaatagagttgatcgtcaaaaccaacattgcgaccgacggtggcgataggcatccgggtggtgc
tcaaaagcagcttcgcctggctgatacgttggtcctcgcgccagcttaagacgctaatccctaactgctggcggaaaagatgtgacagac
gcgacggcgacaagcaaacatgctgtgcgacgctggcgatatcaaaattgctgtctgccaggtgatcgctgatgtactgacaagcctcgc
gtacccgattatccatcggtggatggagcgactcgttaatcgcttccatgcgccgcagtaacaattgctcaagcagatttatcgccagcag
ctccgaatagcgcccttcccccttgcccggcgttaatgatttgcccaaacaggtcgctgaaatgcggctggtgcgcttcatccgggcgaaa
gaaccccgtattggcaaatattgacggccagttaagccattcatgccagtaggcgcgcggacgaaagtaaacccactggtgataccattc
gcgagcctccggatgacgaccgtagtgatgaatctctcctggcgggaacagcaaaatatcactcggtcggcaaacaaattctcgtccctg
atttttcaccaccccctgaccgcgaatggtgagattgagaatataacctttcattcccagcggtcggtcgataaaaaaatcgagataaccgt
tggcctcaatcggcgttaaacccgccaccagatgggcattaaacgagtatcccggcagcaggggatcattttgcgcttcagccatactttt

catactcccgccattcagagaagaaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactggctcttctcgctaaccaa
accggtaaccccgcttattaaaagcattctgtaacaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagtgtctataatcacgg
cagaaaagtccacattgattatttgcacggcgtcacactttgctatgccatagcatttttatccataagattagcggatcctacctgacgctttt
tatcgcaactctctactgtttctccatacccgttttttgggctaaAgaaataattttgtttaactttaagaaggagaatacatcaactagtacg
caagttcacgtaaaaagggtatctagaggttgaggtgatttatgaaaaagaatatcgcatttcttcttgctagcatgttcgttttttctattgct
acaaacgcatacgctgacatccagatgacccagtctccatcctccctgtctgcatctgtaggagacagagtcaccatcacttgccgggcaa
gtcaggatgtgaataccgcggtcgcatggtatcagcagaaaccagggaaagcccctaagctcctgatctattctgcatccttcttgtatagt
ggggtcccatcaaggttcagtggcagtagatctgggacagatttcactctcaccatcagcagtctgcaacctgaagattttgcaacttacta
ctgtcaacagcattacactacccctccgacgttcggccaaggtaccaagcttgagatcaaacgaactgtggctgcaccatctgtcttcatct
tcccgccatctgatgagcagttgaaatctggaactgcctctgtcgtgtgcctgctgaataacttctatcccagagaggccaaagtacagtgg
aaggtggataacgccctccaatcgggtaactcccaggagagtgtcacagagcaggacagcaaggacagcacctacagcctcagcagca
ccctgacgctgagcaaagcagactacgagaaacacaaagtctacgcctgcgaagtcacccatcagggcctgtcctcgcccgtcacaaag
agcttcaacaggggagagtgttaagcctgcaggggatcctctagaggttgaggtgatttatgaaaaagaatatcgcatttcttcttgcatct
atgttcgttttttctattgctacaaacgcgtacgctgaggtgcagctggtggagtctggaggaggcttggtccagcctgggggtccctgag
actctcctgtgcagcctctgggttcaatattaaggacacttacatccactgggtccgccaggctccagggaaggggctggagtgggtcgc
acgtatttatcctaccaatggttacacacgctacgcagactccgtgaagggccgattcaccatctccgcagacacttccaagaacacggcg
tatcttcaaatgaacagcctgagagccgaggacacggccgtgtattactgttcgagatggggcggtgacggcttctatgccatggactact
ggggccaaggaaccctggtcaccgtctcctcagcctccaccaagggcccatcggtcttcccccctggcaccctcctccaagagcacctct
gggggcacagcggccctgggctgcctggtcaaggactacttccccgaaccggtgacggtgtcgtggaactcaggcgccctgaccagcg
gcgtgcacaccttcccggctgtcctacagtcctcaggactctactccctcagcagcgtggtgactgtgccctctagcagcttgggcaccca
gacctacatctgcaacgtgaatcacaagcccagcaacaccaaggtggacaagaaagttgagcccaaatcttgtgacaaaactcacacaT
AAaccgatgcGGccGCttgagagtcagctccttccggtgggcgtgcctggcggcagtagcgcggtggtcccacctgaccccatg
ccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagggaactgccaggcatcaaataaaac
gaaaggctcagtcgaaagactgggccttGTCGACcgaatttctgccattcatccgcttattatcacttattcaggcgtagcaaccagg
cgtttaagggcac