Fall 11-15-2016

DEVELOPMENT AND APPLICATION OF AN EXPANDED STREPTOMYCES GENETIC CODE

Jingxuan He
UNM Department of Chemistry and Chemical Biology

Jingxuan He
Candidate

Chemistry and Chemical Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Charles E. Melançon III, Chairperson

Debra Dunaway-Mariano

Jeremy S. Edwards

Fu-sen Liang

DEVELOPMENT AND APPLICATION OF AN EXPANDED *STREPTOMYCES* GENETIC CODE

by

JINGXUAN HE

B.S., Biotechnology, Beijing University of Chemical Technology, 2011

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Chemistry

The University of New Mexico
Albuquerque, New Mexico

December 2016

ACKNOLEDGMENTS

Foremost, I would like to express my sincere appreciation and gratitude to my advisor Prof. Charles E. Melançon III for the continuous support and encouragement of my Ph.D. study and research, for his patience, motivation, enthusiasm, and immense knowledge. Dr. Melançon is an awesome person, serious scientist, patient teacher, and helpful Ph.D. advisor. It is my great honor to explore the chemical biology research area under his guidance.

Besides my advisor, I would like to thank the rest of my committee members: Dr. Debra Dunaway-Mariano, Dr. Jeremy S. Edwards, and Dr. Fu-sen Liang, for their precious time and insightful comments.

My sincere thanks also goes to my previous and present lab mates; Dr. Yasushi Ogasawara, Dr. Ben Yackley, Briana Van Treeck, Hân B. Nguyễn, Xuechen Zhu, Shijie Huang, and Wubin Gao. I really appreciate all of their help and support of my research.

Last but not the least, I want deliver my thanks to my parents, my wife, my sisters, and my aunt. It is your love, support and encouragement that help me go through the whole Ph.D. study.

DEVELOPMENT AND APPLICATION OF AN EXPANDED *STREPTOMYCES* GENETIC CODE

by

Jingxuan He

B.S., Biotechnology, Beijing University of Chemical Technology, 2011

Ph.D., Chemistry, University of New Mexico, 2016

ABSTRACT

Actinobacteria, especially those of genus *Streptomyces*, are a prominent source of bioactive natural products. The ability to site-specifically incorporate unnatural amino acids (UAAs) into natural product biosynthetic enzymes and ribosomally derived peptides in these organisms would constitute a valuable tool for drug discovery and development. The work described in this dissertation focuses on development and application of an expanded *Streptomyces* genetic code, including development of UAA incorporation systems based on amber suppression and sense codon reassignment, structural diversification of the model thiopeptide natural product thiostrepton using UAAs, and mapping protein-protein interactions in type II polyketide biosynthetic enzymes using photocrosslinking UAAs.

First, we developed an amber suppression-based system of site-specific incorporation of *p*-iodo-L-phenylalanine (pIPhe) and *p*-azido-L-phenylalanine (pAzPhe) into superfolder GFP (sfGFP) in the model natural product producer *Streptomyces venezuelae* ATCC 15439.

Next, the rare leucine codon TTA was reassigned to encode pIPhe and *p*-benzoyl-L-phenylalanine (pBpa) in *S. coelicolor* J1681 (*ΔbldA*), in which the unique $tRNA^{Leu}_{UAA}$ (*bldA*) that recognizes the TTA codon was deleted. In the *S. venezuelae ΔbldA* strain, we achieved 20-fold higher yields of UAA containing protein using the TTA reassignment system compared to the amber suppression-based system; and were able to incorporate up to 10 scattered or 5 tandem UAAs in a single protein using TTA reassignment.

Finally, we have carried out preliminary work on two applications. In the first, we constructed and tested functionality of a system designed to incorporate pAzPhe into the actinorhodin ketosynthase β (KSβ) in *S. coelicolor* J1681 to interrogate protein-protein interactions in actinorhodin biosynthesis. In the second, we have begun developing a system for incorporation of UAAs into thiostrepton in the native producer *Streptomyces laurentii* ATCC 31255. Preliminary results confirm the functionality of amber suppression system in *S. laurentii*; and demonstrated development of a TipA-based fluorescent biosensor for detecting thiopeptide antibiotics in *S. venezuelae*. Work on these two applications has laid the foundation for development of tools to structurally diversify the ribosomally synthesized peptides and to address questions related to natural product biosynthesis and mechanism of action that are relevant to drug discovery and development.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

5’-UTR	5' untranslated region
aaRS	Aminoacyl-tRNA synthetase
CRISPR	Clustered regularly interspaced short palindromic repeats
DBCO-biotin	dibenzocyclooctyne-PEG4-biotin
eGFP	enhanced GFP
ESI-MS	Electrospray ionization mass spectrometry
LB	Luria-Bertani
MAGE	Multiplex automated genomic engineering
Mb-pyrRS	*M. bakeri* pyrrolysyl-tRNA synthetase
$MjtRNA^{Tyr}_{CUA}$	*M. jannaschii* tyrosyl amber suppressor tRNA
$MjtRNA^{Tyr}_{UAA}$	*M. jannaschii* tyrosyl TTA suppressor tRNA
MjTyrRS	*M. jannaschii* tyrosyl-tRNA synthetase
ORF	Open reading frame
pAzPhe	*p*-azido-L-phenylalanine
pAzPheRS	*p*-azido-L-phenylalaninyl-tRNA synthetase
pBpa	*p*-benzoyl-L-phenylalanine
pBpaRS	*p*-benzoyl-L-phenylalaninyl-tRNA synthetase
PCR	Polymerase chain reaction
pIPhe	*p*-iodo-L-phenylalanine
pIPheRS	*p*-iodo-L-phenylalaninyl-tRNA synthetase
RBS	Ribosome binding site
RiPPs	Ribosomally synthesized and post-translationally modified

sfGFP	superfolder GFP
$tRNA_{CUA}$	Amber suppressor tRNA
TSB	Tryptic soy broth
UAA	Unnatural amino acid

Chapter 1. Background and Significance

1. Introduction

Streptomyces, most notably high G+C Gram positive Actinobacteria, are well known for their ability to produce structurally diverse bioactive natural products, including ribosomally synthesized and post-translationally modified peptides (RIPPs), nonribosomal peptides (NRP), polyketides, terpenoids, and other classes of molecules. Natural products are structurally complex, naturally occurring metabolites that are not required for host survival, and are made by complex enzymatic systems.

Biomolecular interactions are critical both to proper functioning of multienzyme complexes involved in natural product biosynthesis, and to natural product bioactivity, which is derived from the ability of a natural product to bind to its target with high affinity. In the former, protein-protein interactions within multienzyme complexes are critical for efficient biosynthesis of natural products in living cells through transfer of biosynthetic intermediates between active sites as part of multi-step reaction pathways. A method to identify and map these protein-protein interactions would provide insights into how natural product multi-enzyme complexes functions. These insights would in turn guide rational engineering of natural product biosynthetic pathways to produce natural product analogs for drug development applications. In the latter, a method to install new chemical functionalities site-specifically into RiPP natural products would facilitate both structural diversification of these compounds for drug development, and interrogation of drug-target interactions to better understand drug mechanisms of action.

In light of the critical role that biomolecular interactions play in proper

functioning of natural product biosynthetic enzymes and in the interactions of natural products with their targets, development of a chemical biology tool to explore these interactions would benefit drug discovery endeavors. Site-specific unnatural amino acid (UAA) incorporation is a valuable tool for the study of protein-protein and peptide-target interactions. Analogous to CRISPR/Cas, a robust and precise genome editing tool, site-specific UAA incorporation allows introduction of novel chemical moieties, including those capable of photcrosslinking and biorthogonal coupling, site-specifically into a protein or peptide.[1,2]

To apply site-specific UAA incorporation technology to *Streptomyces* natural product research, two routes can be envisioned. The first is heterologous expression of natural product biosynthetic gene clusters in a host for which an UAA incorporation system is already available, such as *E. coli*. However, due to differences in codon usage and cell physiology between *Streptomyces* and *E. coli* and complex and poorly understood regulatory mechanisms imbedded in natural product biosynthetic gene clusters, expression of entire *Streptomyces* natural product biosynthetic pathways in *E. coli* is extremely challenging and rarely successful.[3,4] The second route would be development of an UAA incorporation system for use in *Streptomyces* species.

To develop an UAA incorporation system in *Streptomyces*, an evolved orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pair must be expressed in functional form in the host. There are two previously developed aaRS/tRNA pairs that are commonly used in bacteria: the *Methanocaldococcus jannaschii* tyrosyl synthetase/tRNA (MjTyrRS/tRNA) pair[5] and the *Methanosarcina barkeri* pyrrolysyl synthetase/tRNA (MbPylRS/tRNA)[6,7] pairs. We have pursued development of *Streptomyces* UAA systems

employing both the MjTyrRS/tRNA and MbPylRS/tRNA pairs.

In the canonical genetic code, 61 codons encode amino acids and 3 nonsense (stop) codons act as translation termination signals. Evolved orthogonal aaRS are able to aminoacylate their cognate tRNAs with specific UAAs; and the resulting aminoacyl tRNAs decode specific mRNA codons during translation via codon-anticodon interactions to incorporate the UAA. Stop codons, most commonly the amber stop codon (TAG), are widely used to incorporate UAAs. In addition to stop codon suppression, sense codon reassignment and quadruplet codon (frameshift) suppression are also employed. [1,2]

In the work described in Chapter 2 of this dissertation, amber suppression based UAA incorporation systems were developed for *Streptomyces* using MjTyrRS/$tRNA^{Tyr}_{CUA}$ and MbPylRS/$tRNA^{Pyl}_{CUA}$ pairs. In addition, a *Streptomyces* TTA sense codon reassignment system using MjTyrRS/$tRNA^{Tyr}_{UAA}$ was also developed, as described in Chapter 3.

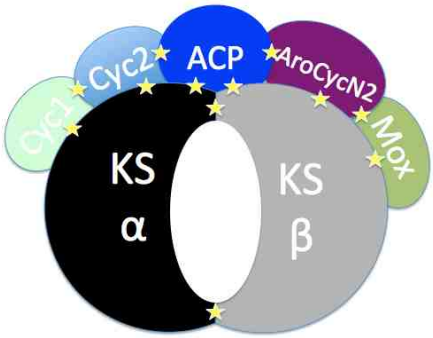


Figure 1-1. Putative type II polyketide biosynthesis multienzyme complex. Stars represent potential sites along protein-protein interfaces for installation of photocrosslinking UAAs represent the photocrosslinker UAAs

Among the common types of natural product biosynthetic systems, our lab is most interested in those responsible for production of type II polyketides. Indirect

evidence from previous studies of type II polyketide systems suggests that interactions exist among the ketosynthase α/β heterodimer and immediate tailoring enzymes (ketoreductase, cyclases) during type II polyketide biosynthesis (Figure 1-1). However, direct evidence for these interactions has not yet been obtained except in the case of interactions between the subunits of the actinorhodin ketosynthase α/β (KSα/β) heterodimer known from the KSα/β co-crystal structure.[8] Development of a *Streptomyces* UAA system, would allow site-specific incorporation of photocrosslinkers into ketosynthase α or β to covalently trap immediate tailoring enzymes *in vivo*. These covalently trapped protein complexes could even be purified and subjected to protein crystallization, which would provide detailed information on the structures of these complexes.

Ribosomally synthesized and post-translationally modified peptides (RiPPs) constitute a large family of natural products made by a variety of organisms and possessing a wide range of bioactivities. Cyanobacterial RiPP analogs containing UAAs have been successfully heterologously produced in *E. coli*. More recently, UAAs were genetically incorporated into the thiopeptide thiocillin through development of suppression UAA incorporation system in the thiocillin producing strain *Bacillus cereus*.[9] However, only 4.5% of sequenced *Bacillus* genomes contain putative thiopeptide biosynthetic gene clusters, whereas 39% of sequenced *Streptomyces* genomes contain such gene clusters.[10] Because of the challenges associated with heterologous expression of *Streptomyces* natural product gene clusters in *E. coli*, development of a *Streptomyces* UAA incorporation system is the most practical strategy to manipulate the large pool of thiopeptides produced by *Streptomyces*.

Several genetically manipulable *Streptomyces* strains, including *Streptomyces coelicolor* A3(2) and its mutants,[11] *Streptomyces lividans*, *Streptomyces albus* J1074, *Streptomyces venezuelae* ATCC 15439,[12] and *Streptomyces avermitilis* SUKA mutants[13] are commonly used as hosts for heterologous expression of natural product enzymes and pathways from diverse Actinobacteria[14-16] and for development of *Streptomyces* genetic tools. *S. venezuelae* ATCC 15439 is among the most rapidly growing model *Streptomyces* hosts (doubling time of ~1 h[12]), grows in a monodispersed manner in liquid media, and can be genetically manipulated via introduction of plasmid DNA using either conjugation or protoplast transformation.

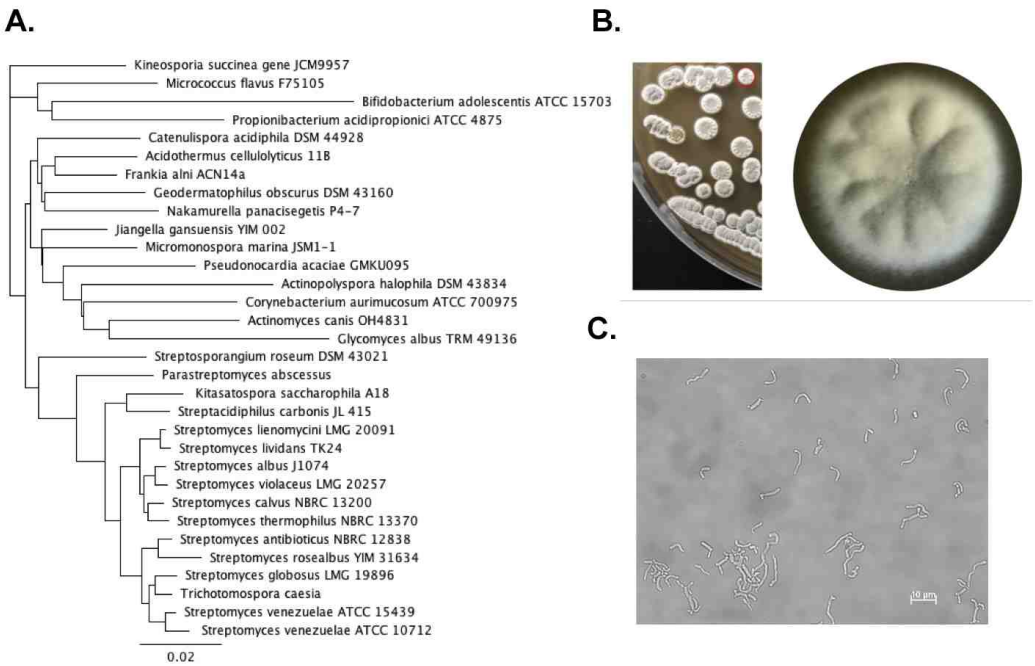


Figure 1-2. (a) Phylogenetic tree of select Actinobacteria showing the position of *S. venezuelae* ATCC 15439; (b) Images of *S. venezuelae* ATCC 15439 colonies; (c) Microscope image of monodispered *S. venezuelae* ATCC 15439 cells.

One unique genetic feature *Streptomyces* possess is that most secondary

metabolites production is regulated by the *bldA* gene using the rare TTA codon. The *bldA* gene encodes the unique $tRNA^{Leu}_{UAA}$[17,18] which decodes the TTA codon. TTA containing genes in *Streptomyces* mostly encode non-essential regulatory proteins or proteins of unknown function. Because of the importance of activating cryptic natural product biosynthetic gene clusters for drug discovery, researchers recently used a constitutively expressed *bldA* gene to globally activate the expression of natural product biosynthetic gene clusters.[19,20] In light of the unique features of *bldA* in *Streptomyces*, the TTA codon is an excellent candidate for reassignment to encode UAAs using the MjTyrRS/$tRNA^{Tyr}_{UAA}$ pair.

2. **Existing genetic tools available in *Streptomyces***

Streptomyces is not nearly as well-researched as *E. coli*. There are only a limited number of well-characterized genetic elements (promoters, ribosome binding sites (RBSs), and terminators) available for use in *Streptomyces* genetic engineering.

The SF14[21] and ermEp[22] promoters are two of the most commonly used constitutive promoters in *Streptomyces*; and are widely used to drive the expression of individual genes. For expression of entire gene clusters, researchers prefer to use native promoters found within the gene cluster or clone uncharacterized intergenic DNA sequences predicted to harbor promoter regions. For example, in the actinorhodin biosynthetic gene cluster, an intergenic region upstream of the *actI-ORF1* (ketosynthase α) gene, and predicted to harbor a bidirectional promoter, was used together with the *actII-ORF4* gene encoding a positive transcriptional regulator, to drive the expression of actinorhodin gene clusters.[23] ActII-ORF4 is a *bldA* dependent regulatory protein harboring a TTA codon at position 5. ActII-ORF4 is expressed after the *bldA* tRNA

transcription and activates the promoters upstream of *actI-ORF1*, resulting in expression of the actinorhodin gene cluster.[24]

A recently employed strategy to discover biotechnologically useful *Streptomyces* promoters was to clone and assay (through use of a reporter gene) a collection of DNA sequences found upstream of housekeeping genes in the *Streptomyces griseus* genome.[25] It was presumed that these sequences were likely to contain strong constitutive promoters to express housekeeping genes and maintain essential cellular functions. Using this approach, a number of new promoters of varying strengths were identified.

The *tipA* promoter (ptipA) from *S. coelicolor* and *S. lividans* is a widely used *Streptomyces* inducible promoter.[26] An operator sequence within ptipA is the binding site for the TipAL protein, which is a negative transcriptional regulator that is responsive to thiopeptides such as thiostrepton. Upon thiopeptide binding, TipAL undergoes a conformational change that unwinds the operator sequence and increases the affinity of RNA polymerase for ptipA more than 10-fold, which activates transcription of downstream genes.[27] ptipA is typically used in hosts such as *S. coelicolor* and *S. lividans* that naturally harbor the *tipAL* gene, and can only be used in hosts lacking *tipAL* if this gene is supplied *in trans*. The ptipA/*tipAL*-based thiopeptide sensor strain described in Chapter 4 is an example of use of this promoter/repressor system in a host that naturally lacks *tipAL*.

In a recent example of *Streptomyces* synthetic biology, libraries of promoters and RBSs were constructed and screened in a high-throughput format to identify a set of constitutive promoters with a range of strengths. Seven of these synthetic promoters with varying strengths were used to activate a cryptic lycopene biosynthetic gene cluster and

overproduce lycopene in *Streptomyces avermitilis*.[28] These findings provide a collection of new, characterized *Streptomyces* promoters.

In addition to constitutive and inducible *Streptomyces* promoters, the availability of *Streptomyces* genome editing technology is another important advance that facilitated the work presented here. Traditional methods for introducing mutations, such as deletions and point mutations, into *Streptomyces* genomes, is based on homologous recombination between plasmid and genomic sequences.[29] The drawback of this method is that it is time consuming and labor intensive due to low and variable recombination efficiency.

The recently developed CRISPR/Cas system has revolutionized genome editing in a wide range of model hosts. This system is derived from a prokaryotic "immune system" that, in its natural context, confers resistance to foreign genetic elements through introduction of double strand breaks in foreign DNA. CRISPR small guide RNAs (sgRNAs) recruit the nuclease Cas9 to specific DNA regions, which are then cut by Cas9. The DNA lesion is then repaired by the host recombination machinery using a DNA template carrying the desired mutation. CRISPR/Cas genome editing systems have been developed in *E. coli*,[30] *Saccharomyces cerevisiae*,[31] and mammalian cells[32]. Recently four groups have reported construction of functional, yet slightly different, CRISPR/Cas systems in *Streptomyces*. Each system has distinct advantages and disadvantages, which are discussed below.

The first reported *Streptomyces* CRISPR/Cas system employed a pSG5 temperature sensitive origin, and codon optimized Cas9 and sgRNA expressed using native promoters derived from the upstream regions of *Streptomyces* housekeeping genes. The spacer sequence, which is the targeting region of the sgRNA, can be easily

introduced into the plasmid using golden gate assembly[33]. The temperature sensitive origin facilitates elimination of plasmid after genome editing is complete. Using this system, genome deletions of 20 bp to 30 kilobase pairs (kb) were obtained in three different *Streptomyces* species with efficiencies of 70% to 100%.[34] The plasmid and strategy employed to perform genome editing of *S. venezuelae* in the work described here are derived from this first reported *Streptomyces* CRISPR/Cas system. The second report[35] uses a similar system as the first, but demonstrates the ability to install a specific point mutation. This was achieved by using an sgRNA with the site of the mutation lying in the protospacer adjacent motif (PAM), a 3 bp sequence immediately 3' to the protospacer (the targeting portion of the sgRNA). The PAM must have sequence NGG for Cas9 to recognize and act on the target DNA. After Cas9-dependent DNA cleavage and homology-dependent repair using the mutant DNA, the mutant sequence is resistant to cleavage by the CRISPR/Cas complex, and the mutation is therefore stably incorporated. In the work presented here, a CRISPR/Cas plasmid was designed and constructed to install a TGA stop codon into the *Streptomyces laurentii* ATCC 31255 genome using this idea with some modifications.

The second and third reports used the inducible promoter ptipA (discussed above) to control the expression of Cas9 gene, minimizing the toxicity of CRISPR/Cas system to transformed cells. In this way, transformation efficiency would be higher than with the system employing a constitutive promoter. However, the requirement of ptipA for the TipAL activator protein, which is not present in the CRISPR/Cas vector, limits the use of this system to *Streptomyces* hosts harboring a genomic copy of.

In the fourth report,[36] the CRISPR/Cas system was further optimized by inclusion

of the counterselectable marker CodA(sm), the D314A mutant of cytosine deaminase, in the vector. The CodA(sm) converts 5-fluorocytosine (5FC) added to the culture into 5-fluorocytosine (5FC), a highly toxic compound.[37] This counterselectable system facilitates identification of mutants that have lost the plasmid after the genome editing process.

3. **Unnatural Amino Acid Incorporation Systems**

Unnatural amino acid (UAA) incorporation systems rely on an evolved orthogonal aminoacyl-tRNA synthetase (aaRS) and its cognate orthogonal tRNA. In presence of the UAA, which are fed to the cells, the orthogonal aaRS can aminoacylate its cognate tRNA with a specific UAA. The UAA charged tRNA is then capable of decoding the reassigned codon during translation, resulting in site-specific incorporation of the UAA into the protein of interest (Figure 1-3).

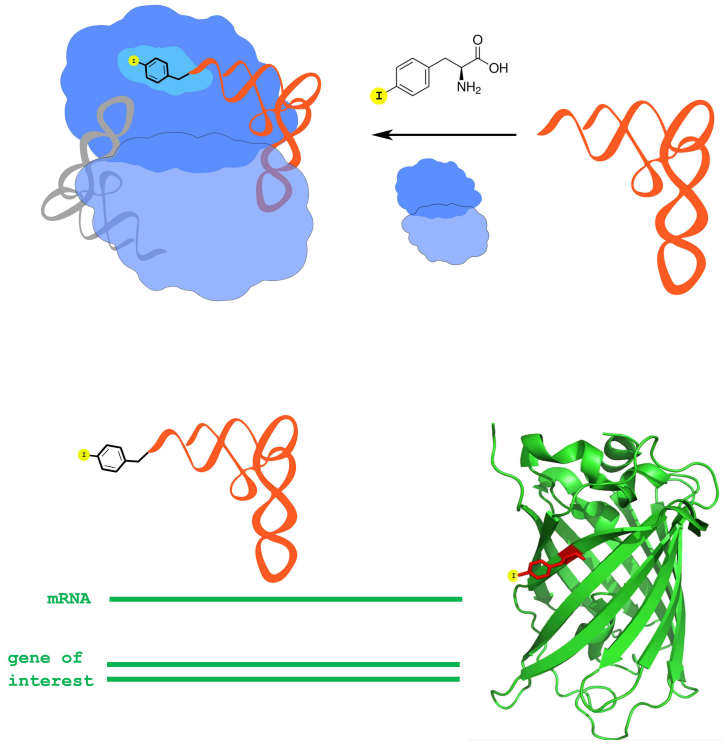


Figure 1-3. Schematic overview of unnatural amino acid incorporation systems

Of the 64 triplet codons, there are three stop codons, which are not normally recognized by tRNA. They are the amber (TAG), ochre (TAA), and opal (TGA) stop codons. Normally protein synthesis is terminated at stop codons by release factors. Co-expression of a nonsense suppressor tRNA (a tRNA with its anticodon mutated to recognize a stop codon) that is charged by the engineered aaRS allows a chosen stop codon to be decoded by the aminoacyl suppressor tRNA during translation. The amber (TAG) stop codon is the most commonly reassigned stop codon in UAA incorporation systems due to the infrequency with which it occurs in the genomes of many model organisms.

Due to competition with release factors, UAA incorporation via nonsense suppression suffers low efficiency and results in production of significant amounts of truncated protein. In a recent report, Release Factor 1 (RF1), which is responsible for translation termination at amber stop codons, was successfully deleted in *E. coli*, and was complemented with a "fixed" mutant copy of Release Factor 2 (RF2) that is capable of recognizing the amber stop efficiently enough to support cell growth.[38] This mutant host showed improved amber suppression efficiency that allowed incorporation of multiple UAAs into a single protein.

In another study, multiplex automated genomic engineering (MAGE) was used to systematically replace all 314 amber stop codons in the *E. coli* genome with alternative stop codons, thereby allowing deletion of RF1 and converting the TAG codon to a "free" codon that could be reassigned for UAA incorporation.[39] This "genome-recoded" organism also showed improved amber suppression efficiency that allowed incorporation of multiple UAAs into a single protein.

Sense codons can also be reassigned to encode UAAs. However, this strategy typically lacks fidelity due to competition with endogenous tRNAs that also recognize the reassigned codon. Sense codon reassignment also causes loss in host fitness due to proteome-wide UAA incorporation. A recent study achieved high fidelity UAA incorporation in response to the rare codon AGG by deleting the $tRNA^{Arg}_{CCU}$ whose function was sufficiently compensated for by the near cognate $tRNA^{Arg}_{UCU}$. However, incorporation of the UAA in response to endogenous AGG codons was sufficiently toxic to the host that expression of recombinant protein could be induced for no more than 3 h.[40]

Frameshift suppression via use of quadruplet codons has also been used in genetic code expansion work. Although the natural ribosome is capable of recognizing tRNAs with quadruplet anticodons, translation of quadruplet codons is relatively inefficient, and results in low UAA incorporation efficiency. To overcome this problem, orthogonal ribosome technology (co-expression of a 16*S* rRNA with mutant anti-Shine-Dalgarno sequence that is complementary to an engineered Shine-Dalgarno sequence controlling translation of the gene of interest) was employed to evolve ribosomes with mutant 16*S* rRNA that were more efficient at decoding quadruplet codons than wild-type ribosomes.[41]

4. **Summary and Thesis Statement**

Because of the important role that biomolecular interactions (protein-protein and peptide-drug interactions) play both in proper functioning of natural product biosynthetic multienzyme complexes and in the interactions of peptide drugs with their targets, systems for site-specific incorporation of UAAs in *Streptomyces* are valuable tools for

studying natural product biosynthesis and mechanism of action.

In the work described here, we employed both the MjTyrRS/$tRNA^{Tyr}_{CUA}$ and MbPylRS/$tRNA^{Pyl}_{CUA}$ pairs to expand the *Streptomyces* genetic code. Due to the distinct genetic differences between *Streptomyces* and *E. coli*, the host in which both aaRS/tRNA pairs were originally evolved, the aaRS genes were codon optimized and *Streptomyces* derived promoters were used to drive the expression of aaRS/tRNA pairs. Through this work, we successfully demonstrated site-specific genetic incorporation of *p*-iodo-L-phenylalanine and *p*-azido-L-phenylalanine into the eGFP and sfGFP proteins in *S. venezuelae* with good to excellent fidelity and with sufficient yields for most applications. Development of this amber suppression-based UAA incorporation system is described in Chapter 2.

In addition, with the help of a unique gift from nature, we are able to reassign the TTA sense codon to encode UAAs with dramatically improved efficiency (~20-fold higher compared to the amber suppression-based system for incorporation of a single UAA). The TTA reassignment system was first demonstrated in *S. coelicolor* J1681 (*ΔbldA*); and was then improved in the faster growing strain *S. venezuelae* using a *bldA* deletion mutant (Sven 624) constructed using the *Streptomyces* CRISPR/Cas system. In Sven 624, up to 10 scattered or 5 tandem TTA codons can be decoded in a single protein using the TTA reassignment system. Such incorporation efficiency had not been previously achieved using nonsense, sense, or frameshift suppression. This advanced *Streptomyces* UAA incorporation system is described in the Chapter 3.

Preliminary work on two applications of interest to our lab is described in Chapter 4. In the first application, the expanded *Streptomyces* genetic code was applied to attempt

to study protein-protein interactions in the actinorhodin biosynthetic multienzyme complex by incorporating photocrosslinking UAA pAzPhe into the target protein ketosynthase β. In the second application, replacement of amino acid residues of thiostrepton, a highly modified RiPP produced by *Streptomyces laurentii* ATCC 31255, with UAAs is ongoing.

Our work on development and application of an expanded *Streptomyces* genetic code will serve as the foundation for using site-specific unnatural amino acid incorporation systems to explore mechanisms of natural product biosynthesis, to structurally diversify peptide natural products, and to carry out other important applications.

5. References

1. Xiao, H., & Schultz, P. G. (2016). At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harbor Perspectives in Biology*, *8*(9), a023945.
2. Dumas, A., Lercher, L., Spicer, C. D., & Davis, B. G. (2015). Designing logical codon reassignment–Expanding the chemistry in biology. *Chemical Science*, *6*(1), 50-69.
3. Li, J., & Neubauer, P. (2014). *Escherichia coli* as a cell factory for heterologous production of nonribosomal peptides and polyketides. *New biotechnology*, *31*(6), 579-585.
4. Yuzawa, S., Kim, W., Katz, L., & Keasling, J. D. (2012). Heterologous production of polyketides by modular type I polyketide synthases in *Escherichia coli*. *Current opinion in biotechnology*, *23*(5), 727-735.

5. Wang, L., Brock, A., Herberich, B., & Schultz, P. G. (2001). Expanding the genetic code of *Escherichia coli*. *Science*, *292*(5516), 498-500.
6. Srinivasan, G., James, C. M., & Krzycki, J. A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science*, *296*(5572), 1459-1462.
7. Wan, W., Tharp, J. M., & Liu, W. R. (2014). Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1844*(6), 1059-1070.
8. Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C., & Stroud, R. M. (2004). An antibiotic factory caught in action. *Nature structural & molecular biology*, *11*(9), 888-893.
9. Luo, X., Zambaldo, C., Liu, T., Zhang, Y., Xuan, W., Wang, C., ... & Schultz, P. G. (2016). Recombinant thiopeptides containing noncanonical amino acids. *Proceedings of the National Academy of Sciences*, 201602733.
10. Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., & Weber, T. (2013). antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic acids research*, gkt449.
11. Gomez-Escribano, J. P., & Bibb, M. J. (2014). Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *Journal of industrial microbiology & biotechnology*, *41*(2), 425-431.

12. Kim, E. J., Yang, I., & Yoon, Y. J. (2015). Developing *Streptomyces venezuelae* as a cell factory for the production of small molecules used in drug discovery. *Archives of pharmacal research*, *38*(9), 1606-1616.
13. Ikeda, H., Shin-ya, K., & Omura, S. (2014). Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *Journal of industrial microbiology & biotechnology*, *41*(2), 233-250.
14. Ongley, S. E., Bian, X., Neilan, B. A., & Müller, R. (2013). Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Natural product reports*, *30*(8), 1121-1138.
15. Zhang, H., Boghigian, B. A., Armando, J., & Pfeifer, B. A. (2011). Methods and options for the heterologous production of complex natural products. *Natural product reports*, *28*(1), 125-151.
16. Baltz, R. H. (2010). *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *Journal of industrial microbiology & biotechnology*, *37*(8), 759-772.
17. Leskiw, B. K., Mah, R., Lawlor, E. J., & Chater, K. F. (1993). Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3 (2). *Journal of bacteriology*, *175*(7), 1995-2005.
18. Pettersson, B. M., & Kirsebom, L. A. (2011). tRNA accumulation and suppression of the *bldA* phenotype during development in *Streptomyces coelicolor*. *Molecular microbiology*, *79*(6), 1602-1614.

19. Kalan, L., Gessner, A., Thaker, M. N., Waglechner, N., Zhu, X., Szawiola, A., ... & Zechel, D. L. (2013). A cryptic polyene biosynthetic gene cluster in *Streptomyces calvus* is expressed upon complementation with a functional *bldA* gene. *Chemistry & biology*, *20*(10), 1214-1224.

20. Gessner, A., Heitzler, T., Zhang, S., Klaus, C., Murillo, R., Zhao, H., ... & Bechthold, A. (2015). Changing biosynthetic profiles by expressing *bldA* in *Streptomyces* strains. *ChemBioChem*, *16*(15), 2244-2252.

21. Labes, G., Bibb, M., & Wohlleben, W. (1997). Isolation and characterization of a strong promoter element from the *Streptomyces ghanaensis* phage 119 using the gentamicin resistance gene (aacC1) of Tn1696 as reporter. *Microbiology*, *143*(5), 1503-1512.

22. Bibb, M. J., Janssen, G. R., & Ward, J. M. (1985). Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of *Streptomyces erythraeus*. *Gene*, *38*(1-3), 215-226.

23. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A., & Khosla, C. (1993). Engineered biosynthesis of novel polyketides. *SCIENCE-NEW YORK THEN WASHINGTON-*, *262*, 1546-1546.

24. Fernández-Moreno, M. A., Caballero, J., Hopwood, D. A., & Malpartida, F. (1991). The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. *Cell*, *66*(4), 769-780.

25. Shao, Z., Rao, G., Li, C., Abil, Z., Luo, Y., & Zhao, H. (2013). Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS synthetic biology*, *2*(11), 662-669.

26. Cone, M. C., Yin, X., Grochowski, L. L., Parker, M. R., & Zabriskie, T. M. (2003). The blasticidin S biosynthesis gene cluster from *Streptomyces griseochromogenes*: sequence analysis, organization, and initial characterization. *ChemBioChem*, *4*(9), 821-828.

27. Chiu, M. L., Viollier, P. H., Katoh, T., Ramsden, J. J., & Thompson, C. J. (2001). Ligand-induced changes in the Streptomyces lividans TipAL protein imply an alternative mechanism of transcriptional activation for MerR-like proteins. *Biochemistry*, *40*(43), 12950-12958.

28. Bai, C., Zhang, Y., Zhao, X., Hu, Y., Xiang, S., Miao, J., ... & Zhang, L. (2015). Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*. *Proceedings of the National Academy of Sciences*, *112*(39), 12181-12186.

29. Kierser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., & Hopwood, D. A. (2000). Practical *Streptomyces* Genetics: John Innes Foundation.

30. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology*, *31*(3), 233-239.

31. Bao, Z., Xiao, H., Liang, J., Zhang, L., Xiong, X., Sun, N., ... & Zhao, H. (2014). Homology-integrated CRISPR–Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS synthetic biology*, *4*(5), 585-594.

32. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819-823.

33. Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PloS one*, *3*(11), e3647.

34. Cobb, R. E., Wang, Y., & Zhao, H. (2014). High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS synthetic biology*, *4*(6), 723-728.

35. Huang, H., Zheng, G., Jiang, W., Hu, H., & Lu, Y. (2015). One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta biochimica et biophysica Sinica*, *47*(4), 231-243.

36. Zeng, H., Wen, S., Xu, W., He, Z., Zhai, G., Liu, Y., ... & Sun, Y. (2015). Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA (sm) combined system. *Applied microbiology and biotechnology*, *99*(24), 10575-10585.

37. Dubeau, M. P., Ghinet, M. G., Jacques, P. É., Clermont, N., Beaulieu, C., & Brzezinski, R. (2009). Cytosine deaminase as a negative selection marker for gene disruption and replacement in the genus *Streptomyces* and other actinobacteria. *Applied and environmental microbiology*, *75*(4), 1211-1214.

38. Johnson, D. B., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., ... & Wang, L. (2011). RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nature chemical biology*, *7*(11), 779-786.

39. Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., ... & Emig, C. J. (2011). Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*, *333*(6040), 348-353.

40. Lee, B. S., Shin, S., Jeon, J. Y., Jang, K. S., Lee, B. Y., Choi, S., & Yoo, T. H. (2015). Incorporation of unnatural amino acids in response to the AGG codon. *ACS chemical biology*, *10*(7), 1648-1653.

41. Wang, K., Neumann, H., Peak-Chew, S. Y., & Chin, J. W. (2007). Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology*, *25*(7), 770-777.

Chapter 2. Development of Amber Suppression Unnatural Amino Acid Incorporation System in *Streptomyces venezuelae* ATCC 15439

1. Introduction

The canonical genetic code specifies 61 sense codons for amino acids and 3 nonsense codons for stop signals in protein translation. In the laboratory, stop codons have been exploited for the incorporation of both natural and unnatural amino acids (UAAs) into proteins. Natural suppressor tRNAs decoding stop codons as proteinogenic amino acids have been identified in *E. coli* and other organisms. Orthogonal aminoacyl-tRNA synthetase/amber suppressor tRNA (aaRS/$tRNA_{CUA}$) pairs have been engineered to incorporate more than 100 various UAAs into protein in response to a stop codon for different applications. With incorporation of novel moieties of UAAs, target proteins or enzymes can be site-specifically modified. UAA incorporation technology is therefore a powerful tool to study and manipulate protein structure and function in living cells.[1,2]

Unnatural amino acid (UAA) incorporation systems have been developed in a variety of hosts, including *E. coli, mammalian cells, S. cerevisiae, C. elegans, P. pastoris, M. smegmatis, and D. melanogaster*.[2] Actinobacteria such as *Streptomyces* are important as a source of bioactive natural products. Due to differences in codon usage and cell physiology, many *Streptomyces* proteins cannot be expressed in *E. coli*. As a result, reconstitution of natural product biosynthetic pathways from *Streptomyces* in *E. coli* is extremely challenging and rarely successful.[3,4] Development of an UAA incorporation system in *Streptomyces* will make possible several important applications, including mapping natural product biosynthesis multienzyme complexes interactions and structurally diversification of ribosomally synthesized and post-translationally modified

peptides (RiPPs).

In *E. coli*, variants of the *M. jannaschii* tyrosyl-tRNA synthetase/amber suppressor tRNA (MjTyrRS/MjtRNA$^{Tyr}_{CUA}$) pair[1,2] and *M. barkeri* pyrrolysyl-tRNA synthetase/amber suppressor tRNA (PylRS/tRNA$^{Pyl}_{CUA}$) pair[5] evolved to incorporate specific UAAs are the commonly used aaRS/tRNA$_{CUA}$ pairs. The photocrosslinker *p*-benzoyl-L-phenylalanine[6] (pBpa), the photocrosslinking and Huisgen cycloaddition partner *p*-azido-L-phenylalanine[7] (pAzPhe), and the cross coupling *p*-iodo-L-phenylalanine[8,9] (pIPhe) are widely used with variants of MjTyrRS/MjtRNA$^{Tyr}_{CUA}$ pair.

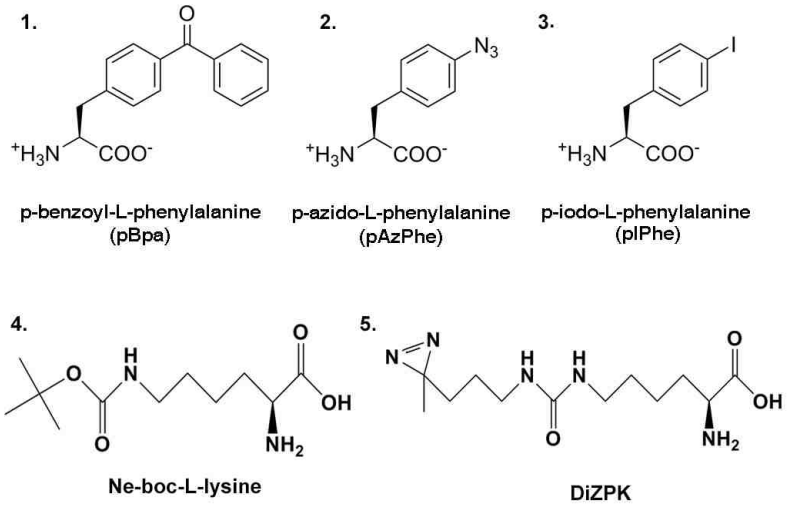


Figure 2-1. Structures of unnatural amino acids used in this study

.

Another efficient photocrosslinker DiZPK[10] is used with variant of PylRS/tRNA$^{Pyl}_{CUA}$ pair. These four UAAs (Figure 2-1) were employed to develop the amber suppression UAA systems in *Streptomyces*.

Streptomyces venezuelae ATCC 15439, well known for its production of the methymycin/pikromycin family of macrolide[11] and for its use in engineered production of macrolide derivatives bearing altered sugar moieties[12], as the host for developing the

UAA incorporation system. *S. venezuelae* ATCC 15439 is among the most rapidly growing model *Streptomyces* hosts (doubling time of ~ 1 h[13]), grows in a monodispersed manner in liquid culture, and can be genetically manipulated easily using either conjugation or protoplast transformation for introduction of plasmid DNA, making it a convenient host for synthetic biology applications.

To develop this system, pSUA5 harboring wild type humanized enhanced GFP (eGFP) was used as positive control to test the expression of the reporter protein. pSUA4 harboring an eGFP gene with a amber stop codon at position 150 (eGFP-150TAG) was designed to test whether *S. venezuelae* has endogenous amber suppression capability. pSUA3 containing eGFP-150TAG and an optimized MjtRNA$^{Tyr}_{CUA}$ Nap1 was designed to test MjtRNA$^{Tyr}_{CUA}$ orthogonality to endogenous aaRS in *S. venezuelae*. Due to the codon usage bias of *Streptomyces*, the MjTyrRS gene was codon optimized and cloned with eGFP-150TAG and MjtRNA$^{Tyr}_{CUA}$ to generate pSUA2, which was designed to test the functionality of MjTyrRS/tRNA$^{Tyr}_{CUA}$ pair in *Streptomyces*.

After function validation of tyrosine incorporation as a whole system, codon optimized pBpaRS, pAzPheRS, and pIPheRS[6-8] were used to replace MjTyrRS to generate pSUA1-pBpaRS, pSUA1-pAzPheRS, and pSUA1-pIPheRS, respectively. In presence of their cognate UAAs, pAzPhe and pIPhe, but not pBpa, were incorporated into sfGFP. In absence of their cognate UAAs, endogenous amino acids were recognized by these three synthetases at varying degrees. Large-scale protein expression and purification was conducted, followed by protein ESI-MS analysis. Complete pIPhe incorporation was obtained in presence of pIPhe from *S. venezuelae* harboring pSUA1-pIPheRS, however, 33% of pAzPhe incorporation from *S. venezuelae* harboring pSUA1-

pAzPheRS in presence of 1 mM pAzPhe. To improve the fidelity of pAzPhe incorporation, other three reported pAzPheRS variants were employed. Besides that, an elongation factor Tu (tufA gene) was codon optimized and cloned into plasmid to facilitate UAA incorporation. Eventually 77% pAzPhe incorporation was obtained by increasing the concentration of pAzPhe from 1 mM to 5 mM.

In addition, PylRS/$tRNA^{Pyl}_{CUA}$ pair was also designed and constructed to incorporate efficient photocrosslinker into sfGFP protein. Preliminary result suggests the functionality of the wild type PylRS/$tRNA^{Pyl}_{CUA}$ pair, not the DZKRS/ $tRNA^{Pyl}_{CUA}$ pair, which is responsible for DiZPK incorporation.

2. Experimental Procedures

General. All general molecular biological and biochemical reagents, including Luria-Bertani (LB) media (Miller), were purchased from VWR (Atlanta, GA) and were used without further purification. Water used for media was obtained from a Barnstead/Thermolyne HN Ultrapure water purification system. Tryptic Soy Broth (TSB) media was purchased from Cole Parmer (Vernon Hills, IL). *p*-iodo-L-phenylalanine, phenylmethylsulfonyl fluoride (PMSF), and dibenzocyclooctyne-PEG4-biotin conjugate were purchased from Sigma-Aldrich (St. Louis, MO). *p*-azido-L-phenylalanine Ne-boc-L-lysine were purchased from Chem-Impex (Wood Dale, IL). *p*-benzoyl-L-phenylalanine was purchased from Bachem (Bubendorf, Switzerland). DiZPK was kindly provided by Dr. Gary A. Sulikowski at the Vanderbilt University. Apramycin sulfate was purchased from Research Products International (Mount Prospect, IL). Restriction enzymes, Phusion DNA polymerase, T4 DNA ligase and calf intestinal alkaline phosphatase were purchased from New England Biolabs (Ipswich, MA). DNA purification and

concentration was performed using the DNA Clean & Concentrator Kit; and agarose gel DNA extraction was performed using the Gel DNA Recovery Kit, both from Zymo Research (Irvine, CA). Plasmid extractions were performed using the QIAprep Spin Miniprep Kit; and proteins were purified by Ni-NTA agarose, both from Qiagen (Valencia, CA). DNA quantification was performed using a Nanodrop 2000 uv-vis spectrophotometer (Thermo Scientific). Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). DNA sequencing was performed by Genewiz (South Plainfield, NJ). PCR reactions were carried out using a Bio-Rad S1000 thermal cycler. Fluorescence measurements were taken using a Molecular Devices SpectraMax M2 Multi-Mode Microplate Reader. Plasmid and DNA sequence design and management was conducted using Vector NTI 10 (Life Technologies). Chemically competent *E. coli* cells were prepared using the rubidium chloride method.[14] Standard molecular biological methods, protocols, reagents, and materials[14] were used for PCR, restriction enzyme digestion, ligation, transformation, selection of *E. coli* transformants, agarose gel electrophoresis, gel extraction, and plasmid isolation unless otherwise specified. Bacteria were cultured in either an Innova 42 or Innova 26 incubator/shaker (New Brunswick Scientific). Lysozyme was purchased from MP Biomedicals (Santa Ana, CA). Cells lysis was performed using a Branson digital sonifier Model SSE-1. Protein purifications were performed using Poly-Prep chromatography columns (Bio-Rad) and HiTrap Phenyl HP columns (HIC, 1 mL size, GE Healthcare). All general SDS-PAGE and Western Blotting equipment and materials, including Mini-PROTEAN Tetra Cell, mini trans-blot module, TGX Stain Free precast polyacrylamide gel, Precision Plus protein standards, Laemmli sample buffer, 10X Tris/Glycine/SDS buffer, Bio-Safe Coomassie G-250 stain, Immun-

blot PVDF membrane, mini trans-blot filter paper, blotting-grade blocker (nonfat dry milk) and Clarity Western ECL substrate were obtained from Bio-Rad (Hercules, CA). Anti-His (C-terminal) primary antibody and goat anti-mouse IgG (H+L)/HRP conjugate secondary antibody were obtained from Life Technologies (Carlsbad, CA). Streptavidin anti-Biotin antibody/HRP conjugate was obtained from Thermo scientific (Rockford, IL). Western Blots were imaged using the molecular imager Gel Doc XR+ (Bio-Rad). Intact protein ESI-MS was performed using an LCT Premier time of flight mass spectrometer (TOF-MS, Waters Instruments).

Bacteria strains. *E. coli DH5α* was used for routine DNA cloning and manipulation. *E. coli* S17-1 was the donor strain used in *S. venezuelae* conjugal transformation experiments. *S. venezuelae* ATCC 15439 was kindly provided by Hung-wen Liu at the University of Texas at Austin, and was used as the host for development and optimization of the amber suppression UAA incorporation system.

Bacterial culture. Routine liquid culture of *E. coli DH5α* for cloning purposes was carried out in 2-5 mL of Luria-Bertani broth in sterile 15 mL conical tubes at 37 °C, 250 rpm overnight (12-16 h). Selection of *E. coli DH5α* transformants was carried out on Luria-Bertani agar plates containing the apramycin at 37°C overnight (12-16 h). *S. venezuelae* ATCC 15439 mutants were routinely grown in 5-10 mL of TSB liquid media containing 50 μg/mL apramycin at 28°C, 200-250 rpm, for 48-72 h in 25×150 mm glass culture tubes with ~fifteen 4 mm glass beads in each tube. *Streptomyces* defined minimal media (MM) was prepared according to a published procedure[15]. Fifty μL of protoplasts and 200 ng of plasmid DNA were used for each transformation. R5 agar was used in place of R2YE agar for plating transformed protoplasts. After 3-4 d, transformant

colonies were observed, and were transferred to 5 mL of TSB medium containing 50 μg/mL apramycin and grown for 72 h. Cells from 1 mL of saturated culture were pelleted by centrifugation, resuspended in 0.5 mL of sterile 20% glycerol, and stored at -80°C.

Preparation of E. coli Competent Cells. Competent cells of *E.coli* were prepared using the RbCl method[14]. A single fresh colony of appropriate *E. coli* strain was used to inoculate 2 mL of the Luria-Bertani (LB) liquid medium culture and the resulting culture was grown overnight at 37 °C with shaking at 250 rpm. A 500 μL of the overnight culture was used to inoculate 50 mL of LB culture in a 250 mL Erlenmeyer flask. When the cell culture reached an OD_{600} of about 0.4, the culture was transferred to a pre-chilled sterile conical tube and incubated on ice for 30 min. After centrifugation at 3,000 g for 5 min, the supernatant was discard and the cell pellet was resuspended in one-third of the original culture volume of ice-cold RF1 solution (100 mM RbCl, 15% glycerol, 50 mM $MnCl_2$, 30 mM potassium acetate, 10 mM $CaCl_2$, pH 5.8, filter sterilized by passage through a 0.22 μm membrane). After incubation on ice for 15 min, the cell suspension was centrifuged at 3,000 g, 4 °C for 5 min and the resulting cell pellet was resuspended in two twenty-fifths of the original culture volume of ice-cold RF2 solution (10 mM RbCl, 10 mM MOPS, 75 mM $CaCl_2$, 15% glycerol, pH 6.8, filter sterilized by passage through a 0.22 μm membrane). The cells were aliquoted into 100 μL portions in pre-chilled microcentrifuge tubes and frozen at -80 °C.

General PCR Conditions. Concentrations of templates, primers, polymerase, dNTPs, and buffer recommended by NEB for Phusion DNA polymerase were used unless otherwise specified. We used two types of PCR protocols to construct all fragments utilized for vector construction: Protocol 1) PCR amplification of a single

fragment with two primers and protocol 2) multiple fragments overlap extension PCR. General PCR programs for each protocol are given below.

Protocol 1		Protocol 2	
98 °C	30 s	98 °C	30 s
98 °C	10 s	98 °C	10 s
Tm-prim−5	30 s	Tm-OE−5	30 s
72 °C	30 s/kb	72 °C	30 s/kb
Repeat 2 times		Repeat 2 times	
98 °C	10 s	98 °C	10 s
Tm-ext−5	30 s	Tm-ext−5	30 s
72 °C	30 s/kb	72 °C	30 s/kb
Repeat 26 times		Repeat 26 times	
72 °C	10 m	72 °C	10 m
4 °C	∞	4 °C	∞

Tm-prim = Tm of the portion of the primer that primes to the sample

Tm-ext = Tm of the entire primer

Tm-OE = Tm of the junction between fragments

Small Scale Protein Expression. Fifty μL of glycerol stock of each mutant was transferred to 10 mL of TSB media containing 50 μg/mL apramycin and 1 or 5 mM UAA (when appropriate); and was grown at 28 °C, 250 rpm, for 72 h. Each cell culture was collected by centrifugation (4,000 g, 10 min, room temperature), resuspended in 1 mL lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) containing 2 mg/mL lysozyme in a 1.5 mL Eppendorf tube, and incubated at 37 °C for 1 h on a mechanical rotator. The resulting mixture was then sonicated on ice for 2 min (20% amplitude, 24 × 5 sec pulses with 10 sec pauses) and cell debris was pelleted by centrifugation (14,000 rpm, 10 min, room temperature). The clarified supernatant was incubated with 50 μL of Ni-NTA slurry (that had been pre-equilibrated with lysis buffer)

at 4 °C for 1 h on a mechanical rotator. Ni-NTA was collected by centrifugation at 300 g for 2 min and washed with 2 × 1 mL wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) before eluting bound protein with 35 µL elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0).

Large Scale Protein Expression and Purification. A 3-step purification protocol (Ni-NTA affinity chromatography followed by 2 successive hydrophobic interaction chromatography steps) was used to obtain high purity sfGFP suitable for intact protein ESI-MS using the LCT Premier ESI-TOF mass spectrometer. First, 50 µL of glycerol stock of each mutant was transferred to 40 mL of TSB or MM media containing 50 µg/mL apramycin and 1 or 5 mM UAA, when appropriate, in a 250 mL Erlenmeyer flask containing ~one hundred 4 mm glass beads and was grown at 28 °C, 250 rpm, for 72 h. Eight mL of the resulting seed culture was used to inoculate each of four 2 L Erlenmeyer flasks containing 400 mL TSB or MM media, 50 µg/mL apramycin, and 1 or 5 mM UAA, when appropriate, with no glass beads (1.6 L total), which were grown at 28 °C, 200 rpm, for 96 h. Cells were pelleted by centrifugation (6,000 rpm, 20 min, room temperature) and the supernatant was decanted. Buffers used in all subsequent steps contained 1 mM PMSF to prevent protease degradation of sfGFP. Cells were resuspended in 24 mL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH to 8.0) containing 2 mg/mL lysozyme and incubated at 37 °C for 1 h on a mechanical rotator. The resulting mixture was then transferred to a 30 mL plastic syringe and passed through a 26-gauge needle to disrupt mycelia. Four equal portions were individually sonicated on ice for 12 min (30% amplitude, 72 × 10 sec pulses with 10 sec pauses), and cell debris was pelleted by centrifugation (17,000 rpm, 45 min, 4 °C).

The pooled supernatants were carefully transferred to a pre-chilled 50 mL conical tube containing 1.6 mL of Ni-NTA slurry (that had been pre-equilibrated with lysis buffer), and incubated at 4 °C for 2 h on a mechanical rotator. The mixture of Ni-NTA and supernatant was transferred to a Poly-Prep chromatography column, flow-through was drained, and the column was washed with 20 mL of lysis buffer followed by 10 mL of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0). Protein was eluted with 5 mL of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 120 mM imidazole, pH 8.0) and was used for the subsequent hydrophobic interaction chromatography (HIC) step.

The Ni-NTA eluate was mixed with the appropriate amount of ammonium sulfate buffer (1.7 M $(NH_4)_2SO_4$, 50 mM NaH_2PO_4, pH 8.0) to give a final concentration of 0.7 M $(NH_4)_2SO_4$; and insoluble proteins were removed by centrifugation (14,000 rpm, 1 min, room temperature). The supernatant was then applied by syringe to an HIC column pre-equilibrated with loading buffer (0.7 M $(NH_4)_2SO_4$, 50 mM NaH_2PO_4, pH 8.0); and the column was washed with 8 mL of loading buffer. Proteins were eluted by step gradient (addition of 7 × 1.5 mL portions of loading buffer containing decreasing concentrations of $(NH_4)_2SO_4$ - 0.6 M, 0.5 M, 0.4 M, 0.3 M, 0.2 M, 0.1, 0 M); and 1.5 mL fractions were collected. SDS-PAGE analysis of these fractions obtained from the first purification procedure revealed that sfGFP begins to elute at 0.3 M $(NH_4)_2SO_4$ and continues to elute as the concentration of $(NH_4)_2SO_4$ decreases; but that a prominent, more hydrophobic non-target protein elutes below 0.2 M $(NH_4)_2SO_4$. Therefore, the 0.3 and 0.2 M $(NH_4)_2SO_4$ fractions were kept for application to the second HIC column, while other fractions were discarded.

The desired fractions from the first HIC column were mixed with the appropriate amount of ammonium sulfate buffer (1.7 M $(NH_4)_2SO_4$, 50 mM NaH_2PO_4, pH 8.0) to give a final concentration of 0.7 M $(NH_4)_2SO_4$; and insoluble proteins were again removed by centrifugation (14,000 rpm, 1 min, room temperature). A similar procedure to that used for the first HIC column was employed for isolation of pure sfGFP. In this case, the same wash and step gradient were used, but 1 mL fractions were collected. SDS-PAGE analysis of fractions revealed that the first 2 × 1 mL fractions eluted with buffer containing no $(NH_4)_2SO_4$ that were obtained at the end of the elution contained pure sfGFP.

Protein Fluorescence Quantification. Partially purified sfGFP proteins obtained from small-scale expression were analyzed by fluorescence spectroscopy to determine the relative yields of sfGFP produced by *S. venezuelae* mutants. Fifteen μL of Ni-NTA eluate obtained from each 10 mL culture was mixed thoroughly with 185 μL of water and transferred to a well of a black 96-well plate. Fluorescence was then measured using a microplate reader (excitation = 485 nm, bandwidth = 9 nm; emission = 520 nm, bandwidth = 15 nm), and background due to buffer was subtracted. A similar method was used to track sfGFP and to estimate the amount of sfGFP present in elution fractions during large scale protein purification. In this method, 50 μL of each fraction was mixed with 150 μL water, and fluorescence was measured using the microplate reader.

A quantitative relationship between the fluorescence signal from the microplate reader assay and the amount of purified sfGFP protein was established. It was employed to calculate the final yields of sfGFP protein obtained from each large scale purification, and to estimate the expressed yield (in mg/L culture) of sfGFP in each mutant. The

quantitative relationship was established as follows. The concentration of sfGFP purified in a large-scale preparation from *S. venezuelae* harboring opt-pSUA2 was determined by constructing a standard curve using the Bradford assay. The fluorescence signal of the same amounts of diluted sfGFP protein used in the Bradford assay were measured using the microplate reader assay, and the two datasets were found to correlate linearly. Through this correlation, we determined that 1×10^6 fluorescence units is equivalent to 92.7 ng of sfGFP protein. This conversion factor was used to directly calculate the final yields of sfGFP proteins obtained after each large scale purification from the fluorescence data, and to estimate expressed yield of sfGFP by fluorescence quantification of Ni-NTA eluate obtained after the first step of the large scale purification. All these protein yield calculations were based on the simplifying assumptions that 100% of expressed sfGFP bound to, and was eluted from, the Ni-NTA resin; and that the brightness of sfGFP mutants bearing UAAs were the same as that of the wild-type protein.

Western Blot Analysis. Thirteen μL of eluate obtained by small-scale protein purification was separated by SDS-PAGE (200 V, 30 min) and transferred to a PVDF membrane (100 V, 350 mA, 60 min). The membrane was washed (3 × 5 min) with TBS (20 mM Tris-HCl, 500 mM NaCl, pH 7.5), blocked with TBS containing 5% w/v milk powder (blocking solution) for 1 h, incubated with blocking solution containing a 1:5000 dilution of the anti-His (C-terminal) primary antibody for 1 h, washed (3 × 5 min) with TBS containing 0.5% (v/v) Tween 20 (TBST), incubated with TBST containing a 1:2000 dilution of the goat anti-mouse IgG (H+L)/HRP conjugate secondary antibody for 1 h, washed (3 × 5 min) with TBST, and washed (1 × 5 min) with TBS. The membrane was then analyzed by incubation with chemiluminescent substrate solution according to the

manufacturer's protocol, and imaged. All membrane handling steps were carried out at room temperature on an orbital rocker. Imaging times varied depending on the intensity of the signals observed, and are reported in the legends of figures that report Western Blot data.

DBCO-biotin Labeling and Streptavidin Western Blot. The procedure employed is based on published work[16] with minor modification. A 100 μM (50 X) stock solution of dibenzocyclooctyne-PEG4-biotin (DBCO-biotin) was prepared in water. Purified sfGFP proteins obtained by large scale preparations from *S. venezuelae* harboring pSUA1-pAzPheRS grown in the presence and in the absence of 1 mM pAzPhe were diluted to final concentrations of 111 nM in 50 mM NaH_2PO_4, pH 8.0 in the presence of 2 μM DBCO-biotin (1:18 molar ratio of protein to DBCO-biotin, 100 μL reaction volumes) and incubated at room temperature for 12 h. The reaction mixtures were then diluted in 900 μL lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and subjected to Ni-NTA batch purification using 30 μL of Ni-NTA slurry (that had been pre-equilibrated with lysis buffer) at 4 °C for 2 h on a mechanical rotator. Ni-NTA was collected by centrifugation at 300 g for 2 min and washed with 1.5 mL lysis buffer before eluting bound protein with 30 μL of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). Thirteen μL of the resulting biotin labeled proteins were separated by SDS-PAGE (200 V, 30 min) and transferred to a PVDF membrane (100 V, 350 mA, 60 min). The membrane was washed (3 × 5 min) with TBS (20 mM Tris-HCl, 500 mM NaCl, pH 7.5), blocked with TBS containing 5% w/v milk powder (blocking solution) for 1 h, incubated with blocking solution containing a 1:5000 dilution of the Streptavidin anti-Biotin/HRP conjugate antibody for 1 h, washed (3 × 5 min) with

TBS containing 0.5% (v/v) Tween 20 (TBST), and washed (1 × 5 min) with TBS. The membrane was then analyzed by incubation with chemiluminescent substrate solution according to the manufacturer's protocol, and imaged. All membrane handling steps were carried out at room temperature on an orbital rocker. Imaging time was 5 s.

Mass spectrometry. Intact mass spectra purified sfGFP proteins were acquired using a TOF-MS equipped with a C_4 reverse phase trap. The buffers used for separation are as follows: Buffer A: 97% H_2O, 3% Acetonitrile, 0.1% Formic acid; Buffer B: 5% H_2O, 95% Acetonitrile, and 0.1% Formic acid. The protein sample was diluted in Solvent A. The samples were held on the C_4 trap for 10 min at 0.2 mL/min with 5% solvent B to attempt to remove contaminants. The HPLC flow was then reduced to 0.1 mL/min and the program was started. The HPLC program, which was optimized for protein analysis using this instrument, was as follows: 0 min, 5% B, flow rate 0.1 mL/min; 1 min, 30% B, flow rate 0.5 mL/min; 4.5 min, 50% B, flow rate 0.025 mL/min; 6 min, 100% B, flow rate 0.025 mL/min; 6.5 min, 100% B, flow rate 0.15 mL/min. sfGFP proteins typically eluted between 3.5 and 6 min. During each run, the distance of the probe from the sample intake orifice was adjusted manually in real time to maintain an ion count as close to the manufacturer recommended count of 136 counts per second. At the end of the program, the trap was rinsed twice with gradients going from low %B to high %B to eliminate carryover. For each sample the instrument was calibrated utilizing myoglobin as a standard. Deconvolution was performed by the MassLynx V4.0 transform program. The transformation result was verified by parallel analysis using MagTran V1.02.

Conjugal Transfer of Vectors into Wild-Type S. venezuelae. This experiment was performed using the methods published by Bierman *et al*[17] with minor modifications. The

S. venezuelae was grown at 28 °C, 250 rpm in the glass tubes with glass beads inside and the *E. coli* S17-1 was cultured at 37 °C, 250 rpm in the cell preparation process for conjugal transfer. The 1 mL of previously prepared wild-type *S. venezuelae* mycelia was used to inoculate into 9 mL of tryptic soy broth (TSB) and the sequenced verified plasmid was transferred into the donor *E. coli* strain S17-1 selected by overnight growth on Luria-Bertani (LB) agar plate containing apramycin (50 μg/mL) and streptomycin (10 μg/mL) at 37 °C. The constructs made for development of UAA system contains apramycin resistance gene and *E. coli* S17-1 is streptomycin resistant. Next, 2 mL of the resulting *S. venezuelae* culture was inoculated into 18 mL of TSB. In parallel, one of the *E. coli* S17-1 transformants was picked up and inoculated into 2 mL TSB containing apramycin (50 μg/mL) and streptomycin (10 μg/mL). 1 mL of the overnight *S. venezuelae* culture was inoculated into 9 mL of TSB and 20 μL of the overnight *E. coli* culture was inoculated into 2 mL TSB containing apramycin (50 μg/mL) and streptomycin (10 μg/mL). In about three hours, the *E. coli* and *S. venezuelae* cells were collected by centrifugation and supernatant was discarded. After two times wash with 2 mL TSB without any antibiotic for cell pellet of *E. coli*, the resulting cell pellets of *E. coli* and *S. venezuelae* were both resuspended in 2 mL TSB. Three 600 μL mixture of *S. venezuelae* and *E. coli* were prepared in 9:1, 1:1, and 1:10 ratios, respectively. Aliquots of 100 μL of each cell mixture were spread onto fresh prepared AS1 plates, which were prepared as follows. A mixture of 1 g of yeast extract, 0.2 g of L-alanine, 0.5 g of L-arginine, 5 g of soluble starch, 2.5 g of NaCl, 10 g of Na_2SO_4, and 20 g of agar were dissolved in 1 L of deionized water, pH was adjusted to 7.5 with H_2SO_4, and the resulting medium was autoclaved. After autoclaving and cooling down, MgCl2 was added to a

final concentration of 10 mM[18]. Three replicate plates were usually made for each cell mixture. Besides that, two plates were spread with *S. venezuelae* only and one plate was spread with *E. coli* only, which were used as controls. All of the plates were incubated at 28 °C for 14-20 h before overlaying appropriate antibiotics. For conjugation plates, 1-2 mL of sterile water containing nalidixic acid (500 μg/mL) and apramycin (500 μg/mL) was overlaid onto the pate. One *S. venezuelae* only plate was overlaid with nalidixic acid (500 μg/mL) and apramycin (500 μg/mL) and one *E. coli* only plate was overlaid with nalidixic acid (500 μg/mL) as negative controls. In addition, the other *S. venezuelae* only pate was overlaid with nalidixic acid (500 μg/mL) as a positive control. All of the plates were incubated at 28 °C for 6 to 10 days, until small *S. venezuelae* transformants can be seen.

Preparation of S. venezuelae Protoplasts. *S. venezuelae* protoplasts were prepared by the standard protocol used for preparation of *S. coelicolor* and *S. lividans* protoplasts[19] with minor modifications. Fifty μL of wild-type *S. venezuelae* spore suspension was inoculated into 5 mL of TSB in a glass tube with glass beads, and grown at 28° C, 250 rpm for 3 days. 1 mL of the resulting culture was used to inoculate into 50 mL YEME medium, which was prepared as follows, and cultured at 28 °C, 250 rpm. A mixture of 0.3 g of yeast extract, 0.5 g of peptone, 0.3 g of malt extract, 1 g of glucose, 34 g of sucrose were dissolved in deionized water to make the final volume of 100 mL and the resulting medium was autoclaved. After autoclaving, 0.2 mL of 2.5 M $MgCl_2$ and 2.5 mL of 20% glycine were added to that autoclaved solution to make the YEME medium[19]. *S. venezuelae* was grown slower in YEME due to high concentration of sucrose. After a 7-10-day culture, the cells were collected by centrifugation, and the supernatant was

discarded carefully. The cell pellet was washed by 30 mL of sterile 10.3% sucrose solution, the cells were then collected by centrifugation, and the supernatant was discarded. The resulting cell pellet was resuspended with 30 mL of lysozyme (2 mg/mL) containing P (protoplast) buffer, which was prepared as follows. 100 mg/mL of lysozyme stock solution was prepared and then filter sterilized by passage through a 0.22 μm membrane. A mixture of 20.6 g of sucrose, 50 mg of K_2SO_4, 0.4 g of $MgCl_2 \cdot 6H_2O$, and 0.4 mL of trace element solution were dissolved in deionized water to make final volume of 160 mL. After autoclaving, 2 mL of 0.5% KH2PO4, 20 mL of 3.68% $CaCl_2 \cdot 2H_2O$, and 20 mL of 5.73%, pH 7.2 TES buffer were added to make the P buffer prior to use. 40 mg of $ZnCl_2$, 200 mg of $FeCl_3 \cdot 6H_2O$, 10 mg of $CuCl_2 \cdot 2H_2O$, 10 mg of $MnCl_2 \cdot 4H_2O$, 10 mg of $Na_2B_4O_7 \cdot 10H_2O$, and 10 mg of $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ were dissolved in 1 L of deionized water to make the metal trace element solution. The cell suspension was incubated at 37 °C on a mechanical rotator for about 1 h and was filtered with loose sterile cotton after small transparent protoplast balls were seen under the microscope. The protoplasts were centrifuged down and the supernatant was removed. The resulting protoplast pellet was washed with 20 mL of P buffer, pelleted by centrifugation, and resuspended in 4 mL of P buffer. The protoplasts were then aliquoted into 300 μL portions in microcentrifuge tubes, frozen at −20 °C overnight and transferred to −80 °C on the second day. Freeze slowly could ensure the viability of protoplasts. In parallel, the rest of P buffer was aliquoted into 1 mL portions in microcentrifuge tubes and stored along with the protoplasts. In this way, the protoplast and P buffer were taken out together for each batch of future protoplast transfer.

Protoplast transfer of Vectors into Wild-Type S. venezuelae. One tube of frozen

protoplasts was taken out and thawed by gently shaking in 37 °C water bath and PEG 1000 was added to the P buffer to a concentration of 25%. Protoplasts suspension was aliquoted into 50 µl portions into eppendorf tubes and each one would be used for one transformation. Two to five µl of plasmid were mixed with 50 µl of protoplast and 200 µl of 25% PEG 1000 containing P buffer was added to the protoplast and plasmids suspension, mixed well by pipetted up and down five times. The resulting solution was quickly transferred to the prepared R5 agar plate and gently spread evenly on the plate by the spreader to avoid destroying the protoplasts. R5 agar plates were prepared as follows. A mixture of 103 g of sucrose, 10 g of glucose, 5 g of yeast extract, 100 mg of casamino acid, 5.73 g of TES, 250 mg of K_2SO_4, 10.12 g of $MgCl_2 \bullet 6H_2O$, 2 mL of metal trace elements, and 3 g of NaOH were dissolved in deionized water to make the final volume of 1 L, adjusted the pH to 7.2 with HCl, 22 g of agar were added, and then autoclaved. After autoclaving, 15 mL of 20% glutamate (Na salt), 15 mL of 2% $NaNO_3$, 10 mL of 0.5% KH_2PO_4, and 4 mL of 5 M $CaCl_2$ were added to it to make the R5 agar medium. The R5 agar plate was made of 21 ml each. For controls, two tubes of 50 µl of protoplasts mixed with PEG 1000 containing P buffer were plated on R5 agar plates. All of the plates were incubated at 28 °C for 14-20 h prior to overlaying antibiotic. Based on the volume of each R5 plate, stock apramycin solution was diluted by sterile 10.3% sucrose to a concentration of 700 µg/mL and 1.5 mL of the apramycin antibiotic (700 µg/mL) was overlaid on the protoplast transfer plates and one of the protoplast only plate as the negative control. The other protoplast only plate was overlaid with pure 10.3% sucrose and would be used as positive control. After drying the plates in the hood, plates were incubated at 28 °C for 3-5 days until colonies appeared.

3. **Vector Design and Construction**

Construction of pSUA2, pSUA3, pSUA4, and pSUA5. pSUA2, pSUA3, pSUA4, and pSUA5 were constructed from the pCM1d[20] vector, which contains the Ppik promoter from the pikromycin gene cluster of *S. venezuelae* ATCC 15439 flanked by unique restriction sites EcoRV and EcoRI at its upstream and downstream ends, respectively; and a multicloning site (MCS) immediately downstream of the EcoRI site with five additional unique restriction sites (NdeI, BamHI, XbaI, BglII, and HindIII, listed in order from 5' to 3'). Among these unique restriction sites, only EcoRV, XbaI, BglII, and HindIII were used in this work. We employed a modular design for pSUA2-5 in which each functional element is flanked by unique restriction sites to facilitate vector construction and optimization.

The amber suppressor tRNA-containing element was designed utilizing natural *Streptomyces coelicolor* A3(2) promoter, 5'-flanking, 3'-flanking, and terminator sequences from a tRNA operon harboring tRNAs that base pair with high frequency codons in the *S. coelicolor* genome. Such elements, when incorporated at the 5' and 3' ends of the optimized MjtRNA$^{Tyr}_{CUA}$ Nap1[21], would be expected to result in high level transcription, proper post-transcriptional processing, and efficient termination in *S. venezuelae*. Bioinformatic analysis of all tRNA operons in the *Streptomyces coelicolor* A3(2) genome resulted in identification of seven candidate tRNA operons containing at least one predicted high usage tRNA. However, only one (nucleotide GI # 24419007, nucleotides 114283-114859) contained easily identifiable promoter and terminator sequences. The promoter-5'-flanking sequence (nucleotides 114283-114351, designated SCOt03p) and 3'-flanking sequence-terminator (nucleotides 114815-114859, designated

SCOt07t) from this operon were chosen for use in the synthetic MjtRNA$^{Tyr}_{CUA}$-containing element. The SCOt03p-MjtRNA$^{Tyr}_{CUA}$-SCOt07t fragment, which possessed an upstream BglII site, artificial PvuI site inserted immediately upstream of the terminator, and downstream HindIII site, was synthesized, cloned, and sequence verified by DNA 2.0 (Menlo Park, CA). Its sequence is shown below. This fragment was excised from the vector using BglII and HindIII, and was ligated into pCM1d digested with the same enzymes, resulting in pCM1d-tRNA. pCM1d-tRNA was used for construction of pSUA2-5.

The humanized eGFP (for use in pSUA5) and eGFP-150TAG (for use in pSUA2-4) genes were amplified from an eGFP-expressing construct (Fu-sen Liang, unpublished). The eGFP gene (sequence given below) was amplified from the original template using primers EGFP-pSUA5-E-up and EGFP-pSUA5-P-dn. This PCR product was used as the template for a second PCR reaction using primers EGFP-pSUA5-E-2-up and EGFP-pSUA5-P-2-dn to construct the full length fragment containing the SF14[22] promoter, synthetic ribosome binding site (RBS), and C-terminal His_6 tag with EcoRV and PvuI restriction sites at the 5'- and 3'-ends, respectively.

Each eGFP-150TAG gene with SF14 promoter, synthetic RBS, and C-terminal His_6 tag was constructed in 2 stages: amplification of the upstream and downstream eGFP gene fragments, and assembly of these fragments using overlap extension PCR. Because of the unique configuration of genetic elements within each of the pSUA2, pSUA3, and pSUA4 vectors, the eGFP-150TAG fragment made for each construct possessed a different set of restriction sites at the 5'- and 3'-ends, and was therefore constructed slightly differently. The upstream fragment of eGFP-150TAG that was used

for construction of pSUA2 and pSUA3 was amplified using primers EGFP-pSUA5-E-up and EGFP-pSUA1-4-N150-dn; and the downstream fragment of eGFP-150TAG that was used for construction of pSUA2 and pSUA3 was amplified using primers EGFP-pSUA1-4-N150-up and EGFP-pSUA1-B-dn. For assembly of eGFP-150TAG used in pSUA2, the fragments were joined by overlap extension PCR using primers EGFP-pSUA1-X-up and EGFP-pSUA1-SCOt03t-dn, resulting in an XbaI restriction site at the 5'-end and a BglII restriction site at the 3'-end. For assembly of eGFP-150TAG used in pSUA3, the fragments were joined by overlap extension PCR using primers EGFP-pSUA3-E-up and EGFP-pSUA1-SCOt03t-dn, resulting in an EcoRV restriction site at the 5'-end and a BglII restriction site at the 3'-end. The upstream fragment of eGFP-150TAG that was used for construction of pSUA4 was amplified using primers EGFP-pSUA5-E-up and EGFP-pSUA1-4-N150-dn; and the downstream fragment of eGFP-150TAG that was used for construction of pSUA4 was amplified using primers EGFP-pSUA1-4-N150-up and EGFP-pSUA5-P-dn. For assembly of eGFP-150TAG used in pSUA4, the fragments were joined by overlap extension PCR using primers EGFP-pSUA5-E-2-up and EGFP-pSUA5-P-2-dn, resulting in an EcoRV restriction site at the 5'-end and a PvuI restriction site at the 3'-end.

Each of the 4 eGFP fragments was digested with the appropriate restriction enzymes and ligated into pCM1d-tRNA digested with these same enzymes. In the case of construction of pSUA4 and pSUA5, which lack the SCOt03p-MjtRNA$^{Tyr}_{CUA}$ fragment, digestion of pCM1d-tRNA with PvuI resulted in excision of the SCOt03p-MjtRNA$^{Tyr}_{CUA}$ fragment and retention of SCOt07t at this stage. This step resulted in construction of pSUA3, pSUA4, and pSUA5; and an intermediate construct lacking MjTyrRS that would

be used to construct pSUA2.

The MjTyrRS fragment with SF14 promoter, synthetic RBS, and C-terminal His6 tag, flanked by an EcoRV restriction site at the 5'-end and an XbaI site at the 3'-end was amplified from vector template pBK-MjYRS-wt (Peter G. Schultz) using primers Syn-w/His-SF14-1-up and Syn-w/His-SF14-1-dn. This PCR product was used as the template for a second PCR reaction using primers EGFP-pSUA3-E-up and Syn-w/His-SF14-2-dn to construct the full length fragment containing the SF14 promoter and C-terminal His6 tag with EcoRV and XbaI restriction sites at the 5'- and 3'-ends, respectively. This fragment was digested with EcoRV and XbaI and ligated into the intermediate construct

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
EGFP-pSUA5-E-up	CGTGAGCTACAATCAATAGTCGATTAGAGGAGAAATTAG AAATGGTGAGCAAGGGCGA	797	EGFP vector
EGFP-pSUA5-P-dn	ATATATCGATCGTTAGTGATGGTGATGGTGATGCAACTC CTTGTACAGCTCGTCCATGC		
EGFP-pSUA5-E-2-up	CGCTAGGATATCGGTTGACCTTGATGAGGCGGCGTGAGC TACAATCAATAGTCGATTAG	844	PCR pdt.
EGFP-pSUA5-P-2-dn	GCCGTCCGCGTCCGCATATATCGATCGTTAGTGATGGTG ATGG		
EGFP-pSUA1-4-N150-dn	GTCGGCCATGATATAGACCTAGTGGCTGTTGTAGTTG (Used with EGFP-pSUA5-E-up)	509	EGFP vector
EGFP-pSUA1-4-N150-up	CAACTACAACAGCCACTAGGTCTATATCATGGCCGAC	326	EGFP vector
EGFP-pSUA1-B-dn	CACAATTAGATCTTTAGTGATGGTGATGGTGATGCAACT CCTTGTACAGCTCGTCCATGC		
EGFP-pSUA1-X-up	TAATGATCTAGAGGTTGACCTTGATGAGGCGGCGTGAGC TACAATCAATAGT	857	PCR pdt.
EGFP-pSUA1-SCOt03t-dn	ATGAACTCTAGCGGATTCCGGGGCCAGCACAATTAGATC TTTAGTGAT		
EGFP-pSUA3-E-up	ATATATGATATCGGTTGACCTTGATGAGGCGGCGTGAGC TACAATCAATAG (Used with EGFP-pSUA1-SCOt03t-dn)	857	PCR pdt.
EGFP-pSUA1-4-N150-up	Used with EGFP-pSUA5-P-dn	325	EGFP vector
Syn-w/His-SF14-1-up	GGCGTGAGCTACAATCAATAGTCGATTAGAGGAATCTCA TATGGACGAATTT	986	pBK-MjYRS-wt
Syn-w/His-SF14-1-dn	ATTAGTGATGGTGATGGTGATGCAACTCTAATCTCTTTC TAATTGGCTCTAAAAT		
Syn-w/His-SF14-2-dn	TGGCTAGGCTCTAGATTAGTGATGGTGATGGTGATGC (Used with EGFP-pSUA3-E-up)	1030	PCR pdt.

Table 2-1. Primer information for constructing pSUA2, pSUA3, pSUA4, and pSUA5

described in the previous paragraph digested with the same restriction enzymes, resulting in pSUA2.

All constructs were verified by restriction mapping and sequencing of each insert in its entirety. Primer information is given in the table. The priming region of each primer is underlined, each restriction site is shown in cyan, each RBS is shown in green, and the N150TAG mutation site is shown in red. Maps of fragments and vectors are given below.

SCOt03p-MjtRNA$^{Tyr}_{CUA}$-SCOt07t fragment sequence. Restriction enzyme sites (BglII, PvuI, and HindIII) are underlined. MjtRNA$^{Tyr}_{CUA}$ sequence is shown in blue and anticodon is underlined. SCOt03p promoter is double-underlined and SCOt07t transcription terminator is dotted underlined.

AGATCTAATTGTGCTGGCCCCGGAATCCGCTAGAGTTCATCACGTCGCCGGGCCGCGCA
AGCGGACCGAACCGACAGCCGGCCGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATC
CGCATGGCATGGGTTCAAATCCCATCGGCCGGACCACACTCACGATCGACGAAGCCCCC
GGCCGCCTCGAGCGGCCGGGGGCTTCGTAAGCTT

The sequence of the humanized eGFP gene is given below. Primer binding sites are underlined, the eGFP coding region is shown in blue with start and stop codons double-underlined, and the His-tag sequence is shown in red.

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGA
CGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT
ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCA
TCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCT

GGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGC

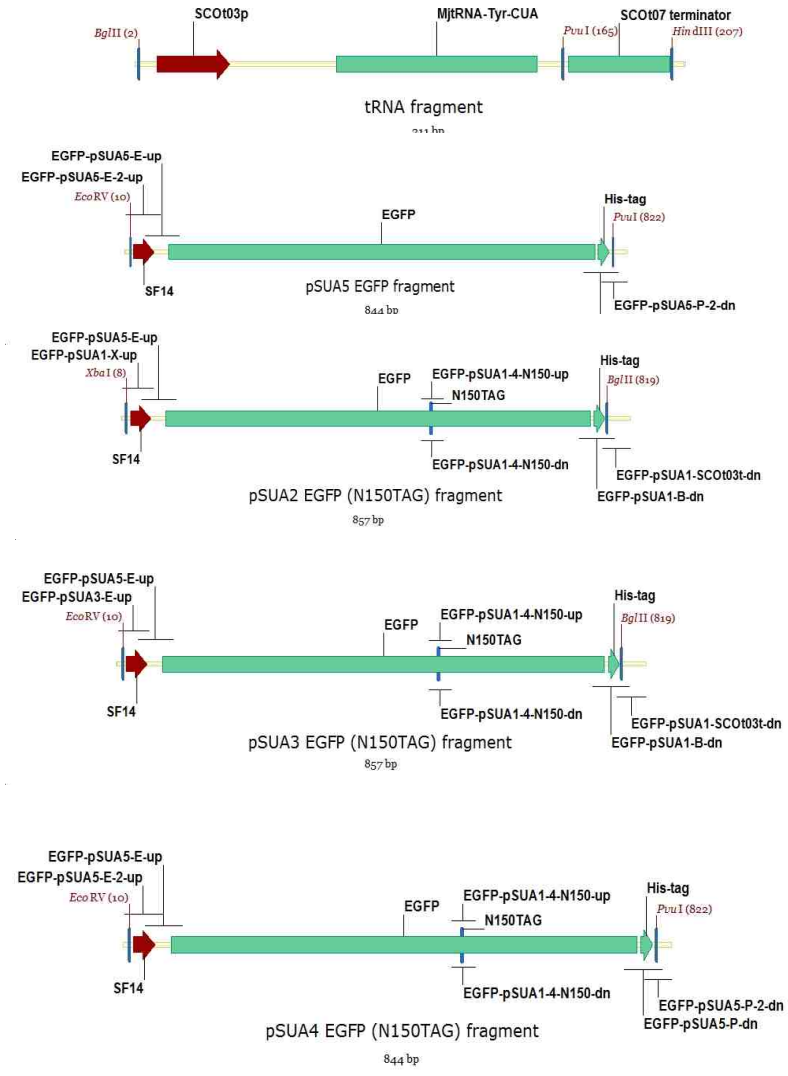
AGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTG

CAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC

CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCG

ATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAG

CTGTACAAGGAGTTGCATCACCATCACCATCACTAA



[redacted]

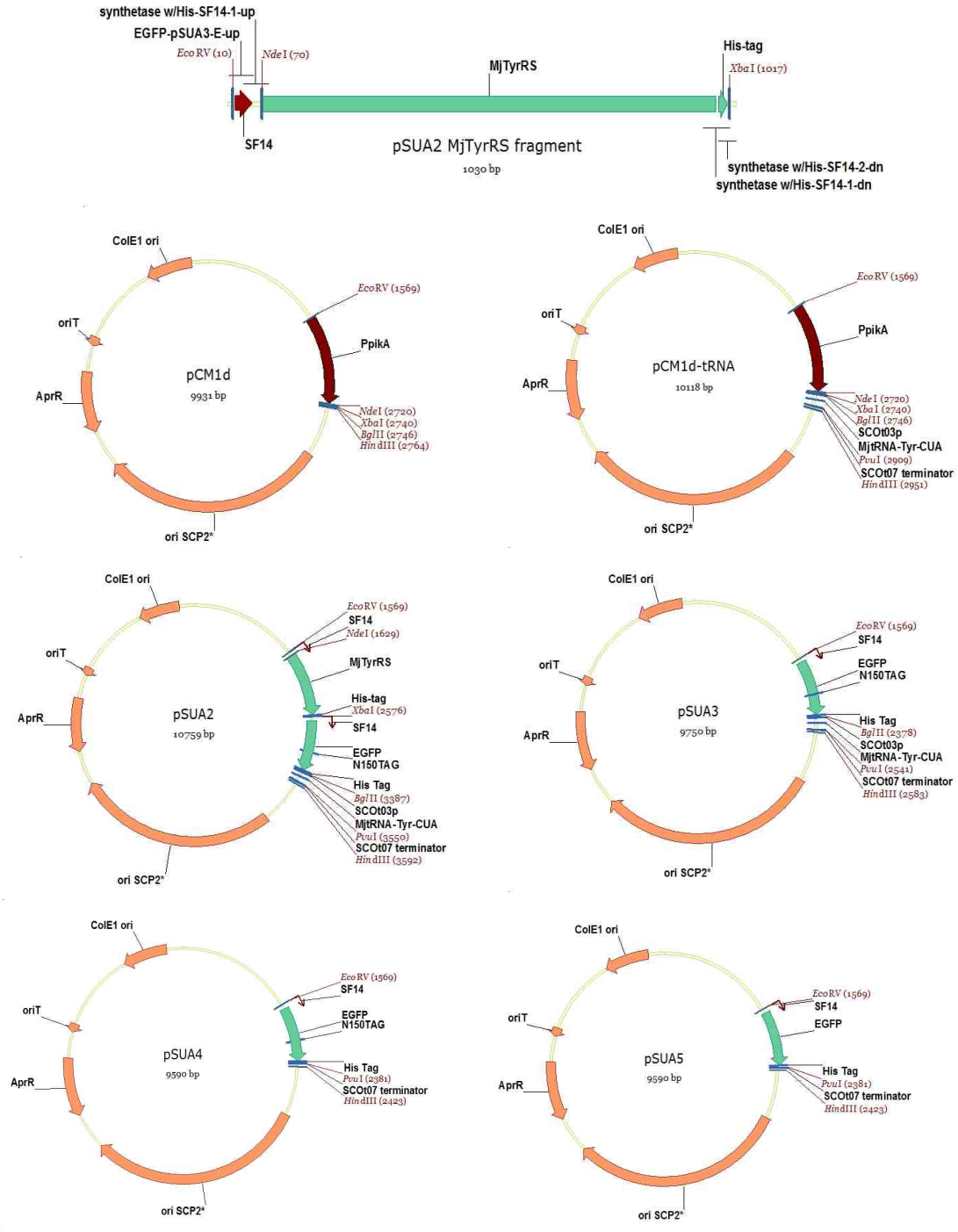


Figure 2-2 (cont.). Map of fragments and vectors of pSUA2, pSUA3, pSUA4, and pSUA5

Construction of opt-pSUA2. opt-pSUA2, which harbors a variant of the MjTyrRS gene that has been codon optimized for expression in Streptomyces (opt-MjTyrRS, sequence given below), was constructed as follows. A 945 bp sequence encoding opt-MjTyrRS with C-terminal His6 tag was designed using the Integrated DNA Technologies (IDT, Coralville, IA) codon optimization tool and Streptomyces coelicolor A3(2) as the target host; and was ordered from IDT as a gBlock. To install the SF14 promoter, synthetic RBS, C-terminal His6 tag, and EcoRV and XbaI restriction sites at the 5'- and 3'-ends respectively, we amplified the opt-MjTyrRS gene using primers opt-synth-1-up and opt-synth-1-dn. This PCR product was used as the template for a second PCR reaction using primers EGFP-pSUA5-E-2-up and opt-synth-1-dn to construct the full length fragment which was digested with EcoRV and XbaI and ligated into pSUA2 digested with these same enzymes, resulting in vector opt-pSUA2. The construct was verified by restriction mapping and sequencing of the insert. Primer information is given in the table below. The priming region of each primer is underlined, the restriction site is shown in cyan, and the RBS is shown in green. A map of the fragment is given below.

Primer Name	**Sequence (5' – 3')**	**amplicon size (bp)**	**template**
opt-synth-1-up	GCTACAATCAATAGTCGATTAGAGGAGAAATTAGAAATGGAC GAGTTCGAGATGATCAAG	996	opt-MjTyrRS
opt-synth-1-dn	TGGCTAGGCTCTAGATCAGTGGTGGTGATGGTGG		
EGFP-pSUA5-E-2-up	Same as used to construct pSUA4 (Used with opt-synth-1-dn)	1033	PCR pdt.

Table 2-2. Primer information for constructing opt-pSUA2

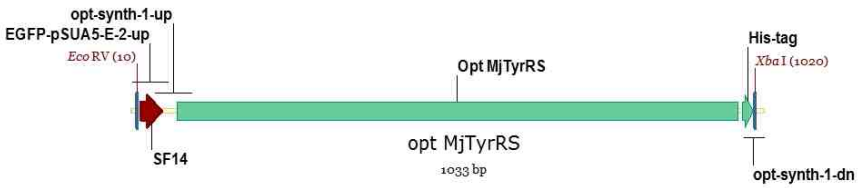


Figure 2-3. Map of fragment for making opt-pSUA2

The sequence of opt-MjTyrRS is given below. Primer binding sites are underlined, the MjTyrRS coding region is shown in blue with start and stop codons double-underlined, and the His-tag sequence is shown in red.

ATGGACGAGTTCGAGATGATCAAGCGGAACACCTCGGAGATCATCAGCGAGGAGGAGCT

GCGCGAGGTCCTCAAGAAGGACGAAAAGAGCGCGTACATCGGCTTCGAACCGTCGGGGA

AGATCCACCTGGGCCACTACCTGCAGATCAAGAAGATGATCGACCTCCAGAACGCGGGC

TTCGACATCATCATCCTCCTGGCGGACCTCCACGCGTACCTGAACCAGAAGGGCGAGCT

CGACGAGATCCGCAAGATCGGCGACTACAACAAGAAGGTCTTCGAAGCCATGGGCCTGA

AGGCCAAGTACGTCTACGGCTCCGAATTCCAGCTGGACAAGGACTACACGCTCAACGTG

TACCGGCTCGCCCTGAAGACCACCCTCAAGCGCGCCCGCCGCTCCATGGAGCTCATCGC

GCGGGAGGACGAGAACCCCAAGGTCGCCGAGGTGATCTACCCGATCATGCAGGTCAACG

ACATCCACTACCTCGGGGTCGACGTGGCCGTCGGGGGCATGGAGCAGCGCAAGATCCAC

ATGCTCGCCCGGGAGCTGCTGCCGAAGAAGGTCGTGTGCATCCACAACCCGGTGCTGAC

GGGGCTGGACGGCGAGGGGAAGATGTCCAGCTCCAAGGGCAACTTCATCGCCGTCGACG

ACTCGCCCGAGGAGATCCGCGCCAAGATCAAGAAGGCGTACTGCCCGGCCGGCGTGGTG

GAGGGCAACCCGATCATGGAGATCGCGAAGTACTTCCTCGAGTACCCGCTGACCATCAA

GCGCCCGGAAAAGTTCGGCGGCGACCTGACCGTCAACTCGTACGAGGAGCTGGAGTCCC

TCTTCAAGAACAAGGAGCTGCACCCGATGGACCTGAAGAACGCGGTCGCGGAGGAGCTC

ATCAAGATCCTGGAGCCGATCCGCAAGCGCCTCGAGCTGCACCACCATCACCACCACTG

A

Construction of opt-pSUA2 (no His). opt-pSUA2 (no His), which is identical to opt-pSUA2 except that the C-terminal His6 tag has been removed, was constructed as follows. opt-MjTyrRS (His tag removed) was amplified from opt-pSUA2 using primers EGFP-pSUA5-E-2-up and Synth-His-rm-dn. This fragment was digested with EcoRV and XbaI and ligated into pSUA2 digested with the same restriction enzymes, resulting in opt-pSUA2 (no His). The construct was verified by restriction mapping and sequencing of the insert. Primer information is given in the table below. The priming region of each primer is underlined and the restriction site is shown in cyan. A map of the fragment is given below.

Primer Name	***Sequence (5' – 3')***	***amplicon size (bp)***	***template***
Synth-His-rm-dn	TGGCTAGGCTCTAGATCAGAGGCGCTTGCGGATCGGCTC (used with EGFP-pSUA5-E-2-up)	1009	opt-pSUA2

Table 2-3. Primer information for constructing opt-pSUA2 (no His*)*

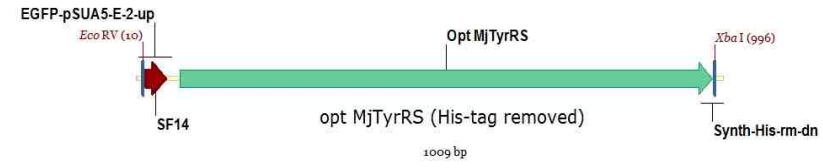


Figure 2-4. Map of fragment for making opt-pSUA2 (no His)

Construction of pSUA1-pBpaRS, pSUA1-pAzPheRS, pSUA1-pIPheRS, opt-pSUA2-sfGFP, and pSUA5-sfGFP. pSUA1-pBpaRS, pSUA1-pAzPheRS, and pSUA1-pIPheRS are derivatives of opt-pSUA2 (no His) in which opt-MjTyrRS (no His) has been replaced with genes encoding evolved MjTyrRS variants pBpaRS, pAzPheRS, and pIPheRS, each with 5 mutations, that are capable of accepting pBpa, pAzPhe, and pIPhe as substrates, respectively; and eGFP-150TAG has been replaced with sfGFP-151TAG,

an sfGFP[23] variant that has been codon optimized for expression in *Streptomyces* and contains an amber stop codon at position 151 and a C-terminal His6 tag. opt-pSUA2-sfGFP is a derivative of opt-pSUA2 in which eGFP-150TAG has been replaced with sfGFP-151TAG. pSUA5-sfGFP is a derivative of pSUA5 in which eGFP has been replaced with sfGFP, which is identical to sfGFP-151TAG except position 151 is a sense codon encoding the native amino acid. To construct these 5 vectors, 3 opt-MjTyrRS variants (pBpaRS, pAzPheRS, and pIPheRS), sfGFP-151TAG, and sfGFP, were generated by PCR.

pBpaRS, pAzPheRS, and pIPheRS were constructed, each from 4 fragments (F1-F4), by one pot multiple fragment overlap extension PCR. pBpaRS F1, pBpaRS F2, pBpaRS F3 and pBpaRS F4 were amplified from opt-pSUA2 (no His) using primer pairs EGFP-pSUA5-E-2-up and opt-synth-32-dn, opt-synth-32-up and opt-synth-103-dn, opt-synth-103-up and opt-synth-158-dn, and opt-synth-158-up and synth-His-rm-dn, respectively. These 4 fragments were assembled and amplified in a one pot overlap extension PCR reaction using primers EGFP-pSUA5-E-2-up and synth-His-rm-dn. pAzPheRS F1, pAzPheRS F2, pAzPheRS F3, and pAzPheRS F4 were amplified from opt-pSUA2 (no His) using primer pairs EGFP-pSUA5-E-2-up and synth-az32-dn, synth-az-io-32-up and synth-io107-dn, synth-az107-up and synth-az158-dn, and synth-az158-up and synth-His-rm-dn, respectively. These 4 fragments were assembled and amplified in a one pot overlap extension PCR reaction using primers EGFP-pSUA5-E-2-up and synth-His-rm-dn. pIPheRS F1, pIPheRS F2, pIPheRS F3, and pIPheRS F4 were amplified from opt-pSUA2 (no His) using primer pairs EGFP-pSUA5-E-2-up and synth-io32-dn, synth-az-io32-up and synth-io107dn, synth-io107-up and synth-io158-dn, and

synth-io158-up and synth-His-rm-dn, respectively. These 4 fragments were assembled and amplified in a one pot overlap extension PCR reaction using primers EGFP-pSUA5-E-2-up and synth-His-rm-dn. pBpaRS, pAzPheRS, and pIPheRS were each digested with EcoRV and XbaI and ligated into pSUA2 digested with the same restriction enzymes to generate 3 intermediate constructs that, upon introduction of sfGFP-151TAG, will generate pSUA1-pBpaRS, pSUA1-pAzPheRS, and pSUA1-pIPheRS.

To construct sfGFP-151TAG and sfGFP, each with the SF14 promoter, synthetic RBS, C-terminal His6 tag, and the appropriate restriction enzyme sites at the 5'- and 3'-ends (XbaI and BglII for sfGFP-151TAG, and EcoRV and PvuI for sfGFP), a 750 bp sequence encoding sfGFP with C-terminal His6 tag was designed using the Integrated DNA Technologies (IDT, Coralville, IA) codon optimization tool and *Streptomyces coelicolor* A3(2) as the target host; and was ordered from IDT as a gBlock. This synthetic sfGFP gene served as the template for amplification of the fragments used to construct sfGFP-151TAG and sfGFP. To construct sfGFP-151TAG, two fragments with the 151TAG codon at their junction were amplified. The upstream fragment was amplified using primers EGFP-pSUA1-X-up and sfGFP-Y151TAG-dn; and the downstream fragment was amplified using primers sfGFP-Y151TAG-up and His-BglII-dn. These were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn to generate sfGFP-151TAG, which was digested with XbaI and BglII. This digested insert was ligated into opt-pSUA2 (no His) digested with the same restriction enzymes to generate opt-pSUA2-sfGFP; and was ligated into the 3 intermediate constructs described in the previous paragraph, pSUA2 digested with the same restriction enzymes, to generate pSUA1-pBpaRS, pSUA1-pAzPheRS, and pSUA1-pIPheRS. To

construct sfGFP, a single fragment was amplified using primers SF14-sfGFP-up and His-PvuI-dn. This PCR product was used as the template for a second PCR reaction using primers EGFP-pSUA5-E-2-up and His-PvuI-dn to generate sfGFP, which was digested with EcoRV and PvuI and ligated into pSUA5 digested with the same restriction enzymes to generate pSUA5-sfGFP.

All constructs were verified by restriction mapping and sequencing of each insert. Primer information is given in the table below. The priming region of each primer is underlined, the restriction site is shown in cyan, mutated codons in aaRS are shown in blue, and the N151TAG mutation site is shown in red. Maps of fragments and representative vector pSUA1-pIPheRS are given below.

Name		position					
		32	103	107	158	159	162
Opt-MjTyrRS	amino acid	Y	V	E	D	I	L
	codon usage	TAC	GTC	GAA	GAC	ATC	CTC
pIPheRS	amino acid	L	—	S	P	L	E
	codon usage	CTC	—	TCC	CCC	CTC	GAG
pAzPheRS	amino acid	T	—	N	P	L	Q
	codon usage	ACC	—	AAC	CCC	CTC	CAG
pBpaRS	amino acid	G	L	P	T	S	—
	codon usage	GGG	CTC	CCC	ACG	AGC	—

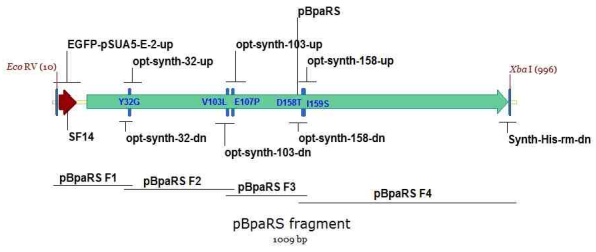
Table 2-4. Mutations in pIPheRS, pAzPheRS, and pBpaRS

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
opt-synth-32-dn	CGATCCCCGCGCTCTTTTCGTCCTTCTTGA (Used with EGFP-pSUA5-E-2-up)	173	opt-pSUA2 (no His)
opt-synth-32-up	CGAAAAGAGCGCGGGGATCGGCTTCGAACCGTCGGGGAAG	241	opt-pSUA2 (no His)
opt-synth-103-dn	GGGGGAGCCGTAGAGGTACTTGGCCTTCAGGCCCATGGCTT		
opt-synth-103-up	GTACCTCTACGGCTCCCCCTTCCAGCTGGACAAGGACTACAC GC	178	opt-pSUA2 (no His)
opt-synth-158-dn	GTGGCTCGTGTTGACCTGCATGATCGGGTAGATCACCTCG		
opt-synth-158-up	TGCAGGTCAACACGAGCCACTACCTCGGGGTCGACGTGGC (Used with Synth-His-rm-dn)	476	opt-pSUA2 (no His)
EGFP-pSUA5-E-2-up	Used with Synth-His-rm-dn	1009	PCR pdts. assembly
synth-az32-dn	CCGACGGTTCGAAGCCGATGGTCGCGCTCTTTTCGTCC (Used with EGFP-pSUA5-E-2-up)	188	opt-pSUA2 (no His)
synth-az-io32-up	ATCGGCTTCGAACCGTCGGGGAAGATCCACC	225	opt-pSUA2 (no His)
synth-az107-dn	GTTGGAGCCGTAGACGTACTTGGCCTTCAGGCC		
synth-az107-up	GTACGTCTACGGCTCCAACTTCCAGCTGGACAAGGACTACAC GC	184	opt-pSUA2 (no His)
synth-az158-dn	CTGGTAGTGGAGGGGGTTGACCTGCATGATCGGGTAGAT		
synth-az158-up	CAACCCCCTCCACTACCAGGGGGTCGACGTGGCCG (Used with Synth-His-rm-dn)	469	opt-pSUA2 (no His)
synth-io32-dn	CCGACGGTTCGAAGCCGATGAGCGCGCTCTTTTCGTCC (used with EGFP-pSUA5-E-2-up)	188	opt-pSUA2 (no His)
synth-io107-dn	GGAGGAGCCGTAGACGTACTTGGCCTTCAGGCC (Used with synth-az-io32-up)	225	opt-pSUA2 (no His)
synth-io107-up	GTACGTCTACGGCTCCTCCTTCCAGCTGGACAAGGACTACAC GC	184	opt-pSUA2 (no His)
synth-io158-dn	CTCGTAGTGGAGGGGGTTGACCTGCATGATCGGGTAGAT		
synth-io158-up	CAACCCCCTCCACTACGAGGGGGTCGACGTGGCCG (Used with Synth-His-rm-dn)	469	opt-pSUA2 (no His)
sfGFP-Y151TAG-dn	CGCGGTGATCTACACGTTGTGGGAGTTGAAGTTGT (Used with EGFP-pSUA1-X-up)	535	synthetic sfGFP
sfGFP-Y151TAG-up	CAACGTGTAGATCACCGCGGACAAGCAGAAGAACG	311	synthetic sfGFP
His-BglII-dn	CACAATTAGATCTTCAGTGGTGGTGATGGTGGTG		
EGFP-pSUA1-X-up	Used with His-BglII-dn	827	PCR pdts. assembly
SF14-sfGFP-up	GGCGTGAGCTACAATCAATAGTCGATTAGAGGAGAAATTAGA AATGAGCAAGGGCGAGGA	798	synthetic sfGFP
His-PvuI-dn	GCATATATCGATCGTCAGTGGTGGTGATGGTGGTG		
EGFP-pSUA5-E-2-up	Used with His-PvuI-dn	828	PCR pdt.

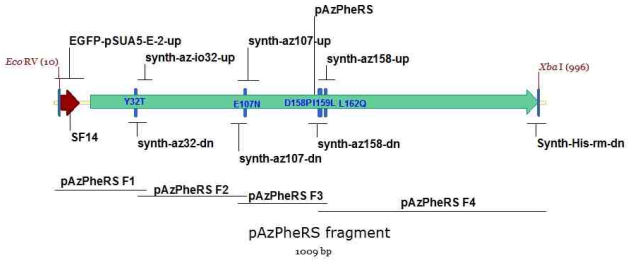
Table 2-5. Prime information for constructing pSUA1-pBpaRS, pSUA1-pAzPheRS, pSUA1-pIPheRS, opt-pSUA2-sfGFP, and pSUA5-sfGFP

The sequence of sfGFP is given below. Primer binding sites are underlined, the sfGFP coding region is shown in blue with start and stop codons double underlined, and the His tag sequence is shown in red.

AAATTAGAAATGAGCAAGGGCGAGGAGCTGTTCACGGGCGTCGTCCCCATCCTCGTCGA
GCTGGACGGCGACGTCAACGGGCACAAGTTCTCCGTCCGCGGCGAGGGCGAGGGGGACG
CCACGAACGGCAAGCTGACGCTGAAGTTCATCTGCACCACGGGCAAGCTCCCCGTGCCC
TGGCCGACCCTGGTCACCACGCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA
CCACATGAAGCGGCACGACTTCTTCAAGTCCGCCATGCCGGAAGGCTACGTGCAGGAGC
GCACGATCTCGTTCAAGGACGACGGCACCTACAAGACGCGGGCGGAAGTCAAGTTCGAG
GGCGACACCCTCGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGCCACAAGCTCGAGTACAACTTCAACTCCCACAACGTGTACATCACCGCGG
ACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCCACAACGTGGAGGACGGG
TCGGTGCAGCTGGCGGACCACTACCAGCAGAACACGCCCATCGGCGACGGCCCGGTGCT
GCTGCCCGACAACCACTACCTCTCCACGCAGTCCGTCCTGTCCAAGGACCCGAACGAGA
AGCGGGACCACATGGTCCTCCTGGAGTTCGTGACCGCCGCCGGCATCACCCACGGGATG
GACGAGCTCTACAAGGGCTCGCACCACCATCACCACCACTGA

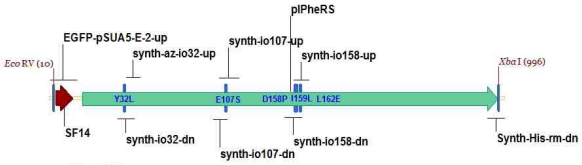


[redacted]



pAzPheRS fragment

1009 bp

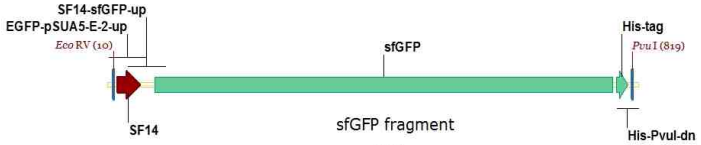


plPheRS fragment

1009 bp

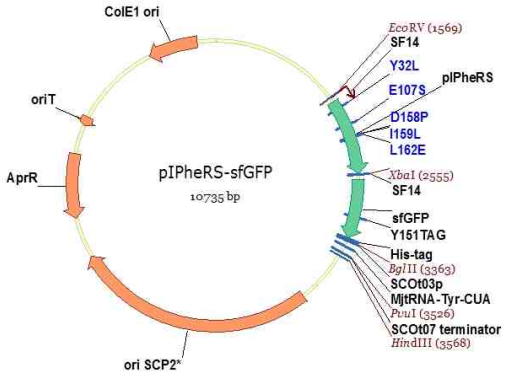
sfGFP (Y151TAG) fragment

827 bp



sfGFP fragment

828 bp



[redacted]

Construction of pSUA1-pAzPheRS-2, pSUA1-pAzPheRS-4, pSUA1-pAzPheRS-5. pSUA1-pAzPheRS-2, pSUA1-pAzPheRS-4, and pSUA1-pAzPheRS-5[7] are derivatives of pSUA1-pAzPheRS in which pAzPheRS is replaced with alternative pAzPheRS sequences pAzPheRS-2, pAzPheRS-4, and pAzPheRS-5, each with 4-5 mutations, that are reported to be capable of accepting pAzPhe as a substrate. To construct these vectors, 3 pAzPheRS variants (pAzPheRS-2, pAzPheRS-4, and pAzPheRS-5), were generated by PCR.

pAzPheRS-2 was constructed from 3 fragments (F1-3) and pAzPheRS-4 and pAzPheRS-5 were each constructed from 4 fragments (F1-4) by one pot multiple fragment overlap extension PCR. pAzPheRS-2 F1, pAzPheRS-2 F2, and pAzPheRS-2 F3 were amplified from pSUA1-pAzPheRS using primer pairs EGFP-pSUA5-E-2-up and synth-io107-dn, synth-io107-up and synth-2az158-dn, and synth-2az158-up and synth-His-rm-dn, respectively. These 3 fragments were assembled and amplified in a one pot overlap extension PCR reaction using primers EGFP-pSUA5-E-2-up and synth-His-rm-dn. pAzPheRS-4 F1, pAzPheRS-4 F2, and pAzPheRS-4 F3, and pAzPheRS-4 F4 were amplified from pSUA1-pAzPheRS using primer pairs EGFP-pSUA5-E-2-up and synth-io32-dn, synth-az-io32-up and synth-com107-dn, synth-af107-up and synth-4az158-dn, and synth-4az158-up and synth-His-rm-dn, respectively. These 4 fragments were assembled and amplified a in one pot overlap extension PCR reaction using primers EGFP-pSUA5-E-2-up and synth-His-rm-dn. pAzPheRS-5 F1, pAzPheRS-5 F2, pAzPheRS-5 F3 and pAzPheRS-5 F4 were amplified from pAzPheRS-sfGFP using primer pairs EGFP-pSUA5-E-2-up and synth-5az32-dn, synth-az-io32-up and synth-com107-dn, synth-5az107-up and synth-5az158-dn, and synth-5az158-up and synth-His-

rm-dn, respectively. These 4 fragments were assembled and amplified in a one pot overlap extension PCR reaction using primers EGFP-pSUA5-E-2-up and synth-His-rm-dn. pAzPheRS-2, pAzPheRS-4, and pAzPheRS-5 were each digested with EcoRV and XbaI and ligated into opt-pSUA2-sfGFP digested with the same restriction enzymes to generate pSUA1-pAzPheRS-2, pSUA1-pAzPheRS-4, and pSUA1-pAzPheRS-5.

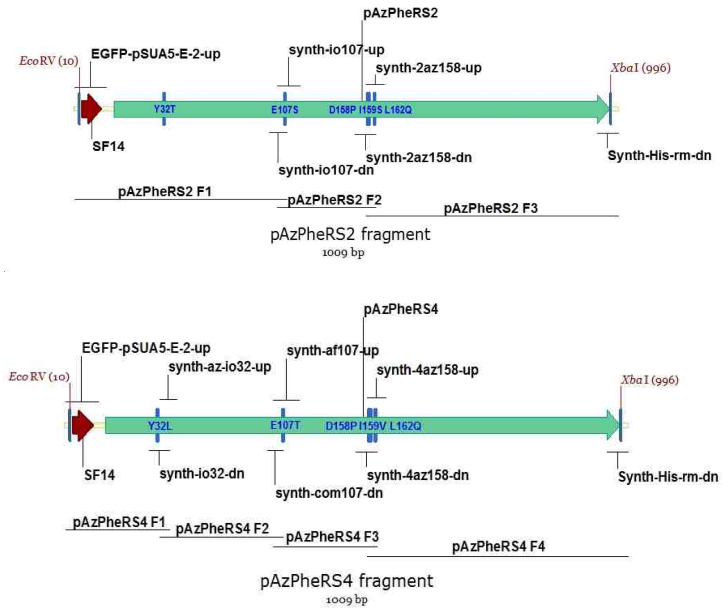
All constructs were verified by restriction mapping and sequencing of each insert. Primer information is given in the table below. The priming region of each primer is underlined and mutated codons in aaRS are shown in blue. Maps of fragments are given below.

Name		position				
		32	107	158	159	162
Opt-MjTyrRS	amino acid	Y	E	D	I	L
	codon usage	TAC	GAA	GAC	ATC	CTC
pAzPheRS2	amino acid	T	S	P	S	Q
	codon usage	ACC	TCC	CCC	TCC	CAG
pAzPheRS4	amino acid	L	T	P	V	Q
	codon usage	CTC	ACC	CCC	GTC	CAG
pAzPheRS5	amino acid	A	R	V	I	D
	codon usage	GCC	CGC	GTC	ATC	GAC

Table 2-6. Mutations in pAzPheRS2, pAzPheRS4, and pAzPheRS5

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
EGFP-pSUA5-E-2-up	Used with synth-io107-dn	394	pSUA1-pAzPheRS
synth-2az158-dn	CTGGTAGTGGGAGGGGTTGACCTGCATGATCGGGTAGAT (Used with synth-io107-up)	184	pSUA1-pAzPheRS
synth-2az158-up	CAACCCCTCCCACTACCAGGGGGTCGACGTGGCCG (Used with Synth-His-rm-dn)	469	pSUA1-pAzPheRS
EGFP-pSUA5-E-2-up	Used with synth-io32-dn	188	pSUA1-pAzPheRS
synth-com107-dn	GGAGCCGTAGACGTACTTGGCCTTCAGGCC (Used with synth-az-io32-up)	222	pSUA1-pAzPheRS
synth-af107-up	CAAGTACGTCTACGGCTCCACCTTCCAGCTGGACAAGGACTACACGC	187	pSUA1-pAzPheRS
synth-4az158-dn	CTGGTAGTGGACGGGGTTGACCTGCATGATCGGGTAGAT		
synth-4az158-up	CAACCCCGTCCACTACCAGGGGGTCGACGTGGCCG (Used with Synth-His-rm-dn)	469	pSUA1-pAzPheRS
synth-5az32-dn	CCGACGGTTCGAAGCCGATGGCCGCGCTCTTTTCGTCC (Used with EGFP-pSUA5-E-2-up)	188	pSUA1-pAzPheRS
synth-5az107-up	CAAGTACGTCTACGGCTCCCGCTTCCAGCTGGACAAGGACTACACGC	187	pSUA1-pAzPheRS
synth-5az158-dn	GTCGTAGTGGATGACGTTGACCTGCATGATCGGGTAGAT		
synth-5az158-up	CAACGTCATCCACTACGACGGGGTCGACGTGGCCG (Used with Synth-His-rm-dn)	469	pSUA1-pAzPheRS

Table 2-7. Prime information for constructing pSUA1-pAzPheRS-2, pSUA1-pAzPheRS-4, and pSUA1-pAzPheRS-5



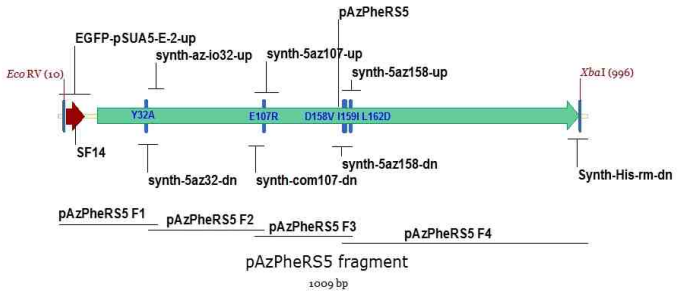


Figure 2-6 (cont.). Map of fragments for constructing pSUA1-pAzPheRS-2, pSUA1-pAzPheRS-4, and pSUA1-pAzPheRS-5

Construction of pSUA1-pBpaRS-tufA, pSUA1-pAzPheRS-tufA, pSUA1-pIPheRS-tufA, and pSUA2-tufA. pSUA1-pBpaRS-tufA, pSUA1-pAzPheRS-tufA, pSUA1-pIPheRS-tufA, and pSUA2-tufA are derivatives of pSUA1-pBpaRS, pSUA1-pAzPheRS, pSUA1-pIPheRS, and opt-pSUA2-sfGFP, respectively, which harbor a codon optimized tufA gene. The tufA gene is derived from *E. coli* and has been codon optimized for expression in *Streptomyces*, and was cloned into the downstream of SCOt07 terminator using HindIII site. A 1264 bp sequence (sequence given below) encoding tufA gene, SF14 promoter, synthetic RBS, and HindIII at both sides was obtained by joining a 500 bp and a 750 bp fragments ordered from IDT as gBlocks using primers HindIII-SF14-up and HindIII-tufA-dn. The PCR product was digested with HindIII and ligated into pSUA1-pBpaRS, pSUA1-pAzPheRS, pSUA1-pIPheRS, and opt-pSUA2-sfGFP digested with HindIII enzyme, resulting in pSUA1-pBpaRS-tufA, pSUA1-pAzPheRS-tufA, pSUA1-pIPheRS-tufA, and pSUA2-tufA. The constructs were verified by restriction mapping and sequencing of the insert. Primer information is given in the table below. The priming region of each primer is underlined, and the restriction site is shown in cyan. A map of the fragment is given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
HindIII-SF14-up	CCCAAGCTTGGTTGACCTTGATGAGGCGGCGTGAGCTACAATCAATAGTC	1264	tufA1 & 2
HindIII-tufA-dn	CCCAAGCTTTCAACCCAGCACCTTCGCCACCACACCGGC		

Table 2-8. Prime information for constructing the tufA fragment

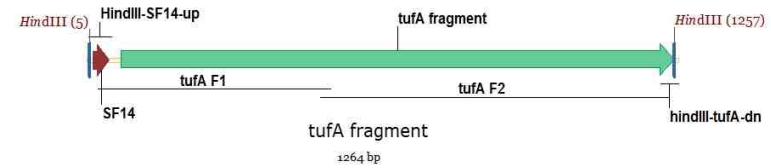


Figure 2-7. Map of tufA fragment

The sequence of SF14-tufA is given below. Primer binding sites are underlined, the SF14 promoter is shown in red, and the tufA coding region is shown in blue with start and stop codons double-underlined.

CCCAAGCTTGGTTGACCTTGATGAGGCGGCGTGAGCTACAATCAATAGTCGATTAGAGG

AGAAATTAGAAATGAGCAAGGAAAAGTTCGAACGCACGAAGCCGCACGTCAACGTCGGC

ACCATCGGGCACGTCGACCACGGCAAGACCACCCTGACGGCCGCCATCACCACGGTCCT

CGCCAAGACCTACGGTGGTGCGGCCCGCGCCTTCGACCAGATCGACAACGCCCCGGAAG

AAAAGGCCCGGGGCATCACGATCAACACCAGCCACGTGGAATACGACACGCCCACCCGG

CACTACGCCCACGTCGACTGCCCGGGCCACGCGGACTACGTCAAGAACATGATCACGGG

GGCCGCGCAGATGGACGGGGCCATCCTCGTGGTCGCCGCCACGGACGGTCCCATGCCGC

AGACCCGCGAGCACATCCTCCTGGGCCGCCAGGTCGGCGTCCCCTACATCATCGTCTTC

CTGAACAAGTGCGACATGGTCGACGACGAAGAGCTGCTCGAGCTGGTCGAGATGGAGGT

CCGGGAGCTCCTGTCCCAGTACGACTTCCCCGGCGACGACACCCCCATCGTGCGCGGCT

CCGCGCTGAAGGCCCTCGAAGGCGACGCGGAGTGGGAGGCCAAGATCCTGGAGCTGGCC

GGGTTCCTCGACTCGTACATCCCGGAGCCCGAGCGCGCCATCGACAAGCCCTTCCTGCT

GCCCATCGAGGACGTGTTCTCCATCAGCGGCCGGGGCACCGTCGTCACGGGCCGCGTGG

AGCGCGGCATCATCAAGGTCGGGGAGGAGGTCGAGATCGTCGGCATCAAGGAGACGCAG

AAGTCCACCTGCACGGGCGTCGAGATGTTCCGCAAGCTGCTGGACGAGGGGCGCGCCGG

CGAGAACGTCGGTGTCCTGCTCCGGGGTATCAAGCGCGAAGAGATCGAGCGGGGGCAGG

TGCTCGCCAAGCCGGGGACCATCAAGCCGCACACCAAGTTCGAGTCCGAAGTGTACATC

CTCAGCAAGGACGAGGGTGGGCGCCACACCCCGTTCTTCAAGGGCTACCGGCCCCAGTT

CTACTTCCGGACCACCGACGTGACGGGCACCATCGAGCTGCCGGAGGGCGTGGAGATGG

TGATGCCCGGCGACAACATCAAGATGGTGGTCACCCTGATCCACCCCATCGCGATGGAC

GACGGCCTGCGGTTCGCCATCCGGGAGGGCGGTCGGACCGTCGGTGCCGGTGTGGTGGC

GAAGGTGCTGGGTTGAAAGCTTGGG

Construction of pSUA1-pBpaRS-Y66TAG. pSUA1-pBpaRS-Y66TAG, which is a derivative of pSUA1-pBpaRS in which sfGFP-151TAG has been replaced by sfGFP-66TAG, was constructed as follows. sfGFP-Y66TAG was constructed by overlap extension PCR from two fragments amplified from pSUA5-sfGFP. The upstream fragment was amplified using primers EGFP-pSUA1-X-up and sfGFP-Y66TAG-dn; and the downstream fragment was amplified using primers sfGFP-Y66TAG-up and His-BglII-dn. These fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn and the resulting PCR product was digested with XbaI and BglII and ligated into pSUA1-pBpaRS digested with the same restriction enzymes, resulting in pSUA1-pBpaRS-Y66TAG. The new construct was verified by restriction mapping and sequencing. Primer information is given in the table below. The priming region of each primer is underlined and the Y66TAG mutation site is shown in red. A map of the fragment is given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
sfGFP-Y66TAG-dn	CCCTAGGTCAGCGTGGTGACCAGGGTCGGCC (Used with EGFP-pSUA1-X-up)	273	pSUA5-sfGFP
sfGFP-Y66TAG-up	TCACCACGCTGACCTAGGGCGTGCAGTGCT (Used with His-BglII-dn)	573	pSUA5-sfGFP
EGFP-pSUA1-X-up	Same as used to construct pSUA2, used with His-BglII-dn	827	PCR pdts. assembly

Table 2-9. Prime information for constructing pSUA1-pBpaRS-Y66TAG

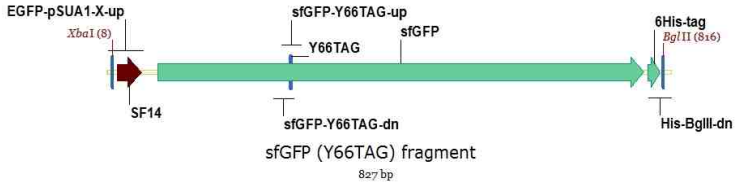


Figure 2-8. Map of sfGFP (Y66TAG) fragment

Construction of pSUA1-pyl, pSUA2-pyl, and pSUA3-pyl. pSUA2-pyl harbors wild-type *M. barkeri* pyrrolysyl-tRNA synthetase, sfGFP/Y151TAG gene, and *M. barkeri* $tRNA^{Pyl}_{CUA}$ with EcoRV and XbaI, XbaI and EcoRI, EcoRI and PvuI at 5'- and 3'- ends, respectively. The *M. barkeri* pyrrolysyl-tRNA synthetase gene was codon optimized for expression in *Streptomyces* and under the control of promoter SF14. pSUA2-pyl is a similar construct as opt-pSUA2-sfGFP, in which MjTyrRS and $tRNA^{Tyr}_{CUA}$ was replaced by *M. barkeri* pyrrolysyl-tRNA synthetase (pylRS) and $tRNA^{Pyl}_{CUA}$ pair, and BglII site was replaced by EcoRI due the existence of BglII in pylRS gene. A 1510 bp sequence encoding pylRS and $tRNA^{Pyl}_{CUA}$ was made by joining a 1000 bp and a 500 bp fragments ordered from IDT as gBlocks using primers EGFP-pSUA5-E-2-up and tRNA-Pyl-PvuI-dn. The resulting PCR product was digested with EcoRV, XbaI, EcoRI, and PvuI to have EcoRV-pylRS-XbaI and EcoRI-$tRNA^{Pyl}_{CUA}$-PvuI fragments. The XbaI-sfGFP/Y151TAG-EcoRI fragment was obtained from digestion of PCR product amplified from opt-pSUA2-sfGFP using primers EGFP-pSUA1-X-up and

His-EcoRI-dn. The resulting three fragments were ligated into opt-pSUA2-sfGFP digested with EcoRV and PvuI by four-way ligation method to generate pSUA2-pyl. The EcoRV-sfGFP/Y151TAG-EcoRI fragment was obtained from digesting the PCR product amplified from opt-pSUA2-sfGFP using primers EGFP-pSUA5-E-2-up and His-EcoRI-dn. The resulting fragment was ligated into pSUA2-pyl digested with EcoRV and EcoRI to generate pSUA3-pyl.

DZKRS was constructed from 4 fragments (F1-F4), by one pot multiple fragment overlap extension PCR. DZKRS F1, DZKRS F2, DZKRS F3, and DZKRS F4 were amplified from pSUA2-pyl using primer pairs EGFP-pSUA5-E-2-up and DZKRS-L274A-dn, DZKRS-L274A-up and DZKRS-C313S-dn, DZKRS-C313S-up and DZKRS-Y349F-dn, DZKRS-Y349F-up and DZKRS-XbaI-dn, respectively. These 4 fragments were assembled and amplified in a one pot overlap extension PCR reaction using primers EGFP-pSUA5-E-2-up and DZKRS-XbaI-dn. The resulting PCR product was digested with EcoRV and XbaI and then ligated into pSUA2-pyl digested with same enzymes to generate pSUA1-pyl. All constructs were verified by restriction mapping and sequencing of each insert. Primer information is given in the table below. The priming region of each primer is underlined, the restriction site is shown in cyan, and mutated codons are shown in blue. Maps of fragments and vectors are given below.

Name		Positions		
		274	313	349
Mb-pylRS	amino acid	Y	C	Y
	codon usage	CTG	TGC	TAC
DZKRS	amino acid	A	S	F
	codon usage	GCC	TCC	TTC

Table 2-10. Mutations in DZKRS

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
tRNA-Pyl-PvuI-dn	GCATATATCGATCGTGAGTGTGGCGGAAACCCCGGGAATCTAAC （Used with EGFP-pSUA5-E-2-up)	1510	pylRS1-2 fragments
His-EcoRI-dn	CCGGAATTCTCAGTGGTGGTGATGGTGGTGCGAGCC (Used with EGFP-pSUA1-X-up)	823	opt-pSUA2-sfGFP
His-EcoRI-dn	(Used with EGFP-pSUA5-E-2-up)	823	opt-pSUA2-sfGFP
DZKRS-L274A-dn	CCGGTCGAGCTTGCGGGCGTAGTTGTACAG （Used with EGFP-pSUA5-E-2-up)	910	pSUA2-pyl
DZKRS-L274A-up	CTGTACAACTACGCCCGCAAGCTCGACCGG	147	pSUA2-pyl
DZKRS-C313S-dn	CCCCGACCCCATCTGGGAGAAGTTGACCATCG		
DZKRS-C313S-up	CGATGGTCAACTTCTCCCAGATGGGGTCGGGG	137	pSUA2-pyl
DZKRS-Y349F-dn	CAGGGTGTCCCCGAAGACCATGCAGCTGT		
DZKRS-Y349F-up	ACAGCTGCATGGTCTTCGGGGACACCCTG	245	pSUA2-pyl
DZKRS-XbaI-dn	TGGCTAGGCTCTAGATCAGAGGTTCGTCGAGATGCCGTTGTAGTAGGA		
DZKRS-XbaI-dn	(Used with EGFP-pSUA5-E-2-up)	1348	PCR pdts. assembly

Table 2-11. Prime information for constructing pSUA1-pyl, pSUA2-pyl, and pSUA3-pyl

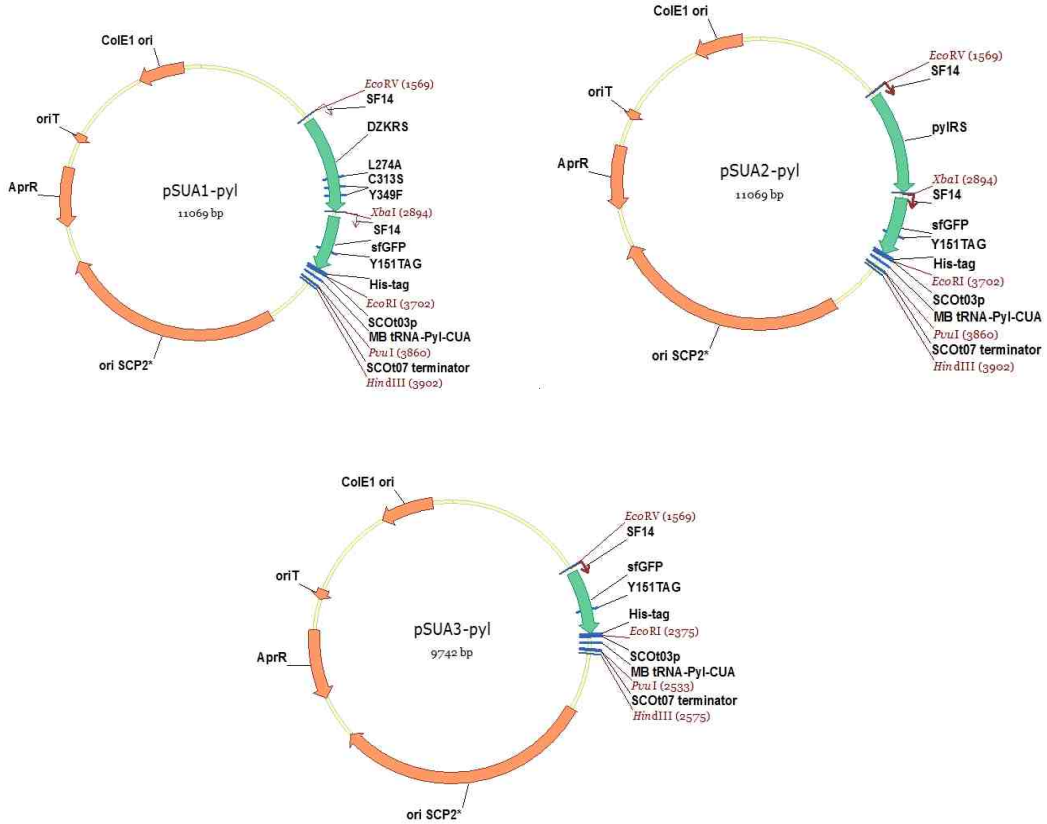


Figure 2-9. Map of vectors pSUA1-pyl, pSUA2-pyl, and pSUA3-pyl.

The sequence of Mb-pylRS is given below with start and stop codons underlined.

ATGGACAAGAAGCCGCTGGACGTCCTGATCAGCGCCACGGGCCTCTGGATGTCGCGGAC

GGGGACGCTCCACAAGATCAAGCACCACGAAGTCAGCCGGTCGAAGATCTACATCGAAA

TGGCCTGCGGGGACCACCTGGTGGTGAACAACAGCCGCAGCTGCCGGACCGCCCGCGCC

TTCCGGCACCACAAGTACCGCAAGACCTGCAAGCGGTGCCGGGTGAGCGACGAAGACAT

CAACAACTTCCTCACCCGCAGCACGGAAAGCAAGAACAGCGTCAAGGTCCGGGTGGTCT

CGGCGCCCAAGGTCAAGAAGGCGATGCCGAAGAGCGTCTCGCGCGCGCCGAAGCCGCTC

GAGAACTCCGTCTCCGCGAAGGCCAGCACGAACACGTCGCGGTCCGTCCCCTCGCCGGC

GAAGTCGACGCCCAACTCCTCGGTGCCCGCCAGCGCGCCGGCGCCCTCCCTGACGCGCT

CGCAGCTGGACCGCGTGGAGGCCCTCCTGTCCCCGGAAGACAAGATCTCCCTGAACATG

GCCAAGCCCTTCCGGGAACTCGAACCCGAACTGGTGACCCGCCGGAAGAACGACTTCCA

GCGGCTCTACACGAACGACCGCGAGGACTACCTCGGGAAGCTCGAACGCGACATCACCA

AGTTCTTCGTGGACCGGGGGTTCCTGGAAATCAAGTCCCCCATCCTGATCCCCGCCGAG

TACGTCGAACGGATGGGGATCAACAACGACACCGAACTGTCGAAGCAGATCTTCCGCGT

GGACAAGAACCTGTGCCTCCGGCCGATGCTCGCGCCCACCCTGTACAACTACCTGCGCA

AGCTCGACCGGATCCTGCCCGGCCCCATCAAGATCTTCGAAGTGGGGCCCTGCTACCGG

AAGGAGAGCGACGGGAAGGAACACCTCGAAGAGTTCACGATGGTCAACTTCTGCCAGAT

GGGGTCGGGGTGCACCCGCGAAAACCTCGAAGCCCTCATCAAGGAGTTCCTCGACTACC

TCGAAATCGACTTCGAAATCGTCGGGGACAGCTGCATGGTCTACGGGGACACCCTGGAC

ATCATGCACGGCGACCTCGAGCTGAGCAGCGCCGTGGTCGGGCCCGTCAGCCTGGACCG

GGAGTGGGGCATCGACAAGCCCTGGATCGGCGCCGGGTTCGGGCTCGAACGCCTGCTCA

AGGTGATGCACGGGTTCAAGAACATCAAGCGCGCCAGCCGCTCCGAATCCTACTACAAC

GGCATCTCGACGAACCTCTGA

The sequence of Mb-tRNA$^{Pyl}_{CUA}$ is given below and anticodon is underlined.

GGGAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCC

GGGGTTTCCGCCA

4. **Results and Discussion**

Function Validation of the Amber Suppression System in Streptomyces. A series of four vectors (Figure 2-10a) were designed and constructed to individually test proper functioning of MjTyrRS, MjtRNA$^{Tyr}_{CUA}$, and the GFP reporter in *S. venezuelae*. Plasmid pSUA5, which contains the wild-type humanized enhanced GFP (eGFP) reporter gene with C-terminal histidine tag (eGFP-His) under control of constitutive *Streptomyces* promoter SF14, was designed to test expression of the reporter gene. Plasmid pSUA4, which contains eGFP-His with an amber stop codon at position 150 (eGFP-150TAG), was designed to test whether *S. venezuelae* has endogenous amber suppression capability. Plasmid pSUA3, which contains eGFP-150TAG and an optimized MjtRNA$^{Tyr}_{CUA}$ Nap1[21] with *Streptomyces coelicolor* tRNA promoter and flanking sequences, was designed to test MjtRNA$^{Tyr}_{CUA}$ orthogonality to endogenous aaRS in *S. venezuelae*, which is critical for system functionality. Plasmid pSUA2, which contains eGFP-150TAG, MjtRNA$^{Tyr}_{CUA}$, and C-terminal histidine tagged MjTyrRS (MjTyrRS-His) under control of the SF14 promoter, was designed to test functionality of the amber suppression system as a whole.

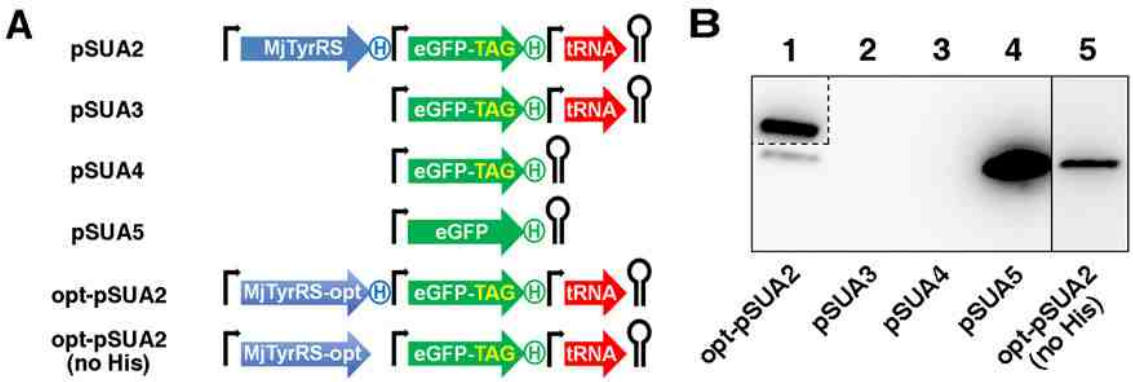


Figure 2-10. (a) Schematics of pSUA2, pSUA3, pSUA4, pSUA5, optpSUA2, and opt-pSUA2 (no His) vector inserts containing elements of the amber suppression reporter system; (b) Anti-His Western blots of Ni-NTA affinity purified eGFP proteins obtained from *S. venezuelae* harboring opt-pSUA2 (lane 1), pSUA3 (lane 2), pSUA4 (lane 3), pSUA5 (lane 4), and opt-pSUA2 lacking the C-terminal His tag on MjTyrRS (lane 5). The molecular weights of MjTyrRS and eGFP are 35 kDa and 27 kDa, respectively. Because of the high expression of MjTyRS-opt relative to eGFP (lane 1), the signal from which interfered with clear visualization of the GFP band, the blot was initially imaged for 5 s, and the upper portion of lane 1 containing MjTyrRS-opt was then excised from the blot (excised portion is marked with dashed lines), and the blot was then imaged for an additional 135 s.

To examine the performance of these constructs, pSUA2, pSUA3, pSUA4, and pSUA5 were individually transferred into *S. venezuelae* by protoplast transformation. The eGFP (or the mixture of eGFP and MjTyrRS in the case of pSUA2) was purified by Ni-NTA affinity. The anti-His western blotting results demonstrat that (Figure 2-10, lane 4) eGFP is expressed in *S. venezuelae* harboring pSUA5, and that (Figure 2-10, lane 3) endogenous amber suppression system is not present in *S. venezuelae*. However, neither amber suppressed eGFP nor MjTyrRS could be deteched in *S. venezuelae* harboring pSUA2. Examination of the MjTyrRS sequence revealed that a large number of codons that occur infrequently in *Streptomyces* genes, including 18 TTA codons, which is the least used codon in *Streptomyces*,[24] which leads to the expression failure of MjTyrRS and eventually expression failure of amber suppressed eGFP protein. To overcome this problem, opt-pSUA2, variant of pSUA2, was design and constructed, which contains the codon optimized MjTyrRS gene (MjTyrRS-opt). The anti-His western blotting result

(Figure 2-10, lane 1) shows a high expression level of MjTyrRS in *S. venezuelae* harboring opt-pSUA2, and more importantly, amber suppressed eGFP is also produced. No protein is detected from *S. venezuelae* harboring pSUA3 (Figure 2-10, lane 2) confirms the MjtRNA$^{Tyr}_{CUA}$ orthogonality to *S. venezuelae* endogenous aaRS. After removal of the C-terminal histidine tag from codon optimized MjTyrRS, the detectable eGFP (Figure 2-10, lane 5) confirms the functionality of this amber suppression system in *Streptomyces* as a whole.

Function Validation of the UAAs Incorporation in *Streptomyces*. Three MjTyrRS-opt variants with the appropriate mutations responsible for specific UAA were designed, constructed and cloned into opt-pSUA2 in the place of MjTyrRS-opt to generate pSUA1-pBpaRS, pSUA1-pAzPheRS, and pSUA1-pIPheRS.

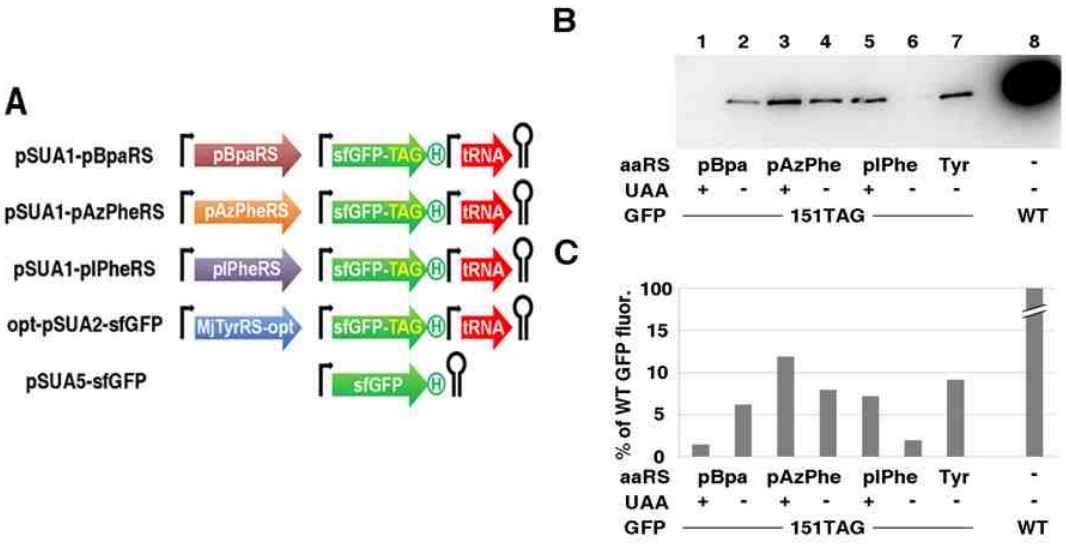


Figure 2-11. (a) Schematics of the pSUA1-pBpaRS, pSUA1-pAzPheRS, pSUA1-pIPheRS, opt-pSUA2-sfGFP, and pSUA5-sfGFP vector inserts. (b) Anti-His Western blot of Ni-NTA affinity purified sfGFP proteins obtained from *S. venezuelae* harboring pSUA1-pBpaRS grown in the presence (lane 1) and the absence (lane 2) of 1 mM pBpa, pSUA1-pAzPheRS grown in the presence (lane 3) and the absence (lane 4) of 1 mM pAzPhe, pSUA1-pIPheRS grown in the presence (lane 5) and the absence (lane 6) of 1 mM pIPhe, opt-pSUA2-sfGFP (lane 7), and pSUA5-sfGFP (lane 8). The blot was imaged for 165 s. (c) Amber suppression efficiencies of the same eight protein samples used for Western blot analysis (b), measured by fluorescence quantification, and normalized to wild-type sfGFP signal. GFP excitation was carried out at 485 nm, bandwidth 9 nm, and GFP emission was detected at 520 nm, bandwidth 15 nm.

These three construct are derivate of opt-pSUA2, in which the MjTyrRS-opt was replaced by pBpaRS, pAzPheRS, and pIPheRS and the amber suppressible eGFP was replaced by a codon optimized improve GFP variant superfolder GFP (sfGFP)[23] with an amber stop codon at the 151 site in the replace of tyrosine. And in the control vector pSUA5, the eGFP was also replaced by the sfGFP gene (Figure 2-11a). These four constructs were individually transferred into *S. venezuelae* and the *S. venezuelae* harboring pSUA1-pBpaRS, pSUA1-pAzPheRS, and pSUA1-pIPheRS were cultured in presence and in absence of its cognate UAAs. The *S. venezuelae* harboring pSUA5 was cultured without adding UAAs. The anti-His western blotting results (Figure 2-11b, lane 1, 3, and 5) suggest that sfGFP is expressed and pAzPhe and pIPhe, but not pBpa, are incorporated into sfGFP protein in the presence of 1 mM cognate UAAs. However, in the absence of UAAs, sfGFP proteins are also detected from *S. venezuelae* harboring pSUA1-pBpaRS, pSUA1-pAzPheRS, and pSUA1-pIPheRS, which suggest that endogenous amino acids are recognized by pBpaRS, pAzPheRS, and pIPheRS to varying degrees (Figure 2-11b, lane 2, 4, and 6). Surprisingly, the sfGFP was not produced in *S. venezuelae* harboring pSUA1-pBpaRS grown in the presence of 1 mM pBpa (Figure 2-11b, lane1), but was produced by the same strain grown without pBpa (Figure 2-11b, lane 2) suggests that pBpa-charged MjtRNA$^{Tyr}_{CUA}$ is produced intracellulary, but that it is unable to be accommodated by the *S. venezuelae* translational machinery to a sufficient extent to allow detection of sfGFP by Western blot. Fluorescence quantification of sfGFP pufified from small scale (Figure 2-11c) suggests that the amber suppression efficiencies range from 7% – 12% compared to wild type sfGFP expressed from *S. venezuelae* harboring pSUA5-sfGFP.

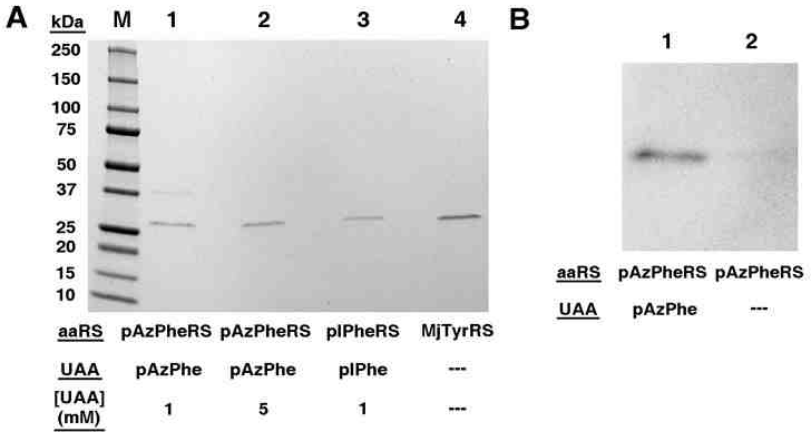


Figure 2-12. (a) SDS-PAGE of purified sfGFP proteins obtained from *S. venezuelae* harboring pSUA1-pAzPheRS grown in the presence of 1 mM pAzPhe (lane 1), pSUA1-pAzPheRS grown in the presence of 5 mM pAzPhe (lane 2), pSUA1-pIPheRS grown in the presence of 1 mM pIPhe (lane 3), and opt-pSUA2-sfGFP (lane 4). (b) Streptavidin Western blot of purified sfGFP proteins obtained from *S. venezuelae* harboring pSUA1-pAzPheRS grown in the presence (lane 1) and the absence (lane 2) of 1 mM pAzPhe, incubated with 2 μM DBCOPEG4- biotin at 25 °C for 12 h, and repurified by Ni-NTA affinity chromatography to remove excess DBCO-PEG4-biotin. The blot was imaged for 5 s.

To confirm the genetic incorporation of pAzPhe by bioorthogonal reaction between pAzPhe and the strained alkyne dibenzocyclooctyne-PEG4-biotin (DBCO-biotin), the proteins was purified from *S. venezuelae* harboring pSUA1-pAzPhe grown either in the presence or in the absence of 1 mM pAzPhe and incubated with the DBCO-biotin, followed by streptavidin Western blotting. The results show that sfGFP produced in the presence of pAzPhe (Figure 2-12b, lane 1), but not that produced in the absence of pAzPhe (Figure 2-12b, lane 2), is labeled by DBCO-biotin.

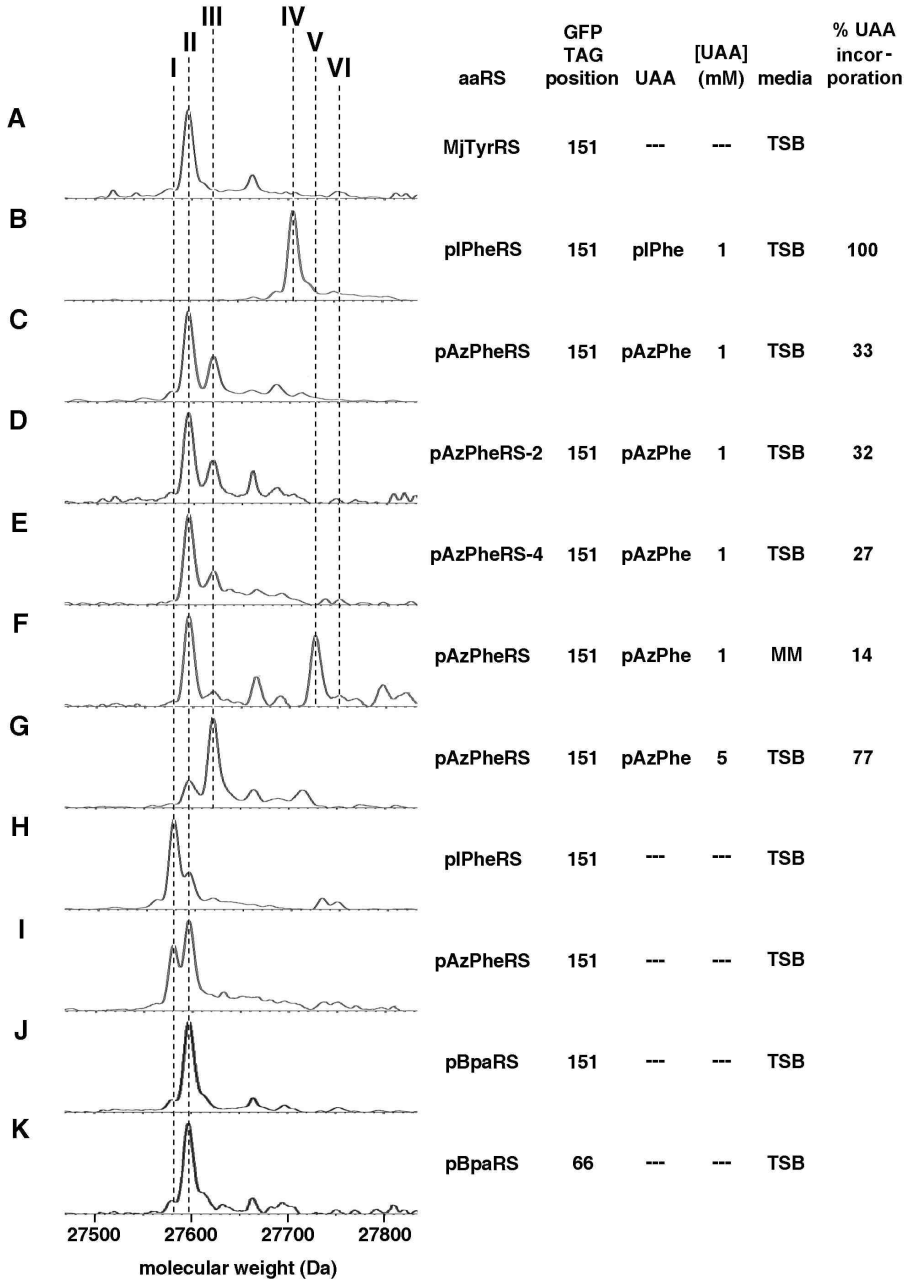


Figure 2-13. Deconvoluted ESI-MS data of GFP proteins purified from *S. venezuelae*. The aaRS, GFP-TAG, UAA, concentration of UAA, and growth media used for each protein expression experiment is shown to the right of each spectrum. The percentage of UAA incorporation, as estimated by the ratios of peak heights, is also shown to the right of each spectrum where appropriate. Peak I corresponds to GFP in which phenylalanine has been incorporated (Spectra H, I), Peaks II and V to GFP with tyrosine or p-amino-L-phenylalanine incorporated (Spectra A, C−K), Peak III and VI to GFP with pAzPhe incorporated (Spectra C−G), and Peak IV to GFP with pIPhe incorporated (Spectrum B). Peaks I−IV correspond to proteins with the N-terminal methionine removed, and Peaks V and VI to proteins with the N-terminal methionine retained (Spectrum F only).

To confirm the incorporation of tyrosine, pIPhe, and pAzPhe and to figure out what endogenous amino acids were recognized by these three aaRS, intact protein electrospray ionization mass spectrometry (ESI-MS) of sfGFP proteins need to be obtained. Large-scale expression (1.6 L each) and purification assay was conducted to obtained pure sfGFP protein subjected to ESI-MS analysis. Mass spectrometry results (Figure 2-13a and b) confirm that homogeneous tyrosine containing protein and homogeneous pIPhe-containing protein were produced by *S. venezuelae* harboring pSUA2-sfGFP and pSUA1-pIPheRS (in presence of 1 mM pIPhe), respectively. However, only 33% incorporation of pAzPhe was observed (Figure 2-13c) from *S. venezuelae* harboring pSUA1-pAzPheRS (in presence of 1 mM pAzPhe), with the remaining protein having a mass consistent with incorporation of either tyrosine or *p*-amino-L-phenylalanine (a known intermediate in the biosynthesis of the *Streptomyces* natural product chloramphenicol,[25] and possible product of enzymatic reduction of pAzPhe[26]), which differ in mass by only one Dalton.

Improvement of pAzPhe incorporation Fidelity. Several parameters, including alternative pAzPheRS variants, *E. coli* elongation factor Tu (EF-Tu, tufA gene) introduction, culturing media, and concentration of pAzPhe, were texted to improve the pAzPhe incorporation fidelity. First, three previous reported alternative pAzPheRS variants (pAzPheR-2, pAzPheR-4, and pAzPheR-5)[7] were constructed and cloned into pSUA1-pAzPheRS in the place of pAzPheRS to generate pSUA1- pAzPheR-2, pSUA1-pAzPheR-4, and pSUA1- pAzPheR-5. The performance was tested by small-scale expression and purification, followed by fluorescence quantification and anti-His western blotting. The results (Figure 2-14) show that differential amounts of protein in the

presence versus the absence of pAzPhe, whereas one (pAzPheRS-5) did not. However, large-scale expression (1.6 L), purification, and ESI-MS analysis of sfGFP from *S. venezuelae* harboring pSUA1- pAzPheR-2 and pSUA1- pAzPheR-4 show (Figure 2-13d, e) that neither variant have improved pAzPhe incorporation fidelity (32% from pAzPheRS-2 and 27% from pAzPheRS-4) with the remaining protein having mass consistent with incorporation of tyrosine or *p*-amino-L-phenylalanine.

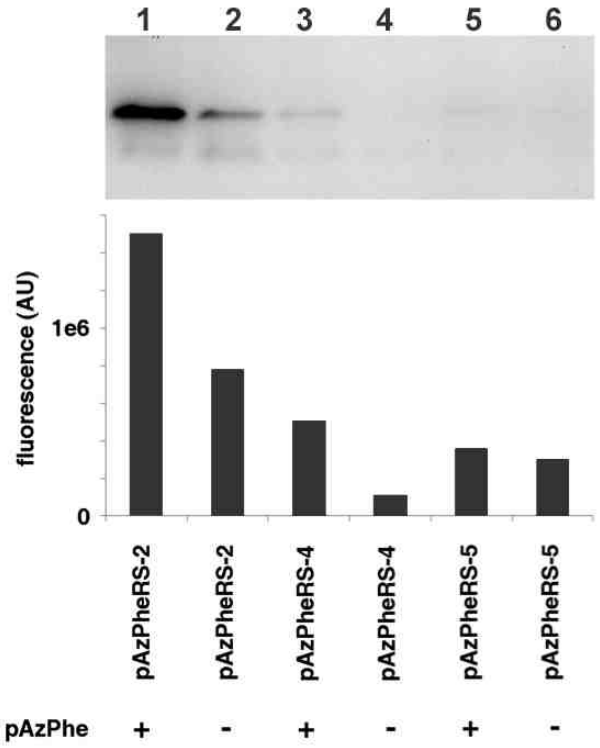


Figure 2-14. Fluorescence quantification and anti-His Western blot of Ni-NTA affinity purified sfGFP proteins obtained from S. venezuelae harboring pSUA1-pAzPheRS-2 grown in the presence (lane 1) and in the absence (lane2) of 1 mM pAzPhe, pSUA1-pAzPheRS-4 grown in the presence (lane 3) and in the absence (lane 4) of 1 mM pAzPhe, and pSUA1-pAzPheRS-5 grown in the presence (lane 5) and in the absence (lane 6) of 1 mM pAzPhe. The blot was imaged for 588 s.

Next, one of *E. coli* elongation factor Tu, tufA gene, was codon optimized and cloned into pSUA1-pBpaRS, pSUA1-pAzPheRS, pSUA1-pIPheRS, and pSUA2-sfGFP to generate pSUA1-pBpaRS-tufA, pSUA1-pAzPheRS-tufA, pSUA1-pIPheRS-tufA, and

pSUA2-tufA to facilitate the UAA incorporation in *Streptomyces*. The performance was test by small-scale expression and fluorescence quantification, followed by anti-His western blotting. The results (Figure 2-15b, lane 2, 4, and 6) show that no sfGFP was produced from *S. venezuelae* harboring pSUA1-pBpaRS-tufA, pSUA1-pAzPheRS-tufA, and pSUA1-pIPheRS-tufA in absence of UAAs, which indicate the tufA gene discriminated the endogenous amino acids incorporation. However, pBpa (Figure 2-15b, lane 1) is still not incorporated into sfGFP from *S. venezuelae* harboring pSUA1-pBpaRS-tufA in presence of 1 mM pBpa. The ESI-MS analysis (Figure 2-15a) of protein obtained from *S. venezuelae* harboring pSUA1-pAzPheRS-tufA shows 50% incorporation of pAzPhe with 17% improvement from the pSUA1-pAzPheRS.

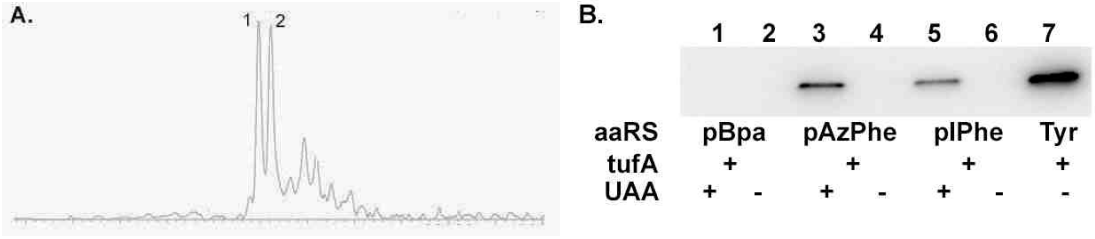


Figure 2-15. (a). Deconvoluted ESI-MS data of GFP proteins purified from *S. venezuelae* harboring pSUA1-pAzPheRS-tufA grown in presence of 1 mM pAzPhe. Peak 1 is tyrosine or p-amino-L-phenylalanine incorporation, peak 2 is pAzPhe incorporation. (b). Anti-His Western blot of Ni-NTA affinity purified sfGFP proteins obtained from *S. venezuelae* harboring pSUA1-pBpaRS-tufA grown in the presence (lane 1) and in the absence (lane 2) of 1 mM pBpa, pSUA1-pAzPhe-tufA grown in the presence (lane 3) and in the absence (lane 4) of 1 mM pAzPhe, pSUA1-pIPheRS-tufA grown in the presence (lane5) and in the absence (lane6) of 1 mM pIPhe, and pSUA2-tufA (lane 7). The blot was imaged for 164 s.

Next, a large-scale (1.6 L) growth in *Streptomyces* minimal media (MM)[15], which contains no amino acids, was tested to attempt to reduce the intracellular concentration of endogenous amino acids that compete with pAzPhe. However, sfGFP isolated from *S. venezuelae* harboring pSUA1-pAzPheRS grown in MM containing 1 mM pAzPhe showed reduced pAzPhe incorporation fidelity (14%) (Figure 2-13f). To compete-out the

endogenous amino acids recognized by pAzPheRS, the concentration of pAzPhe was increased from 1 mM to 5 mM for a large-scale (1.6 L) culture for *S. venezuelae* harboring pSUA1-pAzPhe. The ESI-MS result (Figure 2-13g) indicated that the isolated protein contained 77% pAzPhe incorporation, which should be suitable for many downstream applications, including *in vitro* and *in vivo* photocrosslinking experiments to investigate protein-protein interactions. Thus, we have developed and optimized engineered *S. venezuelae* hosts capable of site-specific incorporation of pIPhe and pAzPhe into target proteins.

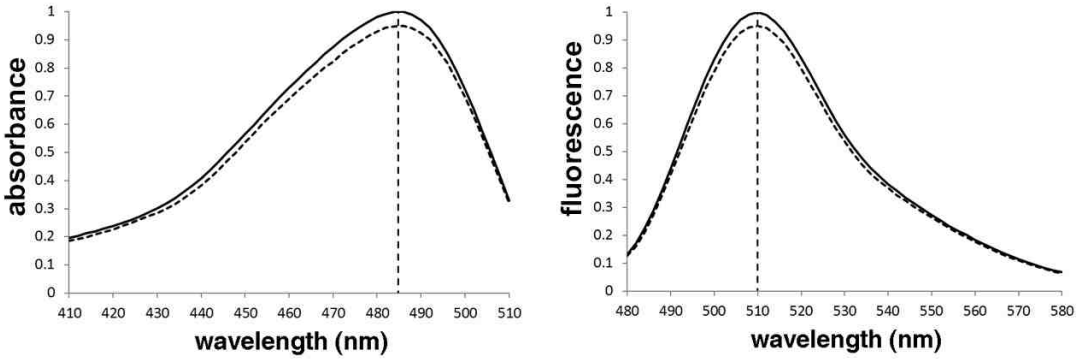


Figure 2-16. Absorbance (left) and emission (right) spectra of GFP proteins purified from *S. venezuelae* harboring opt-pSUA2-sfGFP (solid line) and pSUA1-pBpaRS-Y66TAG (dashed line). The absorbance spectrum was obtained by excitation over the 410−510 nm range and detection of emission at 540 nm, and the emission spectrum was obtained by excitation at 446 nm and detection of emission over the 480−580 nm range. To aid in comparison of peak shapes, both spectra were normalized, with the intensity of each dashed line spectrum adjusted to 95% of the intensity of the solid line spectrum at the absorbance and emission maxima.

Identification of Endogenous Amino Acid Incorporated into Proteins. Because of the observed incorporation of endogenous amino acids in *S. venezuelae* when pBpaRS, pAzPheRS, and pIPheRS were employed in the absence of cognate UAAs, and the observed incorporation of noncognate amino acids when pAzPheRS, pAzPheRS-2, and pAzPheRS-4 were employed in the presence of pAzPhe, neither of which happened in the *E. coli* host in which these aaRS were originally developed, additionally experiments

were carried out to identify the amino acids being incorporated. The results of these experiments are important for future optimization and expansion of UAA incorporation technology in *S. venezuelae* through development of engineered aaRS with improved selectivity for cognate UAAs and metabolic engineering to reduce or eliminate endogenous production of amino acids that compete with cognate UAAs as substrate for aaRS.

First, large-scale expression, purification, and ESI-MS of sfGFP from *S. venezuelae* harboring pSUA1-pIPheRS, pSUA1-pAzPheRS, and pSUA1-pBpaRS grown in the absence of UAAs were conducted. The results showed that a mixture of phenylalanine (71%) and either tyrosine or *p*-amino-L-phenylalanine (29%) was incorporated when pIPheRS was employed (Figure 2-13h); a mixture of phenylalanine (42%) and either tyrosine or *p*-amino-L-phenylalanine (58%) was incorporated when pAzPheRS was employed (Figure 2-13i), and homogeneous tyrosine or *p*-amino-L-phenylalanine was incorporated when pBpaRS was employed (Figure 2-13j). Thus, in the absence of UAAs, each aaRS is capable of charging MjtRNA$^{Tyr}_{CUA}$ with endogenous amino acids, and each aaRS displays a distinct degree of catalytic competence for each of these amino acids.

To distinguish tyrosine or *p*-amino-L-phenylalanine being incorporated into sfGFP, genomic sequence[27] of *S. venezuelae* ATCC 15439 was analyzed and there is no homologues *p*-amino-L-phenylalanine biosynthesis genes PapA, PapB, and PapC,[25] which indicate that tyrosine, not *p*-amino-L-phenylalanine, was incorporated by aaRS. According to the knowledge of mutant fluorophore, replacement of tyrosine at position 66 with various UAAs[28-30], including *p*-amino-L-phenylalanine, will result in fluorescent

spectrum shifts. Among the three aaRS, only pBpaRS gave homogenous tyrosine or *p*-amino-L-phenylalanine incorporation in the absence of pBpa. Then sfGFP gene, with an amber stop codon in the place of tyrosine at position 66, was constructed and cloned into pSUA1-pBpaRS in the place of sfGFP-Y151TAG to generate construct pSUA1-pBpaRS-Y66TAG. Then large-scale expression, purification and ESI-MS analysis were subjected to *S. venezuelae* harboring pSUA1-pBpaRS-Y66TAG grown in the absence of pBpa. The ESI-MS (Figure 2-13k) again showed homogenous tyrosine or *p*-amino-L-phenylalanine incorporation at position 66. The purified mutant sfGFP protein was then subjected to spectrum scan along with the protein purified from *S. venezuelae* harboring pSUA2-sfGFP, which produced sfGFP with the wild type fluorophore. The two (Figure 2-16) were found to be indistinguishable, with excitation and emission maxima of 485 and 510 nm, respectively, which indicates a tyrosine-derived fluorophore. sfGFP containing *p*-amino-L-phenylalanine at position 66 has been reported to possess excitation and emission maxima of 446 and 498 nm, respectively.[30] Therefore, the spectra of sfGFP purified from *S. venezuelae* harboring pSUA1-pBpaRS-Y66TAG are consistent with incorporation of tyrosine, not *p*-amino-L-phenylalanine, by pBpaRS. However, because of the decreased brightness of sfGFP with *p*-amino-L-phenylalanine at position 66 (11% of sfGFP with tyrosine at this position),[30] the possibility of a low level of *p*-amino-L-phenylalanine cannot be excluded. Since this experiment was conducted using pBpaRS, *p*-amino-L-phenylalanine incorporation by pIPheRS and pAzPheRS cannot be excluded. However, taken together, the genomic, fluorescence, and mass spectral data suggest that tyrosine, not *p*-amino-L-phenylalanine, is incorporated not only by pBpaRS in the absence of pBpa, but also by pAzPheRS in the presence and absence of pAzPhe, and by

pIPheRS in the absence of pIPhe. Therefore, we plan to focus *S. venezuelae* UAA incorporation system optimization efforts on decreasing the catalytic competence of engineered aaRS toward tyrosine.

Development of Pyrrolysyl Amber Suppression System in Stretomyces. Besides tyrosyl amber suppression system, pyrrolysyl amber suppression system is another widely used aaRS/tRNA pair for UAA incorporation, which is derived from *Methanosarcina barkeri* (*M. barkeri*) and *Methanosarcina mazei* (*M. mazei*).[5] The *M. barkeri* pyrrolysly-tRNA synthetase/$tRNA^{Pyl}_{CUA}$ (Mb pylRS/$tRNA^{Pyl}_{CUA}$) pair was chosen to develop this UAA incorporation system in *S. venezuelae*. pSUA2-pyl harboring a codon optimized pylRS gene driven by SF14 promoter, $tRNA^{Pyl}_{CUA}$ driven by promoter SCOt03p and terminated by terminator SCOt07t, and C-terminal histidine tagged sfGFP with an amber stop codon in the place of tyrosine at position 151, was designed to test the functionality of Mb pylRS/$tRNA^{Pyl}_{CUA}$ pair in *S. venezuelae*. pSUA1-pyl is a derivate of pSUA2-pyl, in which the pylRS gene was replaced by $DZKRS^{10}$, a variant of pylRS, which is capable of charging $tRNA^{Pyl}_{CUA}$ with DiZPK. DZKRS contains three mutations (L274A, C313S, and Y349F) compared to pylRS. pSUA3-pyl harboring the amber suppressible sfGFP and SCOt03p- $tRNA^{Pyl}_{CUA}$-SCOt07t cassette, was designed to test the $tRNA^{Pyl}_{CUA}$ orthogonality toward endogenous aaRS in *S. venezuelae*. Because of the observation of successfully expression (Figure 2-10, lane 3 and 5) of sfGFP and absence of endogenous amber suppression system in *S. venezuelae*, there is no need to construct pSUA4-pyl and pSUA5-pyl.

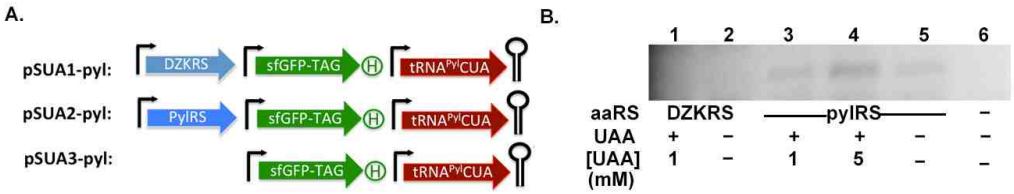


Figure 2-17. (a) Schematic of pSUA1-pyl, pSUA2-pyl, and pSUA3-pyl vector inserts. (b) Anti-His Western blot of Ni-NTA affinity purified sfGFP proteins obtained from *S. venezuelae* harboring pSUA1-pyl grown in the presence (lane 1) and in the absence (lane 2) of 1 mM DiZPK, pSUA2-pyl grown in the presence of 1 mM (lane 3) and 5 mM (4) Ne-boc-L-lysine and in the absence (5) of Ne-boc-L-lysine, and pSUA3-pyl (lane 6). The blot was imaged for 502 s.

pSUA1-pyl, pSUA2-pyl, and pSUA3-pyl were protoplast transferred into *S. venezuelae*. Small-scale expression, purification, and anti-His western blotting were carried out. The western blot result (Figure 2-17b, lane 6) obtained from *S. venezuelae* harboring pSUA3-pyl indicates that the $tRNA^{Pyl}_{CUA}$ is orthogonal to endogenous aaRS in *S. venezuelae*. *S. venezuelae* harboring pSUA2-pyl produced detectable sfGFP in presence of Ne-boc-L-lysine[31] and there was more production from the mutant grown with 5 mM (Figure 2-17b, lane 4) than 1 mM (Figure 2-17b, lane 3) Ne-boc-L-lysine. These results suggest that Ne-boc-L-lysine is incorporated into sfGFP. And the western blotting (Figure 2-17b, lane 3) obtained from the same mutant grown in absence of Ne-boc-L-lysine indicates that endogenous amino acid(s) are recognized by pylRS, which was seen from MjTyrRS and its variants. In addition, there is no detectable sfGFP (Figure 2-17b, lane 1 and 2) from *S. venezuelae* harboring pSUA1-pyl in the presence and in the absence of DiZPK possibly because the low protein yield is beyond the western blotting detection limit. A large-scale expression, purification and anti-His western blotting need to be carried out in the near future to figure out the inconclusive result obtained from pSUA1-pyl.

5. Conclusion

The study described in this chapter, we have designed, constructed, and

optimized amber suppression UAA incorporation systems capable of installing pIPhe, pAzPhe, and Ne-boc-L-lysine site-specifically into proteins in *Streptomyces venezuelae* ATCC 15439 with good to excellent fidelity and with sufficient yield for most applications. Among the many possible applications, we envision using this technology to identify and map protein-protein interactions in natural product biosynthetic multienzyme complexes via UAA mediated photocrosslinking, and to incorporate novel chemical moieties into ribosomally synthesized and post-translationally peptides (RiPPs) to facilitate their detection, isolation, and studies of their mechanism of action. More details towards application were described in Chapter 4.

In the process of developing and optimizing the *S. venezuelae* UAA system, we observed two notable differences in system performance between *S. venezuelae* and previously employed hosts: an inability to incorporate a detectable level of pBpa into the target eGFP or sfGFP protein, and a decreased selectivity of engineered MjTyrRS variants for their cognate UAAs over endogenous proteinogeic amino acids.

One unique feature we learned about *Streptomyces* is the TTA codon, which is contained only in non-essential genes and recognized by its unique deletable tRNA, *bldA*.[32,33] In light of this, we were able to develop a TTA sense codon reassignment UAA incorporation system in J1681[32] (*S. coelicolor* A3(2) *ΔbldA* strain) and Sven 624 (*S. venezuelae* ATCC 15439 *ΔbldA* strain), which were described in the Chapter 3.

6. References

1. Liu, C. C., & Schultz, P. G. (2010). Adding new chemistries to the genetic code. *Annual review of biochemistry*, 79, 413-444.

2. Chin, J. W. (2014). Expanding and reprogramming the genetic code of cells and animals. *Annual review of biochemistry*, *83*, 379-408.
3. Li, J., & Neubauer, P. (2014). *Escherichia coli* as a cell factory for heterologous production of nonribosomal peptides and polyketides. *New biotechnology*, *31*(6), 579-585.
4. Yuzawa, S., Kim, W., Katz, L., & Keasling, J. D. (2012). Heterologous production of polyketides by modular type I polyketide synthases in *Escherichia coli*. *Current opinion in biotechnology*, *23*(5), 727-735.
5. Wan, W., Tharp, J. M., & Liu, W. R. (2014). Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1844*(6), 1059-1070.
6. Chin, J. W., Martin, A. B., King, D. S., Wang, L., & Schultz, P. G. (2002). Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *99*(17), 11020-11024.
7. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., & Schultz, P. G. (2002). Addition of p-Azido-l-phenylalanine to the Genetic Code of *Escherichia coli*. *Journal of the American Chemical Society*, *124*(31), 9026-9027.
8. Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G., & Schultz, P. G. (2004). The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nature biotechnology*, *22*(10), 1297-1301.
9. Yang, M., Li, J., & Chen, P. R. (2014). Transition metal-mediated bioorthogonal protein chemistry in living cells. *Chemical Society Reviews*, *43*(18), 6511-6526.

10. Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., ... & Chen, P. R. (2011). A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nature chemical biology*, *7*(10), 671-677.
11. Xue, Y., Zhao, L., Liu, H. W., & Sherman, D. H. (1998). A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *Proceedings of the National Academy of Sciences*, *95*(21), 12111-12116.
12. Thibodeaux, C. J., Melançon, C. E., & Liu, H. W. (2008). Natural-Product Sugar Biosynthesis and Enzymatic Glycodiversification. *Angewandte Chemie International Edition*, *47*(51), 9814-9859.
13. Kim, E. J., Yang, I., & Yoon, Y. J. (2015). Developing *Streptomyces venezuelae* as a cell factory for the production of small molecules used in drug discovery. *Archives of pharmacal research*, *38*(9), 1606-1616.
14. Sambrook, J., and Russell, D. W. (2001) *Molecular cloning: a laboratory manual 3rd edition*; Cold Spring Harbor Press, Cold Spring Harbor, NY.
15. Borodina, I., Siebring, J., Zhang, J., Smith, C. P., van Keulen, G., Dijkhuizen, L., & Nielsen, J. (2008). Antibiotic overproduction in *Streptomyces coelicolor* A3 (2) mediated by phosphofructokinase deletion. *Journal of Biological Chemistry*, *283*(37), 25186-25199.
16. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A., & Bertozzi, C. R. (2006). A comparative study of bioorthogonal reactions with azides. *ACS Chemical Biology*, *1*(10), 644-648.

17. Bierman, M., Logan, R., O'brien, K., Seno, E. T., Rao, R. N., & Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces spp*. *Gene*, *116*(1), 43-49.

18. Baltz, R. H. (1999). Genetic recombination by protoplast fusion in *Streptomyces*, (Volume 21). *Journal of industrial microbiology & biotechnology*, *22*(4), 460-471.

19. Kierser, T., Bibb, M. J., Buttner, M. J., Chater, K. F., & Hopwood, D. A. (2000). Practical *Streptomyces* Genetics: John Innes Foundation.

20. Melançon, C. E., & Liu, H. W. (2007). Engineered biosynthesis of macrolide derivatives bearing the non-natural deoxysugars 4-epi-D-mycaminose and 3-N-monomethylamino-3-deoxy-D-fucose. *Journal of the American Chemical Society*, *129*(16), 4896-4897.

21. Guo, J., Melançon, C. E., Lee, H. S., Groff, D., & Schultz, P. G. (2009). Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angewandte Chemie International Edition*, *48*(48), 9148-9151.

22. Labes, G., Bibb, M., & Wohlleben, W. (1997). Isolation and characterization of a strong promoter element from the *Streptomyces ghanaensis* phage 119 using the gentamicin resistance gene (aacC1) of Tn1696 as reporter. *Microbiology*, *143*(5), 1503-1512.

23. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, *24*(1), 79-88.

24. Chater, K. F., & Chandra, G. (2008). The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. *The Journal of Microbiology*, *46*(1), 1-11.

25. Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M., & Murakami, T. (2004). Para-position derivatives of fungal anthelmintic cyclodepsipeptides engineered with *Streptomyces venezuelae* antibiotic biosynthetic genes. *Nature biotechnology*, *22*(7), 848-855.

26. Mancuso, L., Jürjens, G., Hermane, J., Harmrolfs, K., Eichner, S., Fohrer, J., ... & Kirschning, A. (2013). Bioreduction of Aryl Azides during Mutasynthesis of New Ansamitocins. *Organic letters*, *15*(17), 4442-4445.

27. He, J., Sundararajan, A., Devitt, N. P., Schilkey, F. D., Ramaraj, T., & Melançon, C. E. (2016). Complete genome sequence of *Streptomyces venezuelae* ATCC 15439, producer of the methymycin/pikromycin family of macrolide antibiotics, using PacBio technology. *Genome announcements*, *4*(3), e00337-16.

28. Wang, L., Xie, J., Deniz, A. A., & Schultz, P. G. (2003). Unnatural amino acid mutagenesis of green fluorescent protein. *The Journal of organic chemistry*, *68*(1), 174-176.

29. Wang, F., Niu, W., Guo, J., & Schultz, P. G. (2012). Unnatural amino acid mutagenesis of fluorescent proteins. *Angewandte Chemie International Edition*, *51*(40), 10132-10135.

30. Reddington, S. C., Rizkallah, P. J., Watson, P. D., Pearson, R., Tippmann, E. M., & Jones, D. D. (2013). Different photochemical events of a genetically encoded phenyl

azide define and modulate GFP fluorescence. *Angewandte Chemie International Edition*, *52*(23), 5974-5977.

31. Luo, X., Zambaldo, C., Liu, T., Zhang, Y., Xuan, W., Wang, C., ... & Schultz, P. G. (2016). Recombinant thiopeptides containing noncanonical amino acids. *Proceedings of the National Academy of Sciences*, 201602733.
32. Leskiw, B. K., Mah, R., Lawlor, E. J., & Chater, K. F. (1993). Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3 (2). *Journal of bacteriology*, *175*(7), 1995-2005.
33. White, J., & Bibb, M. (1997). *bldA* dependence of undecylprodigiosin production in *Streptomyces coelicolor* A3 (2) involves a pathway-specific regulatory cascade. *Journal of bacteriology*, *179*(3), 627-633.

Chapter 3. Development of TTA Sense Codon Reassignment Unnatural Amino Acid Incorporation System in *Streptomyces*.

1. Introduction

Genetic code expansion via nonsense codon (i.e stop codon and four-base codon) suppression is limited to low efficiency because of competition with protein translation termination mechanism or imperfect translational machinery for reading through four-base codon.[1,2] Sense codon reassignment is another route to bring novel moieties into proteins, which has drawn more and more attentions.[3-7] However, the orthogonal aminoacyl-tRNA charged with UAAs need to compete with a large pool of its endogenous tRNA, which leads to incomplete UAA incorporation.[4] There are two examples of complete sense codon reassignment. The first one is to reassign 58 of the 64 codons to selenocysteine.[6] This system is limited to this particular amino acid analogue. In the second example,[3] researchers reassigned the rare codon AGG in *E. coli* with deletion of its corresponding $tRNA^{Arg}_{CCU}$ to obtain complete different UAAs incorporation and even achieve three tandem AGG codons read through. However, the AGG codons in its proteome are also unavoidably translated by UAAs. They had to limit the protein expression to 3 h because of the toxicity of that system to the *E. coli* cell.

Streptomyces, as discussed in Chapter 1, are rich natural products producer and have smartly evolved codon usage bias to regulate their secondary metabolite productions. TTA codon, the rarest one for *Streptomyces*, is only employed in non-essential genes (secondary metabolites regulatory proteins), and is recognized by its unique tRNA gene *bldA*, which is low abundant in *Streptomyces* and delayed transcribed during the cell growth.[8,9] All these features make TTA codon a good candidate to be

reassigned to UAAs. In addition to the fact that *bldA* gene can be deleted and has been deleted without cell fitness loss[8] and reassignment of TTA codons in the secondary metabolites regulatory proteins should not have any toxicity to the cell.[10]

To develop this system, a *bldA* knockout *Streptomyces* strain and a tRNA synthetase/$tRNA_{UAA}$ pair, which can facilitate the UAA charging onto the $tRNA_{UAA}$ and the $tRNA_{UAA}$ can decode TTA codon. *S. coelicolor* A3(2) J1681 is a *bldA* deleted train and has been widely used to research *bldA* regulation mechanisms.[8] And the MjTyrRS is promiscuous toward the anticodon based on previous work.[11] Therefore, simply replacing the anticodon CUA by UAA in the $MjtRNA^{Tyr}_{CUA}$ would allow TTA codon reassigned to UAA in the *ΔbldA* strain.

To utilize this unique gift given from nature and test our idea, pSUA2-TTA was constructed, in which anticodon was changed to UAA for decoding the TTA codon in sfGFP. pSUA4-TTA was also constructed as a negative control to test if the TTA codon can be translated in the *ΔbldA* strain. During waiting for the J1681 strain, pSUA2-TTA and pSUA4-TTA were transferred into the wild type *S. venezuelae* and anti-his Western blotting result showed more proteins were produced by pSUA2-TTA than pSUA4-TTA, which could be a hint of functionality of this design. These two constructs were then tested in the *ΔbldA* strain J1681 and anti-his Western blotting result confirmed that the TTA codon was reassigned to tyrosine with the MjTyrRS/$tRNA^{Tyr}_{UAA}$ pair.

To follow that, pIPheRS-Y151TTA and pBpaRS-S147TTA was constructed to reassign TTA codon to pIPhe and pBpa, respectively. Both Western blotting and protein mass spectrometry results confirmed that pIPhe and pBpa were incorporated into sfGFP protein using the TTA codon reassignment. Encouragement by the observation that pBpa

was first time incorporated into protein in *Streptomyces*, a more challenging goal was set up for the TTA codon reassignment. pIPheRS-TTA5s was constructed to test if this new system is capable of multiple UAAs incorporation in a single protein. According to the structure of sfGFP,[12] all surface exposed tyrosine and polar, non-ionizable resides (ser, Thr, Gln, and Asn) were screened and 15 of them, which do not interact with their nearby residues, were chosen to be reassigned to UAAs. Then 15 single TTA codon mutant sfGFP were constructed and the TTA reassignment efficiency at different sites was tested by fluorescence quantification. pIPheRS-TTA5s-2, pIPheRS-TTA10s, and pIPheRS-TTA15s were designed and constructed with combination the top 5 , top 10, and all of the 15 sites tested by fluorescence quantification. In addition, different tandem TTA codons constructs were also made to test if this system is capable of translating consecutive TTA codons.

Toxicity and protein truncation are two inherent drawbacks of existing unnatural amino acid methodology.[13] Growth curve data was obtained for J1681 with the TTA codon reassignment system, which indicates this system is completely non-toxic to the cell. For protein truncation examination, 11 N-terminal histidine tagged sfGFP constructs were designed and made with direct comparison of amber suppression and TTA reassignment.

The J1681 strain grows slower compared to *S. venezuelae* and easily forms cell clumps, which limits the biomass and makes it difficult for quantitative analysis. To overcome these obstacles, *S. venezuelae ΔbldA* strain is needed. First, the high quality genomic DNA was isolated from *S. venezuelae* and sequenced using PacBio technology.[14] With its genome sequence (GenBank: LN881739) information, *bldA* gene

was located (5,403,465-5,403,551 in its genome) by aligned with the *bldA* gene from *S. coelicolor* A3(2). Then pCRISP-bldA was designed and constructed with the knowledge of *bldA* sequence and recently reported *Streptomyces* CRISPR/Cas system.[15-17]

The *S. venezuelae ΔbldA* strain was constructed using pCRISP-bldA and named as Sven 624 and then most of the TTA reassignment constructs were moved into Sven 624. Higher protein yield and more reliable comparison results between amber suppression system and TTA sense codon reassignment system were obtained. Protein yield is 20 fold higher with TTA reassignment than TAG suppression, which demonstrated this small change had made big improvement. In addition, up to 5 tandem or 10 scattered TTA codons were decoded in a single protein with this TTA reassignment system in Sven 624. And more importantly, there is completely no truncated protein produced in the cell.

All these features and results demonstrate that TTA codon is a free codon for reassignment to UAAs in *Streptomyces* with higher incorporation efficiency and without toxicity. Sven 624 could be a useful cell line for making multiple UAAs containing proteins in the future.

2. **Experimental procedures**

General. Most materials used for work described in this chapter have already been mentioned in the Experimental Procedures section of Chapter 2. *p*-iodo-L-phenylalanine (pIPhe) and *p*-benzoyl-L-phenylalanine (pBpa) were purchased from Chem-Impex (Wood Dale, IL) from then on.

Bacterial Strain. *E. coli DH5α* and *S. venezuelae* ATCC 15439 described in this Chapter have already been mentioned in the Experimental Procedures section of Chapter

2. *E. coli* ER2925 was used to propagate unmethylated DNA for introduction into *S. coelicolor* A 3(2) J1681 by protoplast transformation. *S. coelicolor* A3(2) J1681 was kindly provided by John Innes Centre in United Kingdom. Sven 624 is a *bldA* in frame deletion mutant of *S. venezuelae* ATCC 15439 and was constructed in this study using the reported *Streptomyces* CRISPR/Cas system.[15-17]

Bacterial culture. The bacterial culture used in this chapter was described in the Experimental Procedures section of Chapter 2.

Preparation of E. coli Competent Cells. The procedure used to prepare *E. coli* competent cells was described in the Experimental Procedure section of Chapter 2.

General PCR Conditions. The general PCR conditions used in this chapter was described in the Experimental Procedures section of Chapter 2

Small Scale Protein Expression. Small-scale protein expression from Sven 624 was the same as from *S. venezuelae* ATCC 15439 and was described in the Experimental Procedures section of Chapter 2. For small-scale protein expression from J1681, same procedures were employed with minor modifications, which were described as follows. First, R5 liquid media was used for J1681 small culture instead of TSB in order to avoid cell clump formation. Second, J1681 is grown slower compared to *S. venezuelae* and was cultured for 7 days before collection by centrifugation. Third, J1681 cells were sonicated on ice for 70 seconds (20% amplitude, 35 × 2 sec pulses with 5 sec pauses) due to its less biomass compared to *S. venezuelae.*

Large Scale Protein Expression and Purification. The large-scale protein expression and purification procedures used in this chapter were described in the Experimental Procedures section of Chapter 2. For sfGFP purification from pIPheRS-

TTA5 in J1681, only one hydrophobic interaction chromatography step was used because the hydrophobic property of sfGFP was dramatically increased after five pIPhe incorporation, which lead to acceptable separation after one hydrophobic interaction chromatography purification, which was described as follows. After step gradient, 14 × 1 mL of buffer (50 mM NaH_2PO_4, pH 8.0) was applied to the column, which followed by 2 × 1 mL of pure water elution. All of these fractions were subjected to fluorescence quantification using a microplate reader (excitation = 485 nm, bandwidth = 9 nm; emission = 520 nm, bandwidth = 15 nm), which indicated that protein was eluted out from the column in the last fraction. SDS-PAGE analysis confirmed the presence and purity of the sfGFP.

Protein Fluorescence Quantification. Same conditions were used in this experiment as described in the Experimental Procedures section of Chapter 2.

Western Blot Analysis. These experiments were performed as described in the Experimental Procedures section of Chapter 2.

Mass spectrometry. Mass spectrometry was performed as described in the Experimental Procedures section of Chapter 2.

Preparation of J1681 and Sven 624 Protoplasts. Same protocol was used to preparation of J1681 and Sven 624 protoplasts as described in the Experimental Procedures section of Chapter 2.

Protoplast Transfer of Vectors into J1681 and Sven 624. Same protocol was applied to protoplast transfer of vectors into J1681 and Sven 624 as described in the Experimental Procedures section of Chapter 2.

Growth of J1681 with TTA Sense Codon Reassignment. To assess the toxicity of TTA sense codon reassignment in *Streptomyces*, growths curves assay were conducted for J1681 with pEmpt, pSUA2-TTA and pIPheRS-Y151TTA constructs. pSUA5-sfGFP was digested with EcoRV and HindIII restriction enzymes and the resulting backbone was klenow-treated and blunt-ligated to generate the circular plasmid pEmpt, which was used as a blank control. Fifty μL of glycerol stock of each mutant was transferred to 5 mL of TSB media supplemented with 50 μg/mL apramycin; and was grown at 30 °C, 250rpm, for 4 days. Five hundred μL of each resulting culture was transferred to 20 mL of TSB media containing 50 μg/mL apramycin; and was grown at the same conditions for another 4 days. Each cell culture was transferred to a sterile 40 mL tissue grinder and homogenized with its tight pestle up and down for 50 times. OD_{600} of each homogenized cell culture was measured and cells were collected by centrifugation, supernatant was removed. Each cell pellet was resuspended with variable amount of fresh TSB media based on OD_{600} reading to normalize the cell density for each mutant. Four mL of each normalized mutant cell suspension was added to 200 mL of TSB media containing 50 μg/mL apramycin in a 1 L Erlenmeyer flask with 200 glass beads. Cell culture for each mutant was performed in triplicate flasks and grown at 30 °C, 200 rpm. OD_{600} of cell culture was taken from beginning and every 6 to 8 h, which was done as follows. At each time points, 2.5 mL of cell culture from each flask was taken and homogenized for 45 seconds using the 40 mL tissue grinder. The resulting samples were applied to OD_{600} analysis.

High Quality Genomic DNA Isolation from S. venezuelae ATCC 15439. Twenty-five mL of saturated TSB culture for *S. venezuelae* ATCC 15439 was prepared. Cells

were collected by centrifugation (4,000 g, 10 min, room temperature) and resuspended in 10 mL of 50 mM Tris-HCl, pH 8.0. The cell suspension was transferred to a 40 ml tissue grinder. After thoroughly homogenization, cells were collected by centrifugation (4,000 g, 10 min, room temperature) and the cell pellet was aliquoted into 200 mg portions in microcentrifuge tubes. Two hundred mg of cells was transferred to a porcelain mortar, flash froze using liquid nitrogen, and grinded by porcelain pestle. Three times of liquid nitrogen treatment and grinding process were applied to the cells. The resulting samples were washed off from the mortar using about 900 µL of solution 1 (10% sucrose, 50 mM Tris-HCl, pH 8.0, 10 mM EDTA) and aliquoted into 450 µL portions in microcentrifuge tubes. Fifty µL of 30 mg/mL of fresh prepared lysozyme containing solution 1 was added to each tube and the cell suspension was incubated at 37 °C, 250 rpm, for 1 h. After lysozyme treatment, ten µL of 5 mg/mL Proteinase K in solution 1 and 50 µL of 10% SDS were added to each tube and incubated in 55 °C water bath for 1 h. After that, five hundred µL of PCl (Phenol: Chloroform: Isoamyl Alcohol in the ratio 25: 24: 1, which is saturated with 10 mM Tris, pH 8.0, and 1 mM EDTA in 100 mL) was added to each tube and each tube was vortexed for 30 seconds, followed by centrifugation (14,000 rpm, 5 min). The resulting supernatant was transferred to a new tube and PCl treatment and centrifugation were repeated. Five hundred of chloroform was added to the transferred supernatant and tubes were shaken 10 times by hands and vortexed for 30 seconds, followed by centrifugation (14,000 rpm, 5 min). The supernatant was transferred to a new tube and 700 µL of pre-chilled isopropanol was added, followed by gently inverting the tube until the appearance of the genomic DNA. The DNA was centrifuged (3,000rpm, 10 sec), supernatant was removed, and 1 mL of 70% pre-chilled ethanol was added to wash

the DNA followed by gently inverting 10 times. The DNA was then collected by centrifugation (3,000rpm, 10 s), supernatant was removed, and the DNA was dried at 37 °C for 15 min with the cap off. Then 500 μL of TE buffer (10 mM Tris-HCl, 1 mM EDTA) containing 0.1 mg/mL RNase A was added to the tube and the DNA sample was stored at 4 °C overnight.

After overnight RNase A treatment, the DNA sample was further purified by one time of PCl and chloroform wash, then isopropanol and ethanol wash as described above. The purified DNA was dried at 37 °C for 15 min with the cap off, dissolved in 100 μL of Tris-HCl (10 mM, pH 8.0) and stored at 4 °C for PacBio sequencing.

bldA in frame deletion by pCRISP-bldA. pCRISP-bldA was constructed and protoplast transferred into *S. venezuelae*. Four transformants were picked to inoculate TSB cultures supplemented with 50 μg/mL apramycin and grown at 28 °C for 3 days. Cells were collected by centrifugation and genomic DNA was isolated by Microbial DNA Isolation Kit (MO BIO Laboratories, Inc). The locus of *bldA* region was amplified using primers binding outside of the homology arm and then digested with stuI restriction enzyme. The digestion result indicated that one out of four was *bldA* knockout strain. The *bldA* knockout strain was restreaked on SPA plate supplemented with 50 μg/mL apramycin and grown at 28 °C for 3 days. One single colony was picked and restreaked on SPA plate without apramycin and grown at 37°C for 3 days. Cells only survived at the dense area at 37°C. Some of the survived cells were collected using sterile toothpick and serial diluted to appropriate density for plating on three SPA plates. After a 3-day growth, several hundredth separated colonies were visible. One hundred and twenty five colonies were picked and replicated on SPA plate containing 50 μg/mL apramycin and SPA plate,

and grown at 28 °C for 3 days. Four colonies restored the apramycin sensitivity, which indicated the loss of pCRISP-bldA plasmid. These four colonies were picked to inoculate TSB culture for genomic DNA isolation. The locus of *bldA* region was again amplified from their genomic DNA. The PCR products were conducted digestion mapping and sequencing. The sequencing result and digestion mapping both confirmed that *bldA* gene was successfully in frame deleted by pCRISP-bldA. The *bldA* knockout *S. venezuelae* mutant was named as Sven 624, which was employed as the host for development of TTA codon reassignment unnatural amino acid incorporation system.

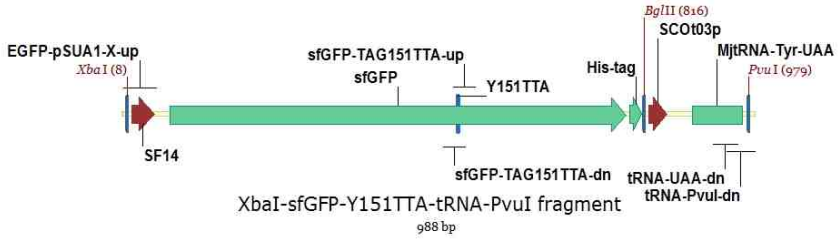
3. **Vector Design and Construction**

Construction of pSUA2-TTA and pSUA4-TTA. pSUA4-TTA is a derivate of pSUA5-sfGFP, in which the TAC codon at 151 site in sfGFP was replaced by the rare leucine codon TTA. pSUA2-TTA is a derivate of opt-pSUA2-sfGFP, in which the TAG codon at 151 site in sfGFP and CUA anticodon in the $tRNA^{Tyr}_{CUA}$ were replaced by TTA codon and UAA anticodon, respectively. To construct sfGFP-151TTA for making the pSUA4-TTA, two fragments with 151TTA codon at their junction were amplified. The upstream fragment was amplified using primers EGFP-pSUA5-E-2-up and sfGFP-TAG151TTA-dn; and the downstream fragment was amplified using primers sfGFP-TAG151TTA-up and His-PvuI-dn. These two fragments were assembled by overlap extension PCR using primers EGFP-pSUA5-E-2-up and His-PvuI-dn to generate sfGFP-Y151TTA, which was digested with EcoRV and PvuI. This digested insert was ligated into pSUA5-sfGFP digested with the same restriction enzymes to generate pSUA4-TTA. To construct the insert for making pSUA2-TTA, two fragments with the 151 TTA codon at their junction and anticodon replacement were amplified. The upstream fragment was

amplified using primers EGFP-pSUA1-X-up and sfGFP-TAG151TTA-dn; and the downstream fragment was amplified using primers sfGFP-TAG151TTA-up and tRNA-UAA-dn. These two fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and tRNA-PvuI-dn to generate the PCR product, which was digested with XbaI and PvuI. This digested insert was ligated into opt-pSUA2-sfGFP digested with the same restriction enzymes to generate pSUA2-TTA. These two constructs were verified by restriction mapping and sequencing of each insert. Primer information is given in the table below. The priming region of each primer is underlined, the restriction site is shown in cyan, mutated codon and anticodon are shown in blue. Maps of fragments and vectors are given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
sfGFP-TAG151TTA-dn	CGCGGTGATTAACACGTTGTGGGAGTTGAAGTTGT (Used with EGFP-pSUA5-E-2-up)	535	Opt-pSUA2-sfGFP
sfGFP-TAG151TTA-up	CAACGTGTTAATCACCGCGGACAAGCAGAAGAACG (Used with His-PvuI-dn)	312	Opt-pSUA2-sfGFP
EGFP-pSUA5-E-2-up	(Used with His-PvuI-dn)	828	PCR pdt.
sfGFP-TAG151TTA-dn	(Used with EGFP-pSUA1-X-up)	535	Opt-pSUA2-sfGFP
tRNA-UAA-dn	GGCCGATGGGATTTGAACCCATGCCATGCGGATTTTAAGTC (Used with sfGFP-TAG151TTA-up)	446	Opt-pSUA2-sfGFP
tRNA-PvuI-dn	GCATATATCGATCGTGAGTGTGGTCCGGCCGATGGGATTTGAACC (Used with EGFP-pSUA1-X-up)	988	PCR pdt.

Table 3-1. Primer information for constructing pSUA2-TTA and pSUA4-TTA



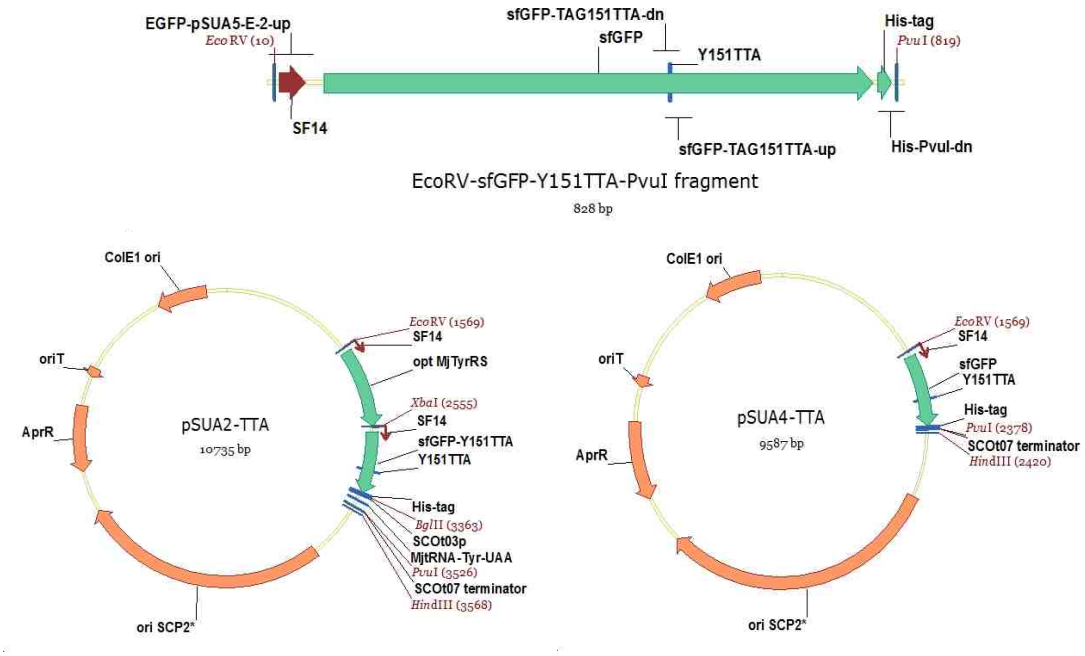


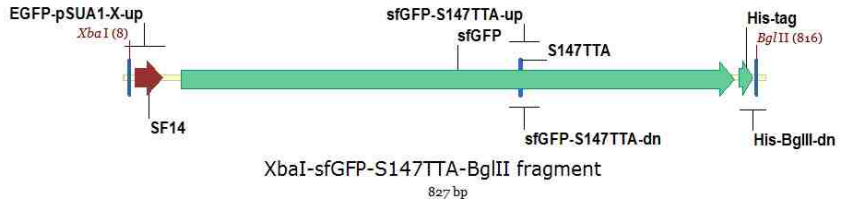
Figure 3-1 (cont.). Maps of fragments for making pSUA2-TTA and pSUA4-TTA, and maps of pSUA2-TTA and pSUA4-TTA

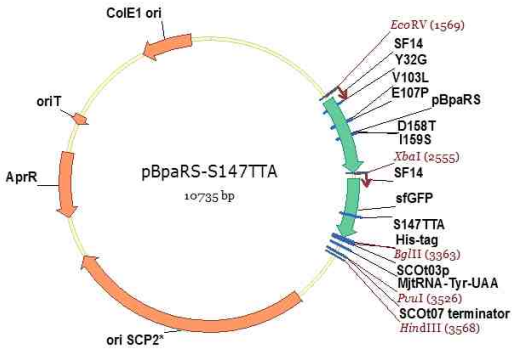
Construction of pIPheRS-Y151TTA and pBpaRS-S147TTA. pIPheRS-Y151TTA is a derivate of pSUA2-TTA, in which the MjTyrRS was replaced by pIPheRS. The pIPheRS gene was digested from pSUA1-pIPheRS-sfGFP, purified, and ligated into pSUA2-TTA digested with same restriction enzymes to generate the pIPheRS-Y151TTA. pBpaRS-S147TTA is another derivate of pSUA2-TTA, in which the MjTyrRS was replaced by pBpaRS and the sfGFP-151TTA gene was replaced by sfGFP-147TTA. To construct sfGFP-147TTA, two fragments with the 147TTA codon at the junction were amplified. The upstream fragment was amplified using primers EGFP-pSUA1-X-up and sfGFP-S147TTA-dn; and the downstream fragment was amplified using primers sfGFP-S147TTA-up and His-BglII-dn. These two fragments were assembled by overlap

extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn to generate the sfGFP-147TTA, which was digested with XbaI and BglII. The pBpaRS gene was directly digested from pSUA1-pBpaRS-sfGFP and purified. The resulting digested two inserts were ligated into pSUA2-TTA digested with EcoRV and BglII restriction enzymes using three-way ligation method to generate the pBpaRS-S147TTA. These constructs were verified by restriction mapping and sequencing of the insert. Primer information is given in the table below. The priming region of each primer is underlined; the restriction site is shown in cyan, mutated codon is shown in blue. Maps of fragment and representative vector are given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
sfGFP-S147TTA-dn	CGCGGTGATGTACACGTTGTGTAAGTTGAAGTTGTACTCG (Used with EGFP-pSUA1-X-up)	535	Opt-pSUA2-sfGFP
sfGFP-S147TTA-up	CGAGTACAACTTCAACTTACACAACGTGTACATCACCGCG (Used with His-BglII-dn)	332	Opt-pSUA2-sfGFP
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	827	PCR pdt.

Table 3-2. Primer information for constructing pBpaRS-S147TTA





Construction of other 14 single TTA mutant sfGFP plasmids. For pIPheRS-S147TTA construct, sfGFP-147TTA was directly digested from pBpaRS-S147TTA and ligated into pIPheRS-Y151TTA to generate the pIPheRS-S147TTA. To construct the other 13 single TTA mutant sfGFP genes, two fragments with the TTA codon at the junction were amplified. Thirteen primer pairs were design and used with primers EGFP-pSUA1-X-up and His-BglII-dn to obtain these 13 single TTA mutant sfGFP. The upstream fragments were amplified using primers EGFP-pSUA1-X-up and the reverse primers with TTA codon; and the downstream fragments were amplified using primes: the corresponding direct primers with TTA codon and His-BglII-dn. The upstream fragment and corresponding downstream fragment were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn to generate these 13 single TTA incorporated sfGFP, which were digested with XbaI and BglII. These digested PCR products were ligated into pIPheRS-Y151TTA digested with same restriction enzymes to generate these 13 single TTA mutant sfGFP plasmids. All of these constructs were verified by restriction mapping and sequencing of the insert. Primer information is given

in the table below. The priming region of each primer is underlined and mutated codon is shown in blue. Map of representative fragment is given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
T38TTA-dn	GCTTGCCGTTTAAGGCGTCCCCCT (Used with EGFP-pSUA1-X-up)	197	pSUA5-sfGFP
T38TTA-up	AGGGGGACGCCTTAAACGGCAAGC (Used with His-BglII-dn)	654	pSUA5-sfGFP
T43TTA-dn	GCAGATGAACTTCAGTAACAGCTTGCCGTTCGT (Used with EGFP-pSUA1-X-up)	217	pSUA5-sfGFP
T43TTA-up	ACGAACGGCAAGCTGTTACTGAAGTTCATCTGC (Used with His-BglII-dn)	643	pSUA5-sfGFP
T97TTA-dn	CCTTGAACGAGATTAAGCGCTCCTGCACGT (Used with EGFP-pSUA1-X-up)	377	pSUA5-sfGFP
T97TTA-up	ACGTGCAGGAGCGCTTAATCTCGTTCAAGG (Used with His-BglII-dn)	480	pSUA5-sfGFP
S99TTA-dn	CCGTCGTCCTTGAATAAGATCGTGCGCTCC (Used with EGFP-pSUA1-X-up)	384	pSUA5-sfGFP
S99TTA-up	GGAGCGCACGATCTTATTCAAGGACGACGG (Used with His-BglII-dn)	473	pSUA5-sfGFP
T105TTA-dn	CCGCGTCTTGTATAAGCCGTCGTCCTTG (Used with EGFP-pSUA1-X-up)	400	pSUA5-sfGFP
T105TTA-up	CAAGGACGACGGCTTATACAAGACGCGG (Used with His-BglII-dn)	455	pSUA5-sfGFP
T118TTA-dn	GCGGTTCACGAGTAAGTCGCCCTCG (Used with EGFP-pSUA1-X-up)	439	pSUA5-sfGFP
T118TTA-up	CGAGGGCGACTTACTCGTGAACCGC (Used with His-BglII-dn)	413	pSUA5-sfGFP
N149TTA-dn	CGCGGTGATGTACACTAAGTGGGAGTTGAAGTTGT (Used with EGFP-pSUA1-X-up)	535	pSUA5-sfGFP
N149TTA-up	ACAACTTCAACTCCCACTTAGTGTACATCACCGCG (Used with His-BglII-dn)	327	pSUA5-sfGFP
N159TTA-dn	TCGCCTTGATGCCTAACTTCTGCTTGTCCG (Used with EGFP-pSUA1-X-up)	563	pSUA5-sfGFP
N159TTA-up	CGGACAAGCAGAAGTTAGGCATCAAGGCGA (Used with His-BglII-dn)	294	pSUA5-sfGFP
N164TTA-dn	GCGGATCTTGAATAACGCCTTGATGCCGT (Used with EGFP-pSUA1-X-up)	577	pSUA5-sfGFP
N164TTA-up	ACGGCATCAAGGCGTTATTCAAGATCCGC (Used with His-BglII-dn)	279	pSUA5-sfGFP
T186TTA-dn	CGTCGCCGATGGGTAAGTTCTGCTGGTAG (Used with EGFP-pSUA1-X-up)	644	pSUA5-sfGFP
T186TTA-up	CTACCAGCAGAACTTACCCATCGGCGACG (Used with His-BglII-dn)	212	pSUA5-sfGFP
Y200TTA-dn	ACTGCGTGGAGAGTAAGTGGTTGTCGGG (Used with EGFP-pSUA1-X-up)	686	pSUA5-sfGFP
Y200TTA-up	CCCGACAACCACTTACTCTCCACGCAGT (Used with His-BglII-dn)	169	pSUA5-sfGFP
S208TTA-dn	CGTTCGGGTCCTTTAACAGGACGGACTGC (Used with EGFP-pSUA1-X-up)	710	pSUA5-sfGFP
S208TTA-up	GCAGTCCGTCCTGTTAAAGGACCCGAACG (Used with His-BglII-dn)	146	pSUA5-sfGFP
N212TTA-dn	CCGCTTCTCTAACGGGTCCTTGGACAGG (Used with EGFP-pSUA1-X-up)	718	pSUA5-sfGFP
N212TTA-up	CCTGTCCAAGGACCCGTTAGAGAAGCGG (Used with His-BglII-dn)	137	pSUA5-sfGFP
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	827	PCR pdts.

Table 3-3. Primer information for constructing 13 single mutant sfGFP plasmids

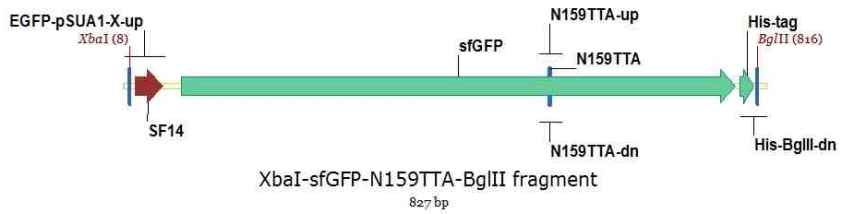


Figure 3-3. Map of representative fragment, which is used for constructing pIPheRS-N159TTA

Construction of pIPheRS-TTA5s. pIPheRS-TTA5s harbors a 5 TTA incorporated sfGFP gene, which was made from assembling 5 fragments with TTA codons at the junction. The pIPheRS-S147TTA was used as the template to amplify these 5 fragments. The five fragments were amplified with primer pairs EGFP-pSUA1-X-up and T43TTA-dn, T43TTA-up and T97TTA-dn, T97TTA-up and N164TTA-dn, N164TTA-up and N212TTA-dn, N212TTA-up and His-BglII-dn, respectively. These fragments were then assembled by overlap extension PCR using primes EGFP-pSUA1-X-up and His-BglII-dn to generate the PCR product, which was digested with XbaI and BglII restriction enzymes. The resulting product was ligated into pIPheRS-Y151TTA digested with same restriction enzymes. The construct was verified by restriction mapping and sequencing of the inset. Primer information is given in the table below. Map of fragment is given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
EGFP-pSUA1-X-up	(Used with T43TTA-dn)	217	pIPheRS-S147TTA
T43TTA-up	(Used with T97TTA-dn)	193	pIPheRS-S147TTA
T97TTA-up	(Used with N164TTA-dn)	230	pIPheRS-S147TTA
N164TTA-up	(Used with N212TTA-dn)	170	pIPheRS-S147TTA
N212TTA-up	(Used with His-BglII-dn)	137	pIPheRS-S147TTA
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	827	PCR pdts.

Table 3-4. Primer information for constructing pIPheRS-TTA5s

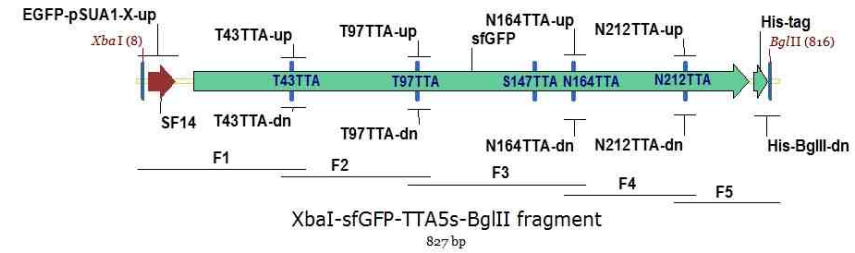


Figure 3-4. Map of fragment for making pIPheRS-TTA5s

Construction of pIPheRS-TTA5s-2, pIPheRS-TTA10s, pIPheRS-TTA15s, pIPheRS-TTA1t, pIPheRS-TTA2t, pIPheRS-TTA3t, pIPheRS-TTA5t, and pIPheRS-TTA10t. After 15 single TTA sites were screened based on the fluorescence analysis, they were ranked from high to low and top 5, top 10 and all of the 15 sites were combined to have the sfGFP-TTA5s-2, the sfGFP-TTA10s, the sfGFP-TTA15s genes. To construct pIPheRS-TTA5s-2, pIPheRS-TTA10s, and pIPheRS-TTA15s, 3 fragments (TTA5s-2 F1-F3), 5 fragments (TTA10s F1-F5), and 5 fragments (TTA15s F1-F5) were amplified, respectively. TTA5s-2 F1-F3 fragments were amplified using primer pairs EGFP-pSUA1-X-up and S147Y151-dn, S147Y151-up and Y200TTA-dn, Y200TTA-up and His-BglII-dn, respectively. These three fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn to generate TTA5s-2 PCR product. TTA10s F1-F5 fragments were amplified using primer pairs EGFP-pSUA1-X-up and T38T43-dn, T38T43-up and T105TTA-dn, T105TTA-up and S147N149Y151-dn, S147N149Y151-up and S208N212-dn, S208N212-up and His-BglII-dn, respectively. These five fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn to generate TTA10s PCR product. TTA15s F1-F5 fragments were amplified using primer pairs EGFP-pSUA1-X-up and T97S99-dn, T97S99-up and T118TTA-dn, T118TTA-up and N159N164-dn, N159N164-

up and T186TTA-dn, T186TTA-up and His-BglII-dn, respectively. These five fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn to generate TTA15s PCR product. These three resulting PCR products were digested with XbaI and BglII restriction enzymes and ligated into pIPheRS-Y151TTA digested with the same restriction enzymes.

	pIPheRS-TTA		
position	5s-2	10s	15s
38		✔	✔
43		✔	✔
97	✔	✔	✔
99			✔
105		✔	✔
118			✔
147	✔	✔	✔
149		✔	✔
151	✔	✔	✔
159			✔
164			✔
186			✔
200	✔	✔	✔
208		✔	✔
212	✔	✔	✔

Table 3-5. Information of codon mutations harbored in these 5, 10, and 15 scattered TTA codons containing constructs

One, two, three, five or ten tandem TTA codons were inserted into sfGFP between its start codon and second codon. To construct these mutant sfGFP genes, two fragments with different number of TTA codons at the junction were amplified. Five primer pairs were design and used with primer EGFP-pSUA1-X-up and His-BglII-dn to obtain these five mutant sfGFP genes. The upstream fragments were amplified using primers EGFP-pSUA1-X-up and the reverse primers with TTA codons; and the downstream fragments were amplified using primes: the corresponding direct primers with TTA codons and His-BglII-dn. The upstream fragment and corresponding downstream fragment were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and His-BglII-dn to generate these 5 mutant sfGFP, which were digested with XbaI and BglII restriction enzymes. These digested PCR products were ligated into pIPheRS-Y151TTA digested with same restriction enzymes to generate pIPheRS-TTA1t, pIPheRS-TTA2t, pIPheRS-TTA3t, pIPheRS-TTA5t, and pIPheRS-TTA10t.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
S147Y151-dn	CGCGGTGATTAACACGTTGTGTAAGTTGAAGTTGTACTCG (Used with EGFP-pSUA1-X-up)	535	pIPheRS-T97
S147Y151-up	CGAGTACAACTTCAACTTACACAACGTGTTAATCACCGCG (Used with Y200TTA-dn)	191	pIPheRS-T97
Y200TTA-up	(Used with His-BglII-dn)	169	pIPheRS-N212
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	827	PCR pdt.
T38T43-dn	GAACTTCAGTAACAGCTTGCCGTTTAAGGCGTCCC (Used with EGFP-pSUA1-X-up)	211	pIPheRS-T38
T38T43-up	GGGACGCCTTAAACGGCAAGCTGTTACTGAAGTTC (Used with T105TTA-dn)	224	pIPheRS-T97
S147N149Y151-dn	CGCGGTGATTAACACTAAGTGTAAGTTGAAGTTGTACTCG (Used with T105TTA-up)	163	pIPheRS-TTA5s-2
S147N149Y151-up	CGAGTACAACTTCAACTTACACTTAGTGTTAATCACCGCG	215	pIPheRS-TTA5s-2
S208N212-dn	CTAACGGGTCCTTTAACAGGACGGACTGC		
S208N212-up	GCAGTCCGTCCTGTTAAAGGACCCGTTAG (Used with His-BglII-dn)	146	pIPheRS-N212
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	827	PCR pdt.
T97S99-dn	CCGTCGTCCTTGAATAAGATTAAGCGCTCC (Used with EGFP-pSUA1-X-up)	384	pIPheRS-TTA10s
T97S99-up	GGAGCGCTTAATCTTATTCAAGGACGACGG (Used with T118TTA-dn)	85	pIPheRS-TTA10s
N159N164-dn	GGATCTTGAATAACGCCTTGATGCCTAACTTCTGCTTG (Used with T118TTA-up)	161	pIPheRS-TTA10s
N159N164-up	CAAGCAGAAGTTAGGCATCAAGGCGTTATTCAAGATCC (Used with T186TTA-dn)	107	pIPheRS-TTA10s
T186TTA-up	(Used with His-BglII-dn)	212	pIPheRS-TTA10s
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	827	PCR pdt.
1TTA-dn	GAACAGCTCCTCGCCCTTGCTTAACATTTCTAATTTCT (Used with EGFP-pSUA1-X-up)	100	pSUA5-sfGFP
1TTA-up	AGAAATTAGAAATGTTAAGCAAGGGCGAGGAGCTGTTC (Used with His-BglII-dn)	768	pSUA5-sfGFP
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	830	PCR pdt.
2TTA-dn	CCTCGCCCTTGCTTAATAACATTTCTAATTTCTCCTCTAATCG (Used with EGFP-pSUA1-X-up)	95	pSUA5-sfGFP
2TTA-up	GAAATGTTATTAAGCAAGGGCGAGGAGCTGTTCAC (Used with His-BglII-dn)	763	pSUA5-sfGFP
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	833	PCR pdt.
3TTA-dn	CTCCTCGCCCTTGCTTAATAATAACATTTCTAATTTCTCCTCTAATCGAC (Used with EGFP-pSUA1-X-up)	100	pSUA5-sfGFP
3TTA-up	GAAATGTTATTATTAAGCAAGGGCGAGGAGCTGTTCAC (Used with His-BglII-dn)	766	pSUA5-sfGFP
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	836	PCR pdt.
5TTA-dn	TCCTCGCCCTTGCTTAATAATAATAATAACATTTCTAATTTCTCCTCTAATCGAC (Used with EGFP-pSUA1-X-up)	105	pIPheRS-TTA3t
5TTA-up	GAAATGTTATTATTATTATTAAGCAAGGGCGAGGAGCTGTTCAC (Used with His-BglII-dn)	772	pIPheRS-TTA3t
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	842	PCR pdt.
10TTA-dn	TAATAATAATAATAATAATAATAATAATAACATTTCTAATTTCTCCTCTAATCGAC (Used with EGFP-pSUA1-X-up)	106	pIPheRS-TTA5t
10TTA-up	TTAGAAATGTTATTATTATTATTATTATTATTATTATTAAGCAAGGGCGAGGAGCTGTTC (Used with His-BglII-dn)	790	pIPheRS-TTA5t
EGFP-pSUA1-X-up	(Used with His-BglII-dn)	857	PCR pdt.

Table 3-6. Primer information for constructing multiple scattered and tandem TTA codons containing constructs

All of these constructs were verified by restriction mapping and sequencing of the insert. Primer information is given in the table. The priming region of each primer is underlined and mutated codon is shown in blue.

Construction of pIPheRS-wtsfGFP, pIPheRS-TTAn1s, pIPheRS-TTAn2s, pIPheRS-TTAn3s, pIPheRS-TTAn5s, pIPheRS-TTAn10s, pIPheRS-TAGn1s, pIPheRS-TAGn2s, pIPheRS-TAGn3s, pIPheRS-TAGn5s, and pIPheRS-TAGn10s. All of these 11 constructs are derivate of pIPheRS-Y151TTA, in which the C-terminal histidine tag was removed and a N-terminal histidine tag plus two amino acids (glycine-serine) was inserted between first and second residues of sfGFP. Besides that, other mutations harbored in these 11 constructs were given in the table below.

pIPheRS											
TTA/TAG	wtsfGFP	TTA	TAG	TTA	TAG	TTA	TAG	TTA	TAG	TTA	TAG
position	none	n1s	n1s	n2s	n2s	n3s	n3s	n5s	n5s	n10s	n10s
38										TTA	TAG
43										TTA	TAG
97		TTA	TAG	TTA	TAG	TTA	TAG	TTA	TAG	TTA	TAG
105										TTA	TAG
147				TTA	TAG	TTA	TAG	TTA	TAG	TTA	TAG
149										TTA	TAG
151								TTA	TAG	TTA	TAG
200						TTA	TAG	TTA	TAG	TTA	TAG
208										TTA	TAG
212								TTA	TAG	TTA	TAG

Table 3-7. Information of codon mutations harbored in these 11 N-terminal histidine tagged sfGFP

Primers EGFP-pSUA1-X-up and SF14-adapter-dn were annealed together, which was used as template to obtain the PCR product using the primers EGFP-pSUA1-X-up and SF14-N-His-dn. Another fragment was amplified using primers N-His-sfGFP-up and sfGFP-BglII-dn. These two fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and sfGFP-BglII-dn, which was digested with XbaI and BglII. This digested PCR product was ligated into pIPheRS-Y151TTA digested with the same restriction enzymes to generate pIPheRS-wtsfGFP, which was used to amplify the

following 7 fragments (TTAns F1-F6, F8) using primer pairs EGFP-pSUA1-X-up and T97TTA-dn, T97TTA-up and sfGFP-BglII-dn, T97TTA-up and sfGFP-S147TTA-dn, sfGFP-S147TTA-up and sfGFP-BglII-dn, sfGFP-S147TTA-up and Y200TTA-dn, Y200TTA-up and sfGFP-BglII-dn, EGFP-pSUA1-X-up and T38T43-dn, respectively. TTAns F7 was amplified from pIPheRS-TTA5s-2 using primers T97TTA-up and sfGFP-BglII-dn; and TTAns F9 was amplified from pIPheRS-TTA10s using primers T38T43-up and sfGFP-BglII-dn. Next, 5 group of fragments (group one: TTAns F1 and F2; group two: TTAns F1, F3 and F4; group three: TTAns F1, F3, F5 and F6; group four: TTAns F1 and F7; group five: TTAns F8 and F9) were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and sfGFP-BglII-dn to generate five mutant sfGFP genes, which were digested with XbaI and BglII restriction enzmyes. The resulting products were ligated into pIPheRS-Y151TTA digested with the same restriction enzymes to generate pIPheRS-TTAn1s, pIPheRS-TTAn2s, pIPheRS-TTAn3s, pIPheRS-TTAn5s, and pIPheRS-TTAn10s, respectively.

For construction of the other five of N-terminal histidine tagged TAG mutant sfGFP constructs, EcoRI restriction site between sfGFP and $tRNA^{Tyl}_{CUA}$ was used instead of BglII due to occurrence of BglII after TAG codon installation at 97 site. Two fragments (TAGns F1-F2) were amplified from pIPheRS-wtsfGFP using primer pairs EGFP-pSUA1-X-up and T97TAG-dn, T97TAG-up and sfGFP-EcoRI-dn, and the resulting fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and sfGFP-EcoRI-dn, which was digested with XbaI and EcoRI restriction enzymes. Another fragment was amplified from opt-pSUA2-sfGFP using primers EcoRI-tRNA-up and tRNA-PvuI-dn, which was digested with EcoRI and PvuI restriction

enzymes. The two digested products were ligated into pIPheRS-Y151TTA digested with XbaI and PvuI restriction enzymes using the three-way ligation method to generate pIPheRS-TTAn1s. Next, four fragments (TAGns F3-F6) were amplified from pIPheRS-wtsfGFP using primer pairs T97TAG-up and S147TAG-dn, S147TAG-up and sfGFP-EcoRI-dn, S147TAG-up and Y200TAG-dn, Y200TAG-up and sfGFP-EcoRI-dn, respectively. Two groups of fragments (group one: TAGns F1, F3, and F4; group two: TAGns F1, F3, F5, and F6) were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and sfGFP-EcoRI-dn to generate two mutant sfGFP genes, which were digested with XbaI and BglII restriction enzmyes. The digested products were ligated into pIPheRS-TAGn1s to generate pIPheRS-TTAn2s and pIPheRS-TTAn3s. TAGns F8-F10 were amplified from pIPheRS-TTAn3s using primer pairs EGFP-pSUA1-X-up and S147Y151TAG-dn, S147Y151TAG-up and N212TAG-dn, N212TAG-up and sfGFP-EcoRI-dn, respectively. These three fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and sfGFP-EcoRI-dn to obtain the PCR product, which was digested with XbaI and EcoRI. The digested product was ligated into pPheRS-TAGn1s digested with the same enzymes to generate pIPheRS-TAGn5s, which was employed as template to amplify five fragments (TAGns F11-F15) using primers EGFP-pSUA1-X-up and T38T43TAG-dn, T38T43TAG-up and T105TAG-dn, T105TAG-up and S147N149Y151TAG-dn, S147N149Y151TAG-up and S208N212TAG-dn, S208N212TAG-up and sfGFP-EcoRI-dn, respectively. These five fragments were assembled by overlap extension PCR using primers EGFP-pSUA1-X-up and sfGFP-EcoRI-dn to obtain the PCR product, which was digested with XbaI and

EcoRI restriction enzymes. The digested product was ligated into pPheRS-TAGn1s digested with the same enzymes to generate pIPheRS-TAGn10s.

All of these constructs were verified by restriction mapping and sequencing of the insert. Primer information is given in the table below. The priming region of each primer is underlined and mutated codon is shown in blue.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
SF14-adapter-dn	TTCTAATTTCTCCTCTAATCGACTATTGATTGTAGCTCACGCC (Used with EGFP-pSUA1-X-up)	73	None
SF14-N-His-dn	GCCCTTGCTCGAGCCGTGGTGGTGATGGTGGTGCATTTCTAAT TTCTCCTCTAATCGACT (Used with EGFP-pSUA1-X-up)	109	PCR pdt.
N-His-sfGFP-up	CCACGGCTCGAGCAAGGGCGAGGAGCTGTTC	737	pSUA5-sfGFP
sfGFP-BglII-dn	CACAATTAGATCTTCACTTGTAGAGCTCGTCCATCCCGTGGGT		
EGFP-pSUA1-X-up	(Used with sfGFP-BglII-dn)	827	PCR pdt.
EGFP-pSUA1-X-up	(Used with T97TTA-dn)	401	pIPheRS-wtsfGFP
T97TTA-up	(Used with sfGFP-BglII-dn)	456	pIPheRS-wtsfGFP
EGFP-pSUA1-X-up	(Used with sfGFP-BglII-dn)	827	PCR pdt.
EGFP-pSUA1-X-up	(Used with T97TTA-dn)	401	pIPheRS-wtsfGFP
T97TTA-up	(Used with sfGFP-S147TTA-dn)	188	pIPheRS-wtsfGFP
sfGFP-S147TTA-up	(Used with sfGFP-BglII-dn)	308	pIPheRS-wtsfGFP
EGFP-pSUA1-X-up	(Used with sfGFP-BglII-dn)	827	PCR pdt.
EGFP-pSUA1-X-up	(Used with T97TTA-dn)	401	pIPheRS-wtsfGFP
T97TTA-up	(Used with sfGFP-S147TTA-dn)	188	pIPheRS-wtsfGFP
sfGFP-S147TTA-up	(Used with Y200TTA-dn)	191	pIPheRS-wtsfGFP
Y200TTA-up	(Used with sfGFP-BglII-dn)	145	pIPheRS-wtsfGFP
EGFP-pSUA1-X-up	(Used with sfGFP-BglII-dn)	827	PCR pdt.
EGFP-pSUA1-X-up	(Used with T97TTA-dn)	401	pIPheRS-wtsfGFP
T97TTA-up	(Used with sfGFP-BglII-dn)	456	pIPheRS-TTA5s-2
EGFP-pSUA1-X-up	(Used with sfGFP-BglII-dn)	827	PCR pdt.
EGFP-pSUA1-X-up	(Used with T38T43-dn)	235	pIPheRS-wtsfGFP
T38T43-up	(Used with sfGFP-BglII-dn)	627	pIPheRS-TTA10s
EGFP-pSUA1-X-up	(Used with sfGFP-BglII-dn)	827	PCR pdt.
T97TAG-dn	CCTTGAACGAGATCTAGCGCTCCTGCACGT (Used with EGFP-pSUA1-X-up)	401	pIPheRS-wtsfGFP
T97TAG-up	ACGTGCAGGAGCGCTAGATCTCGTTCAAGG	452	pIPheRS-wtsfGFP
sfGFP-EcoRI-dn	CCGGAATTCTCACTTGTAGAGCTCGTCCATCCCGTGGG		
EGFP-pSUA1-X-up	(Used with sfGFP-EcoRI-dn)	823	PCR pdt.
EcoRI-tRNA-up	CCGGAATTCAATTGTGCTGGCCCCGGAATCCGCTAG (Used with tRNA-PvuI-dn)	177	opt-pSUA2-sfGFP
EGFP-pSUA1-X-up	(Used with T97TAG-dn)	401	pIPheRS-wtsfGFP
S147TAG-dn	CGCGGTGATGTACACGTTGTGCTAGTTGAAGTTGTACTCG (Used with T97TAG-up)	188	pIPheRS-wtsfGFP
S147TAG-up	CGAGTACAACTTCAACTAGCACAACGTGTACATCACCGCG (Used with sfGFP-EcoRI-dn)	304	pIPheRS-wtsfGFP
EGFP-pSUA1-X-up	(Used with sfGFP-EcoRI-dn)	823	PCR pdt.
EGFP-pSUA1-X-up	(Used with T97TAG-dn)	401	pIPheRS-wtsfGFP
T97TAG-up	(Used with S147TAG-dn)	188	pIPheRS-wtsfGFP

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
Y200TAG-dn	ACTGCGTGGAGAGCTAGTGGTTGTCGGG (Used with S147TAG-up)	191	pIPheRS-wtsfGFP
Y200TAG-up	CCCGACAACCACTAGCTCTCCACGCAGT (Used with sfGFP-EcoRI-dn)	141	pIPheRS-wtsfGFP
EGFP-pSUA1-X-up	(Used with sfGFP-EcoRI-dn)	823	PCR pdt.
S147Y151TAG-dn	CGCGGTGATCTACACGTTGTGCTAGTTGAAGTTGTACTCG (Used with EGFP-pSUA1-X-up)	559	pIPheRS-TAGn3s
S147Y151TAG-up	CGAGTACAACTTCAACTAGCACAACGTGTAGATCACCGCG	223	pIPheRS-TAGn3s
N212TAG-dn	CCGCTTCTCCTACGGGTCCTTGGACAGG		
N212TAG-up	CCTGTCCAAGGACCCGTAGGAGAAGCGG (Used with sfGFP-EcoRI-dn)	109	pIPheRS-TAGn3s
EGFP-pSUA1-X-up	(Used with sfGFP-EcoRI-dn)	823	PCR pdt.
T38T43TAG-dn	GAACTTCAGCTACAGCTTGCCGTTCTAGGCGTCCC (Used with EGFP-pSUA1-X-up)	235	pIPheRS-TAGn5s
T38T43TAG-up	GGGACGCCTAGAACGGCAAGCTGTAGCTGAAGTTC	224	pIPheRS-TAGn5s
T105TAG-dn	CCGCGTCTTGTACTAGCCGTCGTCCTTG		
T105TAG-up	CAAGGACGACGGCTAGTACAAGACGCGG	163	pIPheRS-TAGn5s
S147N149Y151TAG-dn	CGCGGTGATCTACACCTAGTGCTAGTTGAAGTTGTACTCG		
S147N149Y151TAG-up	CGAGTACAACTTCAACTAGCACTAGGTGTAGATCACCGCG	215	pIPheRS-TAGn5s
S208N212TAG-dn	CCTACGGGTCCTTCTACAGGACGGACTGC		
S208N212TAG-up	GCAGTCCGTCCTGTAGAAGGACCCGTAGG (Used with sfGFP-EcoRI-dn)	118	pIPheRS-TAGn5s
EGFP-pSUA1-X-up	(Used with sfGFP-EcoRI-dn)	823	PCR pdt.

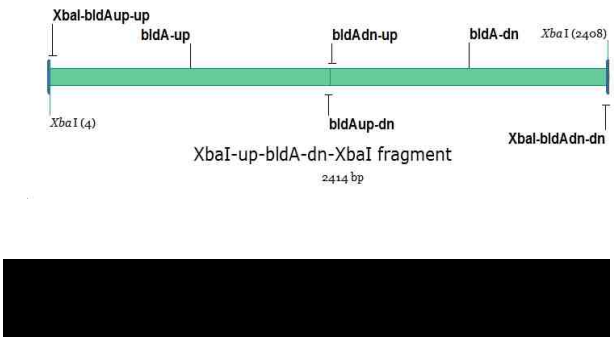
Table 3-8 (cont.). Primer information for constructing 11 N-terminally histidine tagged sfGFP constructs

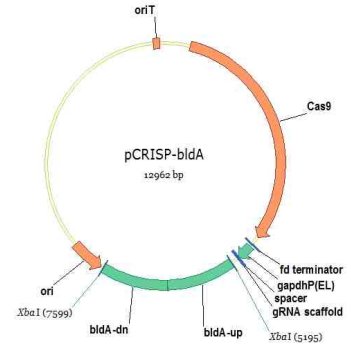
Construction of pCRISP-bldA. pCRISP-bldA is a derivate of pCRISPomyces-2,[15] which was requested from Addgene (Plasmid #61737). To construct pCRISP-bldA, the spacer sequences and homology arms were sequentially inserted into pCRISPomyces-2. Based on the genome sequence of wild type *S. venezuelae* ATCC 15439, fourteen potential spacer sequences were screened by analyzing CRIPSR/Cas off-target effect[18] against its genome and this sequence (5′-ACGGCGAGCTTAAACCTCGC-3′) has the least off-target, which was chosen to construct pCRISP-bldA. Two complementary primers spacer-up and spacer-dn with specific sticky ends on both side were designed and ordered. These two primers were mixed together in one to one ratio with fianl concentration of 5 μM each in 50 μL soulution. The primers solution was heated up, kept

at 95 °C for 5 minutes, and slowly cooled down to the room temperature. One μL of 10 fold diluted of this resulting primers solution was ligated into pCRISPomyces-2 digested with BbsI restriction enzyme to generate the intermediate construct. Next, two fragments were amplified from the genomic DNA of wild type *S. venezuelae* ATCC 15439 using primer pairs XbaI-bldAup-up and bldAup-dn, bldAdn-up and XbaI-bldAdn-dn, respectively. These two fragments were spliced by overlap extension PCR using primers XbaI-bldAup-up and XbaI-bldAdn-dn to generate the homology arm with XbaI site on both ends. This PCR product, containing 1204 bp upstream and 1194 bp downstream of *bldA* gene, was digested with XbaI restriction enzyme , and then ligated into the intermediate construction digested with the same restriction enzyme to generate the final constrution pCRISP-bldA. Either orientation of the homology arm is acceptable. This construct was verified by restriction mapping and sequencing of the inserts. Primer information is given in the table below. The priming region of each primer is underlined and the restriction enzyme site is shown in cyan. Map of the vector is also given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
XbaI-bldAup-up	GCTCTAGATGTGTCCTGGAGGAGGGTGGGGCGGACG	1220	genomic DNA of *S.venezuelae*
bldAup-dn	GCCCGGGTGTTGCGCGAAGCTCCGCGTTGGCAC		
bldAdn-up	CTTCGCGCAACACCCGGGCCGCAGGCGAGTGCCTGG	1213	genomic DNA of *S.venezuelae*
XbaI-bldAdn-dn	GCTCTAGAGGTAGGCCCTGTCCGGGAACGCGGAAAGG		
XbaI-bldAup-up	(Used with XbaI-bldAdn-dn)	2414	PCR pdt.
spacer-up	ACGCACGGCGAGCTTAAACCTCGC	28	None
spacer-dn	AAACGCGAGGTTTAAGCTCGCCGT		

Table 3-9. Primer information for constructing pCRISP-bldA





4. **Results and Discussion**

Western Blotting Analysis of pSUA2-TTA and pSUA4-TTA in J1681. pSUA2-TTA , which contains MjTyrRS/tRNA$^{Tyr}_{UAA}$ pair and sfGFP with TTA codon at the 151 site; pSUA4, which only contains sfGFP with TTA codon at the 151 site, were constructed and transferred into J1681 (*S. coelicolor ΔbldA*). The unique tRNA gene *bldA*, which can recognized TTA codon in *Streptomyces*, is knocked out from its genome in J1681 and the MjTyrRS/tRNA$^{Tyr}_{UAA}$ pair harbored in pSUA2-TTA is capable of decoding the TTA codon with MjTyrRS. The result (Figure 3-6) indicate that in presence of the MjTyrRS/tRNA$^{Tyr}_{UAA}$ pair (lane 1), the TTA codon was decoded. In absence of the MjTyrRS/tRNA$^{Tyr}_{UAA}$ pair (lane 2), TTA codon was not translated through. This result suggests that the TTA sense codon reassignment system works in J1681.

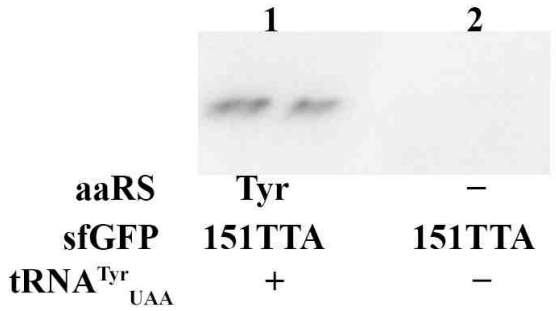


Figure 3-6. Anti-His Western blot of Ni-NTA affinity purified sfGFP proteins obtained from J1681 harboring pSUA2-TTA (lane 1) and pSUA4-TTA (lane 2). The blot was imaged for 258 s.

Western Blotting and ESI-MS Analysis of pIPheRS-Y151TTA and pBpaRS-S147TTA in J1681. After the function validation of MjTyrRS/tRNA$^{Tyr}_{UAA}$ pair in J1681 for reassigning the TTA codon to tyrosine, pIPheRS-Y151TTA and pBpaRS-S147TTA were constructed. pIPheRS-Y151TTA harboring the pIPheRS and TTA containing sfGFP was designed to incorporate pIPhe into TTA codon site in presence of pIPhe. Western blotting result of sfGFP obtained from J1681 harboring pIPheRS-Y151TTA grown in

presence of 1 mM pIPhe (Figure 3-7, A-3, lane 1) suggests that pIPhe is incorporated into sfGFP. For in absence of pIPhe, similar result (Figure 3-7, A-3, lane 2) is observed as in *S. venezuelae* that endogenous amino acids are recognized by pIPheRS. pBpa cannot be incorporated into sfGFP by amber suppression so far.[19] This TTA codon reassignment system should have higher UAA incorporation efficiency due to lack of competing with translational termination signals and endogenous tRNAs. Then the pBpaRS-S147TTA was constructed to test if TTA reassignment system could incorporate pBpa into sfGFP. Western blotting results of sfGFP proteins obtained from J1681 harboring pBpaRS-S147TTA in the presence (Figure 3-7, B-3, lane3) and the absence (Figure 3-7, B-3, lane 4) of 1 mM pBpa suggest that even though endogenous amino acids are still recognized by pBpaRS, pBpa is first time genetically incorporated into sfGFP protein in *Streptomyces*. To further validate the homogenous UAAs incorporation, large-scale

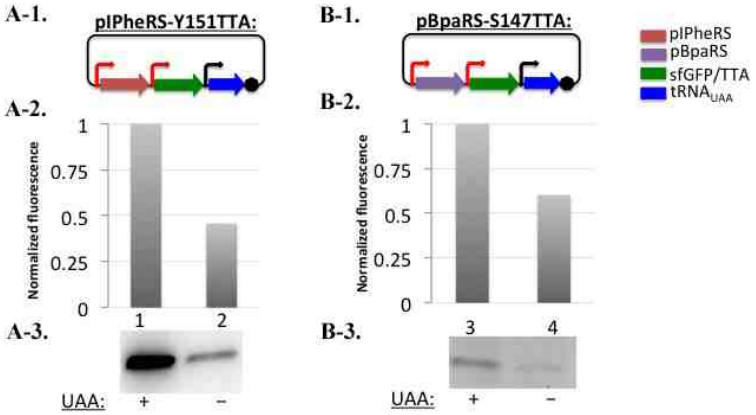


Figure 3-7. A-1. Schematic of pIPheRS-Y151TTA. B-1. Schematic of pBpaRS-S147TTA. A-2. Fluorescence quantification of proteins obtained from J1681 harboring pIPheRS-Y151TTA in the presence (lane 1) and the absence (lane 2) of pIPhe. B-2. Fluorescence quantification of proteins obtained from J1681 harboring pBpaRS-S147TTA in the presence (lane 3) and the absence (lane 4) of pBpa. A-3. Anti-His western blot of Ni-NTA affinity purified sfGFP proteins obtained from J1681 harboring pIPheRS-Y151TTA in the presence (lane 1) and the absence (lane 2) of pIPhe. The blot was imaged for 273 s. B-3. Anti-His western blot of Ni-NTA affinity purified sfGFP proteins obtained from J1681 harboring pBpaRS-S147TTA in the presence (lane 3) and the absence (lane 4) of pBpa. The blot was imaged for 563 s.

expression and purification were conducted for pIPheRS-Y151TTA and pBpaRS-S147TTA, followed by ESI-MS analysis. The results (Figure 3-8) demonstrate that complete pIPhe (Figure 3-8, A) and pBpa (Figure 3-8, B) were incorporated into sfGFP by the novel TTA codon reassignment system in *Streptomyces*. The pBpa incorporation directly confirms that in *Streptomyces* TTA codon reassignment system is more advanced and efficient than the amber suppression described in Chapter 2.

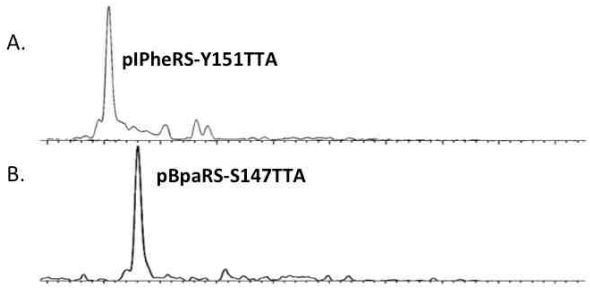


Figure 3-8. Deconvoluted ESI-MS data of sfGFP proteins purified from J1681 harboring A. pIPheRS-Y151TTA; B. pBpaRS-S147TTA

Screening of 15 Sites in sfGFP for UAA Incorporation. TTA codon reassignment was successfully developed and it can achieve quantitative UAA incorporation. To further characterize this new system, multiple UAAs incorporation would be the next step. Instead of choosing different sites and combining them to make multiple sites constructs, thoroughly screening analysis was made to find 15 ideal sites for UAA incorporation in sfGFP. To fully understand the performance of each site, 15 constructs were prepared with a single TTA codon at 15 different sites. Fluorescence quantification result (Figure 3-9) obtained from these 15 single TTA mutant sfGFP constructs grown in presence of 1 mM pIPhe suggests the context of TTA codon has affect on the UAA

incorporation efficiency. These fluorescence quantification result would facilitate us to rational choose sites for multiple UAAs incorporation construct design.

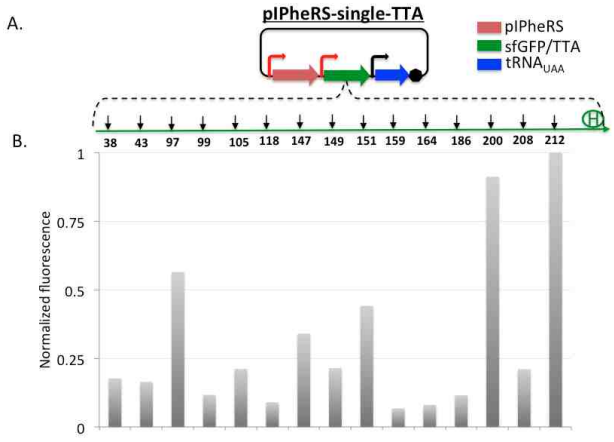


Figure 3-9. A. Schematics of positions of 15 single TTA codons. B. Fluorescence quantification of proteins obtained from J1681 harboring constructs containing these 15 different single TTA codon

Western Blotting and ESI-MS Analysis of pIPheRS-TTA5s. Five (T43, T97 S147, N164, and N212) of those 15 sites were combined because these five sites are distant from each other in sfGFP structure, which can minimize self-interference from these five pIPhe incorporation. pIPheRS-TTA5s was constructed and transferred into J1681. The western blotting result (Figure 3-10, B) obtained from J1681 harboring pIPheRS-TTA5s grown in the presence and the absence of 1 mM pIPhe indicates that five TTA codons in a single sfGFP protein were decoded by the pIPheRS/tRNA$^{Tyr}_{UAA}$ pair in presence of pIPhe and the endogenous amino acids can not be incorporated into 5 TTA codons in a single sfGFP. These results confirm that lack of competition with translational termination signals for stop codon suppression or a large pool of tRNA for other sense codon reassignment makes this TTA codon reassignment unique and highly efficient.

To obtain the molecular weight of mutant sfGFP protein, large-scale expression and purification was conducted. During protein purification, notable different was observed that sfGFP produced by J1681 harboring pIPheRS-TTA5s was too hydrophobic to be eluted out from the HIC column with even low concentration of salt buffer. Eventually, only 13% of total protein applied to the HIC column was eluted out using pure deionized water according to the fluorescence quantification analysis, which suggest that five pIPhe incorporation altered sfGFP protein properties. The pure mutant sfGFP was then subjected to ESI-MS analysis (Figure 3-10, C), which demonstrate five pIPhe were quantitatively incorporated into a single sfGFP protein. The unique peak (Figure 3-10, C) suggests the aminoacyl-tRNA charged by pIPhe can compete out the one charged by endogenous amino acids.

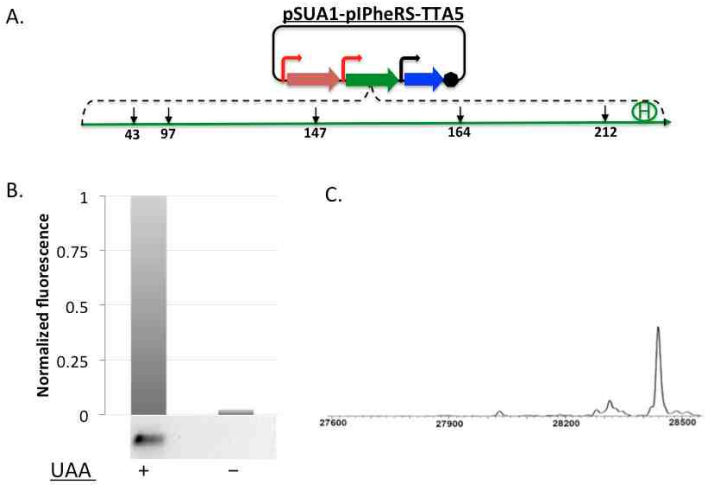


Figure 3-10. A. Schematics of pIPheRS-TTA5s. B. Fluorescence quantification and anti-His western blotting of proteins obtained from J1681 harboring pIPheRS-TTA5s. C. Deconvoluted ESI-MS of protein purified from J1681 harboring pIPheRS-TTA5s in presence of 1 mM pAzPhe. The blot was imaged for 295s.

Growth of J1681 with TTA Sense Codon Reassignment. Undiscriminating UAAs incorporation into proteome is the inherent drawback of sense codon reassignment, which could lead to significant cell fitness loss or even death.[3] The unique feature of

Streptomyces genetic makes TTA a completely free codon for reassignment when its recognizing tRNA gene is deleted. The free TTA codon should not only afford high efficient but also cell friendly UAA incorporation system. To assess the effect of TTA codon reassignment on the host cell, growth curves of J1681 harboring pEmt (empty plasmid, used as control), pSUA2-TTA, and pIPheRS-Y151TTA were obtained. 1 mM pIPhe was added to the J1681 harboring pIPheRS-Y151TTA. The growth results (Figure 3-11) indicate that there are no detectable different between pEmpt and the other two TTA codon reassignment constructs. They started their exponential phase and reached to stationary phase about the same time and they had the similar OD_{600} reading when they were in the stationary phase. This result (Figure 3-11) confirms that TTA codon reassignment is non-toxic in J1681.

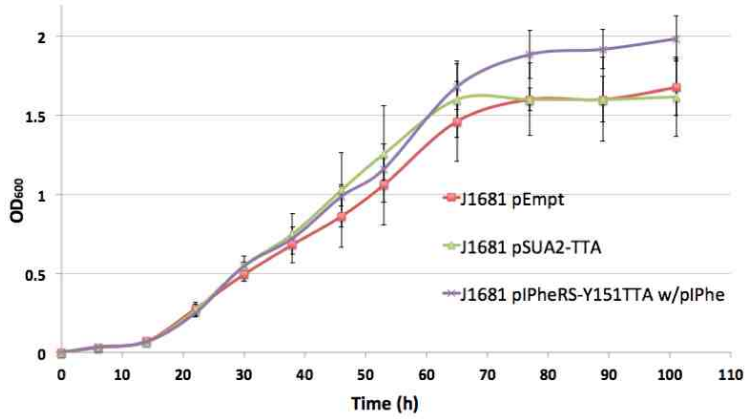


Figure 3-11. Growths of J1681 harboring pEmpt (orange line), pSUA2-TTA (green line), and pIPheRS-Y151TTA with pIPhe (purple line).

Genome Sequencing of S. venezuelae ATCC 15439 using PacBio Technology. High quality genomic DNA (most of DNA pieces are 50 kb) was isolated from *S. venezuelae* ATCC 15439 using the procedures described in the Experimental Procedures of this chapter. This high quality genomic DNA can also facilitate large size and high G-

C content DNA amplification. The resulting genome sample was sent to National Center for Genome Resources (NCGR) in Santa Fe for sequencing using PacBio next-generation technology. This genome sequence was deposited in EMBL/GenBank with the accession no. LN881739.[14]

Construction of S. venezuelae ΔbldA Mutant (Sven 624) using CRISPR/Cas. Strain J1681 grows slowly and forms cell clumps in liquid culture, which result in less biomass and difficulty of quantitative analysis between mutants. To avoid these difficuties, *S. venezuelae ΔbldA* mutant would be a good host to work with. With the genome sequence of wild-type *S. venezuelae*, pCRISP-bldA was constructed and transferred into *S. venezuelae*. The transformants took longer time to become visible than the other plasmids due to the toxicity of CRISPR/Cas9. The transformants were screened by digestion and sequencing of the amplified DNA fragment from their genome. One of four transformants showed about 60% of genome was *bldA* deletion. The strain was cultured at 37 °C and plated on SPA after serial dilution. More than one hundred colonies were screened and 4 clones showed the apramycin sensitive restoration. Genomic DNA

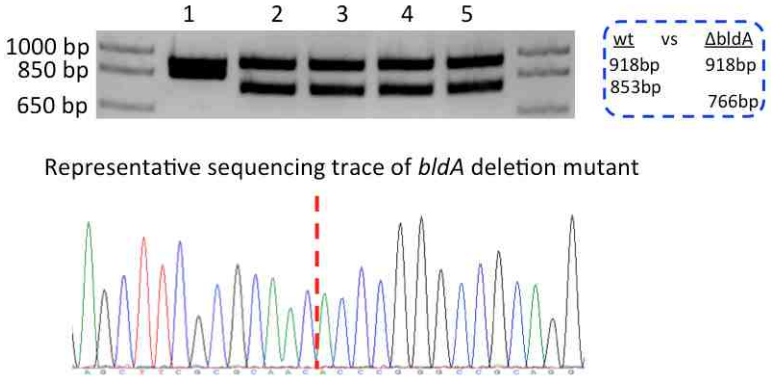


Figure 3-12. Digestion results of PCR products amplified from wild type *S. venezuelae* (lane 1), mutant 1 (lane 2), mutant 2 (lane 3), mutant 3 (lane 4), and mutant 4 (lane5) and the representative sequencing trace of *bldA* deletion mutant

was isolated from those 4 clones and the *bldA* region was amplified. Then the PCR products were digested and sequenced. The PCR digestion result (Figure 3-12) suggests that all of the 4 clones are *ΔbldA* mutants, which is further confirmed by the sequence result. Until then, *bldA* gene was in-frame deleted by CRISPR/Cas9 technique from *S. venezuelae* ATCC 15439 to generate *S. venezuelae ΔbldA* strain, which is named as Sven 624.

Western Blotting Analysis of pSUA2-TTA and pSUA4-TTA and Comparison of TAG Suppression and TTA Reassignment in Sven 624. After construction of Sven 624 strain, pSUA2-TTA and pSUA4-TTA were transferred into Sven 624 to examine functionality of TTA codon reassignment in the new host. Western blotting result confirms (Figure 3-13, A) that TTA codon in sfGFP was only decoded in presence of MjTyrRS/$tRNA^{Tyr}_{UAA}$ harbored in pSUA2-TTA, not in pSUA4-TTA. Since Sven 624 grows in a monodispersed manner, quantitative comparison could be done between TAG suppression and TTA reassignment in Sven 624. Sven 624 harboring pIPheRS-Y151TTA and pIPheRS-Y151TAG were cultured in triplicate. Protein was purified and taken

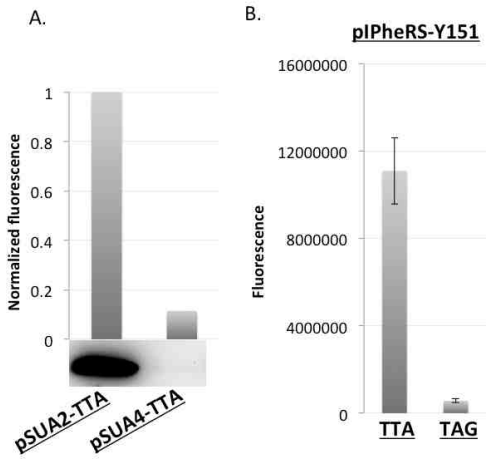


Figure 3-13. A. Fluorescence quantification and anti-His western blotting of Sven 624 harboring pSUA2-TTA and pSUA4-TTA. B. Fluorescence quantification of protein purified from Sven624 harboring pIPheRS-Y151TTA and pIPheRS-Y151TAG in triplicate. The blot was imaged for 204 s.

fluorescence quantification using microplate reader. The fluorescence result (Figure 3-13, B) indicates that protein is about 20 fold higher with TTA reassignment than TAG suppression.

Western Blotting Analysis of Multiple Scattered and Tandem TTA Codon Constructs in Sven 624. Based on the 15 single TTA mutant sfGFP screening results, the fluorescence of each construct was ranked, top 5 sites, top 10 sites, and all of the 15 sites were combined to generate the multiple scattered TTA containing constructs pIPheRS-TTA5s-2, pIPheRS-TTA10s, and pIPheRS-TTA15s, respectively. Western blotting results (Figture 3-14, B, lane 2) show that up to 10 scattered TTA codons in sfGFP was incorporated with pIPhe using the TTA codon reassignment, which is never achieved before using the traditional amber suppression or sense codon reassignment. Tandem UAAs incorporation would be more challenging. To test the TTA codon reassignment

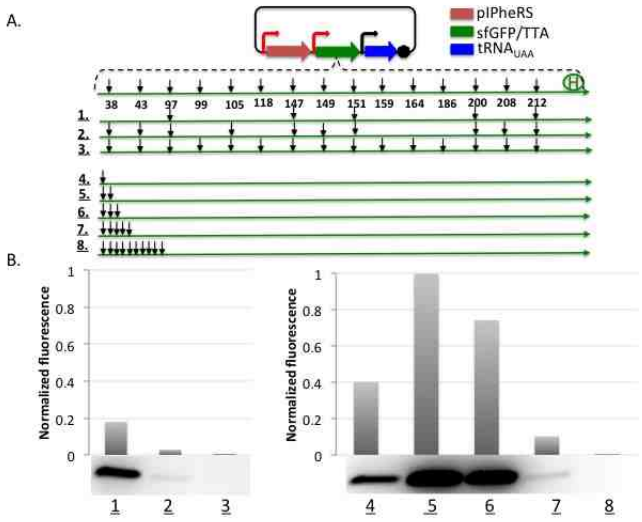


Figure 3-14. A. Illustration of positions of TTA codons in constructs. B. Fluorescence quantification and anti-His western blotting of proteins obtained from Sven 624 harboring pIPheRS-TTA5s-2 (lane 1), pIPheRS-TTA10s (lane 2), pIPheRS-TTA15s (lane 3), pIPheRS-TTA1t (lane 4), pIPheRS-TTA2t (lane 5), pIPheRS-TTA3t (lane 6), pIPheRS-TTA5t (lane 7), and pIPheRS-TTA10t (lane 8). The blot for lane 1-3 was imaged for 360 s and the blot for lane 4-8 was imaged for 44 s.

system, one, two, three, five, and ten consecutive TTA codons were inserted between the start codon and the second codon to generate constructs pIPheRS-TTA1t, pIPheRS-TTA2t, pIPheRS-TTA3t, pIPheRS-TTA5t, and pIPheRS-TTA10t, respectively. Western blotting result (Figure 3-14, B, lane 7) shows that up to five consecutive TTA codons were translated through using the TTA codon reassignment. Surprisingly, single TTA codon incorporation (Figure 3-14, B, lane 4) is less efficient than two and three tandem TTA codons incorporation (Figure 3-14, B, lane 5 and 6) possibly due to the protein stability or protein folding issues.

Western Blotting Analysis of Truncated Proteins Produced from TAG Suppression and TTA Reassignment. One, two, three, five, and ten scattered TTA and TAG codons were installed into sfGFP with N-terminal histidine tag. And wild type sfGFP with N-terminal histidine tag was also constructed. These 11 sfGFP genes were combined with pIPheRS/$tRNA^{Tyr}_{UAA}$ to generate these 11 new constructs. Western blot results (Figure 3-15) indicate that there is completely no truncated protein produced by TTA codon reassignment as we expect. Surprisingly there are only small amount truncated proteins detected from TAG suppression system (Figure 3-15, B, lane 3, 5, and 7) probably because most of the truncated proteins are degraded. This result again confirms 10 pIPhe (Figure 3-15, B, lane 10) are incorporated into sfGFP with TTA codon reassignment, however, TAG suppression system can only suppress one TAG codon (Figure 3-15, B, lane 3).

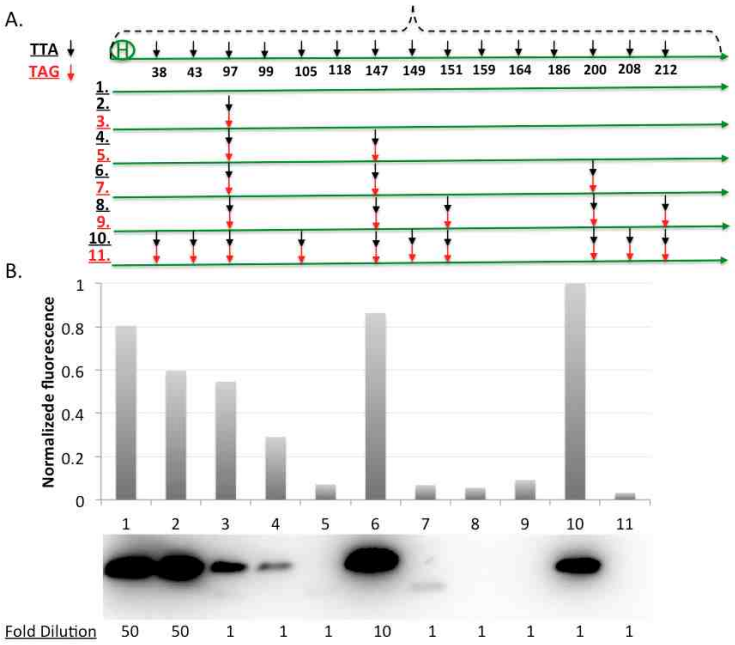


Figure 3-15. A. Illustration of positions of TTA or TAG codons in constructs. B. Fluorescence quantification and anti-His western blotting of proteins obtained from Sven 624 harboring pIPheRS-wtsfGFP (lane 1), pIPheRS-TTAn1s (lane 2), pIPheRS-TAGn1s (lane 3), pIPheRS-TTAn2s (lane 4), pIPheRS-TAGn2s (lane 5), pIPheRS-TTAn3s (lane 6), pIPheRS-TAGn3s (lane 7), pIPheRS-TTAn5s (lane 8), pIPheRS-TAGn5s (lane 9), pIPheRS-TTAn10s (lane 10), and pIPheRS-TAGn10s (lane 11). Samples of lane 1 and lane 3 were 50 diluted and sample of lane 6 was 10 fold diluted prior to fluorescence quantification and ant-His western blotting. The blot was imaged for 104 s.

5. **Conclusion**

A TTA sense codon reassignment system for UAA incorporation was successfully developed in *S. coelicolor* J1681. It is non-toxic and 20 fold more efficient than amber suppression for single UAA incorporation. To increase the UAA containing protein yield and simplify cell culture process, Sven 624, *S. venezuelae ΔbldA* strain, was constructed using *Streptomyces* CRISPR/Cas system. In Sven 624, up to 10 scattered or 5 tandem TTA codons were reassigned to pIPhe and no truncated proteins were produced with TTA codon reassignment system. This system could be useful tool to enhance the UAA containing protein production and to genetically incorporate multiple UAAs into a single protein.

6. References

1. Anderson, J. C., Wu, N., Santoro, S. W., Lakshman, V., King, D. S., & Schultz, P. G. (2004). An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(20), 7566-7571.
2. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M., & Chin, J. W. (2010). Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, *464*(7287), 441-444.
3. Lee, B. S., Shin, S., Jeon, J. Y., Jang, K. S., Lee, B. Y., Choi, S., & Yoo, T. H. (2015). Incorporation of unnatural amino acids in response to the AGG codon. *ACS chemical biology*, *10*(7), 1648-1653.
4. Ho, J. M., Reynolds, N. M., Rivera, K., Connolly, M., Guo, L. T., Ling, J., ... & Söll, D. (2015). Efficient reassignment of a frequent serine codon in wild-type *Escherichia coli*. *ACS synthetic biology*, *5*(2), 163-171.
5. Biddle, W., Schmitt, M. A., & Fisk, J. D. (2015). Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen. *Biochemistry*, *54*(50), 7355-7364.
6. Bröcker, M. J., Ho, J. M., Church, G. M., Söll, D., & O'Donoghue, P. (2014). Recoding the genetic code with selenocysteine. *Angewandte Chemie International Edition*, *53*(1), 319-323.
7. Kwon, I., Kirshenbaum, K., & Tirrell, D. A. (2003). Breaking the degeneracy of the genetic code. *Journal of the American Chemical Society*, *125*(25), 7512.

8. Leskiw, B. K., Mah, R., Lawlor, E. J., & Chater, K. F. (1993). Accumulation of *bldA*-specified tRNA is temporally regulated in *Streptomyces coelicolor* A3 (2). *Journal of bacteriology*, *175*(7), 1995-2005.
9. White, J., & Bibb, M. (1997). *bldA* dependence of undecylprodigiosin production in *Streptomyces coelicolor* A3(2) involves a pathway-specific regulatory cascade. *Journal of bacteriology*, *179*(3), 627-633.
10. Pettersson, B. M., & Kirsebom, L. A. (2011). tRNA accumulation and suppression of the *bldA* phenotype during development in *Streptomyces coelicolor*. *Molecular microbiology*, *79*(6), 1602-1614.
11. Wang, L., Brock, A., Herberich, B., & Schultz, P. G. (2001). Expanding the genetic code of *Escherichia coli*. *Science*, *292*(5516), 498-500.
12. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, *24*(1), 79-88.
13. Dumas, A., Lercher, L., Spicer, C. D., & Davis, B. G. (2015). Designing logical codon reassignment–Expanding the chemistry in biology. *Chemical Science*, *6*(1), 50-69.
14. He, J., Sundararajan, A., Devitt, N. P., Schilkey, F. D., Ramaraj, T., & Melançon, C. E. (2016). Complete genome sequence of *Streptomyces venezuelae* ATCC 15439, producer of the methymycin/pikromycin family of macrolide antibiotics, using PacBio technology. *Genome announcements*, *4*(3), e00337-16.

15. Cobb, R. E., Wang, Y., & Zhao, H. (2014). High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS synthetic biology*, *4*(6), 723-728.
16. Tong, Y., Charusanti, P., Zhang, L., Weber, T., & Lee, S. Y. (2015). CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS synthetic biology*, *4*(9), 1020-1029.
17. Huang, H., Zheng, G., Jiang, W., Hu, H., & Lu, Y. (2015). One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta biochimica et biophysica Sinica*, *47*(4), 231-243.
18. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819-823.
19. He, J., Van Treeck, B., Nguyen, H. B., & Melançon III, C. E. (2015). Development of an unnatural amino acid incorporation system in the actinobacterial natural product producer *Streptomyces venezuelae* ATCC 15439. *ACS synthetic biology*, *5*(2), 125-132.

Chapter 4. Towards applications

1. Introduction

As described in Chapter 1, expansion of the molecular and chemical biology tool set available in *Streptomyces* would greatly facilitate natural product biosynthetic and bioengineering work. Among these useful chemical biology tools, genetically unnatural amino acid (UAA) incorporation systems afford broad applications for manipulation of biosynthetic enzymatic pathways, ribosomally derived peptide drug diversification, antibody conjugate drug discovery, and protein engineering. In this study, amber suppression and TTA sense codon reassignment UAA incorporation systems were successfully developed in model *Streptomyces* hosts, which were described in Chapter 2 and Chapter 3, respectively. These systems would provide the ability to investigate natural products multistep enzymatic pathways and structurally diversifying ribosomally synthesized and post-translationally modified peptides (RiPPs) in *Streptomyces*.

Natural products produced in *Streptomyces* are made from multistep enzymatic pathways and the functionality of these multienzymes often relies on specific and correct protein-protein interactions between domains or other functionally related proteins. pAzPhe and pBpa have been successfully used to map the protein-protein interaction in *E. coli* and mammalian cell to facilitate understanding of their enzymatic mechanisms.[1,2] In this study, the UAA system was applied to map the type II polyketide enzymatic protein-protein interactions. The enzymes for making actinorhodin, the most well studied type II polyketide compound and well characterized pathway, were the most suitable candidate to conduct this study because co-crystalized of acthinorhodin ketosynthase α/β (KSα/β) proteins for making actinorhodin was available[3] and the interaction between

KSα/β and immediate tailoring enzymes responsible for synthesis of the actinorhodin core structure, which could be implied from literatures have not yet been confirmed.

To start this application project, demonstration of KSα/β interaction by incorporation of UAA into KSα/β would be first step, which could confirm the ability of this method for mapping protein-protein interactions in *Streptomyces*. AztufA-KSαβ was then constructed, in which the sfGFP-Y151TTA was replaced by the wild type KSα/β driven by the promoter SF14. The C-terminal histidine tag was attached to the KSβ protein. All of the genetic elements used for KSαβ were the same as used for eGFP and sfGFP, which have been successfully expressed in *S. venezuelae*.

Az-act1 and Az-act2 were designed to multiple pAzPhe incorporation into the interface between KSα/β proteins according to their crystal structure information using the TTA sense codon reassignment. In addition, KSα/β genes were driven by its native promoter, which is regulated by ActII-ORF4 gene.[4] The ActII-ORF4 gene is a regulator protein and bearing a TTA codon at position 5, which makes it *bldA* dependent. To eliminate its *bldA* dependence, the TTA codon was replaced by the most frequently used leucine codon CTC in *Streptomyces*.[5] Az-act1 and Az-act2 were designed to work in J1681 strain. The *bldA* independent ActII-ORF4 regulator protein was expressed and employed to turn on the actinorhodin gene cluster in the genome. In this way, after evidences of KSα/β interaction are obtained, other protein complexes could be pursued by incorporating more pAzPhe into the proposed interaction region between KSα/β and other functionally related enzymes. That is also the advantage of the more efficient TTA sense codon reassignment.

In addition, structurally diversified RiPPs with unnatural amino acid methodology

will afford potentially new or improved bioactivities and facilitate study of mechanism of action/target identification studies for RiPPs.[6,7] Thiopeptides, a class of RiPPs, are highly modified sulfur-rich heterocyclic peptide with more than 100 entities.[8] Thiostrepton, a typical thiopeptide, was isolated from *Streptomyces azureus* ATCC 14921 and *Streptomyces larurentii* ATCC 31255, and is an antibacterial agent against Gram-positive pathogens.[9] There are three residues (Ala2[10], Ala4[11], and Thr7[12]) in the 17 residue core peptide structure are not subjected to post-translational modification during the thiostrepton biosynthesis and these three residues were substituted with other amino acids by Dr. Kelly's group at Georgia Tech.

To further expand the residue substitution with UAAs, a tyrosyl amber suppression UAA system was designed to genetically install pIPhe into the non-modified positions in thiostrepton. To achieve this goal, pSUA1-pIPheRS-Y151TAG was protoplast transferred into a wild-type thiostrepton producer *S. laurentii* ATCC 31255 and it was used to test its functionality in this strain.

According reported work, in frame deletion of *tsrA* gene from genome and reconstitution of the whole *tsr* gene cluster in a replicable plasmid would not work for thiostrepton biosynthesis gene clusters due to potential polar effects from *tsrA* in frame deletion.[13] However, substitution of GGA codon (Gly-30 in the leader peptide) by TGA stop codon in the genome was shown to work with the *tsrA* reconstitution in a replicable plasmid.[14] To follow this idea, pCRISP-tsrA and pOJ-tsrA were designed to deliver this point mutation into the *S. laurentii* genome. In the pCRISP-tsrA construct, a spacer sequence was designed to bind to the leader peptide of *tsrA* gene to induce the double strand DNA broken. After that, a homology arm harbored in pCRISP-tsrA would fix the

genome to install two point mutations for TGA stop codon incorporation and eliminating the corresponding PAM sequence.[15] After loss of the PAM sequence, the genome would be compatible with the CRISP/Cas system. pOJ-tsrA is a traditional way to make point mutation in *Streptomyces* genome by two times of sequential single crossover between plasmid and genome.

Besides that, a thiopeptides sensor development was also described in this chapter. It was designed based on the interaction between TipAL protein and thiopeptides.[16] TipAL is a *Streptomyces* transcriptional activator protein. After covalently binding to thiopeptides on its C-terminal ligand-binding domain, protein conformation changes and then binds to ptipA promoter on its N-terminal DNA binding domain to activate the DNA transcription.[17,18] TipAL and TipAS are both in-frame translated from the same gene (*tipA* gene) and TipAS is translated from the internal start codon of *tipA* and more abundant in the cell than TipAL protein. However, only TipAL protein has the capability of binding to thiopeptides and activating ptipA promoter.[17,18]

First, wild type sfGFP gene was cloned into pXY200[19] in place of the *redD* gene. The wild type TipAL and mTipAL (loss of the internal start codon for TipAS protein translation) were then cloned into that construct to have the pTS and pMTS, respectively. pMTS will produce TipAL protein, which is capable of binding thiopeptides and then activate the transcription of sfGFP. To further optimize this construct, the N-terminal histidine tag on the backbone was removed to have pMTS2, which is capable of sensing thiopeptides and capturing thiopeptides out by purifying mTipAL protein. The C-terminal histidine tag was removed from pMTS2 to have pMTS3, in which the sfGFP was then replaced by actinorhodin KSαβ to have pMTS3- KSαβ construct. In this way, KSαβ was

driven by the inducible promoter ptipA.

2. Experimental procedures

General. Most materials used for work described in this chapter have already been mentioned in the Experimental Procedures section of Chapter 2. Thiostrepton was purchased from Calbiochem.

Bacterial Strain. *E. coli* DH5α and *S. venezuelae* ATCC 15439 described in this Chapter have already been mentioned in the Experimental Procedures section of Chapter 2. *S. laurentii* ATCC 31255 was purchased from Leihniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures in Germany. *E. coli* ETZ12567 was used to propagate unmethylated DNA for conjugal transformation into *S. laurentii* ATCC 31255.

Bacterial culture. The bacterial culture used in this chapter was described in the Experimental Procedures section of Chapter 2. *S. laurentii* is similar as *S. venezuelae* in growth speed and dispersed manner. Therefore, same culture conditions were applied to *S. laurentii*.

Preparation of E. coli Competent Cells. The procedure used to prepare *E. coli* competent cells was described in the Experimental Procedure section of Chapter 2.

General PCR Conditions. The general PCR conditions used in this chapter was described in the Experimental Procedures section of Chapter 2

Small Scale Protein Expression. Small scale protein expression in *S. laurentii* was the same as in *S. venezuelae* and described in the Experimental Procedures section of Chapter 2.

Protein Fluorescence Quantification. Same condition was used in this

experiment as described in the Experimental Procedures section of Chapter 2.

Western Blot Analysis. These experiments were performed as described in the Experimental Procedures section of Chapter 2.

Conjugal Transfer of Vectors into S. laurentii. Same protocol used for *S. venezuelae* was applied to conjugal transfer of vectors into *S. laurentii* as described in the Experimental Procedures section of Chapter 2

Preparation of S. venezuelae Protoplasts. Same protocol was used for preparation of *S. venezuelae* protoplasts as described in the Experimental Procedures section of Chapter 2.

Protoplast transfer of Vectors into S. venezuelae. Same protocol was applied to protoplast transfer of vectors into *S. venezuelae* as described in the Experimental Procedures section of Chapter 2.

3. **Vector Design and Construction**

Construction of AztufA-KSαβ. For construction of AztufA-KSαβ, six point mutations were installed into KSαβ gene to avoid the occurrence of BglII restriction enzyme and four fragments were amplified using primers EGFP-pSUA1-X-up and SF14-adapter-dn, SF14-actI-ORF1-up and actI-G-A-dn, actI-G-A-up and actI-TC-AG-dn, actI-TC-AG-up and His-actI-ORF2-dn, respectively. These four fragments were spliced by overlap extension PCR using primer EGFP-pSUA1-X-up and His-BglII-dn to generate the KSαβ fragment with SF14 promoter, synthetic RBS and C-terminal histidine tag. The resulting fragment was digested with XbaI and BglII restriction enzymes and ligated into AztufA-sfGFP digested with the same restriction enzymes to make AztufA-KSαβ. This construct was verfied by restriction mapping and sequencing of the inserts. Primer

information is given in the table below. The priming region of each primer is underlined and the restriction enzyme site is shown in cyan. Map of the vector is also given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
EGFP-pSUA1-X-up	(Used with SF14-adapter-dn)	73	none
SF14-actI-ORF1-up	ATCAATAGTCGATTAGAGGAGAAATTAGAATTGAAGCGC AGAGTCGTCATCACGGGC	827	genomic DNA of *S. coelicolor*
actI-G-A-dn	TACCCCGAGATTTCGGCGTGGATGCGGGCT		
actI-G-A-up	CGCCGAAATCTCGGGGTACGCGACGCGC	935	genomic DNA of *S. coelicolor*
actI-TC-AG-dn	TGCCGGATGCTGATCTGGCCGGTGTTCACCGC		
actI-TC-AG-up	CCAGATCAGCATCCGGCACGGTATGCGCGGT	820	genomic DNA of *S. coelicolor*
His-actI-ORF2-dn	GTGGTGGTGATGGTGGTGCGAGCCCGGGGTCGGTGCGAA ACGGC		
EGFP-pSUA1-X-up	Used with His-BglII-dn	2605	PCR pdt.

Table 4-1. Primer information for constructing AztufA-KSαβ

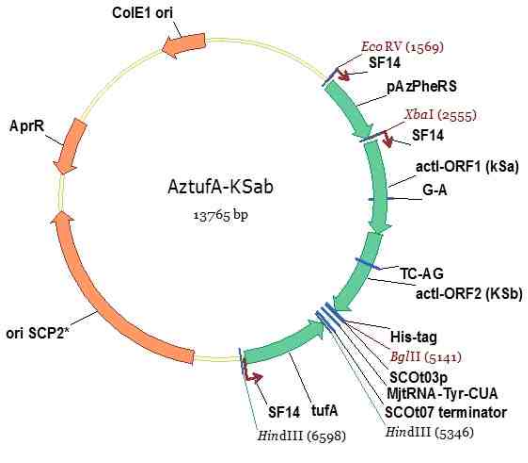


Figure 4-1. Map of AztufA-KSαβ

Construction of Az-act1 and Az-act2. Az-act1 and Az-act2 are derivate of pAzPheRS-Y151TTA, in which the sfGFP-151TTA was removed and KSαβ driven by its mutant native promoter was inserted into downstream of the SCOt07 terminator. First, pAzPheRS-Y151TTA was digested with XbaI and BglII and the resulting backbone was klenow treated, and self ligated to make the circular intermediate plasmid pAzPheRS-nsfGFP. The TTA codon in ActII-ORF4 was replaced by the most frequently used

leucine codon CTC. Two fragments were amplified using primer pairs EcoRI-ORF4-up and ORF4-5CTC-dn, ORF4-5CTC-up and BsrGI-ORF4-dn, respectively and were then assembled by overlap extension PCR using primers EcoRI-ORF4-up and BsrGI-ORF4-dn to generate the final mutant ActII-ORF4 fragment with EcoRI at the 5' –end and BsrGI at the 3' –end. The C-terminal histidine tag was installed at the 3' –end of KSβ. Two fragments were amplified using primer pairs KSb-SfiI-up and His-actI-ORF2-dn, His-act-ACP-up and ACP-HindIII-dn, respectively and were then spliced by overlap extension PCR using primers KSb-SfiI-up and ACP-HindIII-dn to install the C-terminal histidine tag with SfiI at the 5' –end and HindIII at the 3' –end. These two fragments were digested with appropriate restriction enzymes to generate the sticky ends. The third fragment was directly digested from pXZ6 construct (made by Xuechen Zhu of this lab, unpublished work) with EcoRI and SfiI restriction enzymes. These three digested DNA products were ligated into pAzPheRS-nsfGFP digested with HindIIi and BsrGI restriction enzymes to generate the Az-act1.

Az-act2 is a derivate of Az-act1, in which the KSβ was replaced by a mutant KSβ containing three TTA codons. Four fragments were amplified using primer pairs KSb-SfiI-up and Y96TTA-dn, Y96TTA-up and Y136TTA-dn, Y136TTA-up and F261TTA-dn, F261TTA-up and ACP-HindIII-dn, KSb-SfiI-up and ACP-HindIII-dn, respectively. These four fragments were spliced by overlap extension PCR using primers KSb-SfiI-up and ACP-HindIII-dn to make the mutant KSβ, which was digested with SfiI and HindIII restriction enzymes. The digested product was ligated into Az-act1 digested with the same restriction enzymes to generate Az-act2. Theses two constructs were verified by restriction mapping and sequencing of the insert. Primer information is given in the table

below. The priming region of each primer is underlined and mutated codon is shown in blue. Map of representative vector is given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
EcoRI-ORF4-up	GCTTATGCATGTTTAAACTCTAGAGAATTCGATGCGTTC GTCCGGTGGTCGTCGTC	849	pXZ6
ORF4-5CTC-dn	ATGGACACGTCCCAAGAGGTTGAATCTCATCTGCGCCC		
ORF4-5CTC-up	GGGCGCAGATGAGATTCAACCTCTTGGGACGTGTCCAT	790	pXZ6
BsrGI-ORF4-dn	ATATAATGTACAAACTACACGAGCACCTTCTCACCGTTG AGAATTTCCATGTGC		
EcoRI-ORF4-up	Used with BsrGI-ORF4-dn	1601	PCR pdt.
KSb-SfiI-up	TACTGGTCGGCGGTGCTGGACGGC (Used with His-actI-ORF2-dn)	1185	pXZ6
His-act-ACP-up	GGCTCGCACCACCATCACCACCACTAACGACCCCGAGAA CTGGAGAAAGCC	341	pXZ6
ACP-HindIII-dn	CCCAAGCTTATGCATGTTTAAACTCTAGATCATGCCGCC TCGGCCAGTGC		
KSb-SfiI-up	Used with ACP-HindIII-dn	1502	PCR pdt.
Y96TTA-dn	GGTCACCACGCCCATGTCTAAGTCGGTCAGCGA (Used with KSb-SfiI-up)	246	pXZ6
Y96TTA-up	TCGCTGACCGACTTAGACATGGGCGTGGTGACC	150	pXZ6
Y136TTA-dn	GGTGTTCACCGCTAAGAACCAGGCGAAGGACTCGTA		
Y136TTA-up	TACGAGTCCTTCGCCTGGTTCTTAGCGGTGAACACC	408	pXZ6
F261TTA-dn	GGGGGCGGGGTCTAAGGTGGAGGC		
F261TTA-up	GCCTCCACCTTAGACCCCGCCCCC (Used with ACP-HindIII-dn)	791	pXZ6
KSb-SfiI-up	Used with ACP-HindIII-dn	1502	PCR pdt.

Table 4-2. Primer information for constructing Az-act1 and Az-act2

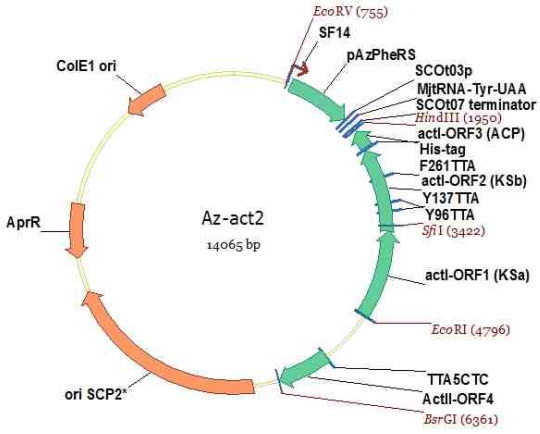


Figure 4-2. Map of Az-act2

Construction of pTS, pMTS, pMTS2, pMTS3 and pMTS3-KSαβ. All of these constructs were made from pXY200[19]. First, an sfGFP fragment was amplified using

primers NdeI-sfGFP-up and BamHI-sfGFP-dn to install NdeI site at the 5' –end and BamHI site at the 3' –end; and it was digested with NdeI and BamHi restriction enzymes and then ligated into pXY200 digested with the same enzymes to generate the intermediate construct pXY-sfGFP. Next, the aph terminator[20] was installed into downstream of sfGFP by BamHI ans EcoRI sites. The aph terminator was made directly by primers. aph-Xbal-up and aph-Xbal-dn were annealed together and the resulting product was used as template to generate the final fragment using primers BamHI-aph-up and EcoRI-aph-dn. The resulting product was digested with BamHI and EcoRI restriction enzymes and the digested product was ligated into pXY-sfGFP digested with the same enzymes to make the second intermediate construct pXY-aph. The wild type TipAL gene with SF14 promoter, synthetic RBS and C-terminal histidine tag installation was amplified from genomic DNA of *S. coelicolor* using primers SF14-TipAL-up and His-TipAL-dn. The PCR product was used as template for a second PCR reaction using primers EcoRI-SF14-up and EcoRI-His-dn to construct the final fragment, which was digested with EcoRI restriction enzyme. The digested product was ligated into pXY-aph digested with the same enzyme to make the final construct. Due to single restriction enzyme digestion and ligation, two orientations of TipAL existed among the plasmids. The plasmid harboring the TipAL gene, which has the same orientation as sfGFP gene was named pTS and used in this study. In the pMTS, the TTA codon in the wild type TipAL was replaced by the most frequently used leucine codon CTC in *Streptomyces*. To construct the mutant TipAL gene, two fragments were amplified with the mutation at their junction using primer pairs EcoRI-SF14-up and TipAL-M110L-dn, TipAL-M110L-up and EcoRI-His-dn, respectively. These two fragments were joined by overlap

extension PCR using primers EcoRI-SF14-up and EcoRI-His-dn to generate the final fragment, which was digested with EcoRI restriction enzyme. The digested product was ligated into pTS digested with the same enzyme to make the constructs. These constructs were screened by sequencing to obtain the pMTS, in which TipAL has the same orientation as in pTS.

To further construct pMTS2, pMTS3, and pMTS3-KSαβ, N-terminal histidine tag derived from backbone of pXY200 and C-terminal histidine tag installed by cloning were sequentially removed to generate pMTS2 and pMTS3. From pMTS3, the sfGFP gene was replaced by KSαβ to make the pMTS3-KSαβ. Due to lack of the full sequence information of pXY200, one primer (aprR-internal-up) was design to anneal to the internal of apramycin resistance gene and its corresponding downstream primer (NdeI-ptipA-dn) was design to remove the N-terminal histidine tag from the backbone of pXY200. Therefore, this fragment was amplified from pXY200 using primers mentioned above and digested with HindIII and NdeI restriction enzyme. pMTS was digested with HindIII and NdeI and the largest DNA fragment and apramycin resistance DNA fragment were recovered from the gel. These two gel purified fragments and the PCR product digested with HindIII and NdeI were mixed and ligated together to generate the pMTS2. The mutant TipAL with C-terminal histidine tag removal was amplified using primers EcoRI-SF14-up and EcoRI-TipA-dn, which was digested with EcoRI restriction enzyme. The digested product was ligated into pMTS2 digested with the same enzyme to generate the pMTS3, which was screened by sequencing to confirm its mTipAL orientation. To construct pMTS3-KSαβ, two fragments were amplified using primer pairs NdeI-KSab-up and KSb-A798T-dn, KSb-A798T-up and BamHI-His-dn, respectively. These two

fragments were spliced by overlap extension PCR using primers NdeI-KSab-up and BamHI-His-dn to construct the final product, which was digested with NdeI and BamHI restriction enzymes. The digested product was ligated into pMTS3 digested with the same restriction enzymes to generate pMTS3-KSαβ.

All of theses constructs were verified by restriction mapping and sequencing of the insert. Primer information is given in the table below. The priming region of each primer is underlined, the restriction enzyme site is shown in cyan, and mutated codon is shown in blue. Maps of fragments are given below.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
NdeI-sfGFP-up	GGAATTGCATATGAGCAAGGGCGAGGAGCTGTT	736	pSUA5-sfGFP
BamHI-sfGFP-dn	CGCGGATCCTCACTTGTAGAGCTCGTCCATCCCGTG		
aph-XbaI-up	CTGCTCTAGAGCCGCCCCGCAGGGCGCTCCGCAGGCCGC TTCCGGACCACTCCG	90	none
aph-XbaI-dn	GTGCTCTAGAGGTACCTCCGACCGCACGGCCGCTTCCGG AGTGGTCCGGAA		
BamHI-aph-up	CGGGATCCGCCGCCCCGCAGGGC	95	PCR pdt.
EcoRI-aph-dn	CCGGAATTCGGCGGCCGGGTACCTCCGACCGCACGGCC		
EGFP-pSUA1-X-up	Used with SF14-adapter-dn	73	none
SF14-TipAL-up	CAATAGTCGATTAGAGGAGAAATTAGAAGTGAGCTACTC CGTGGGACAGGTGGCCGG	808	genomic DNA of *S. coelicolor*
His-TipAL-dn	GTGGTGATGGTGGTGCGAGCCGGGGGTGTGCCGGACGGC GTTGGC		
EcoRI-SF14-up	CCGGAATTCGGTTGACCTTGATGAGGCG	865	PCR pdt.
EcoRI-His-dn	CCGGAATTCTCAGTGGTGGTGATGGTGGTGCG		
TipAL-M110L-dn	GGGTGAGGTTGATTCCGAGGCTGCGTGC (Used with EcoRI-SF14-up)	416	pTS
TipAL-M110L-up	GCACGCAGCCTCGGAATCAACCTCACCC (Used with EcoRI-His-dn)	477	pTS
EcoRI-SF14-up	Used with EcoRI-His-dn	865	PCR pdt.
aprR-internal-up	GAACTCGATGGGCAGGTACTTCTCCTCGGC	unknown	pXY200
NdeI-ptipA-dn	GGAATTGCATATGTGTCCGCTCCCTTCTCTGACGCCG		
EcoRI-TipA-dn	CCGGAATTCTCAGGGGGTGTGCCGGACG (Used with EcoRI-SF14-up)	841	pMTS
NdeI-KSab-up	GGAATTGCATATGAAGCGCAGAGTCGTCATCACGGGCGT CGGCG	2086	AztufA-KSαβ
KSb-A798T-dn	GTCCGGAACCGGGGGCGGGGTCGAA		
KSb-A798T-up	TTCGACCCCGCCCCCGGTTCCGGAC	477	AztufA-KSαβ
BamHI-His-dn	CGCGGATCCTCAGTGGTGGTGATGGTGGTGCGAGCC		
NdeI-KSab-up	Used with BamHI-His-dn	2538	PCR pdt.

Table 4-3. Primer information for constructing pTS, pMTS, pMTS2, pMTS3, and pMTS-KS

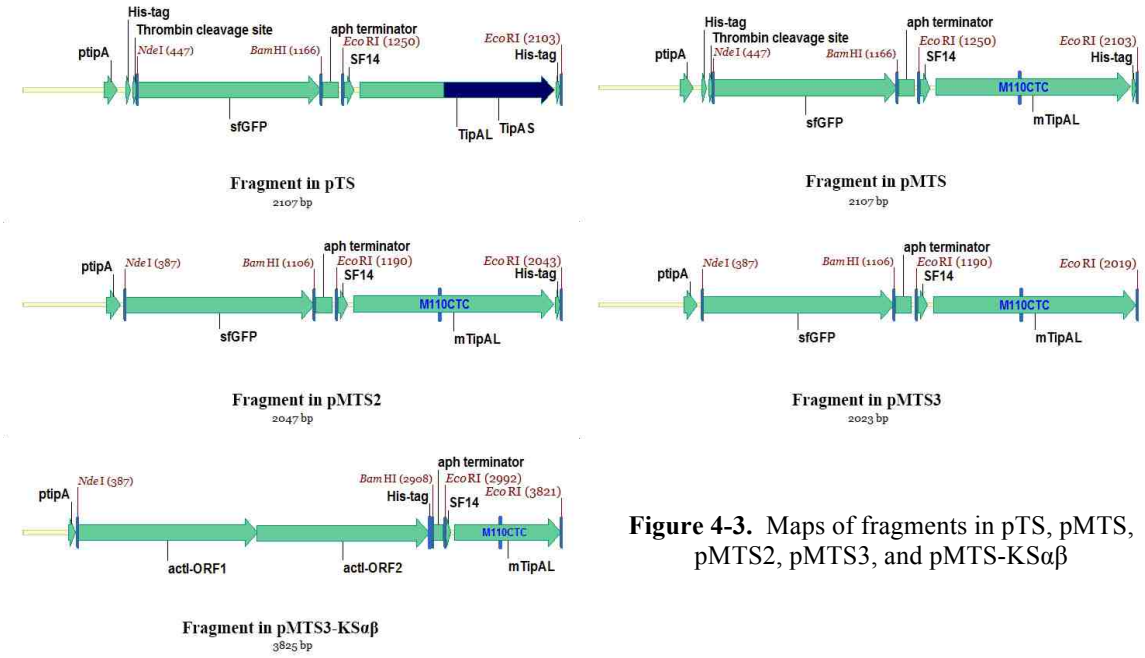


Figure 4-3. Maps of fragments in pTS, pMTS, pMTS2, pMTS3, and pMTS-KSαβ

Construction of pOJ-tsrA and pCRISP-tsrA. pOJ-tsrA was derivate of a suicide vector pOJ260[21]. To construct it, two fragments were amplified from int-3A10[13] with the TGA codon at the their junction using primer pairs tsr-HindIII-up and tsrA-30TGA-dn, tsrA-30TGA-up and tsr-XbaI-dn, respectively. These two fragments were joined together by overlap extension PCR using primers tsr-HindIII-up and tsr-XbaI-dn, which was digested with HindIII and XbaI restriction enzymes. The digested product was ligated into pOJ260 digested with the same restriction enzymes to generate pOJ-tsrA.

pCRISP-tsrA was also used to incorporate TGA stop codon into *tsrA* gene in the genome of *S. laurentii* ATCC 31255. To construct it, the spacer sequences and homology arm were sequentially inserted into pCRISPomyces-2[22]. Based on the genome sequence of wild type *S. laurentii ATCC 31255*, eighteen potential spacer sequences were screened by analyzing CRIPSR/Cas off-target effect against its genome and this sequence (5′-TGTCGAGGGACTCACGGGTC-3′) has the least off-target, which was chosen to

construct pCRISP-bldA. Two complementary primers spacer4-up and spacer4-dn with specific sticky ends on both side were annealed together as described in construction of pCRISP-bldA in Chapter 3. The annealed primer solution was ligated into pCRISPomyces-2 digested with BbsI restriction enzyme to generate the intermediate construct. Next, two fragments were amplified from the genomic DNA of wild type *S. laurentii ATCC 31255* using primer pairs tsr-tp-up and tsrA-tp-dn, tsrA-tp-up and tsr-tp-dn, respectively. These two fragments were spiced by overlap extension PCR using primers tsr-tp-up and tsr-tp-dn to generate the homology arm with XbaI site on both ends. This PCR product was digested with XbaI restriction enzyme , and then ligated into the intermediate construction digested with the same restriction enzyme to generate the final constrution pCRISP-tsrA. Either orientation of the homology arm were acceptable.

These two constructs was verfied by restriction mapping and sequencing of the inserts. Primer information is given in the table below. The priming region of each primer is underlined, mutated codon is shown in blue, and the restriction enzyme site is shown in cyan.

Primer Name	Sequence (5' – 3')	amplicon size (bp)	template
tsr-HindIII-up	CCCAAGCTTCGCATCTGCCGCAGCCAGTCCCGCATCCAAG	1740	int-3A10
tsrA-30TGA-dn	ACGTCCAGACCCGTGAGTCACTCGACACCGATCTCCAGGGCG		
tsrA-30TGA-up	CGCCCTGGAGATCGGTGTCGAGTGACTCACGGGTCTGGACGT	1769	int-3A10
tsr-XbaI-dn	AGGCTCTAGATACGCGAGGTTCAGGACGTCGAGGTCGGCGG		
tsr-HindIII-up	Used with tsr-XbaI-dn	3467	PCR pdt.
tsr-tp-up	GCTCTAGACTCGATCAAACTCGACATGGGCTCTTCTCGCA	1128	pOJ-tsrA
tsrA-tp-dn	AGGGTGTCGACGTCAAGACCCGTGAGTCACTC		
tsrA-tp-up	GAGTGACTCACGGGTCTTGACGTCGACACCCT	1113	pOJ-tsrA
tsr-tp-dn	GCTCTAGAACAGCTCGGCGAAGGTCGCGGTGAG		
tsr-tp-up	Used with tsr-tp-dn	2209	PCR pdt.
spacer4-up	ACGCTGTCGAGGGACTCACGGGTC	28	none
spacer4-dn	AAACGACCCGTGAGTCCCTCGACA		

Table 4-4. Primer information for constructing pOJ-tsrA and pCRISP-tsrA

4. **Results and Discussion**

Heterologous Expression of Actinorhodin KSαβ in S. venezuelae and J1681. After successfully development of this UAA system in *S. venezuelae*, mapping type II polyketides multienzyme protein-protein interactions would be the next step for applying this new tool to natural product biosynthetic pathways research in *Streptomyces*. As described in Chapter 2, pAzPhe incorporation rate was increased from 33% to 50% after heterologous expression of the *E. coli* tufA gene in *S. venezuelae*. Therefore, pSUA1-pAzPhe-tufA was chosen to harbor the KSα/β genes in replace of sfGFP gene. In this way, KSα/β has the same genetic elements, including promoter, RBS, 5' UTR, and C-terminal histidine tag, as for sfGFP to ensure its successful heterologous expression. AztufA-KSαβ was protoplast transferred into *S. venezuelae* and the mutant was cultured. Proteins were purified by Ni-NTA affinity and analyzed by anti-his Western blotting. Unfortunately, no expected band was detected. Several reasons may lead to this negative result, including promoter is insufficient to drive KSα/β expression and there are some possible unknown regulation mechanisms inside of KSα/β genes.

This application project was on hold for a while until the actinorhodin core structure was made from a plasmid, in which the enzymes for making actinorhodin core structure were driven by its native promoter. The plasmid, pXZ6, was designed and constructed by Xuechen Zhu of this lab. Meanwhile, the TTA sense codon reassignment UAA incorporation system was proved to work with higher efficiency to incorporate multiple UAAs in a single protein in *Streptomyces*. The ActII-ORF4, promoter, KSα/β, and ACP cassette was cloned from pXZ6. The TTA codon in the regulator protein, ActII-ORF4, was replaced by the most frequently used leucine codon CTC and histidine tag

was installed at the C-terminus of KSβ. Through analysis of actinorhodin KSα/β co-crystallized structure, three interface residues (Y96, Y136, and F261) were chosen to incorporate pAzPhe for crosslinking KSα/β proteins because in these three sites, there is enough space for holding the pAzPhe and the distance is close enough for the azido group to react with nearby residues after activation by UV lights. These two KSα/β expression cassettes were designed to work with the pAzPheRS/MjtRNA$^{Tyr}_{UAA}$ pair in Az-act1 and Az-act2. These two constructs were protoplast transferred into *J1681* and the mutants were culture. Proteins were purified by Ni-NTA affinity chromatography and analyzed by anti-his western blotting. Unfortunately, there was still no expected band detected. pXZ6 construct was used in *S. coelicolor* M1152, in which the whole actinorhodin gene cluster was deleted. However, the whole actinorhodin gene cluster is still in J1681 strain, which means there are more than 4 kb identical sequence in its genome and these two constructs. This long identical sequence could lead to DNA recombination between genome and plasmids. That is the possible reason for this negative result.

Besides promoter SF14 and the native promoter, the inducible promoter ptipA was also used to drive the expression of KSα/β. The pMTS3-Ksαβ was constructed as described in the Experimental Procedure section in this chapter. There is a mutant TipAL protein in the construct driven by SF14. After adding thiostrepton to the cell culture, the mTipAL would bind to thiostrepton and the protein-thiostrepton complex would target the ptipA[17,18] and activate the downstream KSα/β transcription. pMTS3-Ksαβ was protoplast transferred into wild type *S. venezuelae* and the mutant was cultured with thiostrepton added 24 hours prior to harvesting cells. Proteins were purified by Ni-NTA

affinity chromatography and were analyzed by anti-his western blotting. No expected band was detected. After that, the pMTS2 mutant was cultured and fluorescence assay was conducted. Background similar fluorescence was detected from the pMTS2 mutant cell culture with thiostrepton added 24 hours prior to harvesting cells. This result suggested that the 5' UTR change during N-terminal histidine tag removal messed up the downstream gene expression.

Function Validation of UAA System in S. laurentii. Another application work interests us long time is to structurally diversify the ribosomally synthesized and post-translationally modified peptides (RiPPs). Thiostrepton, a typical thiopeptides, is our target RiPP to genetically incorporate the UAA. *S. laurentii* ATCC 31255 is one of the thiostrepton producers.[9] Therefore, function validation of the UAA incorporation system in *S. laurentii* would be the first step. pSUA1-pIPheRS-Y151TAG was conjugal transferred into wild type *S. laurentii* and the mutant was cultured in 10 mL of TSB media. Protein was purified using Ni-NTA affinity and analyzed by anti-his Western blotting. Western blotting result (Figure 4-4, lane 1) of sfGFP obtained from this mutant in the presence of 1 mM pIPhe suggests that pIPhe is incorporated into sfGFP. However, the mutant grown in the absence of pIPhe (Figure 4-4, lane 2) indicates that endogenous amino acid(s) are recognized by pIPheRS, which was also seen when this system was first developed in *S. venezuelae*. From the study in Chapter 2, tyrosine and phenylalanine are the endogenous amino acid recognized by pIPheRS possibly because of the higher concentrations of these aromatic amino acids in *Streptomyces* than in *E. coli*, where the variants of MjTyrRS evolved. More proteins from mutant grown in absence of pIPhe than in presence of pIPhe indicate that the concentration of tyrosine and phenylalanine in *S.*

laurentii is higher than both *E. coli and S. venezuelae*. In summary, this preliminary result suggests the tyrosyl amber suppression UAA system works in *S. laurentii*. And further protein mass spectrometry analysis need to be done to confirm the fidelity of pIPheRS in *S. laurentii*.

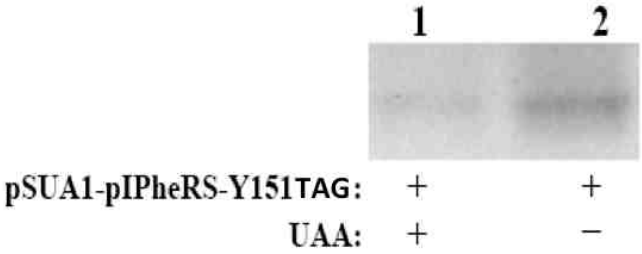


Figure 4-4. Anti-His Western blot of Ni-NTA affinity purified sfGFP proteins obtained from *S. laurentii* harboring pSUA1-pIPheRS-Y151TAG in the presence (lane 1) and the absence (lane 2) of 1 mM pIPhe. The blot was imaged for 503.8 s.

Construction of S. laurentii Mutant. The mature thiostrepton antibiotic is produced and modified from the *tsr* gene cluster containing 21 genes. And previous work showed that heterologous expression of the whole cluster in a fosmid could produce thiostrepton at varying extents, significantly lower than the level of thiostrepton production in *S. laurentii*.[13] To incorporate UAA into thiostrepton, two strategies can be used as described in the introduction section of this chapter. Harboring the amber suppressible *tsr* gene cluster along with pIPheRS/$tRNA^{Tyr}_{CUA}$ pairs in a fosmid and integrating the fosmid into the genome of *S. laurentii* is not conveniently manipulable and has low copy number of pIPheRS/$tRNA^{Tyr}_{CUA}$ pairs, which will affect the UAA incorporation efficiency. Incorporation of stop codon into the leader peptide sequence in *tsrA* gene and reconstitution of amber suppressible *tsrA* gene along with pIPheRS/$tRNA^{Tyr}_{CUA}$ pairs in a replicable multiple copy number plasmid would be better strategy. To pursue this strategy, installing the stop codon in the genome of *S. laurentii*

will be the essential step.

Both pOJ-tsrA and pCRISP-tsrA were constructed and conjugal transferred into wild type *S. laurentii* strain. *S. laurentii* harboring pOJ-tsrA transformant was picked and re-streaked out onto SPA plate supplemented with 50 μg/mL apramycin. The grown up colonies were *S. laurentii*, whose genome was integrated by pOJ-tsrA via single crossover. One fresh single colony was picked to inoculate 10 mL of TSB without antibiotic. The cells were passed through 8 generations with 100-fold dilution for starting each new generation, 24 h per generation, allowing the second single crossover to eliminate the plasmid from its genome. The resulting genome is either wild type sequence or stop codon incorporated sequence in the *tsrA* gene, which have the same probability of occurrence. After 8 generations culture, the cells were serial diluted to an appropriate density and then plated on SPA plate. Fifty-eight colonies from the SPA plate were replicated on both SPA plate containing 50 μg/mL apramycin and SPA plate without antibiotic to screen for apramycin sensitive restoration strains, which fulfilled the second single crossover. More than 50% of the colonies were apramycin sensitive and four colonies were picked and cultured for genomic DNA isolation. The *tsrA* region needs to be amplified and sequenced to test if the mutant strain is made.

CRISPR/Cas system is another route to make the genome point mutation in *Streptomyces*.[15] pCRISP-tsrA was constructed and conjugal transferred into wild type *S. laurentii* strain. The unique tansformant was grown to be visible after 7 days, which is longer than transformations of pOJ-tsrA and pSUA1-pIPheRS-Y151TAG, due to the toxicity of pCRISP-tsrA. Genomic DNA isolation from this unique transformant, locus of *tsrA* gene amplification from its genome and PCR product sequencing work need to be

pursued in the near future.

Construction and Function Validation of Thiopeptides Sensor in S.venezuelae. pTS and pMTS were constructed and protoplast transferred into wild type *S. venezuelae*. The *S. venezuelae* harboring pTS and pMTS were culture separately in 5 mL of TSB liquid media, and 1.5 mL of the resulting cell cultures was used to inoculate two 150 mL of cultures, which were culture at 28 °C, 250 rpm for 48 h. After 48 h pre-culture, the 150 mL cell culture was aliquoted into 10 mL portions into the sterile glass tubes with adding different amount of thiostrepton. The tube cultures were cultured at 28 °C, 250 rpm for 24 h and the resulting cultures were diluted to appropriate cell density and then subjected to fluorescence (excitation = 485 nm, bandwidth = 9 nm; emission = 520 nm, bandwidth = 15 nm) and OD_{600} reading using a microplate reader. The curve of normalized fluorescence over concentration of thiostrepton results (Figure 4-5) indicate that sfGFP production is proportional to the concentration of thiostrepton added in the media. The similarity of pTS and pMTS suggests that sufficient TipAL is produced from pTS during the 48 h pre-culture to react with thiostrepton to activate the transcription of sfGFP protein and higher concentration of TipAL protein in *S. venezuelae* harboring pMTS has no affect on sfGFP transcriptional activation process compared to the mutant containing pTS. The results also show that the fluorescence increases from 5 % to 100% when the concentration of thiostrepton increased from 97.65 nM to1.562 μM, which suggests that they have a good dynamic range to react with thiostrepton. This feature could be employed to monitor thiostrepton production when researchers need to optimize the thiostrepton producer fermentation conditions or to engineer the thiostrepton gene cluster. The fluorescence at the lowest concentration indicates that *S. venezuelae* harboring pTS

and pMTS has good threshold sensitivity as low as about 0.1 μM.

Construction and Function Validation of pMTS2. The ligand binding domain of TipAL protein reacts with thiostrepton or other thiopeptides to form a covalent bond to make this irreversible interaction.[16] In light of that, the mutant TipAL protein was designed not only to detect thiopeptides and activate sfGFP expression as a convenient detectable output, but also to capture the thiopeptides and purify them out during the protein purification. In order to achieve this goal, the N-terminal histidine tag was

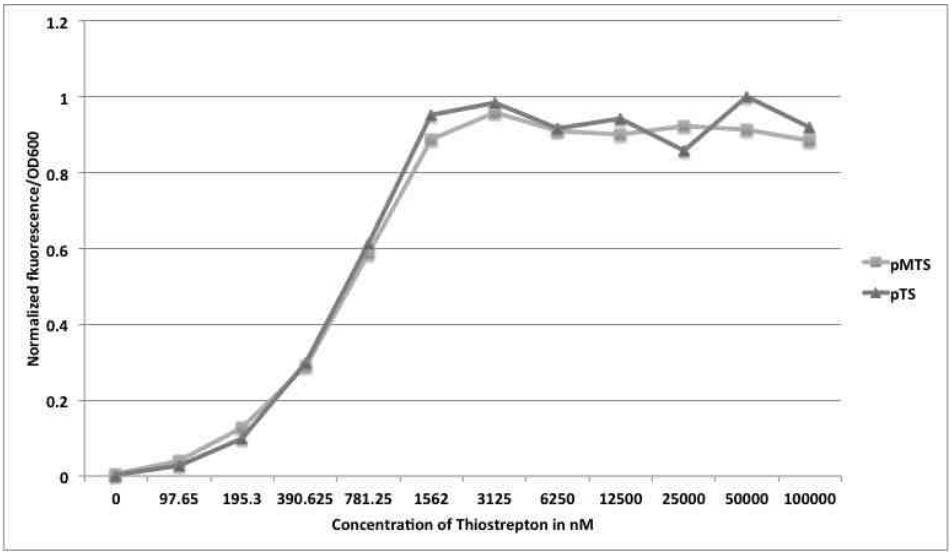


Figure 4-5. Normalized fluorescence/OD_{600} taken from *S. venezuelae* harboring pTS (triangle) and pMTS (square) supplemented with varying concentrations of thiostrepton, grown for 24 h.

removed from pMTS to generate pMTS2, which was introduced into *S. venezuelae* by protoplast transformation. The protein expression was analyzed by anti-his Western blotting. The results (Figure 4-6) indicate that TipAL is the only C-terminally tagged protein expressed from pMTS2 in presence and in absence of thiostrepton, which means that fluorescence signal of sfGFP is proportional to the concentration of thiostrepton in the media without interference of TipAS protein, and the mutant in absence of

thiostrepton (Figure 4-6, lane 2) produced less protein possibly because mutant TipAL and thiostrepton complex could enhance the protein stability, which will benefit mutant TipAL and thiopeptide complexes purification in the following research.

Figure 4-6. Anti-His Western blot of Ni-NTA affinity purified mutant TipAL protein obtained from *S. laurentii* harboring pMTS2 in the presence (lane 1) and the absence (lane 2) of 2 μM thiostrepton. The blot was imaged for 8 s.

5. **Conclusion**

Streptomyces native constitutive and inducible promoters, and synthetic promoter were employed to drive the expression of actinorhodin KSα/β. The negative results confirm again the complexities of natural product biosynthetic gene clusters and conducting synthetic biology research in *Streptomyces*. Encouragement by more knowledge of *Streptomyces* genetics and available genetic and chemical biology tools, new designs and constructs will be made and *Streptomyces* unnatural amino acid methodology will afford a useful tool to research protein-protein interactions of natural product biosynthetic multienzymes in the near future.

For structurally diversification of RiPPs, the preliminary results prove the

functionality of UAA system in *S. laurentii* and *S. laurentii* mutant are under construction using two different strategies, including the powerful genome-editing tool, CRISPR/Cas system. After the mutant strain is made and confirmed, the UAA system will be transferred into the *S. laurentii* mutant strain to deliver novel structural moieties into thiostrepton. The thiopeptides sensor could be used to monitor the production of novel thiostrepton analogues.

6. References

1. Wang, W., Li, T., Felsovalyi, K., Chen, C., Cardozo, T., & Krogsgaard, M. (2014). Quantitative Analysis of T Cell Receptor Complex Interaction Sites Using Genetically Encoded Photo-Cross-Linkers. *ACS chemical biology*, *9*(9), 2165-2172.
2. Grunbeck, A., Huber, T., Sachdev, P., & Sakmar, T. P. (2011). Mapping the ligand-binding site on a G protein-coupled receptor (GPCR) using genetically encoded photocrosslinkers. *Biochemistry*, *50*(17), 3411-3413.
3. Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C., & Stroud, R. M. (2004). An antibiotic factory caught in action. *Nature structural & molecular biology*, *11*(9), 888-893.
4. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A., & Khosla, C. (1993). Engineered biosynthesis of novel polyketides. *SCIENCE-NEW YORK THEN WASHINGTON, 262*, 1546-1546.
5. Fernández-Moreno, M. A., Caballero, J., Hopwood, D. A., & Malpartida, F. (1991). The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. *Cell*, *66*(4), 769-780.

6. Bindman, N. A., Bobeica, S. C., Liu, W. R., & van der Donk, W. A. (2015). Facile removal of leader peptides from lanthipeptides by incorporation of a hydroxy acid. *Journal of the American Chemical Society*, *137*(22), 6975-6978.
7. Piscotta, F. J., Tharp, J. M., Liu, W. R., & Link, A. J. (2015). Expanding the chemical diversity of lasso peptide MccJ25 with genetically encoded noncanonical amino acids. *Chemical Communications*, *51*(2), 409-412.
8. Just-Baringo, X., Albericio, F., & Álvarez, M. (2014). Thiopeptide antibiotics: retrospective and recent advances. *Marine drugs*, *12*(1), 317-351.
9. Kelly, W. L., Pan, L., & Li, C. (2009). Thiostrepton biosynthesis: prototype for a new family of bacteriocins. *Journal of the American Chemical Society*, *131*(12), 4327-4334.
10. Zhang, F., Li, C., & Kelly, W. L. (2015). Thiostrepton variants containing a contracted quinaldic acid macrocycle result from mutagenesis of the second residue. *ACS chemical biology*, *11*(2), 415-424.
11. Zhang, F., & Kelly, W. L. (2015). Saturation mutagenesis of *TsrA* Ala4 unveils a highly mutable residue of thiostrepton A. *ACS chemical biology*, *10*(4), 998-1009.
12. Li, C., Zhang, F., & Kelly, W. L. (2011). Mutagenesis of the thiostrepton precursor peptide at Thr7 impacts both biosynthesis and function. *Chemical Communications*, *48*(4), 558-560.
13. Li, C., Zhang, F., & Kelly, W. L. (2011). Heterologous production of thiostrepton A and biosynthetic engineering of thiostrepton analogs. *Molecular BioSystems*, *7*(1), 82-90.

14. Guo, H., Wang, J., Li, Y., Yu, Y., Zheng, Q., Wu, J., & Liu, W. (2014). Insight into bicyclic thiopeptide biosynthesis benefited from development of a uniform approach for molecular engineering and production improvement. *Chemical Science*, *5*(1), 240-246.

15. Huang, H., Zheng, G., Jiang, W., Hu, H., & Lu, Y. (2015). One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta biochimica et biophysica Sinica*, *47*(4), 231-243.

16. Chiu, M. L., Folcher, M., Griffin, P., Holt, T., Klatt, T., & Thompson, C. J. (1996). Characterization of the covalent binding of thiostrepton to a thiostrepton-induced protein from *Streptomyces lividans*. *Biochemistry*, *35*(7), 2332-2341.

17. Holmes, D. J., Caso, J. L., & Thompson, C. J. (1993). Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*. *The EMBO journal*, *12*(8), 3183.

18. Chiu, M. L., Viollier, P. H., Katoh, T., Ramsden, J. J., & Thompson, C. J. (2001). Ligand-induced changes in the *Streptomyces lividans* TipAL protein imply an alternative mechanism of transcriptional activation for MerR-like proteins. *Biochemistry*, *40*(43), 12950-12958.

19. Cone, M. C., Yin, X., Grochowski, L. L., Parker, M. R., & Zabriskie, T. M. (2003). The blasticidin S biosynthesis gene cluster from *Streptomyces griseochromogenes*: sequence analysis, organization, and initial characterization. *ChemBioChem*, *4*(9), 821-828.

20. Pulido, D., & Jimenez, A. (1987). Optimization of gene expression in *Streptomyces lividans* by a transcription terminator. *Nucleic acids research*, *15*(10), 4227-4240.

21. Bierman, M., Logan, R., O'brien, K., Seno, E. T., Rao, R. N., & Schoner, B. E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces spp*. *Gene*, *116*(1), 43-49.

22. Cobb, R. E., Wang, Y., & Zhao, H. (2014). High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS synthetic biology*, *4*(6), 723-728.