

PNNL-38310

Efficient genetic code expansion tools enable *in vivo* study of lysine acetylation in non-model bacteria

September 2025

1 Elise M. Van Fossen
2 Molly W. Stephenson
3 Andrew Frank
4 Soujanya Akella
5 Rowan Wooldridge
6 Andrew Wilson
7 Youngki You
8 Ernesto Nakayasu
9 Robert Egbert
10 Joshua Elmore



U.S. DEPARTMENT
of **ENERGY**

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062

www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: <http://www.ntis.gov>

Efficient genetic code expansion tools enable in vivo study of lysine acetylation in non-model bacteria

September 2025

1 Elise M. Van Fossen
2 Molly W. Stephenson
3 Andrew Frank
4 Soujanya Akella
5 Rowan Wooldridge
6 Andrew Wilson
7 Youngki You
8 Ernesto Nakayasu
9 Robert Egbert
10 Joshua Elmore

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

Recent proteomic advancements have revealed widespread N ϵ -lysine acetylation in pathways governing pathogenicity, metabolism, and antibiotic resistance in bacteria. The spontaneous, non-specific nature of this modification in prokaryotes obscures its biological role, necessitating prokaryotic specific *in vivo* interrogation systems. Genetic Code Expansion (GCE) offers a powerful method to investigate the roles and regulation dynamics of acetyl-lysine *in vivo* with the precise incorporation of a suite of non-canonical amino acids, including acetyl-lysine analogs. However, its use has been largely restricted to *E. coli* strains due to challenges associated with implementation and optimization of the technology in more diverse bacterial strains. Here, we present a bacterial host-agnostic, readily optimizable GCE platform designed to site-specifically incorporate non-canonical amino acids into target proteins within living bacteria. We further demonstrate the versatility of this technology by showcasing, for the first time, the successful incorporation of acetyl-lysine in a non-*E. coli* bacterium.

Summary

Genetic code expansion (GCE) has been extensively developed in *Escherichia coli*, where it has enabled site-specific incorporation of noncanonical amino acids (ncAAs) to interrogate and engineer protein function. However, the transfer of these systems into non-model bacteria is hindered by genetic tractability, restricting access to GCE in microbial species that otherwise offer advantageous metabolic and physiological traits for biotechnological applications. Serine recombinase-Assisted Genome Engineering (SAGE), is a chromosomal integration platform that permits stable, site-specific incorporation of multiple genes site-specifically into the genomes of diverse bacteria. By porting the orthogonal translation components required for GCE into microbes with SAGE, we can systematically optimize the expression architecture directly in diverse bacterial hosts.

Using this system, we implemented and optimized GCE in *Pseudomonas putida* KT2440 and subsequently ported the platform into multiple additional *Pseudomonas* species and the actinomycete *Rhodococcus jostii*. We demonstrated that orthogonal tRNA copy number is a key determinant of ncAA incorporation efficiency and employed a pooled promoter/RBS library screen to empirically tune aminoacyl-tRNA synthetase expression to host-appropriate levels. Notably, we achieved efficient incorporation of lysine analogs, including N ϵ -acetyl-lysine, where increasing tRNA^{Pyl} copy number enhanced incorporation by 25-fold, enabling production of site-specifically acetylated enolase at conserved lysine residues. Collectively, these results establish a generalizable and extensible framework for stable GCE deployment across phylogenetically diverse bacteria, providing a scalable route to interrogate post-translational modifications, engineer ncAA-dependent biocontainment systems, and expand the chemical and functional repertoire of microbial synthetic biology.

Acknowledgments

This research was supported by the Predictive Phenomics Initiative, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Acronyms and Abbreviations

GCE – genetic code expansion

GFP – green fluorescent protein

trGFP – truncated green fluorescence protein

SAGE – Serine recombinase-Assisted Genome Engineering

ncAA – non-canonical amino acid

tRNA – transfer RNA

pAzF – L-para-azidophenylalanine

sfGFP – super folder green fluorescent protein

aaRS – aminoacyl-tRNA synthetase

AcK – acetyl lysine

Contents

Abstract..... ii

Summary iii

Acknowledgments..... iv

Acronyms and Abbreviations..... v

1.0 Introduction 1

2.0 Results 4

3.0 Discussion 15

4.0 References..... 16

Appendix A – Supplemental Materials..... A.1

1.0 Introduction

Over the past decade, it has become apparent that PTMs play crucial physiological roles in microbial organisms(1, 2). The establishment of a comprehensive biological framework that encompasses the characteristics, regulatory mechanisms, and functions of bacterial PTMs, will unveil new opportunities for bioengineering and the treatment of infectious diseases(2). Yet, PTMs remain one of the largest black boxes in our understanding of bacterial physiology(1-4).

Nε-lysine acetylation is one such modification that has recently emerged as a potentially significant player in bacterial physiology(5-7). Once considered rare in bacteria, mass spectrometry advances have now enabled high-sensitivity analyses of bacterial proteomes, revealing extensive lysine acetylation (up to 40% of all bacterial proteins) in more than 30 different bacterial species(8-11). Proteomic(5, 10, 12-15) and *in vitro* studies(5, 16-21) have connected lysine acetylation to key biological functions, including the regulation of central metabolism(5, 6, 22), transcription(6, 7, 23) and translation(24) as well as modulating pathogeneticity(25) and antibiotic resistance(26) (Fig. 1a). The pervasive presence of acetyl-lysine in these systems indicates that microbial bioengineering efforts and antibacterial designs that are naïve to endogenous acetylation mechanisms may exhibit diminished efficiency and efficacy upon implementation. Further, the systems that govern acetylation themselves are likely to be powerful targets for engineering enhancement and therapeutic intervention.

Our understanding of lysine acetylation in bacterial systems remains limited due to several complicating factors. Firstly, lysine acetylation systems in bacteria differ greatly from eukaryotic systems and even vary widely among bacterial species, making it difficult to create generalized models for PTM mechanisms(2, 7). Secondly, lysine acetylation occurs at very low levels on numerous proteins and often in response to specific environmental conditions that can be challenging to consistently produce in model strains in typical laboratory conditions(6). Additionally, because a significant proportion of lysine acetylation in bacteria occurs non-enzymatically and bacterial acetylases themselves are promiscuous, specialized biochemical tools are required to create acetyl-lysine mimics for analysis(7, 8). In summary, little is known about acetyl-lysine regulatory mechanisms and interaction systems outside of *E. coli*, and what is known in *E. coli* may not translate to other bacteria.

Consequently, identifying the functional roles of lysine acylation and pinpointing the mechanisms that drive it is among the main challenges in bacterial biochemistry today(7). Doing so depends on the establishment of two key capabilities: 1) to model the presence/absence of the PTM on a target protein, and 2) to reliably characterize the interacting entities or phenotypic fluctuations associated with each state(27). Historically, most acetylation studies have used site-directed mutagenesis to substitute the lysine residue of interest with glutamine or arginine. These substitutions mimic acetyl-lysine and non-acetylated lysine, respectively(27).

While these studies have formed the basis of our understanding of lysine acetylation, this method has severe limitations as the substitutions replicate the electrostatic properties of the modifications but not their steric attributes. Consequently,

there are numerous instances where this mutation does not accurately recapitulate the functional effects of acetyl-lysine(28, 29).

Advancements in synthetic and chemical biology have provided new technologies with which to augment lysine interrogation toolkits. Genetic code expansion (GCE) is one such technology, wherein engineered organisms can incorporate non-canonical amino acids (ncAAs), including PTM mimics, site-specifically into proteins, allowing the precise recapitulation of lysine acetylation *in vivo* without relying on enzymatic action or site-directed mutagenesis. GCE technology has also provided ncAAs with unobtrusive crosslinking functionalities that facilitate the enrichment of and identification of PTM-driven protein-protein interactions (Fig. 1a)(27).

As acetyl-lysine occurs in response to specific environmental conditions and is strain-specific, bacterial acetyl-lysine functional investigations should proceed in living bacterial organisms to achieve a more accurate and biologically relevant understanding(7, 23, 30). *In vivo* application of GCE to study bacterial PTMs is nascent but has already been impactful. In *Salmonella typhimurium*, the genetically encoded incorporation of a non-hydrolysable butyryl lysine analogue into HilA (an important transcriptional regulator of *Salmonella* pathogenicity) demonstrated the consequences of specific butyrylation on infectivity(30) and, very recently, *in vivo* GCE was employed in *E. coli* to determine how acetylation modulates transcription factor DNA binding(23).

These two examples highlight the significant potential that *in vivo* genetic code expansion could offer for advancing our understanding of bacterial biology, however the technology still faces a critical limitation; most development and optimization of GCE has been constrained to laboratory strains of *E. coli*. As a result, when working with more diverse bacteria, the efficacy of GCE elements must be determined through trial and error and there are few, if any, universal methods for optimizing GCE in non-*E. coli* strains. This limitation disproportionately impacts systems aimed at incorporating non-canonical amino acids with more subtle modifications, such as acetyl-lysine, as the incorporation efficiency for these smaller modifications tends to be lower compared to ncAAs with bulkier side chains(31). Consequently, optimizing these systems to perform robust science in non-model bacterial hosts remain technically challenging and laborious.

The lack of high-throughput genetic tools for the rapid optimization of expression levels of GCE machinery is a major bottleneck to onboarding complex GCE systems in non-model microbes. Common methods for transferring heterologous DNA into non-model bacteria—replicating plasmids, homologous recombination-based allelic exchange, and transposon mutagenesis—have limitations that restrict their utility. Replicating plasmids are often unstable, impose a fitness cost, and have a limited number of compatible plasmid options(32, 33). While stable integration of heterologous DNA via transposon-based and homologous recombination-based technologies can bypass some of these problems, these methods are not suitable for high-throughput genetic engineering(34). Transposon-based tools have unpredictable integration sites, risking overestimation of GCE efficiency due to the influence of local environment on machinery expression(35). Homologous recombination, being low-efficiency and labor-intensive, is unsuitable for high-throughput assessments of translational machinery expression variants(36).

However, the genetic engineering platform SAGE (Serine recombinase-assisted genome engineering) and similar phase recombinase-based systems offer the potential to overcome these limitations by combining the high-efficiency transformation of replicating plasmids, with the site specificity and stability of homologous recombination. SAGE is a robust and extensible technology that enables site-specific genome integration of multiple DNA constructs, often with efficiency on par with or superior to replicating plasmids(35). The toolkit leverages high-efficiency serine recombinases, each transiently expressed from a non-replicating plasmid, to facilitate efficient, iterative integration of constructs or libraries into diverse bacterial genomes at unique attB sites (Fig. 1c). By utilizing non-replicating plasmids, SAGE is theoretically usable in any bacterial species and has been demonstrated in eight taxonomically distinct bacteria thus far. As such, in this study we apply SAGE to implement GCE in five non-model bacteria to demonstrate a generalizable workflow for onboarding and optimizing genetic code expansion in organisms that are unrelated to *E. coli*. Using this workflow, we dramatically enhanced the efficiency of acetyl-lysine incorporation in non-model bacteria—from nearly negligible levels to a robust system capable of encoding acetyl-lysine at two distinct sites within a biologically relevant protein.

2.0 Results

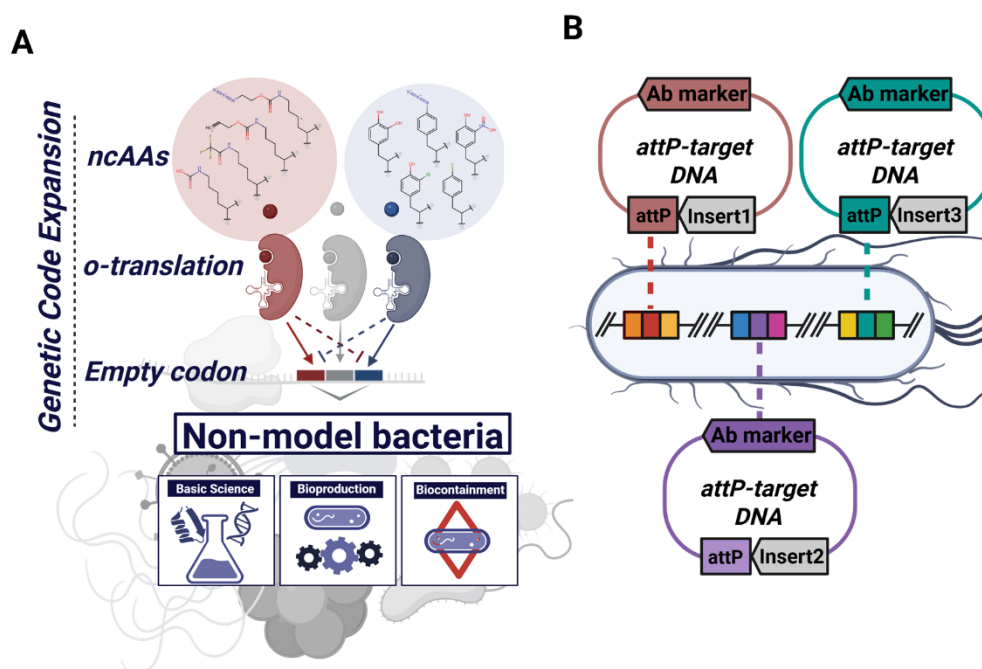


Fig. 1. Overview of systems employed in this work. A Overview of Genetic Code Expansion (GCE). Elements in red and blue represent engineered translational components that act orthogonally to endogenous translation (elements in gray) to produce site-specifically modified protein. **B** Overview of Serine-recombinase Assisted Genome Engineering (SAGE) where serine-recombinases encoded on non-replicating plasmids facilitate the insertion of *attP*-containing target plasmids into cognate *attB*-sequences that have been installed into the host chromosome.

Implementation of plasmid-free genetic code expansion requires the installation of an orthogonal translation system composed of, at minimum, three genetic elements into the chromosome of the host organism. These components include the orthogonal aminoacyl tRNA synthetase (o-aaRS), its cognate orthogonal tRNA (o-tRNA) and a target protein. This target protein contains a 'blank' codon (typically a stop codon) at the desired site of ncAA incorporation (Fig. 1A). The orthogonal elements have been engineered to recognize and "suppress" the stop codon with the ncAA to create full length protein(33). For our initial evaluation of SAGE-GCE, we decided to pursue the incorporation of the ncAA para-azido-L-phenylalanine (pAzF) with the *M. jannaschii* tyrosyl-tRNA synthetase/tRNATyr (MjTyrRS/tRNATyr) system, specifically using the para-cyano-L-phenylalanine aminoacyl-tRNA synthetase (pCNF-RS)(34). This system is a standard for GCE implementation in bacterial systems, as it is highly efficient and polyspecific with the ability to recognize at least 18 ncAAs, including pAzF, pCNF, and other para-substituted phenylalanine analogues. Additionally, pAzF is a valuable ncAA to encode as it contains a chemical handle compatible with crosslinking experiments, applicable for the identification of PTM-dependent protein-protein interactions(35). Genetically incorporating each component into individual attB sites

provides the ability to independently optimize each component. For this, we utilized a variation of SAGE to incorporate multiple distinct ‘target plasmids’ into the chromosome without the need to excise the antibiotic selection marker between each incorporation as is required with the original SAGE implementation (Fig. 1B). For this, each component is linked with one of three antibiotic selection and serine recombinase combinations, which enables incorporation of each element into the chromosome in any order. For SAGE-GCE, we designed target plasmids containing each of the three required components with unique antibiotic resistance markers. These plasmids also included distinct attP attachment sites for integration into three distinct attB sites in the bacterial chromosome by a set of three serine recombinases. Specifically, the MjTyrRS element was designed for insertion at the TG1 attB site, the tRNATyr element at the Bxb1 attB site, and the protein of interest, sfGFP in this case, at the R4 attB site (Fig. 2A). Our initial test platform for the SAGE-based GCE system was the bacterium *Pseudomonas putida* KT2440(36, 37). *P. putida* has become increasingly popular for industrial and environmental applications(38) due to its robust redox metabolism, high tolerance to diverse physicochemical stresses, rapid growth, versatile metabolism, and nonpathogenic nature. Recently, successful replicating plasmid-based GCE was reported in *P. putida* for the first time (39) encouraging our efforts to use this system as a testbed for our SAGE-based GCE system. For this, we used a SAGE-compatible strain of *P. putida* KT2440 (AG5577) for the basis of all further GCE development and application in *P. putida*. This strain contains a collection of 9 distinct heterologous attB sequences that are each recognized by a distinct serine recombinase from the SAGE toolkit(32). Multiple studies focused on implementing GCE in non-*E. coli* strains have reported that it is critical to adjust the regulatory elements (promoters, ribosome binding sites, etc.) for each new organism(16, 17, 40). To maintain GCE components at physiologically relevant expression levels in *P. putida*, we followed a previously reported strategy(39, 40) wherein the most abundant codon in the *P. putida* genome was identified (CUG, encoding leucine (Leu)) along with the associated aaRS and tRNA. Native promoters and terminators flanking tRNA^{Leu} and leucyl-RS (LeuRS) were then used to control transcription of the Mj tRNATyr and MjTyrRS, respectively to create v1.0 (Fig. 2A). Once all SAGE-GCE components were successfully ported into *P. putida* (see Methods), ncAA incorporation efficiency was evaluated using a stop codon readthrough assay with super-folder green fluorescent protein (sfGFP) as a reporter. In this assay, sfGFP contains an amber stop codon at position N150 (sfGFPTAG150). In the absence of ncAA incorporation, expression of this reporter will lead to the synthesis of a truncated, non-fluorescent sfGFP protein. If the ncAA is incorporated into the protein at the amber codon, full length fluorescent sfGFP protein will be synthesized and the cellular fluorescence can be directly correlated with production of ncAA-containing sfGFP protein. Consistent with what was previously reported in *P. putida*, we observed that the pCNF-RS system enabled efficient incorporation of pAzF as indicated by moderate expression of full-length sfGFP only in the presence of the ncAA (Fig. 2B). Further top-down protein mass spectrometry (MS) analysis of the full-length purified protein confirmed the incorporation of pAzF (Fig. S1). Although this initial GCE system exhibited successful incorporation of pAzF, we noted that the efficiency could much improved. As a high concentration of o-tRNA is required for optimal incorporation efficiency in other GCE systems(41), we hypothesized that increasing the copies of o-

tRNA in our system would enhance ncAA suppression. To test this hypothesis, we designed a tRNA cassette to mimic an endogenous *P. putida* tRNA operon, which encodes two identical tRNA^{Leu} in a single pre-tRNA transcript (Fig. 2A, v2.0). This cassette utilizes the endogenous tRNA^{Leu} promoter, the spacer sequence between the tRNAs, and the native downstream terminators, but replaces each tRNA^{Leu} with a MjtRNATyr. We also observed a *P. putida* tRNA operon containing three tRNAs positioned upstream of a gene encoding an elongation factor protein (EF-Tu), an enzyme responsible for facilitating the transfer of aminoacyl-tRNA to the ribosome. As native EF-Tus can exhibit decreased efficiencies when operating in conjunction with ncAAs and orthogonal elements, we drew inspiration from the endogenous *P. putida* operon and added an engineered EF-Tu—designed specifically for the incorporation of pAzF(42)—downstream of the dual MjtRNATyr operon described above, creating v2.5. Lastly, we streamlined v1.0 by combining the sfGFPTAG150 reporter and MjTyrRS on a single construct resulting in a dual plasmid system (Fig. 2A). We evaluated the incorporation efficiencies of the two new GCE systems (v2.0 and v2.5) and found that by adding a second copy of MjtRNATyr (v2.0), we achieved a five-fold improvement compared to the original single tRNA system. The addition of the engineered EF-Tu in v2.5 led to a modest further 1.5-fold improvement (Fig. 2B). We further tested the efficiency of v2.0 by designing a sfGFP construct with two sites for ncAA incorporation (sfGFP2XpAzF, sites D134TAG and N150TAG) to direct the dual incorporation of pAzF into sfGFP (Fig. 2C). We observed reasonable efficiency for dual incorporation with v2.0, highlighting the feasibility of the system for incorporating multiple ncAAs (Fig. 2C). Top-down protein MS analysis confirmed the dual incorporation of pAzF, although the protein experienced some slight degradation during purification which was accounted for in our mass calculation (Fig. S2).

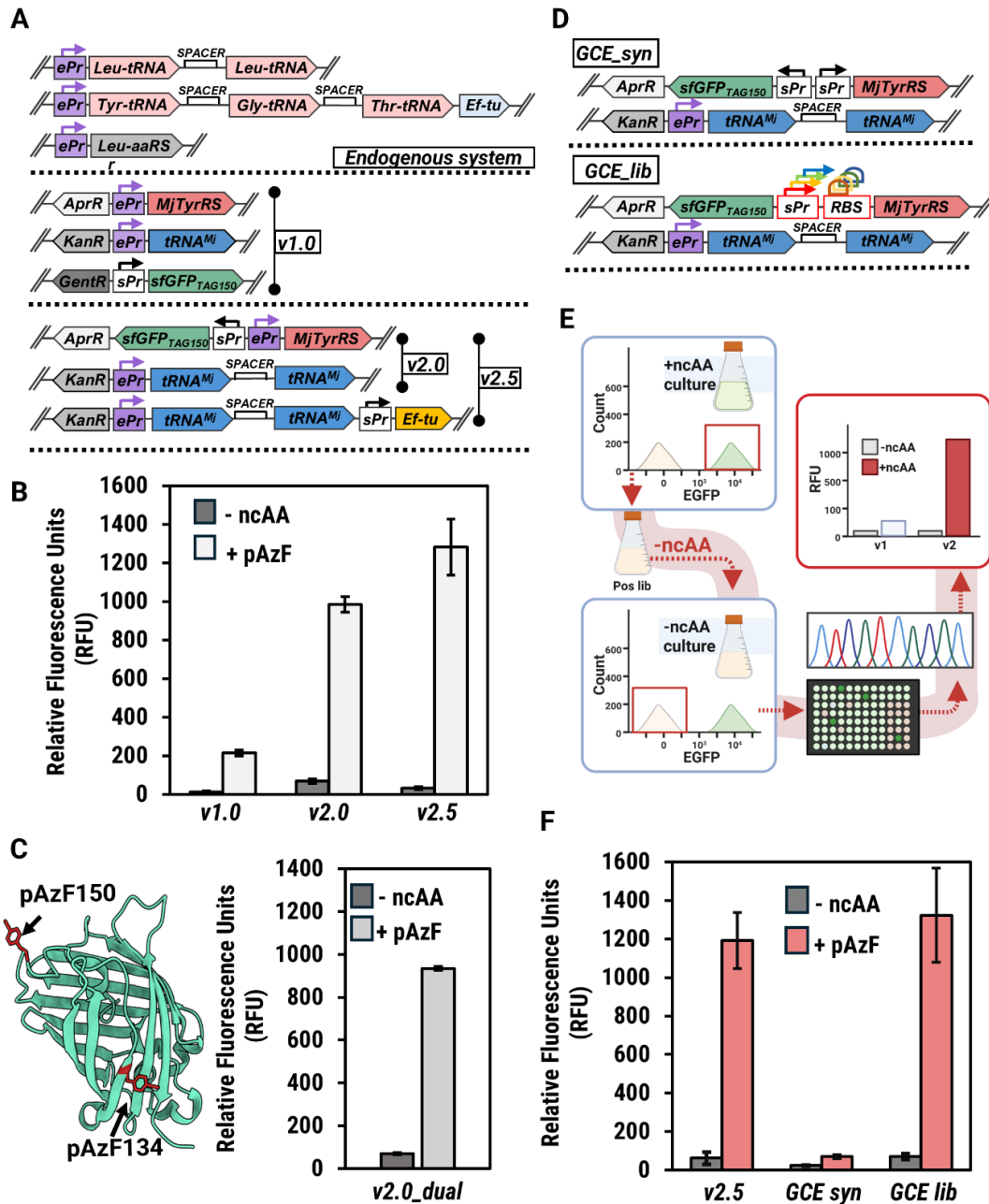


Fig. 2. GCE optimization strategies in *P. putida*. **A** Overview of endogenous and initial GCE translational systems utilized in this work. **B** GCE efficiency assays for KT2440 GCE-SAGE systems from four biologically independent cultures. The GCE system used is indicated below each column set on the x-axis. The ncAA condition for each system is disclosed in a legend in the upper left of the figure. **C** Overview of dual incorporation system. Structure of sfGFP is indicated on the left with each site indicated in red. GCE incorporation efficiency for the dual incorporation system is shown on the right from large scale expression cultures of KT2440. SDS-PAGE of 2X pAzF shown side-by-side with WT GFP, protein band of interest is indicated by an arrow. **D** Overview of GCE synthetic translational systems and GCE library scheme. **E** Scheme of FACS-based library selection for promoter optimization. After transformation into KT2440, the promoter library is cultured in the presence of ncAA to elicit a fluorescent signal. Fluorescent events are sorted via FACS and cultured in the absence of ncAA. Low fluorescent events are then plate sorted for sequencing and final validation. **F** GCE efficiency assays for KT2440 GCE synthetic translational system and optimized GCE library system from four biologically independent cultures. The ncAA condition for each system is disclosed in a legend at the top of the figure.

Multi-component systems often require the coordinated optimization of each component for high performance(37). A common strategy is to improve transcription and translational rates by using engineered promoters for enhanced RNA polymerase recruitment or by altering the 5' UTR to include efficient ribosomal binding sites (RBSs). While the LeuRS promoter enabled robust performance in v2.0, the absence of a clear RNA polymerase binding motif posed a challenge for engineering increased transcription. The synthetic *tac* promoter is broadly used in bacteria, provides strong expression, and is easily tunable due to its well characterized core promoter elements(38). Therefore, we evaluated the *tac* promoter as a replacement for the LeuRS promoter to support *MjTyrRS* expression for efficient ncAA incorporation in *P. putida* (GCE syn system, Figures 2D and 2F). Unexpectedly, sfGFP_{TAG150} expression significantly decreased when *MjTyrRS* transcription was driven by the *tac* promoter compared to the endogenous LeuRS promoter (Fig. 2F). However, this outcome aligns with observations by others that tuning the expression of GCE machinery that was originally developed for use in *E. coli* is crucial for robust GCE application in non-*E. coli* strains(39).

Rather than laboriously evaluating many promoter and ribosomal binding site variants individually to identify alternative promoters that enable improved performance, we leveraged the high transformation efficiency afforded by SAGE and the amenability of the sfGFP reporter assay to evaluate a large collection of *MjTyrRS* expression variants in a pooled assay. We previously demonstrated the efficacy of SAGE for high-throughput sequencing-based methods to assess large promoter libraries in pooled assays(35). Adopting a similar approach, we aimed to optimize *MjTyrRS* production by identifying regulatory elements that ensure sufficient *MjTyrRS* levels for robust ncAA incorporation without causing non-specific amino acid incorporation.

We constructed a pooled *MjTyrRS* expression library using 449 synthetic and natural promoter sequences. To achieve a diverse range of expression levels, we employed a cloning strategy to randomly incorporate one of 106 ribosomal binding site elements downstream of the promoter sequences. These RBS elements were designed by a RBS calculator(40) to span approximately a 441-fold range of translational efficiencies. Additionally, each regulatory sequence included the riboJ insulator sequence(41) between the promoter and ribosomal binding site to stabilize gene expression. The resulting ~209,000 member *MjTyrRS* expression plasmid library was integrated into *Pseudomonas putida* using SAGE, resulting in a pooled collection of ~10,000,000 transformants, each containing a distinct promoter-RBS combination driving *MjTyrRS* expression. Each transformant also contained the reporter construct and the dual o-tRNA cassette from v2.0 enabling GCE (Fig. 2D).

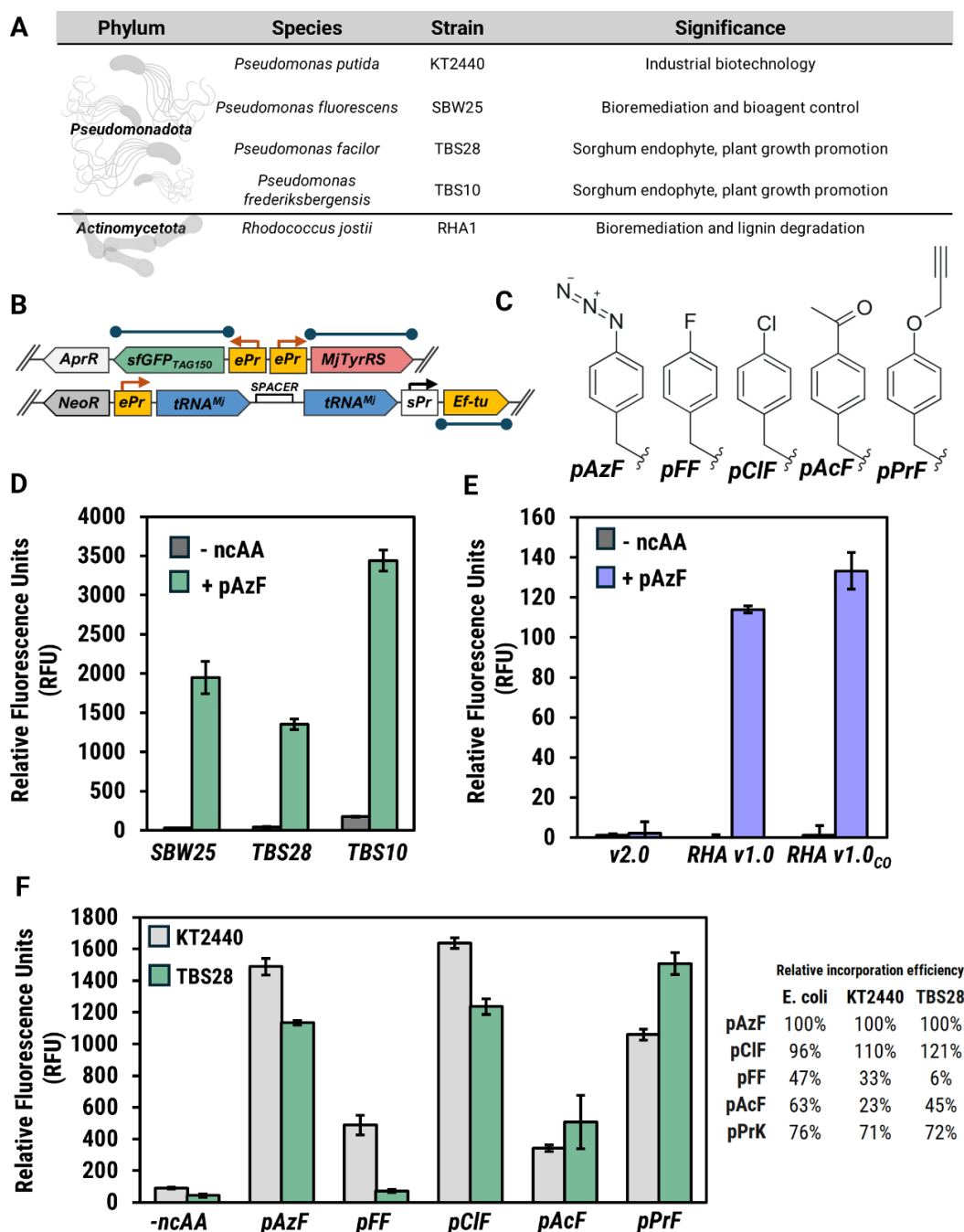


Fig. 3. GCE-SAGE technology in non-model bacteria. **A** Overview of bacterial species used in this study. **B** Overview of RHA1 GCE translational systems. Elements that were codon optimized are indicated by a blue line. **C** Structures of ncAAs screened in KT2440 and TBS28 strains. **D** GCE efficiency assays for GCE v2.0 in three pseudomonads. The pseudomonad used is indicated below each column set on the x-axis. The ncAA condition for each system is disclosed in a legend in the upper left of the figure. **E** GCE efficiency assays for GCE v2.0 vs GCE RHA in RHA1. The GCE system used is indicated below each column set on the x-axis. The ncAA condition for each system is disclosed in a legend in the upper left of the figure. **F** ncAA incorporation efficiency assay for KT2440 and TBS28 with GCE v2.0. The ncAA system used is indicated below each column set on the x-axis. The strain (either KT2440 or TBS28) disclosed in a legend in the upper left of the figure. The incorporation efficiency of each ncAA relative to pAzF are shown for E. coli, KT2440 and TBS28.

As the read-out for GCE efficiency in our system is sfGFP, fluorescence activated cell sorting (FACS) was performed on the strain library to identify efficient and orthogonal library members. This was done by first culturing the strain library in the presence of the ncAA to allow for ncAA-dependent sfGFP production.

Members that exhibit efficient ncAA incorporation and thus generate a fluorescent output are enriched via FACS (positive sort). Enriched members are then cultured in the absence of ncAA to identify members with low background incorporation of native amino acids, ensuring that orthogonality is maintained with the new regulatory elements (Fig. 2E).

For the initial positive selection, we cultivated the library in media containing pAzF and sorted the top 1.5% most fluorescent library members into a pooled library of strains. This pool of fluorescent cells was sub-cultured into media lacking pAzF and after overnight incubation, was sorted again via FACS. This time, members that exhibited no fluorescent signal above background were plate-sorted individually into microtiter plates, cultivated overnight, sequenced and further validated in scaled up GCE efficiency assays (Fig. S3, Table S6).

The top performing strain displayed ncAA incorporation efficiency and orthogonality that was on par with v2.5 (Fig. 2F) despite lacking an engineered EF-Tu. Surprisingly, a *Lactobacillus*_39770 promoter with relatively low transcriptional activity in other *Pseudomonas* sp. was identified in the top performing strain(35). Of note, this promoter was found to have between ~24 to 92-fold lower transcriptional activity than the *tac* promoter in our prior work(35), suggesting that poor performance with the *tac* promoter may be a consequence of excessive *MjTyrRS* expression. The *Lactobacillus*_39770 promoter was coupled with a ribosomal binding site element with a relatively high predicted RBS translation initiation rate of 67067.6, on par with the RBS from our *tac* system. Thus, this library selection method, in one round of positive and negative sorting, produced a promoter/RBS pair with ncAA incorporation efficiency equivalent to the endogenous promoters, permitting the use of non-endogenous promoters to drive GCE technology.

SAGE facilitates the genetic code expansion of phylogenetically distant microbes

After the successful implementation of GCE in *P. putida* KT2440, we evaluated the ability to easily transfer our optimized GCE systems into related *Pseudomonas* sp. (Fig. 3A). These bacteria include the plant growth promoting rhizobacterium *P. fluorescens* strain SBW25(42) and two pseudomonads that were isolated from the endosphere of *Sorghum bicolor* under drought conditions(35, 43). Each of the two, *P. frederiksbergensis* (TBS10)(35) and *P. facilis* (TBS28)(43), have been previously engineered to be compatible with SAGE via the incorporation of a poly-attB landing site(35, 43).

We directly transferred the v2.0 machinery developed in *P. putida* into the chromosomes of the three *Pseudomonads* and observed pAzF incorporation on par with what was observed in *P. putida* demonstrating that SAGE-based GCE systems can be shared across genetically similar bacteria (Figures 3D and 3F). Interestingly, even though the GCE efficiency was similar across the pseudomonads, the fluorescent output did vary across a ~2-fold range. It remains to be seen whether these differences are a

consequence of gene expression differences, or of another physiological trait such as ncAA transport.

As mentioned above, the pCNF-RS used here is polyspecific, meaning it has the capacity to charge the o-tRNA with several distinctive *para*-substituted phenylalanine analogs. However, each organism has its own distinct set of metabolite transporters, sensors, and other physiological regulators that control expression of the proteins that enable ncAA uptake and it remains unclear how much of a role this plays in GCE efficiency. To examine this, we utilized the permissivity of the pCNF-RS employed in our GCE system(44) to conduct a small ncAA screen, testing four additional *para*-substituted phenylalanine derivatives in both *P. putida* and the environmental isolate *P. facilis* (Fig. 3C). We then compared their incorporation efficiencies relative to pAzF to reported values from *E. coli*(44) (Fig. 3F).

While overall the relative incorporation efficiencies among the three organisms were similar, we observed a few notable differences. Specifically, while both pseudomonads exhibited lower incorporation of 4-fluoro-L-phenylalanine relative to *E. coli*, incorporation of this ncAA into sfGFP was almost non-existent in *P. facilis* (Fig. 3F). These differences are unlikely to be due to the activity of the pCNF-RS, as the other ncAAs showed very similar efficiencies. It is more likely that differences in amino acid transport between the organisms are responsible, again underscoring the importance of considering ncAA import mechanisms when integrating new bacterial strains.

A benefit of utilizing SAGE-based tools is the ability to easily transfer materials developed in SAGE-compatible hosts with phylogenetically distant SAGE-compatible bacteria. To demonstrate this benefit, we tested our GCE 2.0 system developed in *Pseudomonas putida* in the actinomycete *Rhodococcus jostii* strain RHA1(45), a representative of a genus with diverse metabolic capabilities ranging from catabolism of cholesterol(46) and petroleum-derived hydrocarbon(47) to conversion of lignin-derived feedstocks into useful chemicals(48). V2.0 was ported into RHA1, yet ncAA-sensitive fluorescence was not observed above background levels (Fig. 3E) suggesting the Pseudomonad-derived machinery was incompatible with RHA1.

To optimize expression, we identified the most abundant codon (Ala) in the RHA1 genome and utilized the endogenous aaRS and tRNA promoters associated with Ala to drive the GCE elements (Fig. 3B). Considering the high GC content of *Rhodococcus* genomes and several strong codon biases, we also created a codon-optimized version of the protein elements to assess the impact of codon optimization on GCE efficiency. The RHA1-tailored system showed a dramatic improvement compared to v2.0, with more subtle differences observed between codon-optimized (RHA v1.0co) and non-codon-optimized (RHA v1.0) GCE-RHA1 systems (Fig. 3B). However, the overall fluorescent output of the RHA1 system is comparable to v1.0 in *P. putida*, suggesting potential for further improvement.

Site specific incorporation of acetyl-L-lysine and other lysine analogs

We were encouraged by our initial success in the implementation of the *Mj*TyrRS-pCNF system and sought to use our optimized platform to broaden the existing lysine PTM interrogation toolkit in non-*E. coli* strains. As the *Mj*TyrRS system primarily incorporates aromatic ncAAs, we chose to onboard the Pyl system, which enables site-

specific incorporation of lysine analogues due to the structural similarity between pyrrolysine—its natural substrate—and lysine(49). Derived from *Methanosarcina* species, the Pyl system is composed of a pyrrolysyl-aminoacyl-tRNA synthetase (PylRS) and its cognate tRNA (tRNA^{Pyl}) and has enabled the incorporation of a suite of ncAAs developed to interrogate lysine PTMs(50). From these, we chose to encode acetyl-lysine (AcK), azido-lysine (AzK) and propargyloxycarbonyl-lysine (PIK) into *Pseudomonas putida* (Fig. 4A)(50). AcK incorporation enables direct interrogation of the influence of lysine acetylation on protein function both *in vivo* and *in vitro*(50) and AzK and PIK contain distinct bioorthogonal click chemistry handles(51) that can be used to probe AcK-dependent protein interactions, facilitate fluorescent labeling and in some cases, to site-specifically model ubiquitin(52).

We integrated the Pyl system into *P. putida* using two engineered PylRS encased in the same genetic framework employed for v2.0. The first, MbPylRS-AF (PylRS1), contains active site mutations that permit the incorporation of bulky lysine analogues(53), suitable for AzK and PIK. The second, chPylRS-IPYE (PylRS2), is a chimeric synthetase known for its high efficiency in incorporating AcK in proteins in *E. coli*(54). These PylRSs were paired with an evolved tRNA^{Pyl}, which has been shown to enhance AcK incorporation into proteins(31) (Fig. 4A). With PylRS1 in GCEv2.0, we observed a high incorporation efficiency of AzK and PIK with no additional alterations to the system (Fig. 4C). However, it was surprising to observe that, despite robust reported performance in *E. coli*, AcK incorporation efficiency was very low in *P. putida* (Fig. 4C). This finding, coupled with very few reported instances of acetyl-lysine incorporation via GCE in organisms other than *E. coli*, suggests that this ncAA poses greater challenges for incorporation than typical ncAAs.

We had previously observed that increasing *Mj*tRNA^{Tyr} copy number enhanced pAzF incorporation efficiency and posited that the tRNA^{Pyl} copy number may be a limiting factor here as well. To test this, we incorporated an additional dual tRNA operon into *P. putida* at a third *attB* site (v3.0, Fig. 4A) bringing the total number of tRNA^{Pyl} copies to four. We repeated this operation for the *Mj*TyrRS-pCNF system then evaluated the ncAA incorporation for both.

Doubling the tRNA copy number improved ncAA with both systems, but to strikingly different extents. Given the already robust incorporation of pAzF with the *Mj*TyrRS system, it was unsurprising to see a substantial but relatively modest 2-fold increase in sfGFP production with the addition of the second tRNA cassette (Fig. 4D, right panel). However, the incorporation efficiency of AcK by the chPylRS(IPYE) system was improved by 25-fold (Fig. 4D, left panel), suggesting tRNA copy number was a limiting factor. This one improvement increased the efficiency of AcK incorporation to be on par with our best existing system and signified that tRNA expression is a critical lever for optimization in each host.

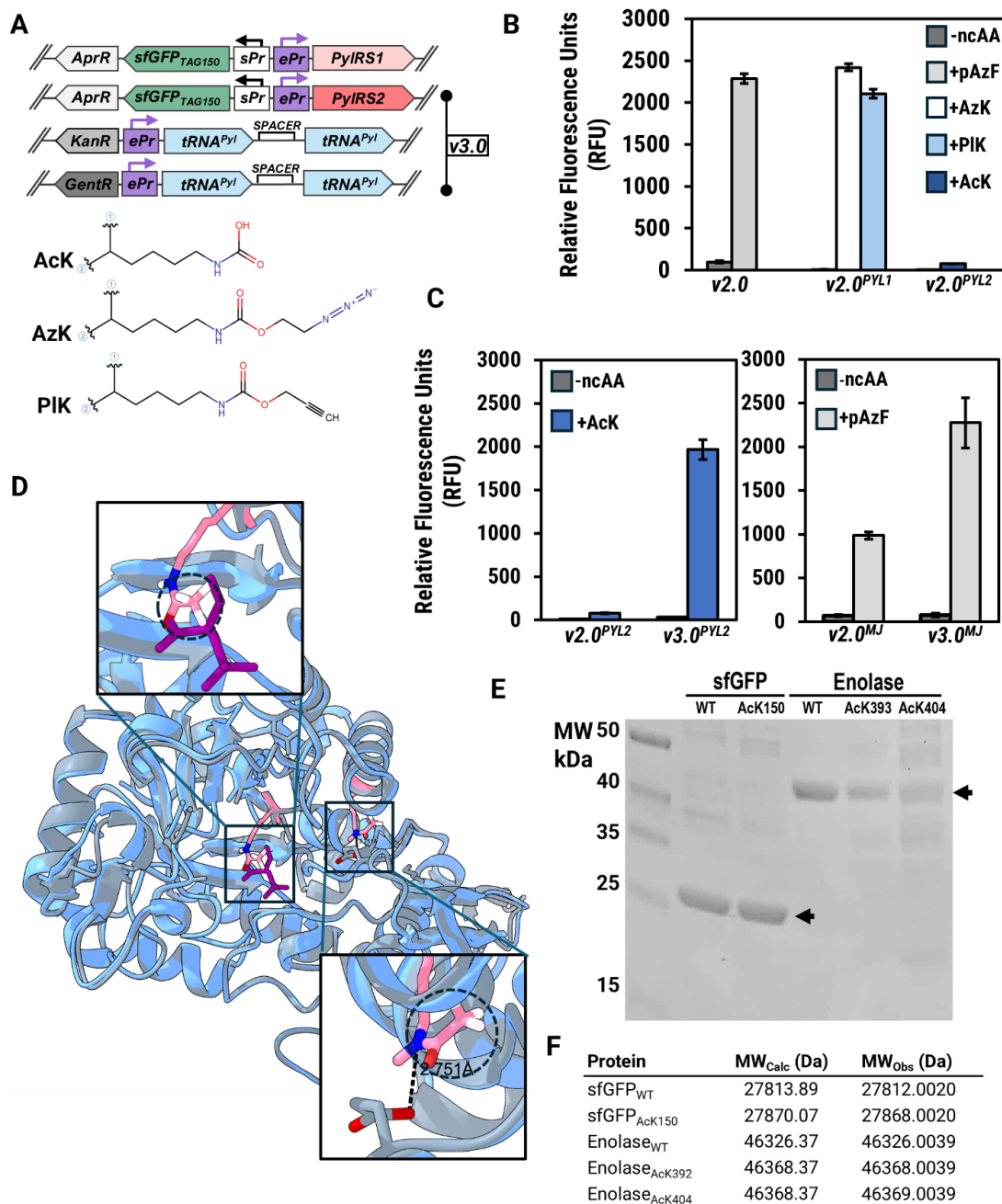


Fig. 4. SAGE-based optimization permits facile acetyl-lysine incorporation. **A** Overview of GCE systems with Pyl components. **B** Structures of ncAAs incorporated with Pyl components. **C** GCE efficiency assays with Pyl systems in KT2440. The GCE system used is indicated below each column set on the x-axis. The ncAA condition for each system is disclosed in a legend on the right of the figure. **D** GCE efficiency assays comparing GCE v2.0 with GCE v3.0. Pyl system and *Mj* systems in KT2440 are compared in left panels and right panels respectively. The GCE system used is indicated below each column set on the x-axis. The ncAA condition for each system is disclosed in a legend in the top left of the figure. **E** Structure of enolase with acetyl lysines and bound substrate. Alpha fold generated structure of KT2440 enolase with acetylations modeled relevant lysine residues (blue ribbon structure) overlaid on PDB structure of enolase from *E. coli* (gray ribbon structure) with bound substrate 2PGA (purple molecule). Acetyl groups are circled. **F** SDS-PAGE of acetylated proteins, purified from KT2440. Arrows indicate the MW for each set of purified proteins. **G** Observed intact masses of purified acetylated proteins with their respective calculated masses.

Production of site-specifically acetylated enolase in Pseudomonas putida with genetic code expansion.

While we were encouraged by the substantial improvement in GCE efficiency driven by additional copies of tRNA, we wanted to vet our system by producing an essential protein with acetyl-lysine encoded at biologically relevant acetylation sites. Previously, Ernesto et al performed a proteomic analysis which identified the acetylation of highly conserved lysines in central metabolic enzymes from diverse bacteria, including *Pseudomonas putida*(5). One essential protein highlighted in the study was the enzyme enolase, which plays a crucial role in glycolysis and gluconeogenesis by catalyzing the reversible conversion of 2-phosphoglycerate to phosphoenolpyruvate, a key step in central carbon metabolism(5). Enolase also performs additional moonlighting functions that contribute to bacterial stress responses and pathogenicity(55). As such, we selected the enolase enzyme as a model protein to test the efficiency of AcK incorporation at multiple highly conserved lysine residues

With respect to site selection, the study identified two conserved lysines of interest in enolase; K392 which is buried within the active site and is involved in phosphoenolpyruvate catalysis and K404 which is located towards the periphery of the protein (Fig. 4E). Under the tested conditions, K404 was found to be acetylated in *Pseudomonas putida*, however, K392 which upon acetylation can ablate enolase activity, was observed to be acetylated in other bacteria but not in *Pseudomonas putida*.

P. putida encodes multiple lysine deacetylases, including those from the CobB and Metal-dependent lysine deacetylase families (NCBI: txid160488). Such deacetylases have a broad substrate range, but they are unable to deacetylate all acetylated lysine residues. We hypothesized that the lack of detectable acetylation at K392 may be due to efficient enzymatic deacetylation of one residue and not the other. By producing the acetylated protein *in vivo* in *Pseudomonas putida*, we could observe whether the modification is maintained at each residue over the course of a typical culturing experiment. If the acetylation remains stable, then that suggests the protein was simply not acetylated under the tested conditions.

For site-specific acetylated enolase production, we replaced the gene for sfGFP_{TAG150} in v3.0 with a series of enolase constructs (enolase_{WT}, enolase_{TAG392} and enolase_{TAG404}) in *P. putida*. Cultures for strains harboring each of these constructs were grown to stationary phase, at which point protein was extracted and affinity-purified to assess expression and acetylation of recombinant enolase. The three variants expressed successfully based on SDS-PAGE analysis (Fig. 4F) and the intact molecular masses of all enolase variants were determined via mass spectrometric analysis. This analysis revealed masses corresponding to acetyl-lysine incorporation for enolase_{TAG392} and enolase_{TAG404} (Figs. 4G and Figs. S4, S5). While the molecular weights for other contaminating proteins were detected in the analysis, the mass associated with enolase_{WT} was not observed in the acetylated enolase samples (Fig. S5). While not exhaustive, this data refutes the hypothesis that intrinsic lysine deacetylase activity during the cultivation was responsible for the lack of acetylation at K392.

3.0 Discussion

Genetic code expansion has revolutionized the analysis of the bacterial acetylome by enabling the precise incorporation of PTMs along with a suite of PTM enrichment and visualization tools, enabling real-time monitoring of acetylation dynamics under physiological conditions. However, inconsistent genetic tractability and unpredictable translational regulatory elements across microbes have hindered the implementation and optimization of this technology in diverse bacteria, particularly those of interest for human health and bioremediation. This is most evident with respect to the ncAA acetyl-lysine itself, as the subtlety of the modification makes it a more challenging incorporation target.

We address this challenge by integrating GCE with serine recombinase-assisted genome engineering (SAGE), a method renowned for its high efficiency across a wide range of bacterial strains. With GCE-SAGE, each GCE component is site-specifically integrated into the bacterial genome at a high enough efficiency for library optimization and with enough landing pads (attB sites) to integrate up to ten constructs genetically. This strategy allowed for the stable integration of GCE components into the genome of *P. putida* and improved efficiency by easily allowing for increased o-tRNA copy number. SAGE also enabled the development of a high-throughput promoter and RBS library selection method using FACS. This approach identified highly efficient promoter/RBS pairs, facilitating the fine-tuning of gene expression in non-model organisms. This technique is crucial for matching GCE output to endogenous production levels, especially when only synthetic promoters or unknown promoter strengths are available.

GCE-SAGE was then successfully transferred to multiple Proteobacteria (*Pseudomonas fluorescens* SBW25, *Pseudomonas frederiksbergensis* TBS10, and *Pseudomonas facilor* TBS28) and the Actinobacterium *Rhodococcus jostii* RHA1. Screening four para-substituted phenylalanine derivatives revealed organism-specific differences in incorporation efficiency, likely due to variations in amino acid transport rather than pCNF-RS activity. Studying and expressing amino acid transporters will be a critical component to enhance ncAA uptake as SAGE enables the application of GCE in new model systems.

We successfully incorporated lysine analogs (AzK, PrK, and AcK) using evolved pyrrolysyl-tRNA synthetases (PylRS). The variable efficiency of incorporation, particularly for AcK, highlighted the need for additional tRNA copies to achieve biologically relevant levels, confirming the critical role of tRNA abundance in GCE performance. It was encouraging that the efficiency of the system could be enhanced from near zero levels by simply adding additional copies of tRNA. As this has been observed in mammalian systems as well, it is likely that this optimization approach is near universal, providing a simple path towards GCE optimization.

The stable incorporation of acetyl-lysine at two distinct sites (K392 and K404) in the essential glycolytic enzyme enolase in *Pseudomonas putida* showcased the practical utility of our optimized GCE platform. In conclusion, we have successfully pioneered the establishment of acetylation in diverse bacterial strains, developed a host-agnostic GCE platform and identified strategies to optimize GCE efficiency in the hosting organism. These advancements herald new opportunities for studying PTMs in bacteria, unlocking numerous possibilities for both fundamental research and industrial biotechnology applications

4.0 References

1. J. A. Cain, N. Solis, S. J. Cordwell, Beyond gene expression: the impact of protein post-translational modifications in bacteria. *J Proteomics* **97**, 265-286 (2014).
2. B. Macek *et al.*, Protein post-translational modifications in bacteria. *Nat Rev Microbiol* **17**, 651-664 (2019).
3. D. G. Christensen *et al.*, Post-translational Protein Acetylation: An Elegant Mechanism for Bacteria to Dynamically Regulate Metabolic Functions. *Front Microbiol* **10**, 1604 (2019).
4. Z. Li *et al.*, Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. *Nat Commun* **5**, 4405 (2014).
5. E. S. Nakayasu *et al.*, Ancient Regulatory Role of Lysine Acetylation in Central Metabolism. *mBio* **8**, (2017).
6. M. Liu *et al.*, Bacterial protein acetylation and its role in cellular physiology and metabolic regulation. *Biotechnol Adv* **53**, 107842 (2021).
7. J. Rizo, S. Encarnación-Guevara, Bacterial protein acetylation: mechanisms, functions, and methods for study. *Front Cell Infect Microbiol* **14**, 1408947 (2024).
8. M. Lammers, Post-translational Lysine Ac(et)ylation in Bacteria: A Biochemical, Structural, and Synthetic Biological Perspective. *Front Microbiol* **12**, 757179 (2021).
9. X. Liu, M. Yang, F. Ge, J. Zhao, Lysine acetylation in cyanobacteria: emerging mechanisms and functions. *Biochem Soc Trans* **53**, (2025).
10. J. Zhang *et al.*, Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol Cell Proteomics* **8**, 215-225 (2009).
11. L. Popova, R. A. Carr, V. J. Carabetta, Recent Contributions of Proteomics to Our Understanding of Reversible N(ε)-Lysine Acylation in Bacteria. *J Proteome Res* **23**, 2733-2749 (2024).
12. B. T. Weinert *et al.*, Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli. *Mol Cell* **51**, 265-272 (2013).
13. B. Schilling *et al.*, Protein acetylation dynamics in response to carbon overflow in Escherichia coli. *Mol Microbiol* **98**, 847-863 (2015).
14. M. L. Kuhn *et al.*, Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. *PLoS One* **9**, e94816 (2014).
15. J. Ren, Y. Sang, J. Lu, Y. F. Yao, Protein Acetylation and Its Role in Bacterial Virulence. *Trends Microbiol* **25**, 768-779 (2017).
16. S. Venkat *et al.*, Characterizing Lysine Acetylation of Isocitrate Dehydrogenase in Escherichia coli. *J Mol Biol* **430**, 1901-1911 (2018).
17. S. Venkat, C. Gregory, K. Meng, Q. Gan, C. Fan, A Facile Protocol to Generate Site-Specifically Acetylated Proteins in Escherichia Coli. *J Vis Exp*, (2017).
18. S. Venkat *et al.*, Genetically encoding thioacetyl-lysine as a non-deacetylatable analog of lysine acetylation in Escherichia coli. *FEBS Open Bio* **7**, 1805-1814 (2017).
19. H. Chen *et al.*, Site-Specifically Studying Lysine Acetylation of Aminoacyl-tRNA Synthetases. *ACS Chem Biol* **14**, 288-295 (2019).
20. J. Araujo, S. Ottinger, S. Venkat, Q. Gan, C. Fan, Studying Acetylation of Aconitase Isozymes by Genetic Code Expansion. *Front Chem* **10**, 862483 (2022).
21. S. Venkat, C. Gregory, J. Sturges, Q. Gan, C. Fan, Studying the Lysine Acetylation of Malate Dehydrogenase. *J Mol Biol* **429**, 1396-1405 (2017).
22. Q. Wang *et al.*, Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science* **327**, 1004-1007 (2010).
23. M. Kremer *et al.*, Bacteria employ lysine acetylation of transcriptional regulators to adapt gene expression to cellular metabolism. *Nat Commun* **15**, 1674 (2024).

24. S. C. Feid *et al.*, Regulation of Translation by Lysine Acetylation in Escherichia coli. *mBio* **13**, e0122422 (2022).
25. J. Luu, V. J. Carabetta, Contribution of N(epsilon)-lysine Acetylation towards Regulation of Bacterial Pathogenesis. *mSystems* **6**, e0042221 (2021).
26. Z. Fang *et al.*, Potential Role of Lysine Acetylation in Antibiotic Resistance of Escherichia coli. *mSystems* **7**, e0064922 (2022).
27. S. Wang, A. O. Osgood, A. Chatterjee, Uncovering post-translational modification-associated protein-protein interactions. *Curr Opin Struct Biol* **74**, 102352 (2022).
28. P. Knyphausen, F. Lang, L. Baldus, A. Extra, M. Lammers, Insights into K-Ras 4B regulation by post-translational lysine acetylation. *Biol Chem* **397**, 1071-1085 (2016).
29. S. Kirchgassner *et al.*, Synthesis, Biochemical Characterization, and Genetic Encoding of a 1,2,4-Triazole Amino Acid as an Acetyllysine Mimic for Bromodomains of the BET Family. *Angew Chem Int Ed Engl* **62**, e202215460 (2023).
30. Z. J. Zhang, V. A. Pedicord, T. Peng, H. C. Hang, Site-specific acylation of a bacterial virulence regulator attenuates infection. *Nat Chem Biol* **16**, 95-103 (2020).
31. C. Fan, H. Xiong, N. M. Reynolds, D. Soll, Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. *Nucleic Acids Res* **43**, e156 (2015).
32. L. De Gelder, J. J. Williams, J. M. Ponciano, M. Sota, E. M. Top, Adaptive plasmid evolution results in host-range expansion of a broad-host-range plasmid. *Genetics* **178**, 2179-2190 (2008).
33. J. Mi *et al.*, Investigation of plasmid-induced growth defect in Pseudomonas putida. *J Biotechnol* **231**, 167-173 (2016).
34. L. Li, X. Liu, K. Wei, Y. Lu, W. Jiang, Synthetic biology approaches for chromosomal integration of genes and pathways in industrial microbial systems. *Biotechnol Adv* **37**, 730-745 (2019).
35. J. R. Elmore *et al.*, High-throughput genetic engineering of nonmodel and undomesticated bacteria via iterative site-specific genome integration. *Sci Adv* **9**, eade1285 (2023).
36. R. D. Arroyo-Olarte, R. Bravo Rodriguez, E. Morales-Rios, Genome Editing in Bacteria: CRISPR-Cas and Beyond. *Microorganisms* **9**, (2021).
37. G. Naseri, M. A. G. Koffas, Application of combinatorial optimization strategies in synthetic biology. *Nat Commun* **11**, 2446 (2020).
38. H. A. de Boer, L. J. Comstock, M. Vasser, The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc Natl Acad Sci U S A* **80**, 21-25 (1983).
39. D. A. Stork *et al.*, Designing efficient genetic code expansion in Bacillus subtilis to gain biological insights. *Nat Commun* **12**, 5429 (2021).
40. H. M. Salis, E. A. Mirsky, C. A. Voigt, Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950 (2009).
41. K. P. Clifton *et al.*, The genetic insulator RiboJ increases expression of insulated genes. *J Biol Eng* **12**, 23 (2018).
42. M. J. Bailey, A. K. Lilley, I. P. Thompson, P. B. Rainey, R. J. Ellis, Site directed chromosomal marking of a fluorescent pseudomonad isolated from the phytosphere of sugar beet; stability and potential for marker gene transfer. *Mol Ecol* **4**, 755-763 (1995).
43. A. Wilson *et al.*, A novel phenylpropanoid methyl esterase enables catabolism of aromatic compounds that inhibit biological nitrification. *bioRxiv*, 2023.2006.2002.543320 (2023).
44. D. D. Young *et al.*, An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. *Biochemistry* **50**, 1894-1900 (2011).
45. M. Seto *et al.*, A Novel Transformation of Polychlorinated Biphenyls by Rhodococcus sp. Strain RHA1. *Appl Environ Microbiol* **61**, 3353-3358 (1995).
46. R. Van der Geize *et al.*, A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into Mycobacterium tuberculosis survival in macrophages. *Proc Natl Acad Sci U S A* **104**, 1947-1952 (2007).

47. L. Martinkova, B. Uhnakova, M. Patek, J. Nesvera, V. Kren, Biodegradation potential of the genus *Rhodococcus*. *Environ Int* **35**, 162-177 (2009).
48. E. M. Spence, L. Calvo-Bado, P. Mines, T. D. H. Bugg, Metabolic engineering of *Rhodococcus jostii* RHA1 for production of pyridine-dicarboxylic acids from lignin. *Microb Cell Fact* **20**, 15 (2021).
49. D. L. Dunkelmann, J. W. Chin, Engineering Pyrrolysine Systems for Genetic Code Expansion and Reprogramming. *Chem Rev* **124**, 11008-11062 (2024).
50. P. Neumann-Staubitz, M. Lammers, H. Neumann, Genetic Code Expansion Tools to Study Lysine Acylation. *Adv Biol (Weinh)* **5**, e2100926 (2021).
51. M. D. Best, Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry* **48**, 6571-6584 (2009).
52. M. A. Shandell, Z. Tan, V. W. Cornish, Genetic Code Expansion: A Brief History and Perspective. *Biochemistry* **60**, 3455-3469 (2021).
53. T. Yanagisawa *et al.*, Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol* **15**, 1187-1197 (2008).
54. D. I. Bryson *et al.*, Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat Chem Biol* **13**, 1253-1260 (2017).
55. W. S. da Cruz Nizer *et al.*, Oxidative Stress Response in *Pseudomonas aeruginosa*. *Pathogens* **10**, (2021).

Appendix A – Supplemental Materials

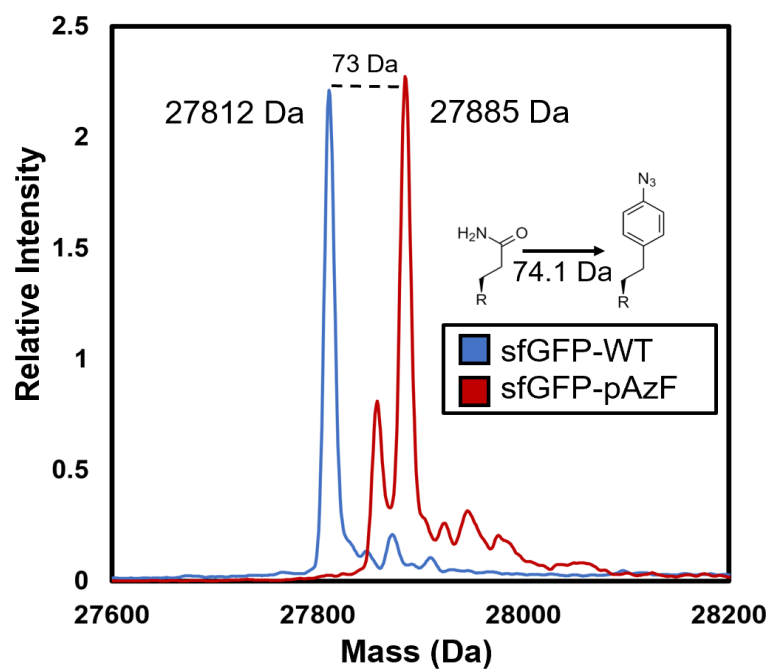


Fig S1. Electrospray ionization mass spectrometry analyses of sfGFP_{WT} (blue trace) and sfGFP_{pAzF} (red trace). Observed masses are shown and correspond well to the expected masses (27811 Da and 27885 Da respectively) and the change in side chain at site 150 (N versus pAzF) is shown with the expected mass shift.

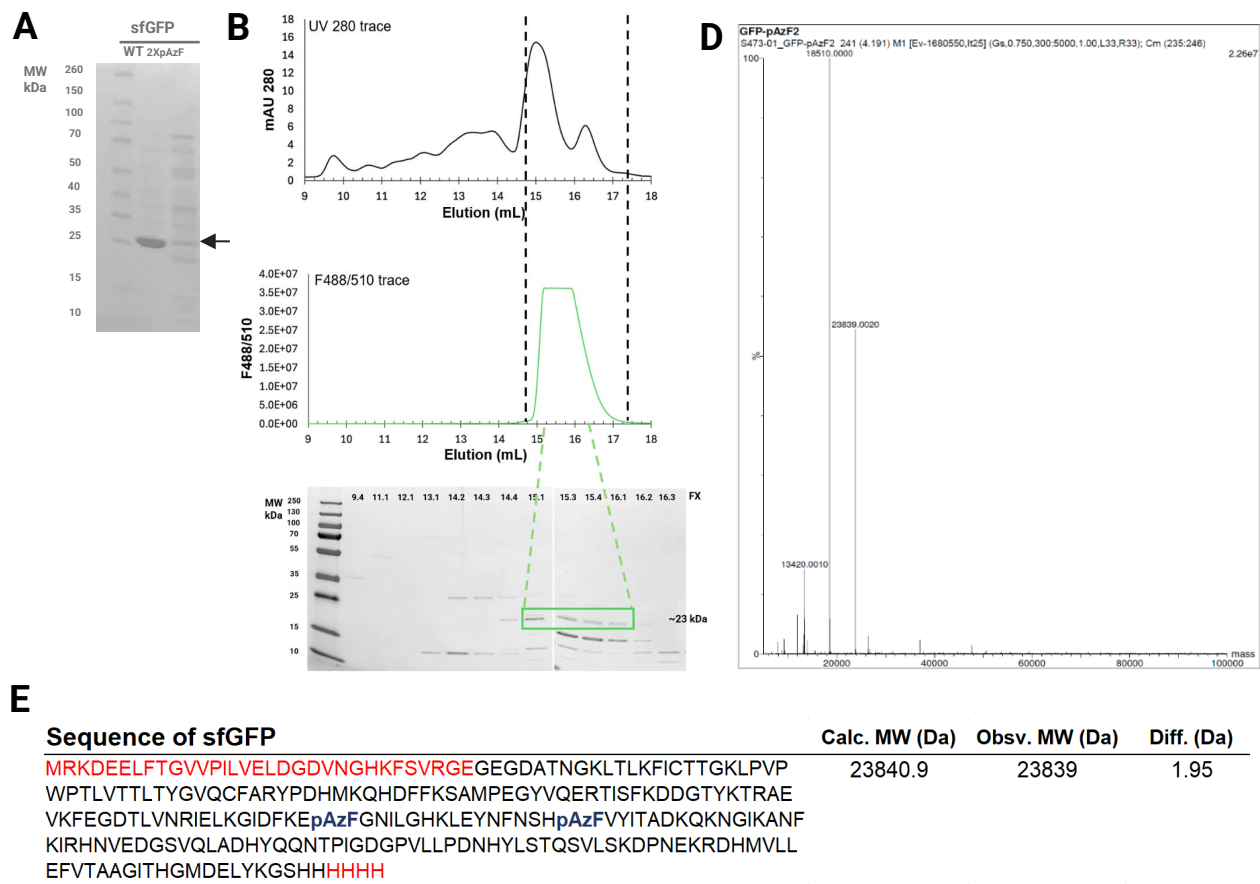


Fig. S2. A. SDS-page of affinity purified sfGFP_{2XpAzF}. Correct MW is indicated by arrow **B** UV 280 and F488/510 traces of size-exclusion chromatography of sfGFP_{2XpAzF} (top and bottom panels respectively) **C** SDS-PAGE analysis of fractions from size-exclusion chromatography. Proteolytically degraded sfGFP_{2XpAzF} with associated fluorescent signature is indicated in green box. **D** Electrospray ionization mass spectrometry analysis of proteolytically degraded sfGFP_{2XpAzF}. **E** Sequence of sfGFP_{2XpAzF}, likely cleaved residues are indicated in red. Calculated mass of degraded sfGFP_{2XpAzF} corresponds well to the observed masses.

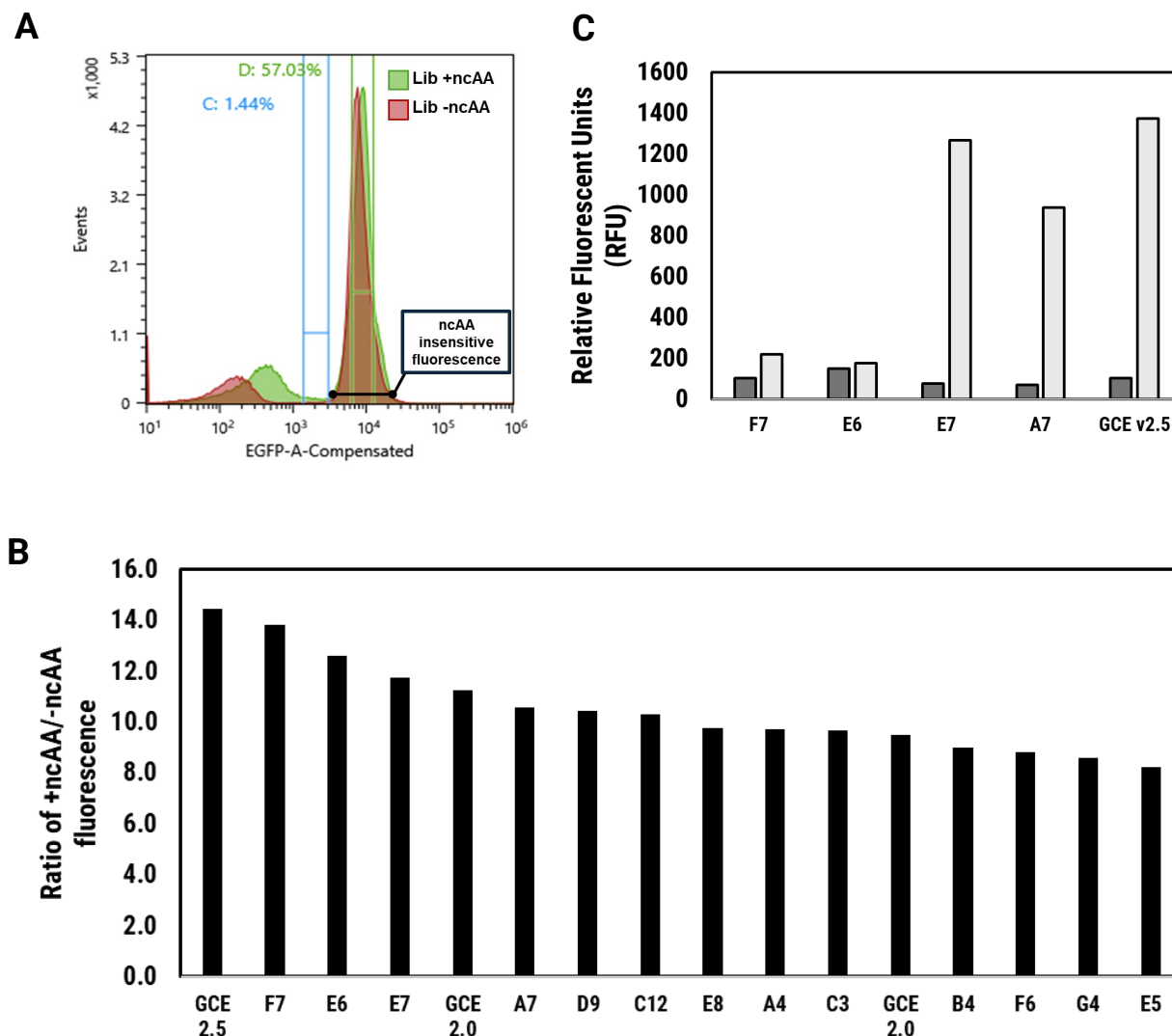
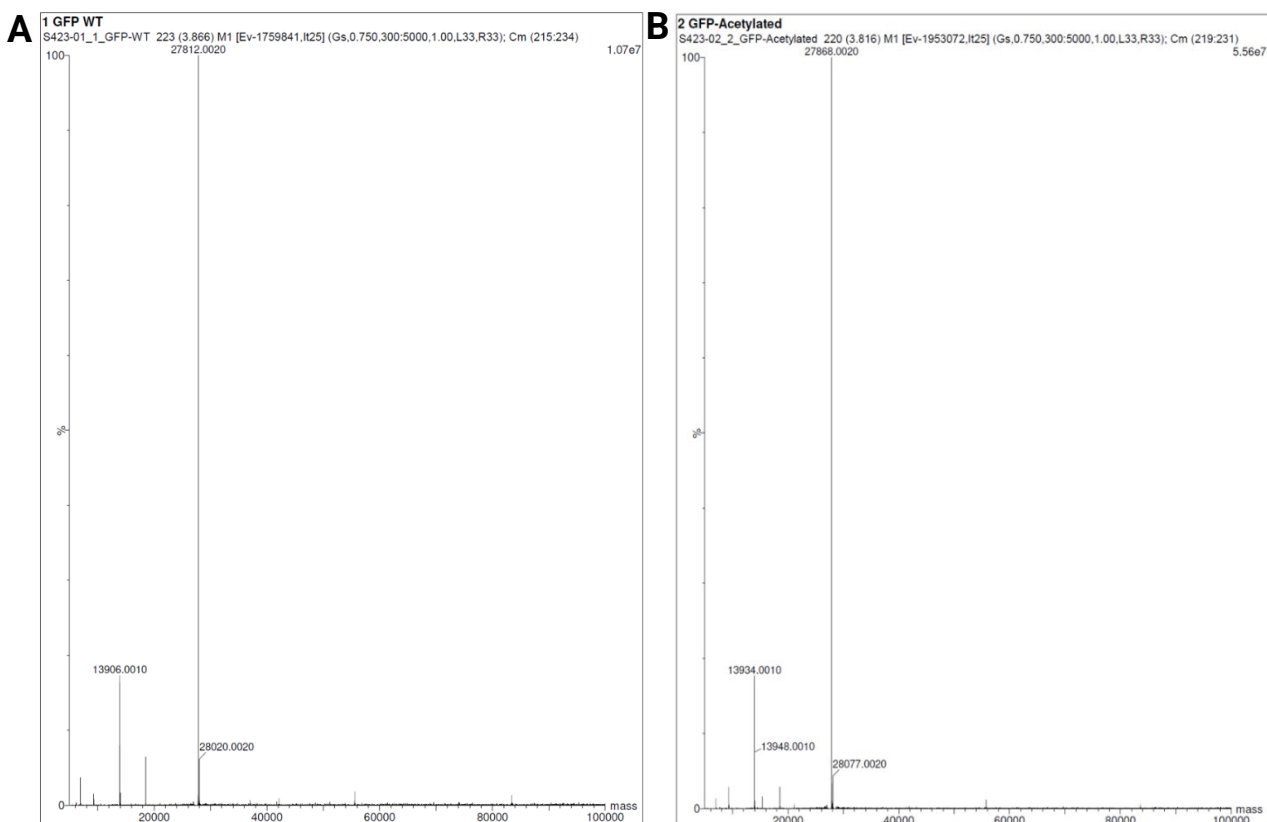
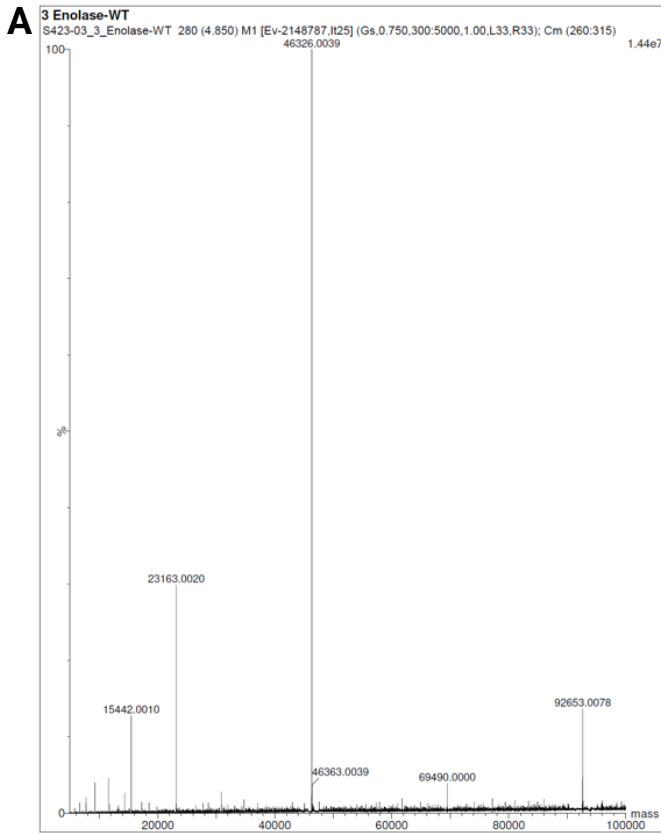


Fig. S3. A Representative histograms of the flow cytometry analysis of the GCE-promoter library cultured with (green trace) and without (red trace) ncAA. The blue gate represents the population that was selected during positive sorts. The green gate represents a highly fluorescent ncAA insensitive population. **B** Results of 0.2 mL GCE incorporation efficiency assays on a subsection of top hits from 96 well single sort. The ratio of +ncAA/-ncAA is given on the y-axis. GCE v2.5 and GCE v2.0 were included on the plate to provide a direct comparison. **C** Results of 2 mL GCE incorporation efficiency assays used to identify final hits.



Supplemental Fig. 4. A Electrospray ionization mass spectrometry analyses of sfGFP. Observed masses are shown and correspond well to the expected mass (27811.3 Da) **B** Electrospray ionization mass spectrometry analyses of sfGFP_{AcK150}. Observed mass is shown and corresponds well to the expected mass (27867.4 Da)



Supplemental Fig. 5. A Electrospray ionization mass spectrometry analyses of enolase_{WT}. Observed mass is shown. Expected mass is 46326.4 Da, including the loss of methionine. **B** Electrospray ionization mass spectrometry analyses of enolase_{AcK342}. Observed mass is shown. Expected mass is 46368.4 Da, including the loss of methionine. **C** Electrospray ionization mass spectrometry analyses of enolase_{AcK404}. Observed mass is shown. Expected mass is 46368.4 Da, including the loss of methionine.

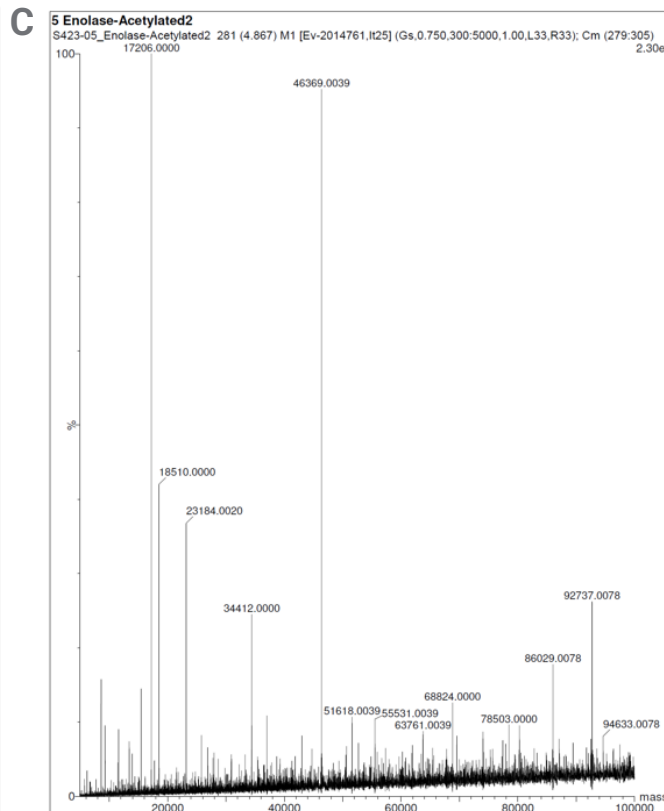
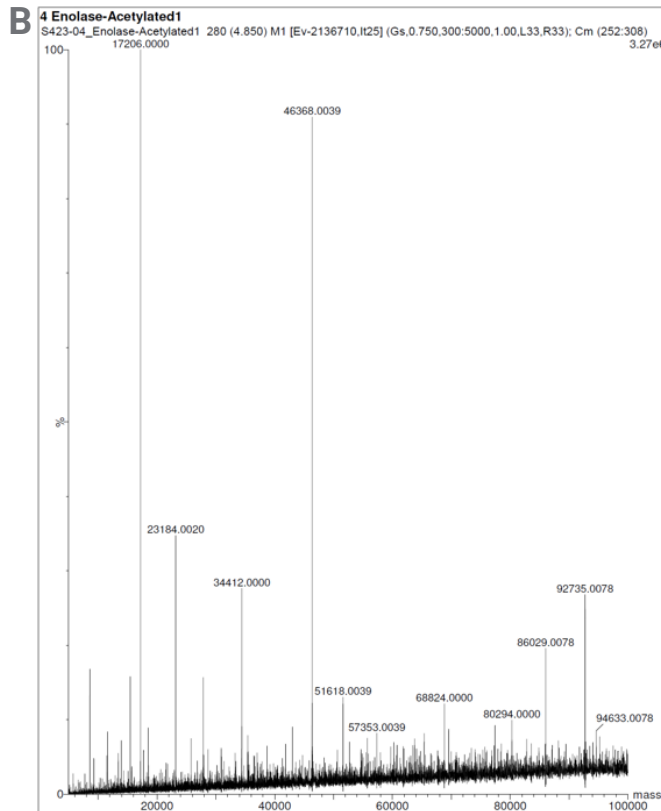


Table S4 - Plasmids used in this study

Name	Genotype	Source
pJH225	CloDF13, <i>aadA1</i> (spec/strep), R4 <i>attP</i>	This work
pJH219	ColA, <i>aac(3)-I</i> (gent), TG1 <i>attP</i>	This work
pJH204	ColE1, <i>nptII</i> (kan/neo), Bxb1 <i>attP</i>	This work
pAW45	CloDF13, <i>aac(3)/IV</i> (apr), R4 <i>attP</i>	This work
pEVF1	pJH0225 JEa3:sfGFP_150	This work
pEVF3	pJH219 Leu-RS:Mj_TyrRS	This work
pEVF7	pJH0204 Leu-tRNA:Mj_tRNA(1X)	This work
pEVF12	pJH0225 JEa3:sfGFP_WT	This work
pEVF45	pAW45 JEa3:sfGFP_WT	This work
pEVF46	pAW45 Leu-RS:Mj_TyrRS, JEa3:sfGFP_150	This work
pEVF75	pJH0204 Leu-tRNA:Mj_tRNA(2X),EF-Tu	This work
pEVF108	pJH0204 Leu-tRNA:Mj_tRNA(2X)	This work
pEVF148	pAW45 Leu-RS:Mj_TyrRS, JEa3:sfGFP_134,150	This work
pMS15	pJH0204, <i>nptII</i> , Ala-tRNA:Mj_tRNA(2X),EF-Tu	This work
pMS3	pAW45 Ala-RS:Mj_TyrRS_CO, Cym:sfGFP_150_CO	This work
pMS16	pAW45 Ala-RS:Mj_TyrRS, Cym:sfGFP_150	This work
pJE2027	pAW45 Ptac:Mj_TyrRS, JEa3:sfGFP_150	This work
pEVF210	pAW45 A7:Mj_TyrRS, JEa3:sfGFP_150	This work
pEVF211	pAW45 E7:Mj_TyrRS, JEa3:sfGFP_150	This work
pEVF212	pAW45 F7:Mj_TyrRS, JEa3:sfGFP_150	This work
pEVF109	pJH0204 Leu-tRNA:Pyl_tRNA(2X)	This work
pEVF145	pAW45 Leu-RS:chimAckK3(IPYE), JEa3:sfGFP_150	This work
pEVF146	pAW45 Leu-RS:MbAckK3(IPYE), JEa3:sfGFP_150	This work
pEVF190	pAW45 Leu-RS:chimAckK3(IPYE), JEa3:enolase_WT	This work
pEVF191	pAW45 Leu-RS:chimAckK3(IPYE), JEa3:enolase_TAG392	This work
pEVF192	pAW45 Leu-RS:chimAckK3(IPYE), JEa3:enolase_TAG404	This work

Table S5 - Strains used in this study

Name	Genotype	Source
NEB 5-alpha F'Iq	<i>Escherichia coli</i> F' proA+B+ lacIq Δ(lacZ)M15 zcf::Tn10 (TetR) / fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England Biolab
KT2440	<i>Pseudomonas putida</i> KT2440	(35)
SBW25	<i>Pseudomonas fluorescens</i> SBW25	(44)
TBS28	<i>Pseudomonas facilor</i> TBS28	(45)
TBS10	<i>Pseudomonas frederiksbergensis</i> TBS10	(31)
RHA1	<i>Rhodococcus jostii</i> RHA1	(46)
AG5577	<i>P. putida</i> KT2440::10x poly-attB	(31)
JE4621	<i>P. fluorescens</i> SBW25 ampC::10x poly-attB	(31)
RS175	<i>P. facilor</i> ::10x poly-attB	(45)
JE5041	<i>P. frederiksbergensis</i> ::10x poly-attB	(31)
AG5879	<i>R. jostii</i> RHA1_RS20555::10x poly-attB	(31)
EVF7	<i>P. putida</i> AG5577 attLR4:pEVF1: attRR4, attLBxb1:pEVF7: attRBxb1, attLTG1:pEVF3: attRTG1	This work
EVF12	<i>P. putida</i> AG5577 attLR4:pEVF45: attRR4	This work
MS5	<i>P. putida</i> AG5577 attLR4:pEVF105: attRR4, attLBxb1:pEVF108: attRBxb1	This work
EVF44	<i>P. putida</i> AG5577 attLR4:pEVF105: attRR4, attLBxb1:pEVF75: attRBxb1	This work
MS44	<i>P. putida</i> AG5577 attLR4:pEVF105: attRR4, attLBxb1:pEVF108: attRBxb1, attLTG1:pEVF141: attRTG1	This work
MS36	<i>R. jostii</i> AG5879 attLR4:pMS3: attRR4, attLTG1:pMS2: attRTG1	This work
MS37	<i>R. jostii</i> AG5879 attLR4:pMS16: attRR4, attLTG1:pMS2: attRTG1	This work
MS61	<i>P. fluorescens</i> JE4621 attLR4:pEVF105: attRR4, attLBxb1:pEVF75: attRBxb1	This work
MS62	<i>P. frederiksbergensis</i> JE5041 attLR4:pEVF105: attRR4, attLBxb1:pEVF75: attRBxb1	This work
MS63	<i>P. facilor</i> RS175 attLR4:pEVF105: attRR4, attLBxb1:pEVF75: attRBxb1	This work
MS52	<i>P. putida</i> AG5577 attLR4:pEVF145: attRR4, attLBxb1:pEVF109: attRBxb1	This work
MS53	<i>P. putida</i> AG5577 attLR4:pEVF145: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work
MS54	<i>P. putida</i> AG5577 attLR4:pEVF146: attRR4, attLBxb1:pEVF109: attRBxb1	This work
MS55	<i>P. putida</i> AG5577 attLR4:pEVF146: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work
MS78	<i>P. putida</i> AG5577 attLR4:pEVF190: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work
MS79	<i>P. putida</i> AG5577 attLR4:pEVF191: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work
MS80	<i>P. putida</i> AG5577 attLR4:pEVF192: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work
MS78	<i>P. putida</i> AG5577 attLR4:pEVF190: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work
MS79	<i>P. putida</i> AG5577 attLR4:pEVF191: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work
MS80	<i>P. putida</i> AG5577 attLR4:pEVF192: attRR4, attLBxb1:pEVF109: attRBxb1, attLTG1:pEVF142: attRTG1	This work

Table S6 – Primers used in this study

Name	Sequence
EVF_GCE_GFP_scrn_F	atgcgtaaagacgaagagctg
EVF_GCE_GFP_scrn_R	cttgtacagttcatccataccatg
EVF_GCE_aaRS_scrn_F	ATGGATGAGTTTGAGATGATTAAACGC
EVF_GCE_aaRS_scrn_R	CAGGCGTTTGCGAATAGG
EVF_GCE_tRNA_scrn_F	GCTGCAGTGCATAAACAGC
EVF_GCE_tRNA_scrn_R	gaAAAGCTTTACATCATCTGCAGAAG
oPNL2125	CGGATTGCAATTGAAGACTTGG
oPNL2126	ATGGATGAGTTTGAGATGATTAAAC
oPNL2127	CCAAGTCTTCAATTGCAATCCG
GCE_lib_v1_aaRS_RBSv1	GTTTAATCATCTCAAACATCATCCATTTATTGHSCCSYTTGCATTATTCGTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv2	GTTTAATCATCTCAAACATCATCCATTATTCGACSTCCTHTACCDACACCTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv3	GTTTAATCATCTCAAACATCATCCATTTATTTTRSCSTTTGCATTATTSCTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv4	GTTTAATCATCTCAAACATCATCCATTAGCAKACHKCCTTAAGTGCAGCCSTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv5	GTTTAATCATCTCAAACATCATCCATTGTAAAACCKBCTTAAGTGHAGCTTTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv6	GTTTAATCATCTCAAACATCATCCATAGATDSCCATCCCTAGTKCCGTGGGTAAACAAAATTATTTGTAGAGGCTGT

Table S6 – Primers used in this study

Name	Sequence
EVF_GCE_GFP_scrn_F	atgcgtaaagacgaagagctg
EVF_GCE_GFP_scrn_R	ctttgtacagttcatccataccatg
EVF_GCE_aaRS_scrn_F	ATGGATGAGTTTGAGATGATTAAACGC
EVF_GCE_aaRS_scrn_R	CAGGCGTTTGCGAATAGG
EVF_GCE_tRNA_scrn_F	GCTGCAGTGCATAAACAGC
EVF_GCE_tRNA_scrn_R	gaAAAGCTTTACATCATCTGCAGAAG
oPNL2125	CGGATTGCAATTGAAGACTTGG
oPNL2126	ATGGATGAGTTTGAGATGATTAAAC
oPNL2127	CCAAGTCTTCAATTGCAATCCG
GCE_lib_v1_aaRS_RBSv1	GTTTAATCATCTCAAACCTCATCCATTTATTGHSCCSYTTGCATTATTCGTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv2	GTTTAATCATCTCAAACCTCATCCATTTATTGACSTCCTHTACCDACACCTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv3	GTTTAATCATCTCAAACCTCATCCATTTATTTTTRSCTCSTTTGCATTATTSCTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv4	GTTTAATCATCTCAAACCTCATCCATTAGCAKACHKCCTTAACTGCAGCCSTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv5	GTTTAATCATCTCAAACCTCATCCATTGTAAAACCKBCTTAACTGHAGCTTTTAAACAAAATTATTTGTAGAGGCTGT
GCE_lib_v1_aaRS_RBSv6	GTTTAATCATCTCAAACCTCATCCATAGATDSCCATCCCTAGTKCCGTGGGTAAACAAAATTATTTGTAGAGGCTGT

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov