LARGE-SCALE LIBRARY GENERATION OF UNNATURAL AMINO ACID BIOSYNTHETIC ENZYMES

Nolan Brown

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Genetics and Molecular Biology.

Chapel Hill
2023

Approved by:

Dorothy Erie

Eric Brustad

Cyrus Vaziri

Nathaniel Hathaway

Kevin Slep

ABSTRACT

Nolan Brown: Large-scale Library Generation of Unnatural Amino Acid Biosynthetic Enzymes
(Under the direction of Dorothy Erie)

Unnatural amino acid (UAA) technology offers powerful and flexible tools for addressing many issues in human health, scientific research, and industry. However, UAA technology is currently limited by prohibitive costs and poor availability. Evolution of UAA biosynthetic enzymes can make UAAs much cheaper and more available, but previous efforts to evolve or design UAA biosynthetic enzymes have only used low-throughput methods which are limiting for general use. This thesis presents a framework by which large libraries of mutant UAA biosynthetic enzymes can be generated for truly high-throughput selections. This thesis also describes the generation of chloramphenicol resistance and GFP selection marker genetic constructs optimized for performing selections on large libraries of UAA biosynthetic enzymes. The combination of this library generation framework and selection markers optimized for finding UAA biosynthetic enzymes should greatly facilitate the evolution of UAA biosynthetic enzymes and drastically improve the cost and availability of UAA technology.

ACKNOWLEDGEMENTS

I want to give a massive thank you to the professors that hosted me in their laboratories. My time in the laboratory of Matthew Hirsch taught me much beyond the biology of AAV vectors, such as independence, integrity, perseverance, and cooperation. I am a much stronger scientist and person due to his mentorship. When I joined the laboratory of Eric Brustad, many things changed for me. I was a pure biology student in a biological chemistry laboratory and I struggled to follow many concepts of their research. However, Eric and all the members of his laboratory were patient, welcoming, and supportive. Eric was not only a great scientific mentor, but also a great life mentor. Under his mentorship, I began one of my two most significant transformations during graduate school. My narrow and self-centered fixation on gene therapy was transformed into a broad appreciation and passion for a diverse array of sciences. When Eric was no longer able to serve as my mentor, Dorothy Erie generously adopted me into her laboratory. Dorothy was an exceptional mentor that also taught me a lot about both science and life, and was very patient and supportive with my struggles to adapt to my new home. Due to her mentorship, I began the second of my two most significant transformations. She taught me to be self-confident and self-compassionate, and that failing doesn't make one a failure. Soon after joining Dorothy's laboratory, I fell very ill. During my years of illness, Dorothy remained incredibly supportive both personally and professionally despite my inability to progress with my

project. There are no words I could ever say or write to express my profound gratitude for her support during that time.

I would also like to thank the other members of my committee: Cyrus Vaziri, Kevin Slep, and Nate Hathaway. Their support and guidance during my turbulent graduate career were invaluable, especially during my illness. I would not have been able to navigate through my difficulties without them.

I want to extend a huge thank you to John Cornett, the student services manager of my curriculum, and Jeff Sekelsky, the director of my curriculum during my graduate career. Both of them provided invaluable guidance and support for the administrative aspects of my degree as well as personal support during hard times. Their help was truly above and beyond.

I am also very grateful for the love and support from my family. They have always supported and encouraged my pursuit of education despite medical professionals claiming that I could never achieve my goals due my impairment. I have learned and grown so much due to my family, and I could never have been who I am today without them.

I have made many friends during my graduate career, all of which I am very fortunate to have made. My friends have supported me in my graduate work, but have been particularly supportive and influential in my personal development. They have brought great happiness into my life and I will forever be thankful for them.

I want to give a special thank you to the medical professionals that have helped me through my illness during the later half of my graduate career, in particular William Pendergraft, Addie Skow, and Cholena Erickson. They listened to me and supported

me in ways that other medical professionals did not, and they never gave up despite the countless dead-ends and setbacks we encountered.

Last but certainly not least, I would like to thank Bram Frohock and Amy Rydeen for providing the unnatural keto acid precursors used in my thesis work, and Tim Schwochert for performing the liquid chromatography-mass spectrometry.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

AND	Anderson promoter
CAT	Chloramphenicol Acetyl Transferase
ecTAT	*Escherichia coli* tyrosine aminotransferase
FRET	Förster resonance energy transfer
GFP	Superfolder Green Fluorescent Protein
MJ	*Methanococcus jannaschii*
mRNA	messenger Ribonucleic Acid
tRNA	Transfer Ribonucleic Acid
UAA	Unnatural Amino Acid
UKA	Unnatural Keto Acid

CHAPTER 1: INTRODUCTION

Unnatural amino acids (UAAs) are proteinogenic amino acids which have not been found in nature. The incorporation of UAAs into proteins provides a powerful and flexible method to site-selectively add novel catalytic activities and chemical properties to proteins in order to address important issues in human health, academic research, and industry. In the clinic, UAAs have been used to great effect for the creation of bispecific antibodies and antibody-drug conjugates for the treatment of cancer [1–7] and for improving the pharmacokinetic properties of therapeutic proteins [8]. In research, UAAs serve a wide host of roles. For example, UAAs have served as single amino acid fluorophores for cellular imaging [9–11]. UAAs also have been used as sites for simple and specific labelling of proteins through various biorthogonal chemistries [12–14], allowing these sites to become a myriad of different chemistries for diverse applications and providing alternative approaches for traditionally difficult experiments such as *in vitro* and *in vivo* single molecule FRET [15–18]. Crosslinking UAAs are another useful tool in research, having been used to study transient and difficult to capture protein interactions and manufacture highly customizable hydrogels [19–24]. Protein activity regulation can also be achieved through UAAs in various ways, although the primary method used so far is obstruction of the protein's active site or allosteric sites [25–30]. In addition to function regulation, enzyme properties such as rate, specificity or promiscuity, and tolerance to extreme conditions (e.g. heat and organic solvents) can be improved with the use of UAAs [31]. The incorporation of UAAs can even be used to

grant enzymes totally novel catalytic activity beyond what has ever been observed in nature[32]. The abilities to modify enzyme properties and create entirely new catalytic activities have great potential for industrial applications as well. These abilities provide a possible avenue to make difficult and expensive reactions widely employed in the synthesis of pharmaceuticals, plastics, cosmetics, and many other materials much more efficient and economical through the use of engineered enzymes, especially for those reactions requiring enantioselective chemistry [31,33–35].

The incorporation of UAAs site selectively into proteins is performed using orthogonal tRNA / tRNA synthetase pairs. These orthogonal pairs act in such a way that the orthogonal tRNA is only bound by its respective orthogonal tRNA synthetase, and likewise the orthogonal tRNA synthetase doesn't bind any of the twenty canonical amino acids. In this way, the orthogonal tRNA is selectively "charged" only by the orthogonal tRNA synthetase and only with UAAs. The orthogonal tRNA contains a specific anticodon which allows it to site specifically bind to the mRNA during translation, leading to site specific incorporation of whatever UAA is linked to the tRNA into the growing peptide chain. The general approach of this orthogonal tRNA / tRNA synthetase pair is how the naturally occurring amino acid pyrrolysine is encoded and incorporated into proteins within methanogenic archaea [36,37], and soon after the discovery of the pyrrolysine tRNA / tRNA synthetase pair they were repurposed for UAA incorporation [38]. Around the same time, the *Saccharomyces cerevisiae* glutamyl-tRNA synthetase pair and the *Methanococcus jannaschii* tyrosyl-tRNA synthetase pair were shown to be orthogonal in *Escherichia coli* and demonstrated to be effective systems for the site specific incorporation of UAAs as well [39–41]. These tRNA / tRNA synthetase pairs

typically use an anticodon complementary to the amber stop codon (genetically encoded as TAG), effectively “reassigning” that stop codon to instead incorporate UAAs [42]. An alternative approach is to use “frameshift suppression”, in which the coding sequence of a protein contains a quadruplet codon at the selected incorporation site instead of a triplet codon. Concomitantly, the orthogonal tRNA is made to contain a complementary quadruplet anticodon to match this site. This causes the incorporation of anything other than a UAA during translation to lead to a frameshift which usually causes early peptide termination [43]. A number of studies have utilized frameshift suppression instead of stop codon suppression due to its greater orthogonality (it doesn’t need to overwrite any of the preexisting triplet codons) and due to its ability to encode multiple UAAs using two or more quadruplet codons simultaneously. However, frameshift suppression remains less common in certain biological systems due to its reduced efficiency and protein yield [44–48]. The tRNA / tRNA synthetase pairs used for UAA incorporation are often inefficient (or totally incapable) of charging a particular UAA of interest without optimization and a large number of studies have been performed to expand the substrate specificity of various synthetases to accommodate more UAAs and to modify tRNAs to increase efficiency and orthogonality [49–54].

While a large body of work has focused on the creation and optimization of tRNA / tRNA synthetase pairs for UAA incorporation, a number of other factors can interfere with efficient UAA incorporation. Ribosomes can act as a limiter on the efficiency of UAA incorporation depending on the particular structure of the UAA and orthogonal tRNA. In particular, orthogonal tRNAs using a quadruplet codon are inefficiently processed by the ribosome. Mutant, orthogonal ribosomes have been generated which

can more efficiently decode quadruplet codons at the cost of translational fidelity. To prevent these less accurate ribosomes from translating the normal proteome, they only recognize mRNAs carrying dedicated orthogonal ribosome binding sites [55]. In the stop codon reassignment strategy, the orthogonal tRNA is constantly in competition with the translation termination factor which recognizes the amber stop codon. This competition reduces the efficiency at which the orthogonal tRNA can bind the mRNA and results in lower incorporation efficiency of the UAA. To address this, multiple strains of *Escherichia coli* and one strain of HEK293T cells have been engineered to either reduce the activity or totally remove the termination factor which recognizes the amber stop codon [42,56–58]. Another study accomplished a similar effect by generating an orthogonal ribosome with higher relative binding of orthogonal tRNAs compared to the amber codon termination factor [59]. A relatively uninvestigated limiting factor of UAA incorporation is the efficiency at which UAAs are taken up by cells. A small number of studies have shown that UAAs demonstrate varying propensities to be taken up by cells, with UAAs more closely resembling canonical amino acids being taken up the most efficiently [39,60]. Only one study has attempted to increase the uptake of UAAs through biological means. Mutation of the leucine binding protein in *Escherichia coli* cells was used to create a UAA specific cell membrane transporter [60]. Similarly, only one study has shown increased uptake through chemical alteration of UAAs, specifically esterification which lead to increased cellular uptake in HEK293T cells [61]. To date, no studies have investigated the intracellular stability of UAAs.

UAAs have been successfully incorporated into proteins within a wide range of biological systems. The most commonly used system is bacteria, with most studies

using *Escherichia coli* [40], but yeast has also been thoroughly used for studies involving UAA incorporation [62–64]. The use of UAAs in human cell lines has become increasingly more frequent, primarily being performed in HEK293 embryonic kidney cells but also in other lines such as HeLa cervical cells, Caco-2 colon cells, A549 lung cells, and HCN-A94 neuronal stem cells [65–69]. A number of non-human mammalian cells have shown successful UAA incorporation as well, including mouse primary neurons, IH3T3 mouse fibroblasts, and CHO K1 hamster ovarian cells [66,69,70]. *Xenopus laevis* oocytes were among the earliest eukaryotic systems to be used, followed by more complex organisms such as *Caenorhabditis elegans, Drosophila melanogaster*, *Mus musculus*, and eventually *Danio rerio* [46,71–74]. Of note, UAAs have be incorporated into proteins successfully in plants, but only in a single study using *Arabidopsis thaliana* [75]. Cell free expression systems, essentially crude cell lysates supplemented with micro and macro nutrients, offer an alternative approach that bypasses concerns of cellular uptake and cytotoxicity of UAAs, but are currently limited due to poor expression efficiency and scalability issues [76,77]. In addition to UAA incorporation in proteins within cellular (and cell derived) systems, UAAs have also been incorporated into viral systems. These viral systems have been used for studying structural and functional elements of viral proteins, creating replication deficient vaccine particles, and also are being applied to gene therapy viral vectors [17,78–81]

While UAAs have great potential for a myriad of uses in human health, research, and industry, UAA technology is currently limited by three major problems. First, the majority of studies performing UAA incorporation rely on overriding naturally occurring codons, usually the amber stop codon. When a codon is overridden, it prevents the

codon from functioning properly which leads to deleterious molecular behavior. In the case of the amber stop codon, open reading frames which are terminated by the amber stop codon no longer have a reliable stop signal which causes the transcription and translation of long stretches of genetic material downstream of the open reading frames. This uncontrolled transcription leads to compromised cell health and reduced protein yields [56]. Second, the uptake efficiency of UAAs remains poorly characterized. Currently, cells are routinely cultured in relatively high concentrations of UAAs (~1 mM) to ensure uptake, but this practice wastes large amounts of UAAs and often leads to compromised cell health and viability which may in turn reduce the production of the protein of interest. Third, the vast majority of UAAs are difficult and time consuming to chemically synthesize, which prohibits research laboratories from synthesizing their own UAAs and makes commercial UAAs very expensive. Some UAAs of potential research interest are even totally impossible to synthesize with current chemical synthesis technologies.

The ability of cells to biochemically synthesize their own UAAs *in vivo* would circumvent the later two limitations of current UAA technology. Previous work has shown that cells are capable of synthesizing UAAs from pre-existing cellular metabolites or from cheap, accessible chemical precursors that are efficiently taken up into cells [82,83] A relatively small set of studies have demonstrated that efficient UAA biosynthesizing enzymes can be evolved from naturally occurring enzymes involved in canonical amino acid metabolism. Phenylalanine ammonia lyase from various organisms has been shown to be capable of generating phenylalanine derived UAAs with much improved activity following rational mutagenesis [84–87]. The ß subunits of tryptophan synthase from

Pyrococcus furiosus and *Thermotoga maritima* demonstrated efficient generation of various tryptophan analogues after evolution via random mutagenesis [88]. While encouraging, these studies relied on low-throughput selection strategies which severely limited the number of mutant enzymes that could be investigated and limited the application of these approaches for accessible, on demand creation of UAA biosynthetic enzymes. In contrast, a multitude of studies have evolved orthogonal tRNA synthetases using high-throughput selection strategies that take advantage of UAA-incorporation dependent expression of selection markers [50,53,54]. It stands to reason that the same high-throughput methods used for the evolution of orthogonal tRNA synthetases could also be applied to the evolution of UAA biosynthetic enzymes. Applying these high-throughput methods to the evolution of UAA biosynthetic enzymes would allow rapid and very flexible creation of *in vivo* UAA biosynthesis strategies which would bypass concerns for UAA uptake and UAA cost and availability. Using an iterative approach, entire *in vivo* biosynthetic pathways for UAAs can be developed by evolving the final enzyme which produces the UAA of interest and then “walking backwards” up the pathway to either an endogenous metabolite or cheap, accessible precursor molecules as illustrated in Figure 1-1. To this end, this thesis describes a universally applicable framework by which large libraries of mutant UAA biosynthetic enzymes can be quickly and easily made to allow for truly high throughput biological selections based on UAA dependent expression of selection markers. As a proof-of-concept, a library of approximately one million mutants of the *Escherichia coli* aromatic amino acid aminotransferase (ecTAT) encoded by the *tyrB* gene was generated. The ecTAT enzyme catalyzes the reversible conversion of the tyrosine keto acid into the tyrosine

amino acid using a nitrogen donor (Figure 1-1). The ecTAT protein was chosen for mutagenesis due to it being well characterized and its activity on tyrosine, the base amino acid from which many UAAs in literature are derived. Tyrosine derived UAAs are also favorable targets due to their unnatural keto acid (UKA) biosynthetic precursors being quick, cheap, and easy to chemically synthesize. In addition to the *tyrB* library, genetic constructs for the UAA dependent expression of chloramphenicol acetyl transferase and GFP were created to serve as selection markers responsive to intracellular UAA abundance. These UAA dependent selection markers can be used to select for the upstream mutant ecTAT activity of converting UKA precursor molecules into UAAs. When cells are treated with a UKA, the production of these selection markers is dependent on the mutant ecTAT activity because their translation depends on the presence of UAA which in turn is dependent on the ability of the ecTAT mutant to convert the UKA precursor into the UAA (Figure 1-1). The dependence of the selection marker expression on UAA presence is achieved by including one or more amber stop codons within the coding sequence of the genes, as shown in Figure 1-2. These codons are only decoded by the translation termination factor and by the ectopically expressed orthogonal tRNA containing a CUA anticodon. A cognate orthogonal tRNA synthetase capable of charging the orthogonal tRNA with UAAs but not natural amino acids is also expressed. Naturally, the tRNA synthetase is only capable of charging the orthogonal tRNA with UAA when UAA is present within the cell. During translation, when the ribosome encounters the amber stop codon, the orthogonal tRNA binds the mRNA. If the orthogonal tRNA has not been charged, this leads to early termination of the selection marker protein because no amino acid with which to extend the peptide chain

is attached to the tRNA. Conversely, if the orthogonal tRNA is charged with UAA, the ribosome is capable of extending the peptide chain with that UAA when it reaches the amber stop codon resulting in the production of functional, full-length protein with the UAA site specifically incorporated. These genetic constructs for UAA dependent selection marker expression and the presented framework for large scale UAA biosynthetic enzyme mutant libraries should allow for the rapid evolution of biosynthetic enzymes for virtually any UAA of interest and greatly improve the accessibility of UAA technology.

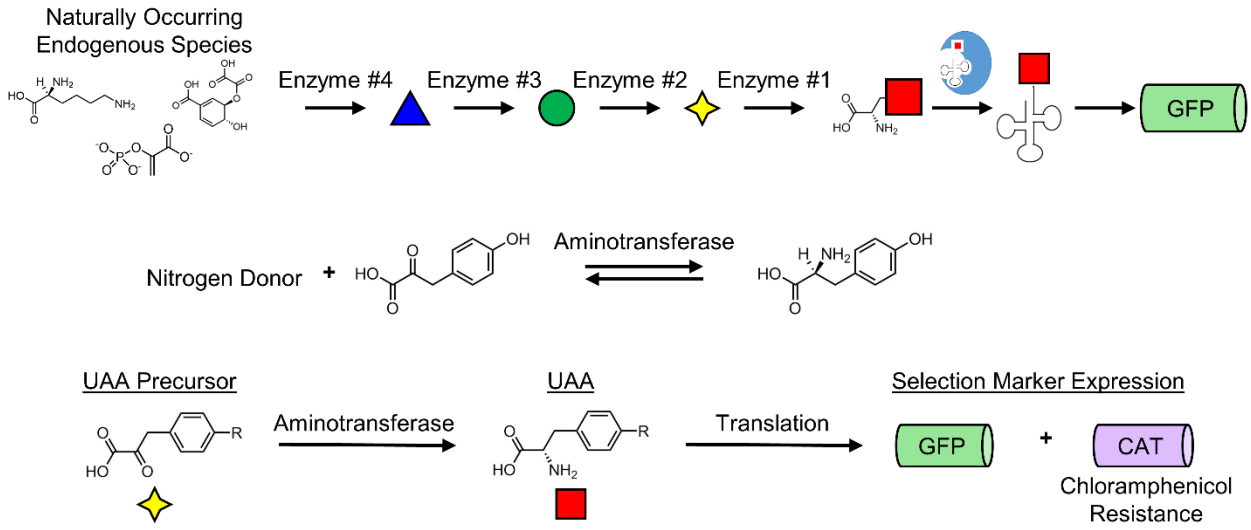

Figure 1-1: Biological selections can be used to evolve enzymes for in vivo unnatural amino acid biosynthesis.
Enzymes can be evolved to biosynthesize UAAs and the cellular expression of UAA biosynthetic enzymes allows cells to make UAAs from either accessible chemical precursors or endogenous metabolites. Iterative evolution of UAA biosynthetic enzymes can be used to generate entire pathways for the cellular production of UAAs (top). ecTAT catalyzes the conversion of the tyrosine keto acid into the tyrosine amino (middle) and was chosen as a proof of concept enzyme for producing a large mutant library (see text). By using selection markers whose translation is dependent on UAA presence, the upstream ecTAT mutant can be selected for efficient conversion of UKA precursor into UAA (bottom).

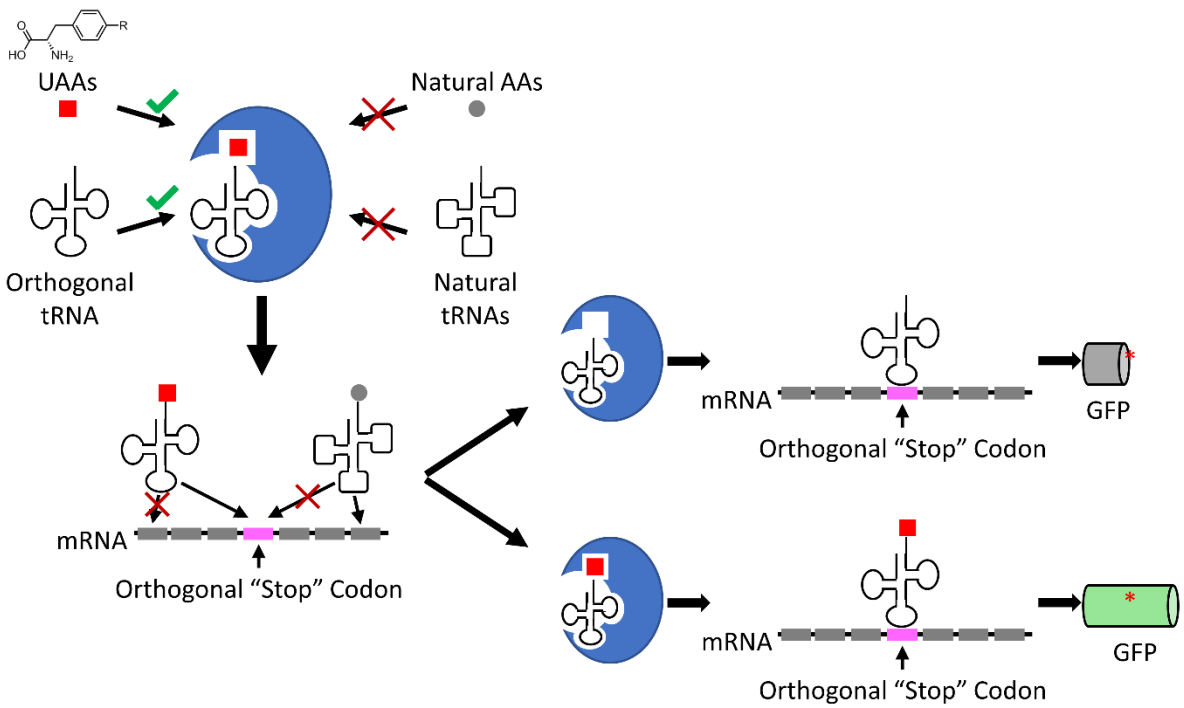

Figure 1-2: Selection marker production is dependent on UAA presence. Orthogonal tRNA / tRNA synthetase pairs only interact with each other. The orthogonal synthetase recognizes neither the natural 20 amino acids nor any of their natural tRNAs (top left). This lets the orthogonal tRNA be specifically charged only with UAA. The selection marker genes contain an amber stop codon which serves as the incorporation signal for the UAA. The orthogonal tRNA only recognizes this orthogonal stop codon and conversely this stop codon is not recognized by any of the natural tRNAs, making this site specifically encode only UAAs (bottom left). During translation, if the orthogonal tRNA which binds the orthogonal codon is uncharged, then translation will stop prematurely leading to a truncated protein product (top right). However, if the orthogonal tRNA is charged during translation, then the UAA is site specifically incorporated and functional, full length protein is made (bottom right). In this way, selection marker expression is dependent on the presence of UAA in the cell.

CHAPTER 2: GENERATION AND VALIDATION OF SELECTION SYSTEM

2.1 Introduction

The ability to perform high-throughput selections for UAA biosynthetic enzymes will greatly increase the availability of UAAs by allowing on-demand creation of UAA biosynthetic enzymes. However, a universal high-throughput selection framework requires a robust and highly flexible method for selecting mutants of interest. The selection methodology presented in this thesis comprises of two parts. First, the chloramphenicol resistance gene provides a primary selection during which only cells capable of producing UAA are able to grow. This growth based selection allows for virtually any size of library to be used. Second, the GFP reporter serves as a secondary selection marker by which improved UAA production compared to the wild type enzyme can be confirmed by simple fluorescence intensity measurement. In addition, the GFP contains a 6x His tag allowing facile collection and purification of the protein. The purified GFP can then be measured by mass spectrometry to directly confirm incorporation of specific UAAs. Both selection markers contain an amber stop codon, making their expression dependent on UAA presence as illustrated in Figure 1-2. Provided that an appropriate tRNA / tRNA synthetase pair is expressed alongside the selection markers, these selection markers will allow the selection of a biosynthetic enzyme for virtually any UAA from a library of millions, possibly billions, of candidate mutants.

2.2 Validation of UAA and UKA incorporation

Before the selection system could be made and optimized, the incorporation of UAAs and UKAs into selection marker proteins needed to be confirmed first to verify that the UAA incorporation machinery was functioning as expected. Successful production of a selection marker dependent on UAA incorporation when cells are treated with a UKA would confirm that all elements of the selection system pathway were functioning as expected including the conversion of UKA into UAA by ecTAT (see Figure 2-1A). Cells expressing the MJ tRNA / tRNA synthetase pair and a GFP with an early TAG codon were grown in the presence of various UAAs and UKAs. O-methyl-tyrosine UAA served as a positive control. After allowing GFP expression, clarified cell lysates were collected and measured for relative fluorescence intensity. The recorded fluorescence intensities are shown in Figure 2-1B. Both the o-methyl-tyrosine UAA and UKA showed high relative GFP fluorescence, showing that the machinery was working including ecTAT conversion of an UKA into an UAA. Based on the crystal structure of ecTAT, naphthyl UKA was not expected to be bound efficiently by the enzyme and thus was expected to be converted inefficiently. Surprisingly, however, both the naphthyl UAA and UKA demonstrated high levels of fluorescence suggesting that ecTAT is more promiscuous than anticipated. The biphenyl UAA showed high fluorescence but its UKA showed only about half as much fluorescence. Similarly, the trifluoro UAA demonstrated moderate-to-high amounts of fluorescence but its UKA showed almost undetectable amounts of fluorescence. This suggests that the biphenyl UKA and especially the trifluoro UKA are not being efficiently converted into UAAs by ecTAT and would serve as promising targets for evolving more efficient ecTAT mutants.

Expression of GFP mediated by the addition of UAA or UKA to the growth medium is highly suggestive of incorporation of the UAA or UKA into the GFP protein, but direct validation of incorporation can be achieved by measuring the GFP protein mass by mass spectrometry. The GFP protein expressed in these experiments includes a 6x His tag on the C-terminus, allowing simple purification using nickel or cobalt beads. Purified GFP taken from the lysates of cells grown in the presence of either o-methyl-tyrosine UKA or naphthyl UKA was analyzed by mass spectrometry, the spectra of which are shown in Figure 2-2. The measured masses of both GFP proteins were within one Dalton of the predicted masses, confirming that these UKAs were both converted into UAAs and incorporated into the GFP protein exactly once.

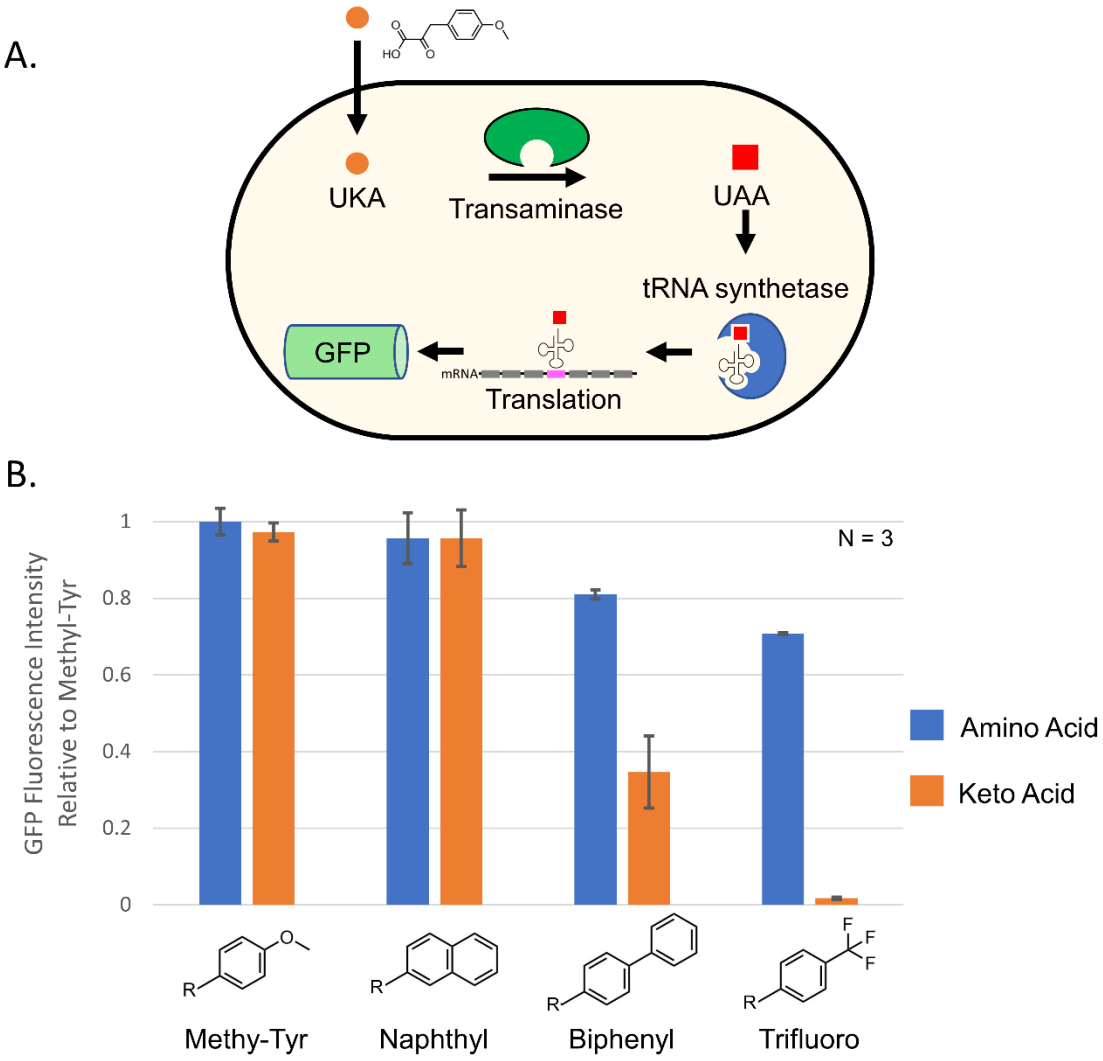

Figure 2-1: Functional incorporation of different UAAs and UKAs into GFP was confirmed.
(A) Incorporation of a UKA as a UAA relies on four major steps. The UKA must be taken up by the cell, converted into a UAA by ecTAT, be charged onto the orthogonal tRNA, and incorporated into the selection marker during translation. A failure at any one these steps prevents incorporation and thus selection marker expression. **(B)** Cells were treated with various UKAs and UAAs and the GFP fluorescence intensities of their lysates were measured.

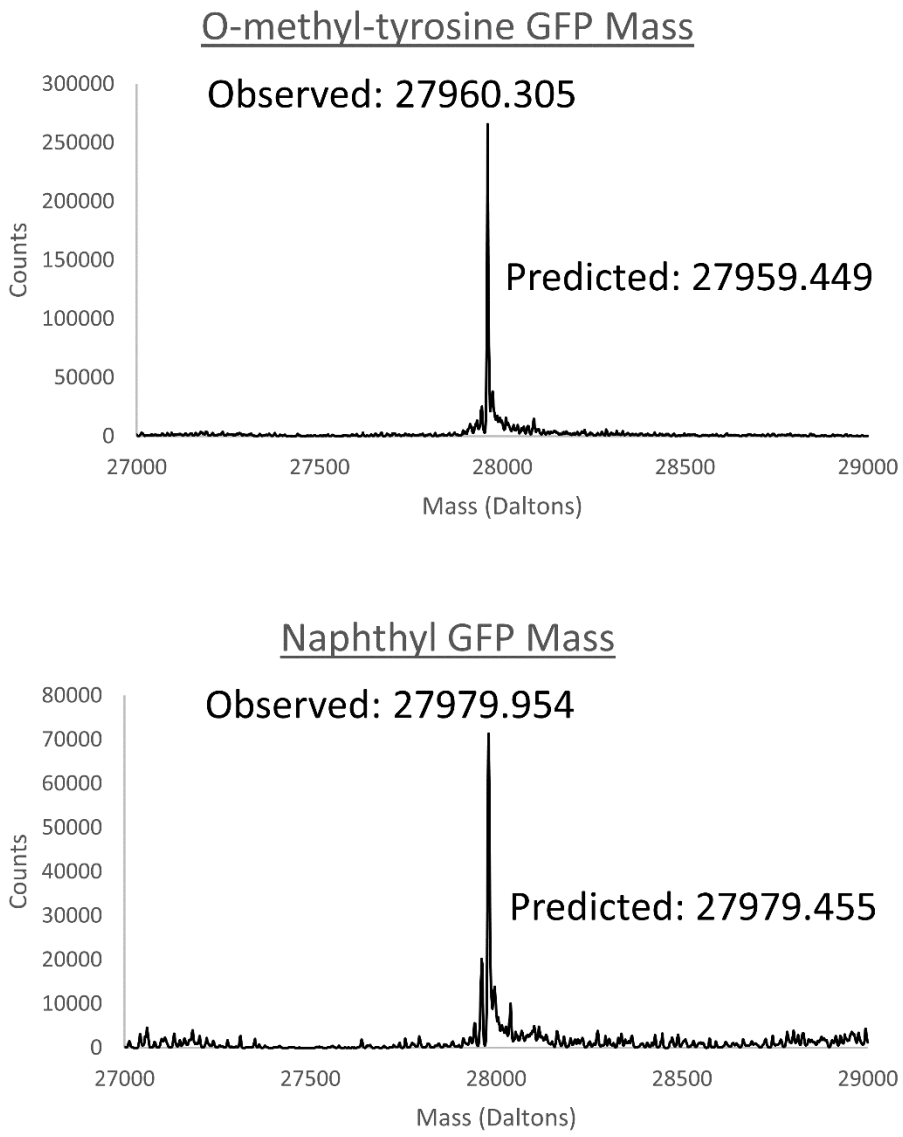

Figure 2-2: Transamination and Incorporation of UKAs into GFP was confirmed by mass spectrometry.
GFP from cells treated with o-methyl-tyrosine UKA or naphthyl UKA was purified and measured by mass spectrometry. The predicted and observed masses are presented along with the mass spectra.

2.3 Optimization of chloramphenicol resistance selection

With the incorporation of both UAAs and UKAs confirmed, the optimization of the selection system could begin. The first step was to verify and optimize the ability to select cells with high UAA abundance based on chloramphenicol resistance. Initially, the resistance of cells against chloramphenicol with and without UAA added was checked in liquid culture, but this method was later changed to colony formation assays. While not ideal, the initial liquid culture comparison revealed that the cells demonstrated a very high chloramphenicol resistance even in the absence of UAA as shown in Figure 2-3. This high basal level of chloramphenicol resistance would make selection using chloramphenicol resistance impossible and so the chloramphenicol genetic construct had to be modified to lower the basal level of chloramphenicol resistance. The first step was to migrate the CamR gene from its high copy number plasmid (at the time, pET-Duet) into the low copy number pULTRA. The second step was to introduce a second TAG codon into the CamR coding sequence to reduce any “read-through” (erroneous incorporation of natural amino acids at the TAG codon instead of UAAs or early termination). The third step was to swap the natural CamR promoter for a panel of promoters with varying strengths of expression. For this, five members of the Anderson promoter collection were chosen and used to create five different CamR constructs. The promoter members and their relative strength were: 23119 (100%), 23110 (33%), 23114 (10%), 23109 (4%), and 23113 (1%). From here on, these promoters will be referred to as “AND X” where X is their percentage of relative strength. These changes to the CamR gene are graphically summarized in Figure 2-4.

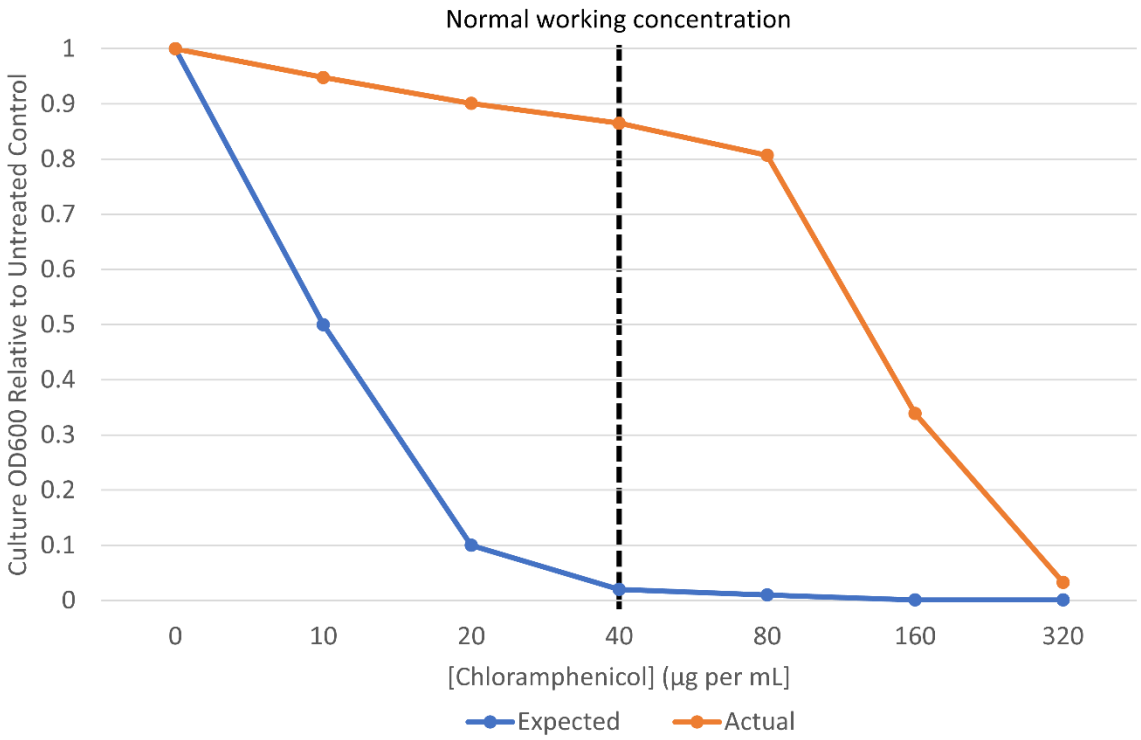

Figure 2-3: Cells demonstrated high levels of basal chloramphenicol resistance. Cell cultures grown overnight with varying levels of chloramphenicol had their OD600 measured. The blue “expected” line shows a typical cellular response to chloramphenicol when cells do not carry the chloramphenicol resistance construct. The dotted line shows the typical working concentration of chloramphenicol (40 µg per mL).

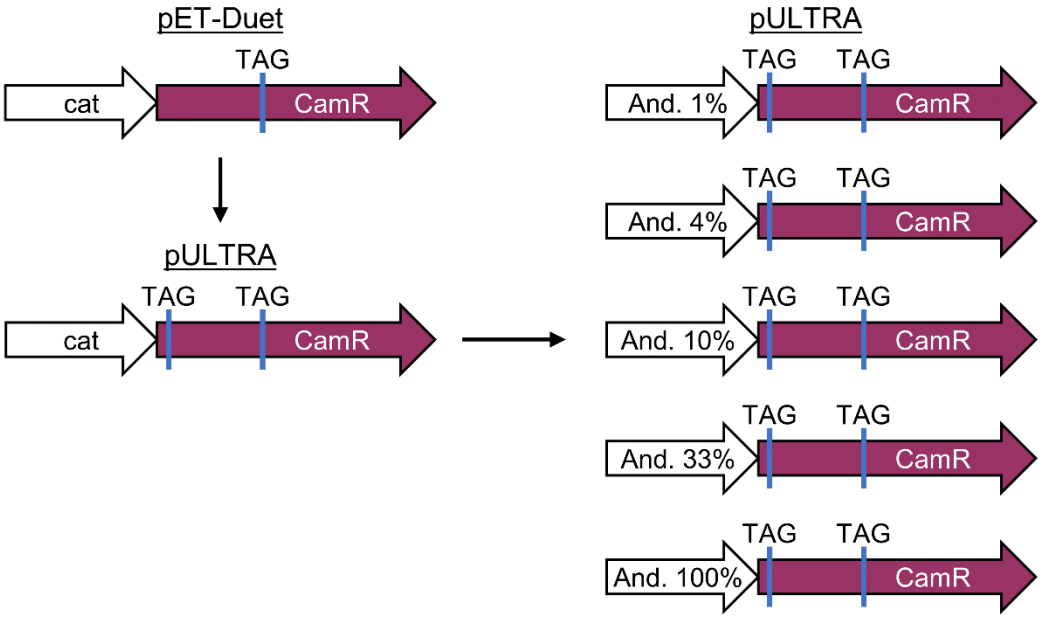

Figure 2-4: Chloramphenicol resistance gene and promoter were altered to reduce basal resistance.
The chloramphenicol resistance gene and promoter were moved from the high copy number plasmid pET-Duet to the low copy number plasmid pULTRA. In addition, a second amber stop codon was introduced near the beginning of the coding sequence to reduce readthrough (left). The natural promoter was replaced with five of the promoters from the Anderson promoter collection, producing five new constructs with promoters of varying expression strength (right).

The five new CamR constructs along with a negative control lacking the CamR gene were then tested for chloramphenicol resistance with and without o-methyl-tyrosine UAA using a colony formation assay. The results of this test are shown in Figure 2-5. AND 1% and AND 3% are not shown as they were indistinguishable from the negative control. Due to chloramphenicol's bacteriostatic nature, the sizes of colonies were measured instead of comparing the number of colonies between treatments. The negative control demonstrated total inhibition of growth even at low chloramphenicol concentrations. The AND 10% showed no difference in colony sizes between the untreated and treated samples except in the absence of chloramphenicol, making the AND 10% unsuitable for selections. The AND 33% construct demonstrated a modest difference in colony size in the intermediate chloramphenicol concentrations. AND 100% showed resistance in both the untreated and treated cells at lower chloramphenicol concentrations, but only the treated cells showed strong resistance at high chloramphenicol concentrations. The AND 100% promoter was chosen for the final CamR construct as it demonstrated the greatest potential for separating cells proficient in UAA incorporation from cells that are not proficient in UAA incorporation based on its difference in colony sizes at high chloramphenicol concentrations. The final pULTRA construct containing the AND 100% promoter is shown in Figure 2-6.

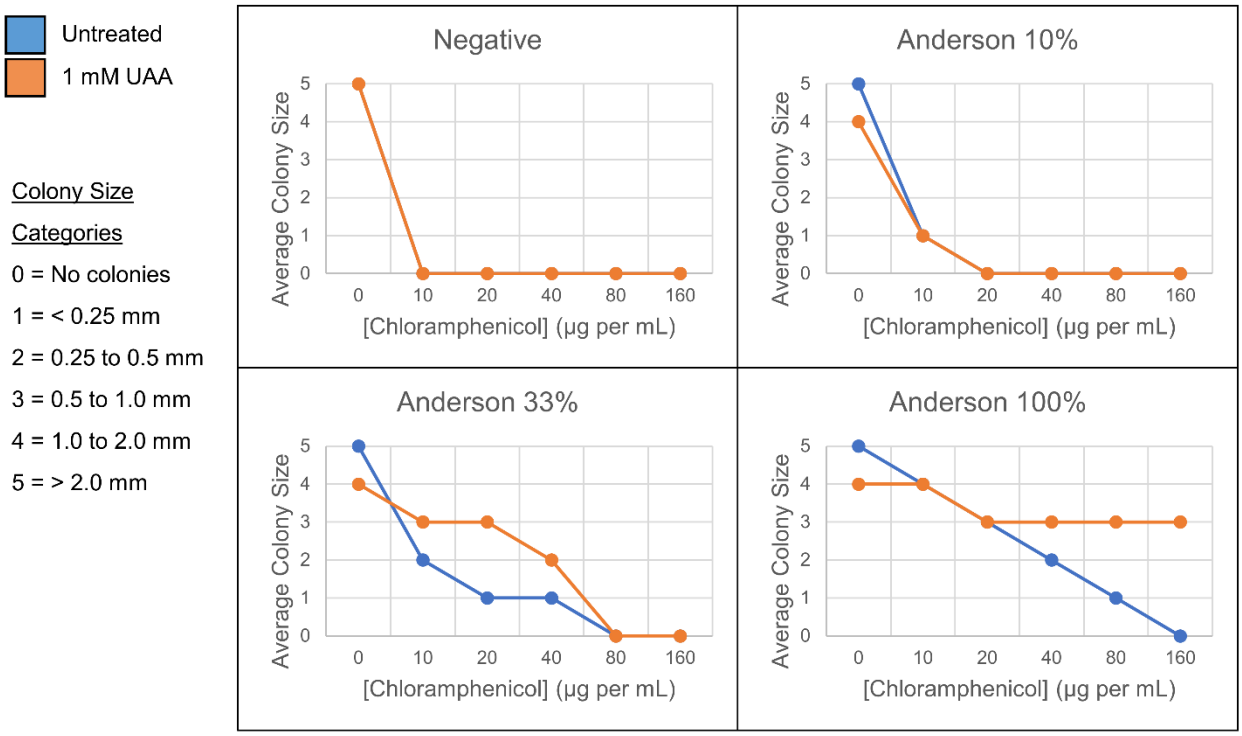

Figure 2-5: Altered chloramphenicol resistance constructs demonstrated reduced basal resistance.
Cells carrying the new chloramphenicol resistance constructs were plated onto LB agar plates with or without o-methyl-tyrosine UAA added. In addition, a negative control construct lacking the chloramphenicol resistance gene was also plated. Plates contained varying amounts of chloramphenicol ranging from 0 to 160 µg per mL. Colony size was measured instead of colony number due to chloramphenicol's bacteriostatic nature. The AND 1% and AND 3% constructs are not shown because they were indistinguishable from the negative control.

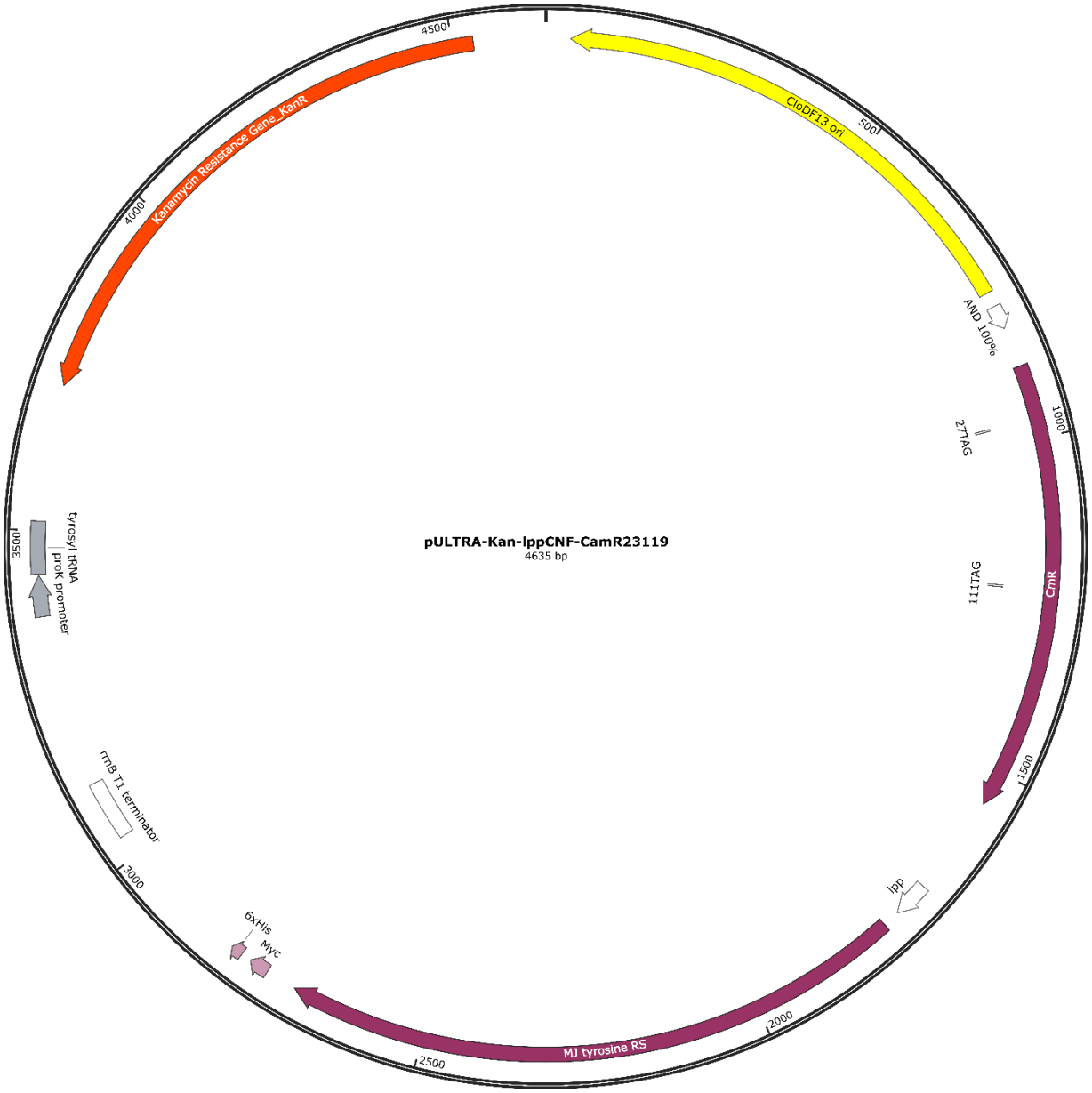

Figure 2-6: Map of final pULTRA construct.
The final pULTRA plasmid contained the chloramphenicol acetyltransferase gene driven by the AND 100% promoter, the orthogonal MJ tyrosine tRNA synthetase gene driven by the lpp promoter, and the orthogonal tRNA driven by the proK promoter.

2.4 Optimization of GFP expression

GFP expression is a quick and convenient method of confirming UAA incorporation and can be used for secondary verification of a mutant ecTAT having greater catalytic activity than the wild type ecTAT. For these purposes, having the highest possible expression of GFP is preferred. To achieve high GFP expression, the GFP construct was cloned into the pET-3a high copy number vector and was driven by the lpp promoter, a very strong constitutive promoter. However, the AND 100% is also a very strong promoter and it was hypothesized that the AND 100% promoter may facilitate greater GFP abundance compared to the lpp promoter. To compare, an identical pET-3a construct was created using mutagenic PCR. Primers flanking the GFP gene and containing the adjacent NotI and SphI restriction sites were designed such that amplification would lead to the replacement of the lpp promoter by the AND 100% promoter. Specifically, the upstream promoter had the AND 100% sequence and the adjacent NotI site appended to it as a 5' end flap. This mutagenesis approach is illustrated in Figure 2-7A. The resulting PCR product was then cloned back into the vector using standard restriction digest and ligation cloning. The AND 100% and lpp constructs were then compared for GFP abundance by growing cells in the presence of o-methyl-tyrosine and measuring relative cell lysate fluorescence. The results are shown in Figure 2-7B. The AND 100% construct showed approximately double the fluorescence signal of the lpp. Based on this, the AND 100% construct was used going forward. A map of the pET-3a plasmid containing the AND 100% GFP construct is shown in Figure 2-8.

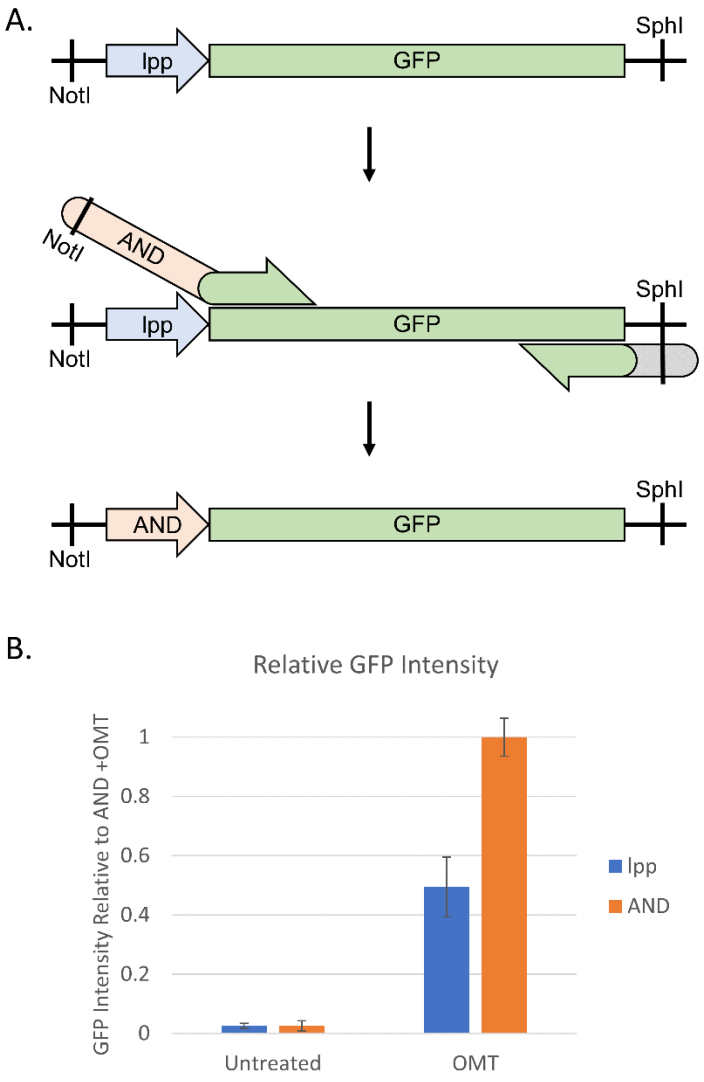

Figure 2-7: AND 100% promoter mediated greater GFP fluorescence than lpp promoter.
(A) The lpp promoter driving the GFP gene was swapped for the AND 100% promoter using PCR. The left primer annealed to the 5' end of the GFP gene immediately downstream of the lpp promoter. The left primer contained the AND 100% promoter sequence appended to its 5' end along with the NotI restriction site sequence. The right primer annealed to the 3' end of the GFP gene and included the SphI restriction site sequence appended to its 5' end. The primer pair generated a GFP gene with the AND 100% promoter instead of the lpp promoter which could then be easily cloned back into the vector using the NotI and SphI sites. **(B)** Cells carrying either the lpp GFP or AND 100% GFP construct were grown in the absence or presence of o-methyl-tyrosine UAA. Cell lysates were collected and the GFP fluorescence of the lysates was measured.

Figure 2-8: Map of final pET-3a construct.
The final pET-3a plasmid contained the GFP gene driven by the AND 100% promoter and the ecTAT gene driven by the lpp promoter.

CHAPTER 3: PRODUCTION OF BIOSYNTHETIC ENYZME MUTANT LIBRARY

3.1 Selection of mutation sites

The first step following the optimization of the selection system was choosing which residues of the ecTAT protein to target for mutagenesis. This was done by analyzing a crystal structure of the *E. coli* aspartate aminotransferase (PDB entry 1AHG). The aspartate aminotransferase crystal structure was chosen because aspartate aminotransferase and ecTAT are incredibly similar in structure (atomic RMSD of 1.615 when aligned in PyMOL), are capable of catalyzing each other's reactions [89], and at the time of analysis no ecTAT crystal structure with a bound intermediate was available but such a crystal structure for aspartate aminotransferase was available. Based on the structural and catalytic similarity, the aspartate aminotransferase crystal structure should serve as a functional proxy for the exTAT crystal structure. Four residues in the substrate binding pocket were chosen based on their proximity and theoretical ability to hydrogen bond with the side chain hydroxyl of the tyrosine intermediate. These residues were aspartic acid 11, asparagine 133, arginine 281, and serine 285. The binding pocket and selected residues are shown in Figure 3-1. In addition, the location of the mutagenesis sites within the *tyrB* gene are shown.

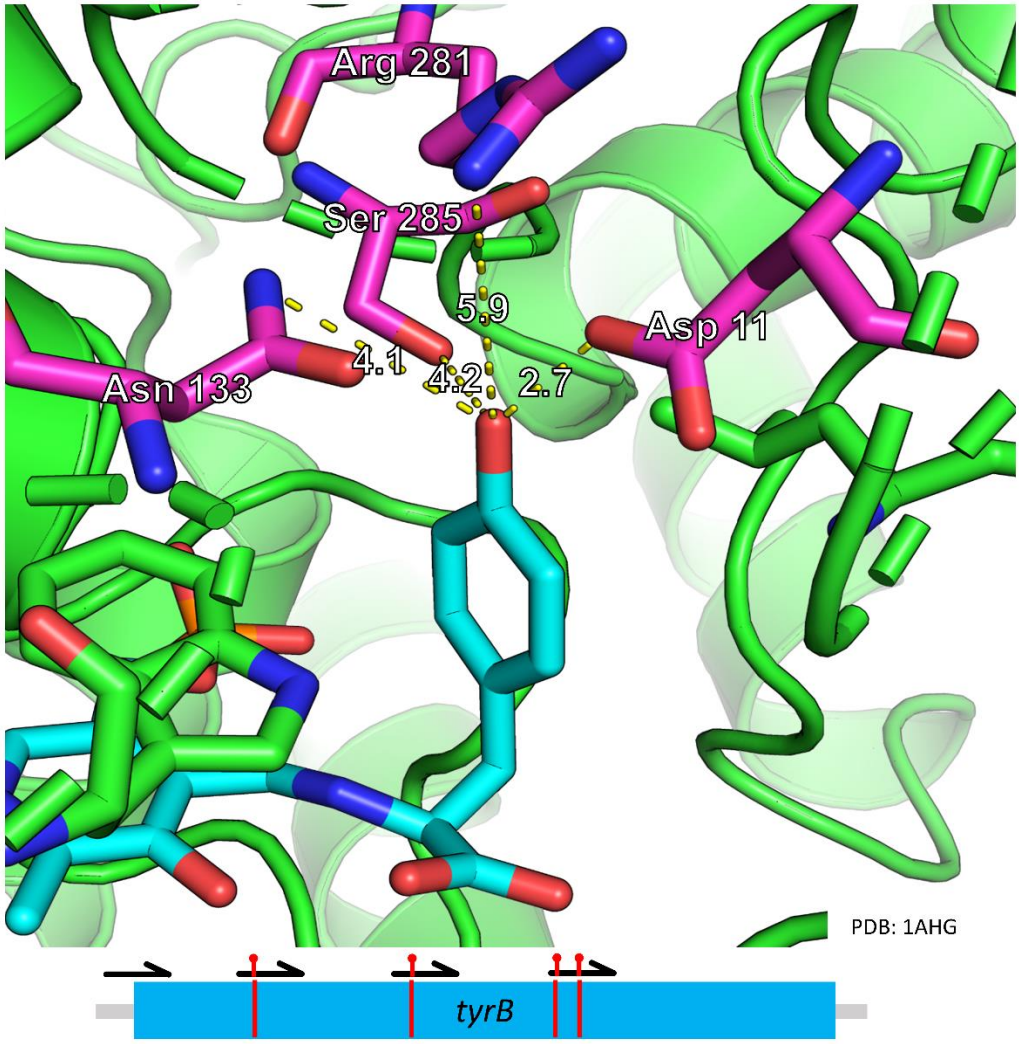

Figure 3-1: Four residues were selected for mutagenesis based on proximity to intermediate's side chain hydroxyl group.
The binding pocket of the Escherichia coli aspartate aminotransferase is shown in green with the bound PLP-tyrosine intermediate shown in cyan. The four magenta residues were selected for mutagenesis based on their proximity to the hydroxyl group of the PLP-tyrosine intermediate's side chain. Yellow dashed lines with numbers are measurements of distance between the selected residues and the PLP-tyrosine intermediate's side chain hydroxyl in angstroms. At the bottom is a cartoon of the tyrB gene showing where the four residues are encoded within the gene (red vertical lines) as well as the approximate locations of mutagenic primer annealing during PCR (black arrows).

3.2 Optimization of library construction

Construction of the *tyrB* library was performed in two steps. First, four fragments containing random mutations were generated via independent PCRs. These PCRs used primers containing homology with the ends of adjacent fragments to allow future overlap PCR resulting in the four fragments being joined together into one complete *tyrB* mutant gene. These initial primers also contained NNK codons at the selected residue positions in order to mutate those residues randomly. NNK codons are codons where the first two nucleotides can be any of the four nucleotides and the third nucleotide can be either guanine or thymine. NNK codons were chosen because they allow for encoding all 20 natural amino acids but require less mutants in the library to reach high statistical coverage compared to totally random codons. The second step after generation of the four fragments is to join them together using overlap PCR which is illustrated in Figure 3-2A. In overlap PCR, the homologous ends of the fragments can anneal to each other, allowing the fragments to serve as primers for one another. During extension, annealed adjacent fragments are permanently spliced together. Repeated cycles of melting, annealing, and extending lead to all four fragments being assembled together into a single functional *tyrB* gene which can then be amplified using standard PCR. Bulk mixture and assembly of the fragments by overlap PCR results in a random joining of adjacent fragments, leading to a completely random mutagenesis of the selected four residues.

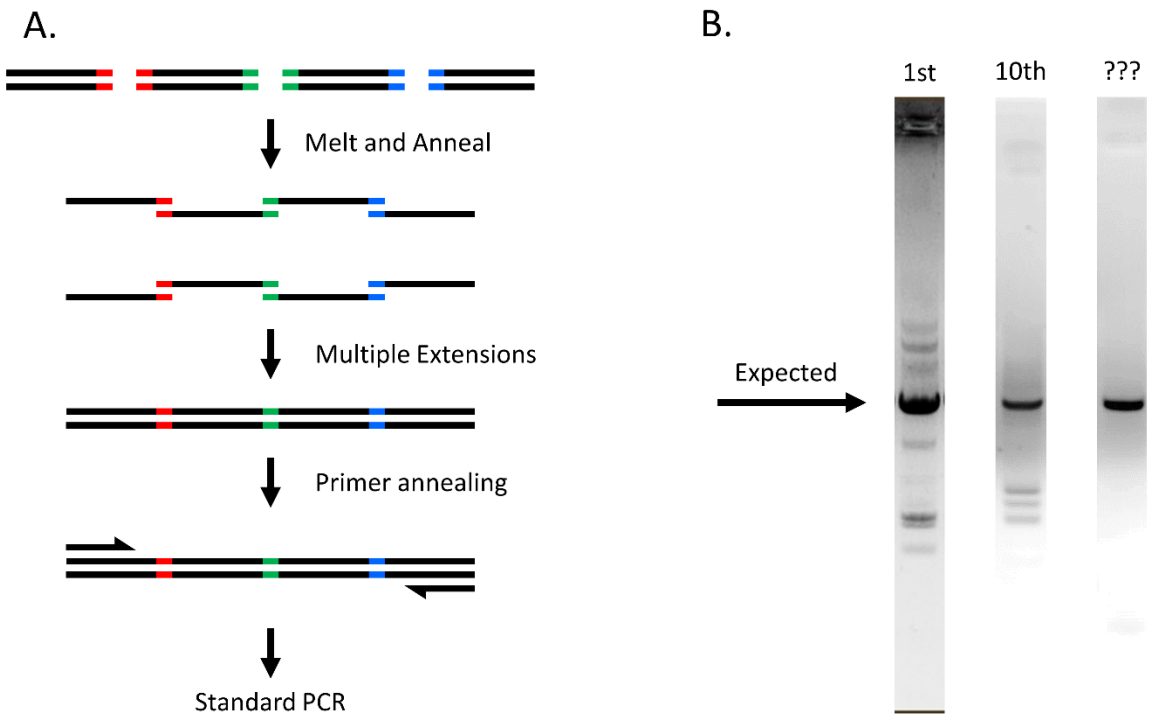

Figure 3-2: Efficient and specific library production was achieved through optimization of mutagenic overlap PCR.
(A) Individual mutagenized fragments of the tyrB gene are stitched together using overlap PCR. Each fragment (black horizontal lines represent each strand of DNA) has homology at its ends (colored blocks) that is shared with adjacent fragments. The fragments are mixed together and cycled through melting, annealing, and extension steps as done in standard PCR. During annealing, some homologous ends will anneal with the ends of adjacent homologous fragments instead of annealing back to a totally complementary strand. Then, these annealed ends will serve as primers during extension, causing the two fragments to become one entirely double stranded fragment of DNA comprising of both original sequences. Through repeated cycles, all fragments will be stitched together in this manner to form full length, functional tyrB genes containing a random assortment of mutations at the four selected sites. These full length genes can then be amplified using standard PCR. **(B)** Gels lanes showing the progression of tyrB gene amplification as optimizations are made to the protocol. The arrow shows where the tyrB gene band should appear on the gel based on DNA length. Bands that are not at the height indicated by the arrow are non-specific products.

While overlap PCR is a relatively quick and simple approach for assembling the *tyrB* mutant fragments together, it is prone to low yields and the generation of unintended products if not well optimized. A plethora of conditions and optimizations were tested in order to develop a consistent protocol with high yields and high specificity. The conditions tested were:

- Single pot" reaction for both assembly and amplification compared to sequential assembly and amplification in separate reactions
- Assembly and amplification annealing temperatures from 50 to 72 degrees
- Assembly and amplification extension times from 30 seconds per kilobase pair to 120 seconds per kilobase pair
- Assembly and amplification initial denaturation time from zero seconds to 300 seconds
- Assembly fragment concentrations at 1 nM, 10 nM, and 100 nM
- Amplification primers sequences
- Amplification primers concentrations at 200 nM, 500 nM, and 1 µM
- Amplification template concentration at 1 µL, 2 µL, 5 µL, and 10 µL of raw assembly reaction
- Flanking sequence length past restriction sites at 6 bp, 50 bp, and 300 bp
- Phusion high fidelity polymerase compared to Q5 high fidelity polymerase
- Phusion HF buffer compared to Phusion GC buffer
- Dimethyl sulfoxide concentration from 0% to 5%
- Betaine concentration from 0 M to 2.5 M
- Final extension time from 0 seconds to 600 seconds

- Amplification and assembly cycle numbers from 15 to 50

The progression of the library construction optimization is shown in Figure 3-2B. Initially, the library generation had high yield but a large number of off-target products. After extensive optimization, the yield was moderately reduced but the library production was entirely specific. The final conditions are detailed in the Methods chapter. The single most important optimization was the switch from a single pot reaction to separate assembly and amplification reactions. This change drastically improved the specificity in exchange for only a modest decrease in yield. The annealing temperature played only a moderate role in specificity, but the initial fragment concentration for assembly played a major role in specificity with the lowest concentration giving much greater specificity with slightly decreased yield. As expected, greater cycle numbers resulted in greater yield and specificity wasn't compromised until cycle numbers increased over 40 for both the assembly and amplification reactions. All other conditions resulted in little to no changes in the specificity or yield of the reactions.

3.3 Optimization of library ligation

Once the library production was optimized, the library needed to ligated into the pET-3a vector and transformed into cells. This was accomplished using traditional restriction digest and ligation cloning using the NotI and SphI sites flanking the *tyrB* gene site on the pET-3a vector. One complication of this approach was the short end fragments that were cleaved from the library PCR product during the restriction digest. These fragments were long enough to be retained during purification of the PCR product and would contain the complementary overhangs that matched the cleaved ends of both the PCR product and cleaved vector. It was hypothesized that these small

contaminating fragments would interfere with the ligation of the PCR product into the vector by being ligated to the ends of the PCR product and vector fragments, effectively capping their ends making them unable to participate in the desired ligation reaction. This hypothesis is illustrated in Figure 3-3. The ends of the PCR product could not be shortened because primers targeting those sites were less efficient and less specific. Thus, a method for determining the impact of these small fragments on ligation and transformation efficiency was required. This method is illustrated in Figure 3-4. The library assembly was performed using two sets of amplification primers: one normal and one that also had biotin covalently linked to the 5’ end. The two libraries were digested and purified as normal. Then, both libraries were mixed with ferrous beads conjugated to streptavidin. This was done in order to bind any library fragments made with biotinylated primers that did not have both ends cleaved off as well as the small end fragments. The beads were pulled down by magnet and the supernatant which should contain only the fully cleaved library was taken and purified. The purified supernatant was then ligated into the pET-3a vector as normal and transformed into NEB-10b cells using electroporation. The transformed cells were plated onto LB agar plates and the number of resulting colonies was counted and compared between the biotinylated and unbiotinylated libraries. The unbiotinylated library showed an average colony count of 28 per plate while the biotinylated library showed an average colony count of 84 per plate (a three-fold increase). This suggests that the contaminating end fragments reduce ligation and transformation efficiency and should be removed during cellular library production.

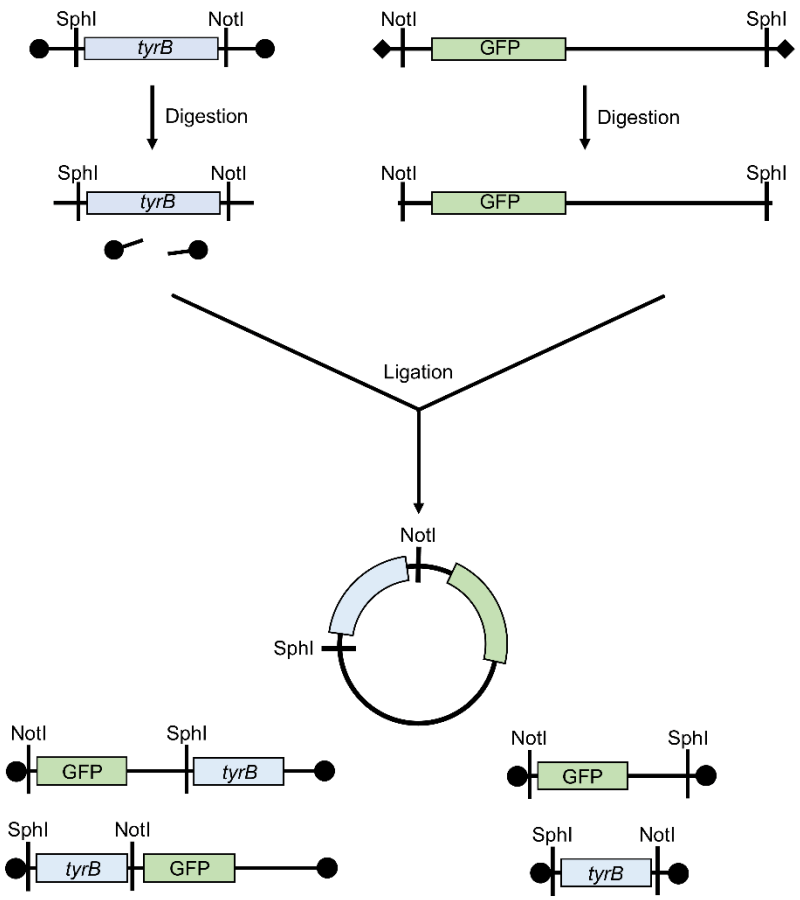

Figure 3-3: Contaminating cleavage fragments may interfere with ligation reaction.
The tyrB gene library (top left) is cloned into the vector (top right) using traditional restriction digest and ligation cloning. The first step is to cleave off the ends of tyrB library PCR product using the NotI and SphI restriction enzymes. The vector is cut with the same enzymes. The fragment produced from the vector cleavage was dephosphorylated and thus shouldn't interfere with ligation. However, the ends of the PCR product could not be dephosphorylated and were large enough to remain after PCR cleanup. Thus, the ends contaminated the ligation reaction. It was hypothesized that these contaminating end fragments would ligate to the vector fragment or back onto the PCR product and prevent full ligation of the PCR product into the vector. Theoretically, in addition to the desired circular plasmid product, a mixture of four possible linear products "capped" with the contaminating end fragments would also be produced and lead to reduced yield of the desired plasmid.

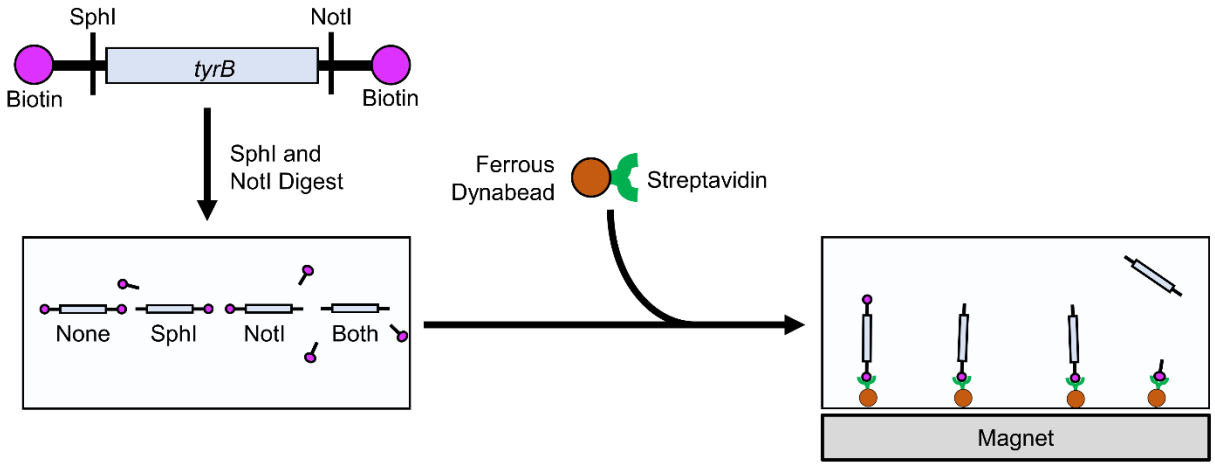

Figure 3-4: Removal of contaminating fragments by streptavidin conjugated bead pulldown.
In order to remove the contaminating end fragments, the tyrB library PCR was performed with primers having biotin covalently linked to their 5' ends (top left). The library was then digested, theoretically producing a mixture of undigested, partially digested, and fully digested tyrB products along with the contaminating end pieces. The restriction digest product was mixed with a molar excess of ferrous streptavidin beads. Any DNA with biotin still attached after the digest was bound by the ferrous streptavidin beads and then pulled down using a magnet. This effectively removes any incompletely digested tyrB fragments as well as any contaminating end fragments, leaving only the fully digested tyrB in the supernatant.

3.4 Verification of library diversity in cells

After the library production protocol was optimized, the ability to transform and recover mutants from cells needed to be confirmed. A large preparation of the *tyrB* library was generated and ligated into vector as described in the previous sections. The plasmid library was transformed into NEB-10b cells using electroporation in ten replicates and the transformed cells were plated onto separate LB agar plates. The total number of transformants across all transformations was estimated by counting the number of colonies on the LB agar plates and multiplying by the dilution factor. The estimated total was 2,400,000 transformants. GLUE is a program for predicting the fraction of all possible mutants that is contained in a mutant library based on the number of possible mutants and the number of mutants contained within a library [90]. The GLUE webserver was used to predict the *tyrB* library coverage, which returned an estimated 90.93% library coverage. To confirm mutation of the *tyrB* gene sequence, seven colonies were selected, grown, and had their plasmid DNA collected using the Qiagen QIAprep Spin Miniprep Kit. The purified plasmids were sent for sequencing to GENEWIZ (now Azenta). Table 3-1 shows the residues encoded by the mutant sequences at the four selected mutagenesis sites as they compare to the wild type sequence. The sequencing demonstrated that the seven mutants had variations at all four residues, indicative of successful mutagenesis and transformation of the library. Of note, the residues encoded by the sequences recovered seem to have biases to particular amino acids, most notably the aspartic acid 11 residue being primarily mutated to serine or proline. These biases may be due to non-uniform sequence

distribution in the mutagenic primers or could be due to biological effects such as certain residues granting growth advantages.

Wild Type	**GAC (Asp 11)**	**AAC (Asn 133)**	**CGC (Arg 281)**	**TCC (Ser 285)**
Mutant #1	TCT (Ser)	AGG (Arg)	GTT (Val)	TCC (Ser)
Mutant #2	TGT (Ser)	TGT (Cys)	GGG (Gly)	TAT (Tyr)
Mutant #3	CCC (Pro)	AAC (Asn)	TGG (Trp)	GGG (Gly)
Mutant #4	TGT (Ser)	GGT (Gly)	GGT (Gly)	GAT (Asp)
Mutant #5	TCT (Ser)	CAT (His)	TGT (Cys)	AAG (Lys)
Mutant #6	CCT (Pro)	GGT (Gly)	GTG (Val)	TCG (Ser)
Mutant #7	TTT (Phe)	TTG (Leu)	GGG (Gly)	AAG (Lys)

Table 3-1: Transformed library demonstrated diversity in recovered mutants. Seven colonies of the transformed tyrB library were picked and sequenced. The wild type tyrB sequence at the four codons encoding the four mutated residues is shown at the top along with the encoded amino acid in parentheses. Mutants are listed below, likewise with the nucleotide sequence at the four mutation sites and the corresponding encoded amino acid in parentheses.

CHAPTER 4: DISCUSSION

4.1 Discussion

UAAs are powerful and flexible tools for a wide host of applications spanning medicine, research, and industry. To date, they have been used to fight cancer, improve the pharmacokinetic properties of therapeutic proteins, and have been used for a myriad of purposes in scientific research. UAAs have been successfully incorporated into functional proteins in a wide host of systems spanning from prokaryotic cells to mammals. However, despite their success over the past two decades, UAAs still have not become nearly as ubiquitous as they could due to prohibitive costs and lack of availability. Previous studies have sought to evolve enzymes for the biosynthesis of UAAs, but these studies have used low through-put and laborious methods which severely limit their ability to find efficient mutant enzymes. This dissertation sought to overcome these limitations of UAAs by creating a large proof-of-concept mutant library of amino acid biosynthetic enzymes from which mutants could be selected to biosynthesize any given tyrosine derived UAA *in vivo*.

First, the ability of the cellular system to convert UKAs into UAAs and incorporate them into functional proteins was validated through UKA dependent GFP expression. Differential GFP production between the biphenyl and trifluoro UKAs and UAAs suggests that the wild type ecTAT is not completely promiscuous and that these UKAs may serve as favorable targets for expanding the substrate specificity of the enzyme. The high basal chloramphenicol resistance granted by the initial genetic constructs was

tuned down by transferring the chloramphenicol resistance gene to a low copy plasmid, introducing a second TAG stop codon to limit read-through, and replacing the natural chloramphenicol resistance promoter with promoters from the Anderson promoter collection. These steps and their converse are generally applicable for tuning antibiotic resistance and may be useful for the optimization of other selection systems. Similarly, replacement of the GFP gene's lpp promoter with the AND 100% lead to greater fluorescence, suggesting that the AND 100% promoter could be suitable as an even higher expression alternative to the very ubiquitous lpp promoter.

After the selection system was finalized, the protocol for the mutant library construction was optimized across numerous parameters. The most notable of these optimizations was using an approach that separated the fragment assembly and library amplification steps into two separate reactions. This is likely due to the presence of primers in the assembly reaction causing the amplification of incompletely assembled fragments or possibly serving as bridge oligos which transiently hold non-adjacent fragments together due to non-specific interactions. The initial fragment concentrations and the annealing temperatures were the only other two major factors influencing amplification specificity, with annealing temperatures showing a relatively minor influence. All other optimization parameters showed little to no improvement on specificity or yield. Which condition optimizations are most important will likely very drastically from system to system based on gene and primer sequences, but the separation of the assembly and amplification reactions is likely universally superior to a single pot approach. Following production of the library, the library PCR product needed to be efficiently ligated into the vector for transformation. It was hypothesized that small

contaminating end fragments left over from restriction digest of the library PCR product would interfere with the ligation by capping the ends of the substrates. Removal of these end fragments by streptavidin conjugated beads lead to a three fold increase in transformants, suggesting that small fragments sharing overhangs with ligation inserts or vectors can reduce ligation efficiency and should be removed or avoided when high ligation efficiency is required. Finally, transformation and recovery of the library from cells demonstrated that the mutagenesis of the *tyrB* gene was successful and maintained *in vitro*. Confirmation of this cellular library serves as evidence for the feasibility of the generation of large mutant libraries of UAA biosynthetic enzymes on the order of millions of mutants, far larger than libraries previously generated.

The work presented in this thesis serves as the foundation for a universal framework by which high throughput selections for UAA biosynthetic enzymes can be performed quickly, cheaply, and easily. Future efforts to generate UAA biosynthetic enzymes can follow the methods presented here-in for generating a cellular library of mutants which then allows the discovery of mutants of interest with simple, traditional biological selection using chloramphenicol. The GFP construct allows confirmation of selected mutants by direct comparison of fluorescence between selected mutants and the wild type enzyme. The GFP protein also can be quickly and easily pulled down via its 6x His tag for confirmation of UAA incorporation using mass spectrometry. These elements make selection and confirmation of mutant UAA biosynthetic enzymes quick and easy, allowing the creation of tailored UAA biosynthetic enzymes on demand for virtually any desired application. The ability to rapidly make on demand UAA biosynthetic enzymes will vastly increase the availability of UAA technology in general

by making UAA acquisition much cheaper and by allowing the creation of UAAs that are not amenable to purely chemical synthesis. This framework is not specific to any particular enzyme or organism, and can be applied to any process which leads to the accumulation of intracellular UAA whether that process be enzymatic or otherwise. The rapid and facile nature of the framework also lends itself to iterative approaches by which upstream enzymes of a particular UAA precursor can be evolved sequentially to produce an entire biosynthetic pathway.

CHAPTER 5: METHODS

5.1 Standard PCR amplification

Unless otherwise noted, all PCRs were performed under the following conditions. Reactions were 50 µL in volume with a final primer concentration of 500 nm (per primer), 200 nm of dNTPs (each) (NEB, Catalogue: N0447S), 1X Phusion HF buffer (NEB, Catalogue: B0518S), 1 unit Phusion HF polymerase (NEB, Catalogue: M0530S), and 1 ng of DNA template. The reactions were heated to 98 degrees for 45 seconds then cycled forty times through the following steps: 98 degrees for 15 seconds, 60 degrees for 30 seconds, and 72 degrees for 120 seconds. A final extension step of 72 degrees for 300 seconds was also performed.

5.2 Library generation

5.2.1 Library overlap PCR

First, four fragments of the *tyrB* gene were individually amplified using mutagenic primers following the standard PCR amplification conditions (primer sequences given below). These fragments were purified using the Qiagen QIAquick PCR Purification Kit (Qiagen, Catalogue: 28104) following the manufacturer's protocol, except that DNAs were eluted in 50 µL of water. Then, these four fragments were combined into a standard PCR reaction at equal molar ratios totaling a combined mass of 7.0 ng (1.0 ng for fragment #1, 1.5 ng for fragment #2, 2.0 ng for fragment #3, and 2.5 ng for fragment #4) except that no primers were added. After the completion of this PCR, 5 µL of the reaction was immediately used as template in a secondary PCR, this time with primers

added which amplify the entire library construct. Other than using the first round PCR as template, the secondary PCR followed the standard PCR methodology. Once the secondary PCR completed cycling, it was immediately purified using the Qiagen QIAquick PCR Purification Kit following the manufacturer's protocol, except the DNA was eluted in 50 µL of water.

Fragment #1 primers:

Forward: CGACTTAAGCATTATGCGGCCGC

Reverse: GCCAGCGTAGGCGTCAACTTTTTG

Fragment #2 primers:

Forward: AAGTTGACGCCTACGCTGGCNNKCCGATTCTTACGCTTATGGAGCGTTTT

Reverse: TTCCCAGGTAGGATCGCTGACC

Fragment #3 primers:

Forward: GGTCAGCGATCCTACCTGGGAANNKCACGTAGCAATATTCGCCGGGG

Reverse: AACTGTTGCTTTCAATTGCCCCAGTAC

Fragment #4 primers:

Forward: ACTGGGGCAATTGAAAGCAACAGTTNNKCGCAACTACNNKAG

CCCGCCGAATTTTGGTG

Reverse: TAGGTTGAGGCCGTTGAGCAC

Complete library amplification primers:

Forward: TGTATACCTCCTTCTTATACTTAACGCTAG

Reverse: CAGACTTTACGAAACACGGAAACC

5.2.2 Restriction digestion of library DNA

The purified library DNA was mixed with NEB's rCutSmart buffer (NEB, Catalogue: B6004) to a final buffer concentration of 1X using water. Then, 1 µL of each NotI-HF (NEB, Catalogue: R3189S) and SphI-HF (NEB, Catalogue: R3182S) restriction enzymes were added and mixed. The reaction was incubated at 37 degrees overnight (approximately 16 hours) then heat inactivated at 80 degrees for 20 minutes. The DNA was then purified using the Qiagen QIAquick PCR Purification Kit following the manufacturer's protocol, except the DNA was eluted in 50 µL of water.

5.2.3 Removal of inhibitory restriction digest fragments

750 µg of M-280 Dynabeads (Invitrogen, Catalogue: 11205D) were washed with 1 mL of 1X Bind and Wash Buffer (5 mM Tris-HCl and 1.0 M NaCl at pH 7.5). The buffer was aspirated from the beads and 50 µL of digested library DNA was mixed with 50 µL of 2x Bind and Wash Buffer, then the combined mixture was added to the beads. The beads and DNA were incubated at room temperature for 30 minutes with occasional gentle mixing by pipette (about every five minutes). Then, the beads were pelleted with a magnet and the supernatant was removed. The supernatant containing the fully digested library was purified using the Qiagen QIAquick PCR Purification Kit following the manufacturer's protocol, except the DNA was eluted in 50 µL of water.

5.3 Plasmid production and digestion

5.3.1 Vector maxiprep

pET3a and pULTRA plasmids were maintained and grown in NEB10b cells (NEB, Catalogue: C3019H). Cells carrying the plasmid were inoculated into 500 mL of Lysogeny Broth [Miller] (LB) in a 2 liter cell culture flask directly from frozen glycerol stocks. The culture was incubated at 37 degrees for 18 hours with shaking at 225 RPMs. DNA was then extracted and purified using the Qiagen HiSpeed Maxiprep Kit (Qiagen, Catalogue: 12662) following the manufacturer's protocol, except that the DNA was collected in 1 mL of water.

5.3.2 Vector digestion

pET3a vector was digested by mixing 1 µg of purified DNA with rCutSmart Buffer to a final concentration of 1X buffer in a total reaction volume of 50 µL. Then, 1 µL of each NotI-HF and SphI-HF were added to the reaction. The reaction was incubated at 37 degrees overnight. Next, the reaction was heat inactivated by incubating at 80 degrees for 20 minutes. Then, 1 µL of recombinant shrimp alkaline phosphatase (NEB, Catalogue: M0371S) was added and mixed. The reaction was incubated for another hour at 37 degrees and then heat inactivated again at 80 degrees for 20 minutes. The digested DNA was then purified using the Qiagen QIAquick PCR Purification Kit following the manufacturer's protocol, except that the DNA was eluted in 50 µL of water.

5.4 Ligation of library into vector

Purified pET-3a vector and library insert DNAs were mixed at a 1:3 molar ratio with a total of approximately 10 ng of DNA per µL in forty 50 µL reactions. Each reaction contained 1X T4 DNA Ligase Buffer (NEB, Catalogue: B0202S) and 400 units of T4

DNA Ligase (NEB, Catalogue: M0202S). The reactions were incubated at 16 degrees for 16 hours, then heat inactivated at 65 degrees for 20 minutes. Then, the reactions were combined and purified using the Qiagen QIAquick PCR Purification Kit following the manufacturer's protocol, except that DNAs were eluted in 60 µL of water. During purification, ten ligation reactions were combined over each single PCR purification column.

5.5 Transformation of ligated DNA

For each transformation, an electroporation cuvette with a 1 mm gap was pre-chilled on ice for 30 minutes. 50 µL of electrocompetent NEB10b cells were mixed with 5 µL of ligation DNA (~1 µg) then placed in the chilled electroporation cuvette. The cells were electroporated at 1800 volts using a Bio-Rad MicroPulser electroporator (Catalogue: 1652100) and immediately recovered by mixing the cells with 1 mL of SOC medium. The cells were incubated for 45 minutes at 37 degrees with shaking at 225 RPMs in a 1.7 mL centrifuge tube. After incubation, cells were plated by taking 15 µL of culture, creating four 1:10 dilutions in 100 µL of LB, then spreading each dilution onto LB agar plates with appropriate antibiotics. Plated cells were used to estimate the number of viable transformants. For library transformations, in addition to plating cells, the remaining recovery cultures were pooled into a single 250 mL LB liquid culture with antibiotics and incubated at 37 degrees and 225 RPMs until the culture reached an OD600 of 1.0 (approximately 10 hours). The library culture was then frozen into glycerol stocks by mixing equal parts culture and 50% glycerol (in water) and aliquoting.

5.6 Preparation of competent cells

NEB10b cells were inoculated into a 5 mL LB culture with 50 µg per mL streptomycin in a 15 mL culture tube and incubated overnight at 37 degrees and 225 RPMs. Then, four 500 mL LB cultures with 50 µg per mL streptomycin in 2 liter culture flasks were each inoculated with 1 mL of overnight culture. The large cultures were incubated at 37 degrees and 225 RPMs until they reached an O.D. 600 of 0.6. The cells were pelleted for 10 minutes at 4,500 RCF at 4 degrees. Pellets were washed in 500 mL of 15% glycerol at room temperature. The cells were pelleted and washed again with 500 mL of 15% glycerol and then pelleted and washed again with 50 mL of 15% glycerol. Then, cells were resuspended in 5 mL of 15% glycerol and frozen at -80 degrees as 50 µL aliquots.

5.7 Cell culture

For liquid cultures, NEB-10b cells were grown in Lysogeny Broth [Miller] (LB) which comprised of 10 grams per liter sodium chloride, 5 grams per liter yeast extract, and 10 grams per liter tryptone dissolved in water. For solid cultures, NEB-10b cells were grown on LB agar plates containing LB plus 15 grams per liter agar at 37 degrees. Cultures also contained 50 µg per mL streptomycin. Unless otherwise specified, cells were grown in 5 mL cultures inside 15 mL culture tubes at 37 degrees with shaking at 225 RPMs. When required for plasmid maintenance, additional antibiotics were added to the following concentrations: ampicillin at 100 µg per mL, chloramphenicol at 40 µg per mL, and kanamycin at 50 µg per mL.

5.7.1 UAA and UKA dependent GFP expression

Cells were grown until an O.D. 600 of approximately 0.5. Then, an UAA or UKA was added to a final concentration of 1 mM. The cells were incubated overnight (16 hours) at 37 degrees and 225 RPMs to allow protein expression.

5.7.2 Measurement of GFP expression by fluorimetry

Grown cell cultures were pelleted by spinning at 4,000 RCF for 15 minutes at 4 degrees. The supernatants were removed and the pellets were each resuspended in 1 mL of bacterial lysis buffer (10 mM Tris-HCl, 50 mM potassium chloride, 1 mM EDTA, 1.0% Triton-X 100, 0.5% IGEPAL-CA-630, and 200 mg per liter of lysozyme; pH 8.0). The cells were incubated at 37 degrees with shaking at 225 RPMs for one hour. The cell debris was pelleted at 12,000 RCF and 200 µL of each supernatant was transferred to a well of a black 96 well plate. The GFP fluorescence was measured using a TECAN M1000 plate reader using the default GFP profile.

5.7.3 Chloramphenicol resistance colony formation assay

Cells were grown overnight in a 5 mL culture as described above. Then, a fresh 5 mL culture was inoculated with 50 µL of the overnight culture and incubated at 37 degrees with shaking at 225 RPMs until an O.D. 600 of 0.5. From this new culture, four 1:10 dilutions (1:10, 1:100, 1:1000, 1:10000) were made in LB in a total volume of 5 mL. These dilutions were then plated onto LB agar plates containing streptomycin (50 µg per mL), ampicillin (100 µg per mL), kanamycin (50 µg per mL), o-methyl-tyrosine UAA or UKA (1 mM), and varying concentrations of chloramphenicol: 0, 10, 20, 40, 80, and 160 µg per mL. For each plate, 100 µL of diluted culture was spread using a cell spreader.

The plates were incubated overnight at 37 degrees. Then, cell colony size was measured manually and averaged.

5.8 Verification of UAA conversion and incorporation by mass spectrometry

Cells were grown in the presence of o-methyl-tyrosine or naphthyl UKA to allow GFP expression as detailed in section 5.7.1. Lysate was then prepared as described in section 5.7.2. 500 µL of clarified lysate was purified using 200 µL of HisPur Cobalt Resin (Thermo Scientific, Catalogue: 89964) at room temperature following the manufacturer's batch method protocol. After purification, the GFPs were buffer exchanged into water by diluting the GFPs to 4 mL with water and spinning in a 10,000 MWCO concentrator at 4500 RPMs at 4 degrees for 15 minutes. The buffer exchange was repeated twice more. GFP concentrations were measured by UV spectrometry and 250 µL samples were prepared at 50 µM in water. Samples were filtered through glass wool before loading on the high performance liquid chromatography instrument. Liquid chromatography-mass spectrometry was performed by collaborator Tim Schwochert.

REFERENCES

(1) Kim, C. H.; Axup, J. Y.; Lawson, B. R.; Yun, H.; Tardif, V.; Choi, S. H.; Zhou, Q.; Dubrovska, A.; Biroc, S. L.; Marsden, R.; Pinstaff, J.; Smider, V. V.; Schultz, P. G. Bispecific Small Molecule–Antibody Conjugate Targeting Prostate Cancer. *Proceedings of the National Academy of Sciences* **2013**, *110* (44), 17796–17801. https://doi.org/10.1073/pnas.1316026110.

(2) Lu, H.; Zhou, Q.; Deshmukh, V.; Phull, H.; Ma, J.; Tardif, V.; Naik, R. R.; Bouvard, C.; Zhang, Y.; Choi, S.; Lawson, B. R.; Zhu, S.; Kim, C. H.; Schultz, P. G. Targeting Human C-Type Lectin-Like Molecule-1 (CLL1) with a Bispecific Antibody for Acute Myeloid Leukemia Immunotherapy. *Angew Chem Int Ed Engl* **2014**, *53* (37), 9841–9845. https://doi.org/10.1002/anie.201405353.

(3) Kim, C. H.; Axup, J. Y.; Dubrovska, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G. Synthesis of Bispecific Antibodies Using Genetically Encoded Unnatural Amino Acids. *J Am Chem Soc* **2012**, *134* (24), 9918–9921. https://doi.org/10.1021/ja303904e.

(4) VanBrunt, M. P.; Shanebeck, K.; Caldwell, Z.; Johnson, J.; Thompson, P.; Martin, T.; Dong, H.; Li, G.; Xu, H.; D'Hooge, F.; Masterson, L.; Bariola, P.; Tiberghien, A.; Ezeadi, E.; Williams, D. G.; Hartley, J. A.; Howard, P. W.; Grabstein, K. H.; Bowen, M. A.; Marelli, M. Genetically Encoded Azide Containing Amino Acid in Mammalian Cells Enables Site-Specific Antibody-Drug Conjugates Using Click Cycloaddition Chemistry. *Bioconjug Chem* **2015**, *26* (11), 2249–2260. https://doi.org/10.1021/acs.bioconjchem.5b00359.

(5) Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; Lu, Y.; Tran, H.; Seller, A. J.; Biroc, S. L.; Szydlik, A.; Pinkstaff, J. K.; Tian, F.; Sinha, S. C.; Felding-Habermann, B.; Smider, V. V.; Schultz, P. G. Synthesis of Site-Specific Antibody-Drug Conjugates Using Unnatural Amino Acids. *Proceedings of the National Academy of Sciences* **2012**, *109* (40), 16101–16106. https://doi.org/10.1073/pnas.1211023109.

(6) Tian, F.; Lu, Y.; Manibusan, A.; Sellers, A.; Tran, H.; Sun, Y.; Phuong, T.; Barnett, R.; Hehli, B.; Song, F.; DeGuzman, M. J.; Ensari, S.; Pinkstaff, J. K.; Sullivan, L. M.; Biroc, S. L.; Cho, H.; Schultz, P. G.; DiJoseph, J.; Dougher, M.; Ma, D.; Dushin, R.; Leal, M.; Tchistiakova, L.; Feyfant, E.; Gerber, H.-P.; Sapra, P. A General Approach to Site-Specific Antibody Drug Conjugates. *Proceedings of the National Academy of Sciences* **2014**, *111* (5), 1766–1771. https://doi.org/10.1073/pnas.1321237111.

(7) Zimmerman, E. S.; Heibeck, T. H.; Gill, A.; Li, X.; Murray, C. J.; Madlansacay, M. R.; Tran, C.; Uter, N. T.; Yin, G.; Rivers, P. J.; Yam, A. Y.; Wang, W. D.; Steiner, A. R.; Bajad, S. U.; Penta, K.; Yang, W.; Hallam, T. J.; Thanos, C. D.; Sato, A. K. Production of Site-Specific Antibody-Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System. *Bioconjug Chem* **2014**, *25* (2), 351–361. https://doi.org/10.1021/bc400490z.

(8) Cho, H.; Daniel, T.; Buechler, Y. J.; Litzinger, D. C.; Maio, Z.; Putnam, A.-M. H.; Kraynov, V. S.; Sim, B.-C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; GC, B.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E. Optimized Clinical Performance of Growth Hormone with an Expanded Genetic Code. *Proceedings of the National Academy of Sciences* **2011**, *108* (22), 9060–9065. https://doi.org/10.1073/pnas.1100387108.

(9) Charbon, G.; Brustad, E.; Scott, K. A.; Wang, J.; Løbner-Olesen, A.; Schultz, P. G.; Jacobs-Wagner, C.; Chapman, E. Subcellular Protein Localization by Using a Genetically Encoded Fluorescent Amino Acid. *Chembiochem* **2011**, *12* (12), 1818–1821. https://doi.org/10.1002/cbic.201100282.

(10) Charbon, G.; Wang, J.; Brustad, E.; Schultz, P. G.; Horwich, A. L.; Jacobs-Wagner, C.; Chapman, E. Localization of GroEL Determined by in Vivo Incorporation of a Fluorescent Amino Acid. *Bioorg Med Chem Lett* **2011**, *21* (20), 6067–6070. https://doi.org/10.1016/j.bmcl.2011.08.057.

(11) Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G. A Genetically Encoded Fluorescent Probe in Mammalian Cells. *J Am Chem Soc* **2013**, *135* (34), 12540–12543. https://doi.org/10.1021/ja4059553.

(12) McKay, C. S.; Finn, M. G. Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation. *Chem Biol* **2014**, *21* (9), 1075–1101. https://doi.org/10.1016/j.chembiol.2014.09.002.

(13) Chen, X.; Wu, Y.-W. Selective Chemical Labeling of Proteins. *Org Biomol Chem* **2016**, *14* (24), 5417–5439. https://doi.org/10.1039/c6ob00126b.

(14) Lee, K. J.; Kang, D.; Park, H.-S. Site-Specific Labeling of Proteins Using Unnatural Amino Acids. *Mol Cells* **2019**, *42* (5), 386–396. https://doi.org/10.14348/molcells.2019.0078.

(15) Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E. A. Genetically Encoded Copper-Free Click Chemistry. *Angew Chem Int Ed Engl* **2011**, *50* (17), 3878–3881. https://doi.org/10.1002/anie.201008178.

(16) Swiderska, K. W.; Szlachcic, A.; Czyrek, A.; Zakrzewska, M.; Otlewski, J. Site-Specific Conjugation of Fibroblast Growth Factor 2 (FGF2) Based on Incorporation of Alkyne-Reactive Unnatural Amino Acid. *Bioorg Med Chem* **2017**, *25* (14), 3685–3693. https://doi.org/10.1016/j.bmc.2017.05.003.

(17) Das, D. K.; Govindan, R.; Nikić-Spiegel, I.; Krammer, F.; Lemke, E. A.; Munro, J. B. Direct Visualization of the Conformational Dynamics of Single Influenza Hemagglutinin Trimers. *Cell* **2018**, *174* (4), 926-937.e12. https://doi.org/10.1016/j.cell.2018.05.050.

(18) Rogers, J. M. G.; Lippert, L. G.; Gai, F. Non-Natural Amino Acid Fluorophores for One- and Two-Step FRET Applications†. *Anal Biochem* **2010**, *399* (2), 182–189. https://doi.org/10.1016/j.ab.2009.12.027.

(19) Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a Photocrosslinking Amino Acid to the Genetic Code of Escherichia Coli. *Proceedings of the National Academy of Sciences* **2002**, *99* (17), 11020–11024. https://doi.org/10.1073/pnas.172226299.

(20) Costa, S. A.; Simon, J. R.; Amiram, M.; Tang, L.; Zauscher, S.; Brustad, E. M.; Isaacs, F. J.; Chilkoti, A. Photocrosslinkable Unnatural Amino Acids Enable Facile Synthesis of Thermoresponsive Nano- to Micro-Gels of Intrinsically Disordered Polypeptides. *Adv Mater* **2018**, *30* (5), 10.1002/adma.201704878. https://doi.org/10.1002/adma.201704878.

(21) Simms, J.; Uddin, R.; Sakmar, T. P.; Gingell, J. J.; Garelja, M. L.; Hay, D. L.; Brimble, M. A.; Harris, P. W.; Reynolds, C. A.; Poyner, D. R. Photoaffinity Cross-Linking and Unnatural Amino Acid Mutagenesis Reveal Insights into Calcitonin Gene-Related Peptide Binding to the Calcitonin Receptor-like Receptor/Receptor Activity-Modifying Protein 1 (CLR/RAMP1) Complex. *Biochemistry* **2018**, *57* (32), 4915–4922. https://doi.org/10.1021/acs.biochem.8b00502.

(22) Murray, C. I.; Westhoff, M.; Eldstrom, J.; Thompson, E.; Emes, R.; Fedida, D. Unnatural Amino Acid Photo-Crosslinking of the IKs Channel Complex Demonstrates a KCNE1:KCNQ1 Stoichiometry of up to 4:4. *eLife* **2016**, *5*, e11815. https://doi.org/10.7554/eLife.11815.

(23) Yang, B.; Tang, S.; Ma, C.; Li, S.-T.; Shao, G.-C.; Dang, B.; DeGrado, W. F.; Dong, M.-Q.; Wang, P. G.; Ding, S.; Wang, L. Spontaneous and Specific Chemical Cross-Linking in Live Cells to Capture and Identify Protein Interactions. *Nat Commun* **2017**, *8* (1), 2240. https://doi.org/10.1038/s41467-017-02409-z.

(24) Chin, J. W.; Schultz, P. G. In Vivo Photocrosslinking with Unnatural Amino Acid Mutagenesis. *ChemBioChem* **2002**, *3* (11), 1135–1137. https://doi.org/10.1002/1439-7633(20021104)3:11<1135::AID-CBIC1135>3.0.CO;2-M.

(25) Kneuttinger, A. C.; Straub, K.; Bittner, P.; Simeth, N. A.; Bruckmann, A.; Busch, F.; Rajendran, C.; Hupfeld, E.; Wysocki, V. H.; Horinek, D.; König, B.; Merkl, R.; Sterner, R. Light Regulation of Enzyme Allostery through Photo-Responsive Unnatural Amino Acids. *Cell Chemical Biology* **2019**, *26* (11), 1501-1514.e9. https://doi.org/10.1016/j.chembiol.2019.08.006.

(26) Liu, C. C.; Qi, L.; Yanofsky, C.; Arkin, A. P. Regulation of Transcription by Unnatural Amino Acids. *Nat Biotechnol* **2011**, *29* (2), 164–168. https://doi.org/10.1038/nbt.1741.

(27) Cao, W.; Qin, X.; Wang, Y.; Dai, Z.; Dai, X.; Wang, H.; Xuan, W.; Zhang, Y.; Liu, Y.; Liu, T. A General Supramolecular Approach to Regulate Protein Functions by Cucurbit[7]Uril and Unnatural Amino Acid Recognition. *Angewandte Chemie International Edition* **2021**, *60* (20), 11196–11200. https://doi.org/10.1002/anie.202100916.

(28) Chou, C.; Young, D. D.; Deiters, A. A Light-Activated DNA Polymerase. *Angewandte Chemie International Edition* **2009**, *48* (32), 5950–5953. https://doi.org/10.1002/anie.200901115.

(29) Lemke, E. A.; Summerer, D.; Geierstanger, B. H.; Brittain, S. M.; Schultz, P. G. Control of Protein Phosphorylation with a Genetically Encoded Photocaged Amino Acid. *Nat Chem Biol* **2007**, *3* (12), 769–772. https://doi.org/10.1038/nchembio.2007.44.

(30) Nödling, A. R.; Spear, L. A.; Williams, T. L.; Luk, L. Y. P.; Tsai, Y.-H. Using Genetically Incorporated Unnatural Amino Acids to Control Protein Functions in Mammalian Cells. *Essays in Biochemistry* **2019**, *63* (2), 237–266. https://doi.org/10.1042/EBC20180042.

(31) Adhikari, A.; Raj Bhattarai, B.; Aryal, A.; Thapa, N.; Kc, P.; Adhikari, A.; Maharjan, S.; B. Chanda, P.; P. Regmi, B.; Parajuli, N. Reprogramming Natural Proteins Using Unnatural Amino Acids. *RSC Advances* **2021**, *11* (60), 38126–38145. https://doi.org/10.1039/D1RA07028B.

(32) Drienovská, I.; Rioz-Martínez, A.; Draksharapu, A.; Roelfes, G. Novel Artificial Metalloenzymes by in Vivo Incorporation of Metal-Binding Unnatural Amino Acids. *Chem. Sci.* **2014**, *6* (1), 770–776. https://doi.org/10.1039/C4SC01525H.

(33) Jäckel, C.; Hilvert, D. Biocatalysts by Evolution. *Curr Opin Biotechnol* **2010**, *21* (6), 753–759. https://doi.org/10.1016/j.copbio.2010.08.008.

(34) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nature* **2012**, *485* (7397), 185–194. https://doi.org/10.1038/nature11117.

(35) Davids, T.; Schmidt, M.; Böttcher, D.; Bornscheuer, U. T. Strategies for the Discovery and Engineering of Enzymes for Biocatalysis. *Curr Opin Chem Biol* **2013**, *17* (2), 215–220. https://doi.org/10.1016/j.cbpa.2013.02.022.

(36) Hao, B.; Gong, W.; Ferguson, T. K.; James, C. M.; Krzycki, J. A.; Chan, M. K. A New UAG-Encoded Residue in the Structure of a Methanogen Methyltransferase. *Science* **2002**, *296* (5572), 1462–1466. https://doi.org/10.1126/science.1069556.

(37) Srinivasan, G.; James, C. M.; Krzycki, J. A. Pyrrolysine Encoded by UAG in Archaea: Charging of a UAG-Decoding Specialized TRNA. *Science* **2002**, *296* (5572), 1459–1462. https://doi.org/10.1126/science.1069588.

(38) Hendrickson, T. L.; de Crécy-Lagard, V.; Schimmel, P. Incorporation of Nonnatural Amino Acids into Proteins. *Annu Rev Biochem* **2004**, *73*, 147–176. https://doi.org/10.1146/annurev.biochem.73.012803.092429.

(39) Liu, D. R.; Schultz, P. G. Progress toward the Evolution of an Organism with an Expanded Genetic Code. *Proceedings of the National Academy of Sciences* **1999**, *96* (9), 4780–4785. https://doi.org/10.1073/pnas.96.9.4780.

(40) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of Escherichia Coli. *Science* **2001**, *292* (5516), 498–500. https://doi.org/10.1126/science.1060077.

(41) Ryu, Y.; Schultz, P. G. Efficient Incorporation of Unnatural Amino Acids into Proteins in Escherichia Coli. *Nat Methods* **2006**, *3* (4), 263–265. https://doi.org/10.1038/nmeth864.

(42) Mukai, T.; Hayashi, A.; Iraha, F.; Sato, A.; Ohtake, K.; Yokoyama, S.; Sakamoto, K. Codon Reassignment in the Escherichia Coli Genetic Code. *Nucleic Acids Res* **2010**, *38* (22), 8188–8195. https://doi.org/10.1093/nar/gkq707.

(43) Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. Incorporation of Nonnatural Amino Acids into Streptavidin through In Vitro Frame-Shift Suppression. *J. Am. Chem. Soc.* **1996**, *118* (40), 9778–9779. https://doi.org/10.1021/ja9614225.

(44) Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. Incorporation of Nonnatural Amino Acids into Proteins by Using Various Four-Base Codons in an Escherichia Coli in Vitro Translation System. *Biochemistry* **2001**, *40* (37), 11060–11064. https://doi.org/10.1021/bi0108204.

(45) Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. An Expanded Genetic Code with a Functional Quadruplet Codon. *Proceedings of the National Academy of Sciences* **2004**, *101* (20), 7566–7571. https://doi.org/10.1073/pnas.0401517101.

(46) Shafer, A. M.; Kálai, T.; Bin Liu, S. Q.; Hideg, K.; Voss, J. C. Site-Specific Insertion of Spin-Labeled l-Amino Acids in Xenopus Oocytes. *Biochemistry* **2004**, *43* (26), 8470–8482. https://doi.org/10.1021/bi035542i.

(47) Rodriguez, E. A.; Lester, H. A.; Dougherty, D. A. In Vivo Incorporation of Multiple Unnatural Amino Acids through Nonsense and Frameshift Suppression. *Proceedings of the National Academy of Sciences* **2006**, *103* (23), 8650–8655. https://doi.org/10.1073/pnas.0510817103.

(48) Ohtsuki, T.; Manabe, T.; Sisido, M. Multiple Incorporation of Non-Natural Amino Acids into a Single Protein Using TRNAs with Non-Standard Structures. *FEBS Letters* **2005**, *579* (30), 6769–6774. https://doi.org/10.1016/j.febslet.2005.11.010.

(49) Wang, L.; Schultz, P. G. A General Approach for the Generation of Orthogonal TRNAs. *Chemistry & Biology* **2001**, *8* (9), 883–890. https://doi.org/10.1016/S1074-5521(01)00063-1.

(50) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. An Efficient System for the Evolution of Aminoacyl-TRNA Synthetase Specificity. *Nat Biotechnol* **2002**, *20* (10), 1044–1048. https://doi.org/10.1038/nbt742.

(51) Young, D. D.; Young, T. S.; Jahnz, M.; Ahmad, I.; Spraggon, G.; Schultz, P. G. An Evolved Aminoacyl-TRNA Synthetase with Atypical Polysubstrate Specificity. *Biochemistry* **2011**, *50* (11), 1894–1900. https://doi.org/10.1021/bi101929e.

(52) Guo, L.-T.; Wang, Y.-S.; Nakamura, A.; Eiler, D.; Kavran, J. M.; Wong, M.; Kiessling, L. L.; Steitz, T. A.; O'Donoghue, P.; Söll, D. Polyspecific Pyrrolysyl-TRNA Synthetases from Directed Evolution. *Proceedings of the National Academy of Sciences* **2014**, *111* (47), 16724–16729. https://doi.org/10.1073/pnas.1419737111.

(53) Chatterjee, A.; Xiao, H.; Schultz, P. G. Evolution of Multiple, Mutually Orthogonal Prolyl-TRNA Synthetase/TRNA Pairs for Unnatural Amino Acid Mutagenesis in Escherichia Coli. *Proceedings of the National Academy of Sciences* **2012**, *109* (37), 14841–14846. https://doi.org/10.1073/pnas.1212454109.

(54) Fan, C.; Xiong, H.; Reynolds, N. M.; Söll, D. Rationally Evolving TRNAPyl for Efficient Incorporation of Noncanonical Amino Acids. *Nucleic Acids Res* **2015**, *43* (22), e156. https://doi.org/10.1093/nar/gkv800.

(55) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* **2010**, *464* (7287), 441–444. https://doi.org/10.1038/nature08817.

(56) Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. RF1 Knockout Allows Ribosomal Incorporation of Unnatural Amino Acids at Multiple Sites. *Nat Chem Biol* **2011**, *7* (11), 779–786. https://doi.org/10.1038/nchembio.657.

(57) Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. Genomically Recoded Organisms Expand Biological Functions. *Science* **2013**, *342* (6156), 357–360. https://doi.org/10.1126/science.1241459.

(58) Schmied, W. H.; Elsässer, S. J.; Uttamapinant, C.; Chin, J. W. Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl TRNA Synthetase/TRNA Expression and Engineered ERF1. *J. Am. Chem. Soc.* **2014**, *136* (44), 15577–15583. https://doi.org/10.1021/ja5069728.

(59) Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. Evolved Orthogonal Ribosomes Enhance the Efficiency of Synthetic Genetic Code Expansion. *Nat Biotechnol* **2007**, *25* (7), 770–777. https://doi.org/10.1038/nbt1314.

(60) Ko, W.; Kumar, R.; Kim, S.; Lee, H. S. Construction of Bacterial Cells with an Active Transport System for Unnatural Amino Acids. *ACS Synth. Biol.* **2019**, *8* (5), 1195–1203. https://doi.org/10.1021/acssynbio.9b00076.

(61) Takimoto, J. K.; Xiang, Z.; Kang, J.-Y.; Wang, L. Esterification of an Unnatural Amino Acid Structurally Deviating from Canonical Amino Acids Promotes Its Uptake and Incorporation into Proteins in Mammalian Cells. *Chembiochem* **2010**, *11* (16), 2268–2272. https://doi.org/10.1002/cbic.201000436.

(62) Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. An Expanded Eukaryotic Genetic Code. *Science* **2003**, *301* (5635), 964–967. https://doi.org/10.1126/science.1084772.

(63) Wang, Q.; Wang, L. New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast. *J. Am. Chem. Soc.* **2008**, *130* (19), 6066–6067. https://doi.org/10.1021/ja800894n.

(64) Wang, Q.; Wang, L. Genetic Incorporation of Unnatural Amino Acids into Proteins in Yeast. In *Unnatural Amino Acids: Methods and Protocols*; Pollegioni, L., Servi, S., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2012; pp 199–213. https://doi.org/10.1007/978-1-61779-331-8_12.

(65) Shen, B.; Xiang, Z.; Miller, B.; Louie, G.; Wang, W.; Noel, J. P.; Gage, F. H.; Wang, L. Genetically Encoding Unnatural Amino Acids in Neural Stem Cells and Optically Reporting Voltage-Sensitive Domain Changes in Differentiated Neurons. *STEM CELLS* **2011**, *29* (8), 1231–1240. https://doi.org/10.1002/stem.679.

(66) Li, J.; Yu, J.; Zhao, J.; Wang, J.; Zheng, S.; Lin, S.; Chen, L.; Yang, M.; Jia, S.; Zhang, X.; Chen, P. R. Palladium-Triggered Deprotection Chemistry for Protein Activation in Living Cells. *Nat Chem* **2014**, *6* (4), 352–361. https://doi.org/10.1038/nchem.1887.

(67) Hemphill, J.; Borchardt, E. K.; Brown, K.; Asokan, A.; Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J Am Chem Soc* **2015**, *137* (17), 5642–5645. https://doi.org/10.1021/ja512664v.

(68) Hemphill, J.; Chou, C.; Chin, J. W.; Deiters, A. Genetically Encoded Light-Activated Transcription for Spatiotemporal Control of Gene Expression and Gene Silencing in Mammalian Cells. *J Am Chem Soc* **2013**, *135* (36), 13433–13439. https://doi.org/10.1021/ja4051026.

(69) Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K.-F.; Slesinger, P. A.; Wang, L. Genetically Encoding Unnatural Amino Acids for Cellular

and Neuronal Studies. *Nat Neurosci* **2007**, *10* (8), 1063–1072. https://doi.org/10.1038/nn1932.

(70) Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A. Genetically Encoded Optochemical Probes for Simultaneous Fluorescence Reporting and Light Activation of Protein Function with Two-Photon Excitation. *J Am Chem Soc* **2014**, *136* (44), 15551–15558. https://doi.org/10.1021/ja5055862.

(71) Greiss, S.; Chin, J. W. Expanding the Genetic Code of an Animal. *J Am Chem Soc* **2011**, *133* (36), 14196–14199. https://doi.org/10.1021/ja2054034.

(72) Parrish, A. R.; She, X.; Xiang, Z.; Coin, I.; Shen, Z.; Briggs, S. P.; Dillin, A.; Wang, L. Expanding the Genetic Code of Caenorhabditis Elegans Using Bacterial Aminoacyl-TRNA Synthetase/TRNA Pairs. *ACS Chem. Biol.* **2012**, *7* (7), 1292–1302. https://doi.org/10.1021/cb200542j.

(73) Chen, Y.; Ma, J.; Lu, W.; Tian, M.; Thauvin, M.; Yuan, C.; Volovitch, M.; Wang, Q.; Holst, J.; Liu, M.; Vriz, S.; Ye, S.; Wang, L.; Li, D. Heritable Expansion of the Genetic Code in Mouse and Zebrafish. *Cell Res* **2017**, *27* (2), 294–297. https://doi.org/10.1038/cr.2016.145.

(74) Kang, J.-Y.; Kawaguchi, D.; Coin, I.; Xiang, Z.; O'Leary, D. D. M.; Slesinger, P. A.; Wang, L. In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids. *Neuron* **2013**, *80* (2), 358–370. https://doi.org/10.1016/j.neuron.2013.08.016.

(75) Li, F.; Zhang, H.; Sun, Y.; Pan, Y.; Zhou, J.; Wang, J. Expanding the Genetic Code for Photoclick Chemistry in E. Coli, Mammalian Cells, and A. Thaliana. *Angew Chem Int Ed Engl* **2013**, *52* (37), 9700–9704. https://doi.org/10.1002/anie.201303477.

(76) Gao, W.; Cho, E.; Liu, Y.; Lu, Y. Advances and Challenges in Cell-Free Incorporation of Unnatural Amino Acids Into Proteins. *Frontiers in Pharmacology* **2019**, *10*.

(77) Shrestha, P.; Smith, M. T.; Bundy, B. C. Cell-Free Unnatural Amino Acid Incorporation with Alternative Energy Systems and Linear Expression Templates. *New Biotechnology* **2014**, *31* (1), 28–34. https://doi.org/10.1016/j.nbt.2013.09.002.

(78) Strable, E.; Prasuhn, D. E. Jr.; Udit, A. K.; Brown, S.; Link, A. J.; Ngo, J. T.; Lander, G.; Quispe, J.; Potter, C. S.; Carragher, B.; Tirrell, D. A.; Finn, M. G. Unnatural Amino Acid Incorporation into Virus-Like Particles. *Bioconjugate Chem.* **2008**, *19* (4), 866–875. https://doi.org/10.1021/bc700390r.

(79) Banerjee, P. S.; Ostapchuk, P.; Hearing, P.; Carrico, I. S. Unnatural Amino Acid Incorporation onto Adenoviral (Ad) Coat Proteins Facilitates Chemoselective Modification and Retargeting of Ad Type 5 Vectors. *Journal of Virology* **2011**, *85* (15), 7546–7554. https://doi.org/10.1128/JVI.00118-11.

(80) Kelemen, R. E.; Erickson, S. B.; Chatterjee, A. Production and Chemoselective Modification of Adeno-Associated Virus Site-Specifically Incorporating an Unnatural Amino Acid Residue into Its Capsid. In *Noncanonical Amino Acids: Methods and Protocols*; Lemke, E. A., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2018; pp 313–326. https://doi.org/10.1007/978-1-4939-7574-7_20.

(81) Kelemen, R. E.; Erickson, S. B.; Chatterjee, A. Synthesis at the Interface of Virology and Genetic Code Expansion. *Curr Opin Chem Biol* **2018**, *46*, 164–171. https://doi.org/10.1016/j.cbpa.2018.07.015.

(82) Zhang, K.; Li, H.; Cho, K. M.; Liao, J. C. Expanding Metabolism for Total Biosynthesis of the Nonnatural Amino Acid L-Homoalanine. *Proceedings of the National Academy of Sciences* **2010**, *107* (14), 6234–6239. https://doi.org/10.1073/pnas.0912903107.

(83) Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. Generation of a Bacterium with a 21 Amino Acid Genetic Code. *J. Am. Chem. Soc.* **2003**, *125* (4), 935–939. https://doi.org/10.1021/ja0284153.

(84) Bartsch, S.; Bornscheuer, U. T. Mutational Analysis of Phenylalanine Ammonia Lyase to Improve Reactions Rates for Various Substrates. *Protein Eng Des Sel* **2010**, *23* (12), 929–933. https://doi.org/10.1093/protein/gzq089.

(85) Rowles, I.; Groenendaal, B.; Binay, B.; Malone, K. J.; Willies, S. C.; Turner, N. J. Engineering of Phenylalanine Ammonia Lyase from Rhodotorula Graminis for the Enhanced Synthesis of Unnatural L-Amino Acids. *Tetrahedron* **2016**, *72* (46), 7343–7347. https://doi.org/10.1016/j.tet.2016.06.026.

(86) Weise, N. J.; Ahmed, S. T.; Parmeggiani, F.; Galman, J. L.; Dunstan, M. S.; Charnock, S. J.; Leys, D.; Turner, N. J. Zymophore Identification Enables the Discovery of Novel Phenylalanine Ammonia Lyase Enzymes. *Sci Rep* **2017**, *7* (1), 13691. https://doi.org/10.1038/s41598-017-13990-0.

(87) Ahmed, S. T.; Parmeggiani, F.; Weise, N. J.; Flitsch, S. L.; Turner, N. J. Engineered Ammonia Lyases for the Production of Challenging Electron-Rich l-Phenylalanines. *ACS Catal.* **2018**, *8* (4), 3129–3132. https://doi.org/10.1021/acscatal.8b00496.

(88) Romney, D. K.; Murciano-Calles, J.; Wehrmüller, J. E.; Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **2017**, *139* (31), 10769–10776. https://doi.org/10.1021/jacs.7b05007.

(89) Powell, J. T.; Morrison, J. F. The Purification and Properties of the Aspartate Aminotransferase and Aromatic-Amino-Acid Aminotransferase from Escherichia Coli. *European Journal of Biochemistry* **1978**, *87* (2), 391–400. https://doi.org/10.1111/j.1432-1033.1978.tb12388.x.

(90) Patrick, W. M.; Firth, A. E.; Blackburn, J. M. User-Friendly Algorithms for Estimating Completeness and Diversity in Randomized Protein-Encoding Libraries. *Protein Eng* **2003**, *16* (6), 451–457. https://doi.org/10.1093/protein/gzg057.