NORTHWESTERN UNIVERSITY

Improving Co-Translational Incorporation of Non-Standard Amino Acids into Recombinant Proteins in *Escherichia coli*

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemical and Biological Engineering

By

Jessica Gabrielle Perez

EVANSTON, ILLINOIS

March 2017

ABSTRACT

Improving Co-Translational Incorporation of Non-Standard Amino Acids into Recombinant Proteins in *Escherichia coli*

Jessica G. Perez

The genetic code, a set of rules by which mRNA codons are translated into the twenty standard amino acids used in protein synthesis, was once thought to be immutable. As we expanded our understanding of molecular biology more variations in this genetic code have been found across the animal kingdom. These findings have spurred interest in utilizing genetic code variations to incorporate non-standard, chemically diverse amino acids into proteins using the ribosome. By synthesizing proteins with novel chemical properties inside cells, we have the opportunity to transform how we synthesize materials and therapeutics, investigate protein structure, and understand the evolution of the translation system. The efficiency of non-standard amino acid (nsAA) incorporation still remains ~1000-fold lower than standard amino acid incorporation. This ultimately affects the total amount of modified protein possible to synthesize, limiting its application and use. In this thesis, I outline several strategies toward improving co-translational incorporation of nsAAs into recombinant proteins using *Escherichia coli*. The majority of my work has focused on developing an improved strain for nsAA incorporation. This has been accomplished by removing negative effectors of protein synthesis to increase modified protein yields up to 17-fold and by introducing a T7 system capable of tuning the expression of recombinant proteins. Additionally, I have demonstrated that independent promoter optimization and translational component evolution can be combined for a synergistic benefit. Specifically, we found that using an optimized promoter plasmid with engineered components resulted in a 2- to 20-fold enhancement of sfGFP expression containing multiple nsAAs. The work outlined here has furthered the development of co-translational incorporation of nsAAs into recombinant proteins and will help produce the next wave of highly functional biomaterials and protein therapeutics.

ACKNOWLEDGEMENTS

First, I would like to thank my entire family, especially my mom, dad, and Daniel. Thank you to my mom for always reminding me that I have nothing to regret if I put my hardest effort forward. Thank you to my dad for setting high academic expectations and teaching me to how create my own opportunities. I would also like to thank my partner, Daniel, for supporting me all through undergraduate and graduate school. His constant encouragement, advice, patience, and moral support has brought me unmeasurable joy.

Next, I would like to thank all of my Jewett lab colleagues for making the lab a wonderful place to work and grow as an individual. Special thanks goes to the Marg Wed crew (Jess, Ashty, Rey, Erik and Anne) and the Texas office (Ashty, Do Soon and Jian) for giving me so many fun, heart-warming memories and providing me the balance that every graduate student needs. I would like to thank Team *In Vivo*, Erik and Gan, for their training, support, and friendship. I would like to thank Gan, Erik, and Jian for their critical review of this dissertation and my thesis committee for their valuable feedback and advice along the way.

Lastly, I would like to thank my advisor, Mike, for allowing me to pursue my interests in *in vivo* synthetic biology, creating an amazingly inclusive, collaborative workplace and supporting me through graduate school and beyond.

LIST OF ABBREVIATIONS

aaRS……aminoacyl-tRNA synthetase

aa-tRNA……aminoacyl-tRNA

aTc……anhydrotetracycline

AzRS……p-azido-phenylalanyl-tRNA synthetase

bla……β-lactamase

bp……base pair

CFPS……cell-free protein synthesis

CHO……Chinese hamster ovary

CmR, cat……chloramphenicol resistance gene

EF-Tu…… elongation factor Tu

ELP……elastin-like polymer

E. coli……*Escherichia coli*

GFP……green fluorescent protein

IPTG……isopropyl β-D-1-thiogalctopyranoside

LacO……Lac operator

MAGE……multiplex automated genome engineering

MASC……multiplex allele-specific colony

*Mj*TyrRS……*M. jannaschii* TyrRS

nsAA……non-standard amino acid

o-aaRS……orthogonal aminoacyl tRNA synthetase

OD……optical density

o-tRNA……orthogonal tRNA

OTS……orthogonal translation system

pAzF……p-azido-phenylalanine

pAzFRS……p-Azido-phenylalanine aminoacyl tRNA synthetase

ProCarb……..….N_6-(propargyloxycarbonyl)-$_L$-Lysine

PTMs……..….post-translational modifications

Pyl……..……..…pyrrolysine

PylRS…….…….....……pyrrolysine aminoacyl tRNA synthetase

RF-1……..…….release factor-1

sfGFP…….……...…....super folded green fluorescent protein

sfGFP-wt……..……wild-type sfGFP

T7RNAP……...…….……T7 RNA polymerase

tRNA…….……..……transfer RNA

TABLE OF CONTENTS

LIST OF FIGURES

be quickly screened for specific characteristics *in vitro* before implementation *in vivo*. First, in the Design stage, the genetic parts or networks are designed, informed by computational models or literature. Next, in the Build stage, several designs are built. In the Test phase, built designs are assayed *in vitro*. If the Test stage does not yield the desired behavior, one reinitiates the cycle *n* number of times until the desired characteristic is achieved. The Test stage can also inform the modeling and allow for better models for the Design stage. Once the desired characteristics are found, top candidates that behave as expected can be implemented *in vivo*, with an increased likeliness of being functional. **(B) Variables for characterization and optimization.** *In vitro* prototyping can occur by characterizing or optimizing various levels on the genetic network "abstraction hierarchy." At the most basic level, transcriptional and translational parts or purified components of a network can be analyzed for functionality. *In vitro* prototyping can also be applied at the device level to assess input-output relationships. Finally, an *in vitro* systems level analysis allows for an isolated study of how multiple genetic devices feed into each other and a preview of the overall network behavior…….…….123

LIST OF TABLES

CHAPTER 1

1 Introduction: Thesis Overview

1.1 Motivation and Objectives

Synthetic biology is an emerging field which seeks to add new functionality to biological systems for the creation of high value technologies and products [13-17]. Efforts in synthetic biology have included synthesis of high value products, such as the anti-malarial drug artemisinin [18] and development of cell-based biosensors and therapeutics [19-21]. Recent efforts in modifying genomes [22, 23] and development of cell-free platforms [24, 25] each represent research areas with a vast potential in the near future.

From synthetic biology has come the field of genetic code expansion which enables non-standard amino acids (nsAAs) to be incorporated into proteins co-translationally in various organisms including *Escherichia coli,* [26-30] yeast [31], Chinese Hamster Ovary cells [32], *Drosophila melanogaster* [33], and *Caenorhabditis elegans* [34]. This technique has a wide range of applications. The introduction of biorthogonal handles for protein tagging [11, 35] enables synthetic mimicry of nature's post-translational modifications and artificial modifications such as PEGylation which is vital for increasing therapeutic protein solubility and stability [36]. The synthesis of new classes of scaffolds and biomaterials are also possible using click-chemistry capable nsAAs [37, 38]. The precise introduction of novel chemistries into proteins has supported studies in protein structure [39] and synthesis of novel enzymes [40].

However, nsAA incorporation into proteins is limited by several obstacles. First, nsAA incorporation efficiencies are ~1000 times lower than native amino acid incorporation [41], which ultimately reduces modified protein yields. Second, target nsAAs occasionally are incompatible with the host's native translation system due to insolubility and cell pentation issues and reduced binding affinities to translational components [41]. Lastly, strains developed for nsAA incorporation [42] have not been optimized for protein production which limits achievable modified protein yields.

This work detailed efforts towards addressing all the nsAA incorporation obstacles listed above. An improved strained for co-transitional incorporation of nsAAs has been developed by removing negative effectors to protein synthesis to increase modified protein yields. The capabilities of this strain have been augmented by the introduction of a T7 RNAP Polymerase (T7RNAP) system capable of tuning the expression of recombinant proteins. Several factors were explored to optimize T7RNAP's inducibility response and reduce basal recombinant protein expression. Additionally, the nsAA incorporation efficiency of p-azido-phenylalanine was improved by evolving several translation components. Lastly, we demonstrated that modified protein expression can be optimized through promoter tuning efforts.

1.2 Thesis Outline

The following chapters provide detailed information on the field of co-translational incorporation of non-standard amino acid (nsAAs) and our efforts for improving this process in *Escherichia coli*. These chapters will include background information, experimental methodology, results, discussion, and future directions for various efforts towards improving nsAA incorporation. Chapter 2 reviews the current state of synthesizing proteins containing nsAAs. Techniques reviewed include chemical synthesis methods, expression of proteins in organisms with post-translational modifications mechanisms and installing nsAA into proteins co-translationally. Chapter 3 focuses on our effort to remove negative effectors of protein synthesis, improving yields of proteins containing nsAAs. Chapter 3 briefly describes the design and construction of an optimized T7 RNA Polymerase expression system and includes work in collaboration with Oliver Weisser and Erik Carlson. Chapter 4 will deeply detail the optimization of a T7 expression system in a genomically recoded strain. In this chapter, we explore various factors involved in increasing protein yields and improving the induction response using the T7 expression system. This chapter will include work in collaboration with Oliver Weisser and Erik Carlson. Chapter 5 will describe our efforts to evolve *Methanococcus jannaschii* p-azido-phenylalanyl-tRNA synthetase and elongation factor Tu to improve nsAA incorporation efficiencies. We also demonstrate the utility of tuning the expression levels of orthogonal translational components. Chapter 5 will include work in collaboration with Rui Gan. Chapter 6 describes the adjacent but complementary field of cell-free protein synthesis. While this field is in its

infancy, there are many opportunities for the nsAA field to be applied. This book chapter was written in collaboration with Jessica Stark. Finally, Chapter 7 details a summary of this work, recommendations for future directions, and a final perspective on the nsAA field.

CHAPTER 2

2 Expanding the chemistry of life

2.1 Abstract

Synthesizing recombinant proteins containing non-standard amino acids (nsAA), as opposed to the standard 20 amino acids, has much utility in numerous fields due to the addition of novel chemical properties obtained when using expanded chemistry to build proteins. For instance, the majority of protein based biopharmaceuticals approved or in clinical trials bear some form of modification (glycosylation, carboxylation, hydroxylation, sulfation, amidation, etc.), which can profoundly affect protein properties relevant to their therapeutic application. Structural biologists often use fluorescent amino acids or probes when studying changes in protein structure. Redox-active or heavy metal containing amino acids can be inserted into an enzyme's active site to change its reactivity or substrate affinity in the area of biocatalysis. Numerous methods have been developed to synthesize proteins containing nsAA in order to accomplish these goals. Here, I will review methods for synthesizing modified proteins as well as highlighting each method's advantages and disadvantages.

2.2 Introduction

The majority of a cell's functions are derived from interactions with proteins. These informational macromolecules consist of a sequence of amino acid residues that direct folding of the native protein into a three-dimensional structure which dictates its function. Proteins can range in length from several amino acids to as large as ~30,000 amino acids (Titin, muscle tissue protein) [43]. Additionally, the number of distinct human protein forms exceed the number of genes in the genome by 10-100 orders of magnitude. This increase in protein diversity after translation stems from post-translational modifications (PTMs). These can include the covalent addition of one or more groups, such as phophoryl, acetyl or glycosyl, to one or more of the amino acids side chains in a particular protein [44]. Protein modifications play a crucial

role in enzyme activity, protein turner, localization, protein-protein interactions, modulation for various signaling cascades, DNA repair, and cell division [44].

Given the pivotal role of PTMs in the regulation of cellular environments, there is a great effort towards synthetically mimicking nature's capacity to install such modifications. The application of modified proteins are many; they are as varied as the *in vivo* tracking of protein-fluorophore conjugates [45] to the polyethylene glyol (PEG)ylation of therapeutic proteins to reduce immunogenicity [36] to the production of material with novel properties [46] to probing the mechanism of enzymes in a pathway [47]. Thus, with the hunger for precise molecular knowledge of protein function, precise protein modification is becoming increasingly more vital.

Here, I will discuss the dominant chemical efforts to synthesize proteins with modified residues. Though these efforts have yielded success and are useful in many applications, they are ultimately limited by short peptide sequences. Alternative methods utilize the cell's biosynthetic potential to precisely synthesize peptides using the ribosome in a template-guided manner on a single-molecule scale. These methods include expressing recombinant proteins in host cells that natively possess PTM mechanisms. This technology is limited to the modification mechanisms available in the host strain. Emerging technologies that exploit the cell's translation system to install nsAAs co-translationally overcome the limitation of native PTM mechanisms. Finally, I will discuss new frontiers utilizing co-translational incorporation of nsAAs.

2.3 Current Methods for Chemical Synthesis of Modified Proteins

The ability to synthesize pure polypeptides has been a goal for many years for therapeutic application and biophysical studies. Prior to the biotechnological advances of the 1960's to recombinantly express a protein of interest in a cell, scientists relied on chemical peptide synthesis.

Chemical methods to form peptide bonds date back as far as 100 years [48]; however it has taken great effort to develop methods to efficiently form amide bonds at low cost using environmentally friendly reagents and is still an ongoing issue. In fact, in 2007 the American Chemical Society Green

Chemistry Institute (comprising members from major pharmaceutical industries worldwide) voted “amide formation avoiding poor atom economy reagents” as the top challenge for organic chemistry. [49]. Current methods to chemically form peptide bonds generally involve solution phase synthesis (SPS) and solid-state peptide synthesis (SSPS).

Classic SPS is based on coupling a single amino acid in solution. This method can reliably synthesize 4-22 amino acids in length through rapid, continuous solution methods using blocking agents Fmoc and Bsmoc [50]. Longer peptides can be formed by combining smaller SPS synthesized peptides together. SPS is an attractive technique due to its relatively low reagent cost and straightforward scale up methods. This method’s downfall is that it requires purification and characterization after every step [51]. Additionally, the solubility of the peptide worsens and become more unpredictable with chain length [51]. Solution synthesis is especially valuable for large-scale manufacturing and for specialized laboratory applications [52]; however SSPS is a more practical part of present day scientific research with the aid of automated instruments.

In SPPS, the chemistry that has been used in solution-phase methods has remained similar, but the covalent attachment step links the nascent peptide chain to an insoluble polymeric support [53]. The anchored peptide is then extended by a series of addition cycles. In this approach it is essential that the reactions are driven to completion by the use of excess soluble reagents, which can then be removed through several simple wash steps [53]. This method, while effective, can lead to an increased cost in the synthesis of longer proteins. On the other hand, SPPS methods tend be very cost effective for short proteins, as compared to recombinant protein synthesis methods [54]. SPPS can also utilize generic chemical and purification process for peptide manufacturing up to the multi-100 kg scale [54].

SPPS methods have led to the development of several innovative pharmaceutical aids. Calcitonin (32 amino acids; Cancer) [55], Desmopressin (9 amino acids; Diabetes insipidus) [56], and Leuprolide (10 amino acids; Cancer) [57] are all made by SPPS methods. Many of these products are made at large scale. For instance, HIV fusion inhibitor (36 amino acids; Fuzeon) and direct thrombin inhibitor (20 amino acids; bivalirudin) are routinely manufactured on a large scale that exceeds 100 kg per year [58]. Despite these successes, large-scale peptide manufacturing presents several challenges that are not well served

by traditional SPPS processes. First, producing highly pure materials devoid of impurities during peptide coupling is difficult. Second, fragment coupling can also lead to poor yields. Third, large excesses of protected amino acid, coupling agents and additives are very expensive. For instance, the synthesis of 1,000 kg of Fuzeon requires roughly 45,000 kg of raw material, not including the solvent used in the synthesis or purification [58].

In sum, SPS and SPPS are amazingly versatile methods that can incorporate any amino acid, including those outside the 20 standard amino acids used in nature. As the field of peptide science grows so will the need for these methods. Many chemically synthesized peptide drugs have already been approved by the FDA and have reached the market. SPS and SPPS also opens the possibility for a wide range of products not available in nature to be synthesized and studied. For instance, SPS and SPPS has been widely used for the synthesis of computer-aided designed novel peptides as well as the screening of peptide libraries for drug targets [59]. While SPS and SPPS are ultimately limited to short peptide lengths due to the diminishing returns given by the production of longer peptides, these methods will continue to innovate and improve as the demand for cheap, pure, short peptides increases.

2.4 Expressing Proteins with Post-Translational Modifications

As the demand for high-quality, large (>100 amino acids) proteins increases, more cost effective strategies are necessary for the commercial synthesis of proteins. Utilizing the biosynthetic potential of the cell, proteins can be synthesized from a DNA template accurately (error of 10^{-4}), quickly (20 amino acids per second) [60] and at high yield (g/L levels) [61] at lengths of 100s-1000s of amino acids. Because proteins containing nsAAs are especially valuable for their therapeutic application (i.e. monoclonal antibodies and fc-fusion proteins), the research and pharmaceutical sectors have utilized a variety of cell lines to take advantage of their native PTM mechanisms.

Numerous prokaryotic- and eukaryotic-based expression systems have been developed to facilitate the production of recombinant proteins. The most prominent being used in industry are based on recombinant *E.coli*, yeast, and mammalian cell lines, along with insect cell lines and transgenic animals [62]. Because many microbial (*E. coli* and yeast) cell lines lack PTMs similar to those in humans,

mammalian cell lines have dominated the pharmaceutical industry. Several rodent- or human-derived cells like Baby Hamster Kindey (BHK), Chinese Hamster Ovary (CHO), HeLa and HepG2 cell lines are frequently used in biomedical research for heterologous protein expression with nearly 70% of all recombinant protein therapeutics being produced in CHO cells [63].

While over 100 PTMs have been characterized in nature [64] only a subset are used in therapeutic proteins including glycosylation, proteolytic processing, disulfide bond formation and (to a more limited extent) carboxylation, hydroxylation, sulfation and amidation [65]. These PTMs in particular are known to influence product stability and biological activity, while other PTMs like acetylation, ADP ribosylation and phosphorylation regulate various intracellular processes like gene expression, endosomal vesicle trafficking and signal transduction, which are less common characteristics in biopharmaceuticals [65]. Additionally, many biopharmaceuticals bear a combination of two or more PTMs. For example, activated protein C (Xigris) is carboxylated, hydroxylated, glycosylated, proteolytically processed and processes a disulfide linkage [66].

Despite the growing understanding of PTM mechanisms, controlling the PTM patterns has proven to be extremely difficult due to its heavy sensitivity to cell culture parameters. For example, the glycosylation pattern of a recombinant glycoprotein can be disrupted by pH [67], the buildup of ammonia primarily as a by-product of glutamine metabolism [68] and oxygen content of the bioreactor [69]. Being able to control the exact PTM and its location is essential, to the secretion, drug efficacy and stability of biopharmaceuticals. Additionally, regulators require a detailed characterization of all PTM profiles and batch-to-batch consistency of biopharmaceuticals before they are approved by the FDA. Thus, controlling exact PTM pathways and patterns of the protein of interest is a huge obstacle to an industry that is still in its infancy.

In sum, there has been considerable success in the development and engineering of production cells to deliver proteins optimized for given PTMs. In particular monoclonal antibodies (Mabs) are the largest and fastest growing sector within the biopharmaceutical industry [70]. They are increasingly being used in therapy because of their high specificity, low toxicity, long half-life, predictable pharmacokinetics

and high dose demand [70]. However, despite these successes, control of PTMs is still an emerging field. While it is possible to optimize a cell line for the production of a particular therapeutic protein the same cell line may be suboptimal or deficient for production of another, extending R&D timeline for these pharmaceuticals. Additionally, these methods are ultimately limited to native PTM mechanisms and constrained to established expression platforms. In order to have more designability in protein expression inside cells alternative routes like modifying the translation system to incorporate nsAAs is needed.

2.5 Installing NsAAs in Proteins Co-Translationally

Utilizing the native translation system to incorporate nsAAs has the potential to enable template-guided protein synthesis with novel chemistries not present in current biological systems and not limited by protein length. Already researchers have demonstrated the incorporation of over 100 nsAAs using these methods. However, this field is still emerging due to limited knowledge of the inner workings of the translation system.

This field is built upon the discovery of exceptions to the standard genetic code. While it was known that 64 codons exist (61 sense codons and 3 nonsense codons) as early as the 1970s it was found that yeast's mitochondria UGA codon (traditionally a stop codon) was assigned to encode tryptophan [71]. Today there are over 20 known variations to the standard genetic code [72]. Selenocysteine (Sec), known as the 21st amino acid, is one such example. Sec is encoded by the UGA codon in many species including humans [73]. Additionally, UAG genetically encodes pyrrolysine (Pyl), which is essential in trimethylamine metabolism in archaeal methanogens, in some organisms (for example, Desulfitobacterium) [74]. Ambiguous decoding in yeast and "open" (unused) codons in some organisms (i.e. Mycoplasma) were other dogma-breaking findings that reshaped our view of protein synthesis and the genetic code. Given these findings many of the basic assumptions underlining the presumed immutability of the code are now known to be false or incomplete. Cells tolerate [75] and can even derive selective advantage from ambiguous decoding [76]. Cells encode more than 20 amino acids and codon reassignment and recoding is possible.

Given this foundational wisdom, researchers have begun to hijack the native translational system to incorporate nsAAs. The process is complex and involves many parts but offers benefits that traditional SPS/SSPS methodologies do not. Expanding the genetic code beyond the 20 standard amino acids requires (i) an open code to encode (often UAG) (ii) a nsAA that can permeate the cell, (iii) an aminoacyl-tRNA synthetase (aaRS) capable of efficiently ligating a desired nsAA, (iv) a tRNA that can decode the open codon and (v) compatible elongation factors and ribosomes (**Figure 2.1**). Developing an efficient orthogonal translation system (OTS) thus demands optimization of all the above components. By definition, the aaRS-tRNA orthogonal pair must not cross react with endogenous aaRS-tRNA pairs and are in this way 'orthogonal' to the translation machinery of the host cell. For genetic code expansion in *E.coli*, the most successful orthogonal pairs are derived from archaea. Tyrosyl-tRNA synthetase from *Methanocaldococcus jannaschii*, pyrrolysl-tRNA synthetase (PylRS) from *Methanosarcina* species and phosphoseryl-tRNA synthetase (SepRS) found in archaeal methanogens, are the main vehicles for code expansion. The tyrosyl-tRNA synthetase system has been used to install a diverse array of tyrosine derivatives [77, 78], whereas PylRS and its engineered variants support translation with lysine, phenylalanine and Pyl derivatives, and click-chemistry-reactive nsAAs [79-81]. Phosphoseryl-tRNA synthase, which has a natural function in cysteine biosynthesis and Cys-tRNA^{Cys} formation in archaea, was instrumental for expanding the genetic code of *E.coli* with o-phosphorserine [41, 82, 83].

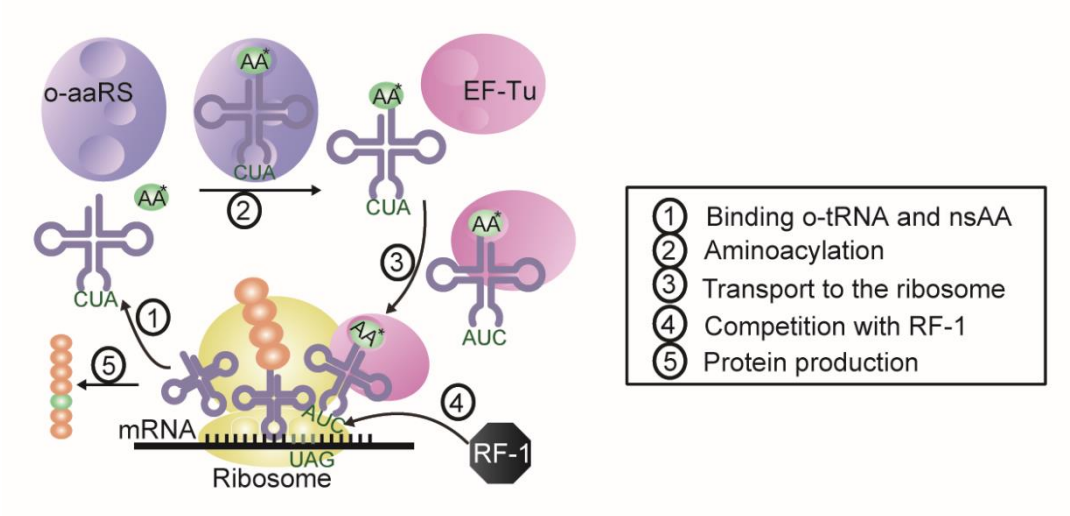


Figure 2.1. Schematic representation of co-translational incorporation of a nonstandard amino acid (nsAA) using an orthogonal translation system and amber suppression. The orthogonal aminoacyl-tRNA synthetase (o-aaRS) first binds its cognate nsAA and cognate o-tRNA. The o-aaRS then catalyzes the aminoacylation of the o-tRNA. The aminoacyl-tRNA (aa-tRNA) is then released from the o-aaRS and transported to the ribosome by the elongation factor-Tu (EF-Tu). The nsAA-o-tRNA associates with the A-site of the ribosome and its anticodon then binds the complementary triplet codon of the messenger RNA (mRNA). The ribosome then ligates the nsAA to the growing peptide chain. When release factor 1 (RF-1) outcompetes the nsAA-o-tRNA for binding at the UAG amber stop codon, the protein is truncated, which results in a decrease of nsAA incorporation efficiency.

Despite the development of numerous OTSs many challenges remain. First, mutagenesis strategies to change the specificity of o-aaRSs drastically decreases the o-aaRS binding affinity to the engineered o-tRNA ~100 fold and reduces aminoacylation efficiencies 800- to 7000-fold [84]. This creates o-aaRSs that aminoacylate nsAAs at nearly the same levels as near-cognate amino acids, drastically increasing non-specific incorporation at UAG codons. Second, depending on the OTS, there can be EF-Tu compatibility issues. For example, o-phosphoserine-tRNA does not bind to the native EF-Tu due to o-phosphoserine's negative charge [41]. This OTS required an additional EF-Tu to be expressed on a plasmid that was mutated to bind the nsAA-o-tRNA and transport it to the ribosome for successful nsAA incorporation. Third, there may also be ribosome compatibility issues as the size of the nsAA increases given that the ribosomal exit tunnel diameter is ~10-28 Å wide [85]. Lastly, because nsAA are often assigned to the amber (UAG) stop codon, there is an inherent competition event between the nsAA-o-tRNA and release factor-1 (RF-1). Endogenous RF-1 recognizes UAG codons and subsequently activates hydrolysis and the peptidyl-tRNA to release the peptide chain. During nsAA incorporation using

amber suppression, RF-1 competes with nsAA-o-tRNA, resulting in a significant amount of truncated product and reduced nsAA incorporation efficiencies. Recently, researchers developed a RF-1 deficient strain in which all 321 amber codons were recoded to TAA, completely freeing the amber codon to encode nsAAs and removing the nsAA-o-tRNA/RF-1 competition event [42].

2.6 Frontiers in Incorporating NsAAs Co-Translationally

Given the progress made in installing nsAA co-translationally, this method has been applied to many novel platforms for a diverse suite of exciting applications. These innovations include (i) growing the number of codons available for recoding using engineered ribosomes, (ii) introducing nsAAs into proteins using eukaryotic cells, and (iii) applying nsAA work in a cell-free environment. These applications, along with others, will heighten progress made in the nsAA field using co-translational incorporation techniques.

To increase the number of codons available for recoding, J. Chin and coworkers developed an evolved orthogonal ribosome (o-ribosome) that operated in parallel to the native ribosome [86]. The o-ribosomes were engineered to use four-base codons to enable more than 200 nsAA combinations to be encoded in a protein [87]. Although this method has an incorporation efficiency of ~20% using two amber codons with p-azido-L- phenylalanine (pAzF) and N6-[(2-propynyloxy)carbonyl]-L-Lysine, this method holds much promise to encode multiple, different nsAAs into one protein.

Additionally, outstanding progress has been made in expanding the genetic code available in eukaryotic cells. This has involved extensive development of new OTSs, considering that the new OTS must be orthogonal to the native translation system which varies greatly between higher-order species. NsAAs have been successfully implemented in yeast [31], CHO cells [32], *Drosophila melanogaster* [33], and *Caenorhabditis elegans* [34]. However, incorporation of nsAAs in higher order organisms runs into unique obstacles such as maintaining stability of the OTS system [88].

Cell-free protein synthesis (CFPS) systems offer another approach for the incorporation of nsAAs. CFPS is the *in vitro* synthesis of proteins without using intact cells. The lack of physical boundaries permits precise manipulation of reaction contents, simplifies product production, and enables

efficient incorporation of bulky or charged nsAA that typically exhibit poor membrane permeability *in vivo* [38]. Production of modified proteins in CFPS has seen radical innovations in the last several years. Albayrak and Swartz demonstrated the production of 1.5 mg/mL of sfGFP containing pAzF (azido; click chemistry) and p-acetylphenylalanine (pAcF; keto, orthogonal reactivity) with a 50% –88% amber suppression efficiency [89]. This success was possible by optimizing the concentration of orthogonal translational components in the open environment of CFPS. Cell-free systems also show great promise as a evolution platform for synthesizing new o-aaRSs. Because the intracellular concentration of nsAAs are often high due to limited nsAA export and catabolism mechanisms, o-aaRS evolved *in vivo* have a higher K_m value for the target nsAA. CFPS systems allow the nsAA concentration to be controlled, permitting for evolved o-aaRSs to reach lower K_m values and thus higher specificity.

In sum, co-translationally incorporating nsAAs holds astounding potential for transforming the way proteins containing nsAAs are synthesized. Due to the high protein synthesis rate of the cell and its ability to accurately synthesis large proteins, co-translational incorporation methods can meet the gap between chemical synthesis methods and eukaryotic expression platforms; however, in order accomplish these goals OTSs need to be optimized for improved aminoacylation efficiency and nsAA specificity. In addition, genome-engineering efforts to remove negative effectors of *in vivo* protein synthesis (e.g. nucleases, proteases) will increase yields of nsAA-containing proteins in chassis strains [90].

2.7 Conclusions and Outlook

An increased interest in producing biopharmaceuticals has resulted in drastic improvements of chemical peptide synthesis, development of recombinant production strains capable of decorating proteins post-translationally and advancements in expanding the genetic code to incorporation nsAAs co-translationally. These complementary methods have shown to be vital to the study of protein structure and cellular biology [47] and the synthesis of new protein-based materials [46]. To increase the yield and purity of all three methods each individual process has several opportunities for improvement.

In the next decade, chemical synthesis methods for production of short peptides will need to increase efficiency of peptide bond formation while reducing production costs, perhaps through using

more environmentally friendly reagents. Protein expression strains will begin to become more specialized for specific PTMs as our understanding of PTM pathways and profiles improve. The addition of heterologous PTM pathways into *E. coli* is one way researchers are approaching this problem, studying pathways in a more simplified, “empty” chassis. Finally, co-translational incorporation of nsAAs holds great potential to produce high-yielding amounts of pure modified proteins; however much more research is need to improve o-aaRS’ aminoacylation efficiency and specificity.

In conclusion, expansion of genetically encoded chemistry is now possible using a variety of highly sophisticated techniques. Each of the three techniques detailed here are complementary and equally important for research development as our ability to synthesize modified proteins improves.

CHAPTER 3

3 Improving Genomically Recoded *Escherichia Coli* for *In Vivo* Production of Proteins Containing Non-Standard Amino Acids

3.1 Abstract

The genetic code, a cipher for translating mRNA codons into the twenty standard amino acids used in protein synthesis, was once thought to be immutable. However, over time more variations in the genetic code were found across many species. Hijacking these mechanisms allow for non-standard, chemically diverse amino acids to be incorporated into proteins by the ribosome *in vivo*. This expansion of the genetic code has the potential to transform how we synthesize materials and therapeutics, investigate protein structure, and understand the evolution of the translation system.

A pioneering effort has recently developed an *Escherichia coli* strain lacking all TAG amber stop codons and release factor 1 which allows for more efficient genetic encoding of an additional, non-standard amino acid (nsAA). However, this strain has not previously been optimized for protein production which is critical for its industrial application and wide spread use. Here, we describe the construction of a series of genomically recoded organisms that are optimized for protein production and have the additional ability to tune protein expression through a T7 RNA polymerase (T7RNAP). We found that reduction of nuclease and protease activity increases wild-type sfGFP production by 260% and sfGFP containing two nsAAs production by 2.2- and 5.6-fold with p-azidophenylalanine and N_6-(propargyloxycarbonyl)-$_L$-Lysine (ProCarb), respectively. Additionally, we constructed several strains containing an IPTG-inducible T7RNAP cassette which shows a 17-fold improvement in production of sfGFP containing two ProCarbs. We envision that our library of strains will provide the community with multiple strain options for expression of proteins containing nsAAs with increased protein yield.

3.2 Introduction

Essential to the central dogma, the genetic code is a universal code that describes how mRNA codons are translated into twenty standard amino acids. Of the 64 possible codons, 61 codons translate

the twenty amino acids and the three remaining codons (UAA, UAG, UGA) are responsible for termination of protein synthesis [91]. This biochemical principle extends through all kingdoms of life. However, by the late 1970's variations in the genetic code were being discovered everywhere. Stop codons and synonymous amino acid codons in particular were found to be reassigned to other amino acids, including those outside the standard twenty [92].

Today over twenty variations to the standard genetic code have been revealed [72]. These findings have spurred interest in utilizing genetic code variations to modify the codon table and encode for non-standard amino acids (nsAAs). Expanding the set of amino acids capable of being incorporated co-translationally by the ribosome opens opportunities to site-specifically introduce new chemistries into proteins and has the potential to transform how we synthesize materials, study protein structure and understand the translation system. For instance, synthesis of high molecular weight, high yielding polypeptides can be achieved inside cells [93] while it can be difficult using other methods [52]. nsAAs have been utilized for biophysical studies [94, 95], creating new biocatalysts [96], and synthesizing proteins containing post-translational modifications [82, 83, 97]. Probing the translation system *in vivo* will also expand our understanding of how translational functions and its evolution over time [98].

Currently over 150 nsAAs have been incorporated into proteins co-translationally [99, 100]. This is most often achieved by introducing an orthogonal translation system (OTS) that genetically encodes nsAAs at the UAG stop codon, the least used stop codon in the *Escherichia coli (E. coli*) genome. This system is orthogonal in that it should not interact with the native translation components. The OTS is expressed off the pEVOL plasmid [101] which includes two copies of an orthogonal aminoacyl-tRNA synthetase (o-aaRS) and orthogonal tRNA (o-tRNA). The o-tRNA has a modified anticodon specific to the UAG stop codon, whereas the o-aaRS is evolved to bind and aminoacylate the nsAA of interest to the o-tRNA [102]. This process, which was pioneered by Schultz and colleagues [103], is called amber suppression as it allows the nsAA to be encoded at the UAG (amber) stop codon.

Despite this progress, a key limitation to OTSs is premature truncation of the recombinant protein at the UAG codon by endogenous release factor-1 (RF-1). RF-1 recognizes and binds at amber codons, subsequently activating hydrolysis of peptidyl-tRNA to release the peptide chain [104]. The inherent

competition between RF-1 and the nsAA-o-tRNA at the amber codon causes inefficient incorporation of nsAAs and reduces the modified protein yield. Fortunately, major advances have removed amber codon competition. An *E. coli* strain lacking all 321 UAG amber stop codons and RF-1, termed *C321.*Δ*A*, frees a codon for total dedication to an additional nsAA [1]. This strain has an increased ability to incorporate multiple nsAAs as compared to other *E. coli* strains, enabling many applications [83, 90, 105, 106] most notably in biocontainment [106].

Despite these advances, *C321.*Δ*A* has not been optimized for protein production. As opposed to standard commercially available protein production strains, like BL21(DE3), *C321.*Δ*A* is directly derived from the K-strain *MG1655* (considered wild-type *E. coli*) and has not undergone strain development for protein production. Given previous work to improve protein production in *E. coli* by removing negative effectors [107, 108], we sought to introduce previously characterized mutations into *C321.*Δ*A* and test their impact on protein production in *C321.*Δ*A* (**Figure 3.1**).

Here we target DNAase *endA* [90, 109], RNAases *rne* [110] and *rnb* [111, 112] and proteases *lon* [113] and *ompT* [113] (**Table 3.1**). We found that reduction of nuclease and protease activity increases sfGFP containing 2TAGs production by 2.3- and 5.6-fold with p-azidophenylalanine (pAzF) (“click” chemistry; photoreactive crosslinker) and N_6-(propargyloxycarbonyl)-$_L$-Lysine (ProCarb) (pyrrolysine analog), respectively. Given the advantages of high recombinant protein expression with the T7 promoter system [114, 115], we also introduced this system into *C321.*Δ*A* to increase its utility. Several strains were constructed containing an isopropyl β-$_D$-1-thiogalactopyranoside (IPTG) inducible T7RNAP cassette which showed a 17-fold improvement in production of sfGFP containing 2TAGs with ProCarb. We envision that our library of strains will provide the community with multiple strain options for expression of proteins containing nsAAs with increased protein yield.

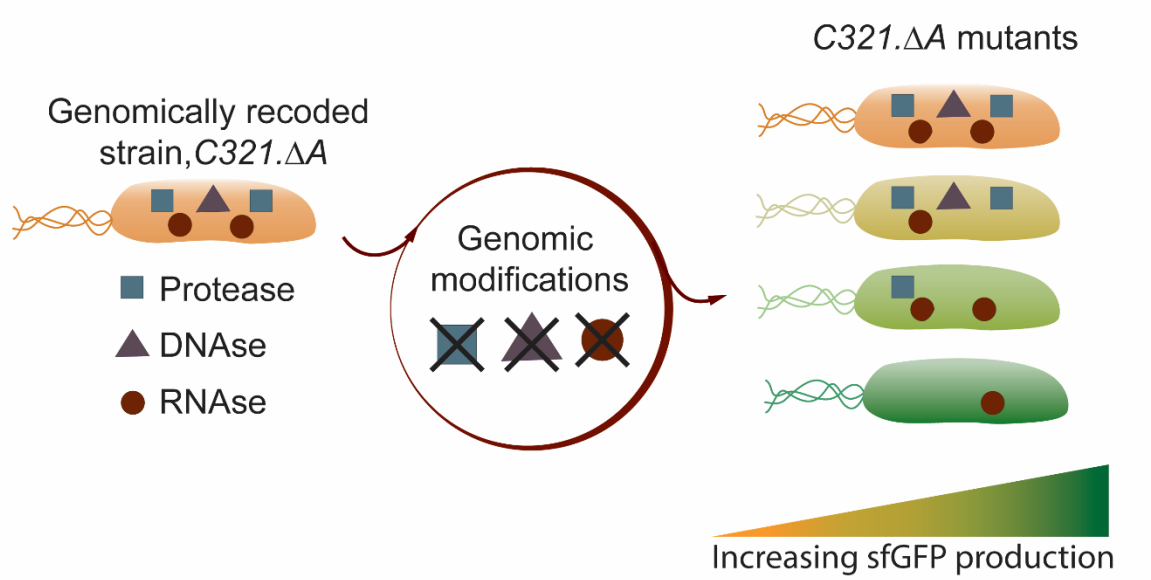


Figure 3.1: Reducing protease and nuclease activity in *C321.ΔA* via genomic engineering to increase protein production capacity. Mutagenic oligonucleotides were introduced into *C321.ΔA*, targeting two proteases (*lon* and *ompT*), two RNAase (*rne* and *rnb*), and a DNAase (*endA*) to reduce protease and nuclease activity. Through multiple rounds of multiplex automated genome engineering (MAGE) several *C321.ΔA* mutants were generated. Changes in protein production capacity were characterized by expressing sfGFP.

Table 3.1: Proteases and nucleases targeted in this study.

General Function	Gene	Specific Function	Mutation	Phenotype	Reference
DNA Stability	endA	Endonuclease I	GGATGT 748 TAACTGA	Truncation at 748 nt	[116]
RNA Stability	rne	RNase B	GGT 584 TAACTGA	Truncation at 584 nt	[110]
	rnb	RNAase E	GACGCC 632 TAACTGA	Truncation at 632 nt	[117]
Protein Stability	lon	ATP-dependent protease	Removal of promoter	Unknown	[118]
	ompT	Outer membrane protease VII	D103A	Elimination of proteolytic activity	[119, 120]

3.3 Results and Discussion

3.3.1 Introducing genetic mutations into proteases and nucleases

The functional inactivation of two proteases, two RNAses and one DNAse was performed using multiplex automated genome engineering (MAGE) [121] (**Table 3.1**). The lon protease was functionally deactivated using a mutagenic MAGE oligonucleotide to remove its promoter. This mutation is similar to the lon protease mutation found in *BL21(DE3)* where a transposable element, IS186, inserted directly into the *lon* promoter preventing lon's expression [122]. The point mutation D103A was introduced into *ompT* to eliminate proteolytic activity while maintaining its structural motifs due to OmpT's possible chaperone function [119, 120]. RNAse E, encoded by *rne*, was truncated by inserting a stop codon at nucleotide 131 (*rne131*), which has been found to increase mRNA half-life [110, 123]. RNAse II, encoded by *rnb* and involved with mRNA degradation [124], and Endonuclease I, encoded by *endA* and generates breaks in double-stranded DNA [125], were truncated by inserting a stop codon followed by a frameshift in the first ¼ portion of the reading frame [121]. We hypothesized that nuclease activity reduction will stabilize both the DNA template and mRNA transcript for the recombinant protein of interest.

Starting with the parental strain, *C321.ΔA*, these mutations were made in single, double and some triple and quadruple combinations. Mutations were screened by multiplex allele-specific colony (MASC) PCR [22] or colony PCR which amplifies mutation regions and confirmed by DNA sequencing. The average doubling time for each MAGE-modified strain was measured in 2x2x YT media and was found to be within 12% of the parental strain (**Figure 3.2**), suggesting that the gene disruptions did not drastically affect cellular fitness.

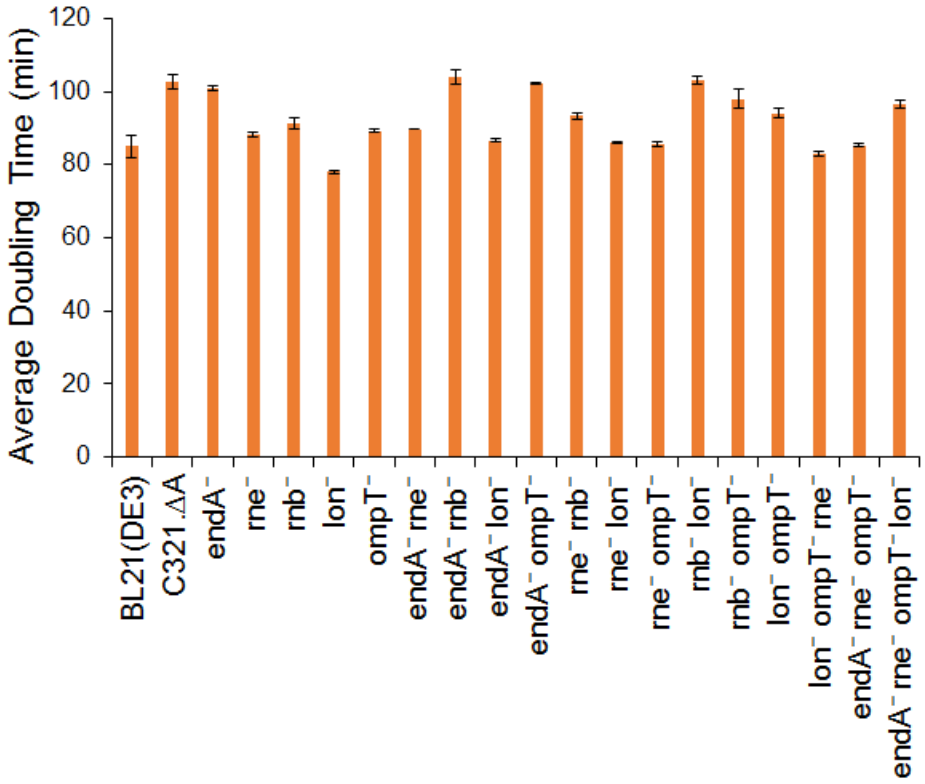


Figure 3.2: Growth rates of *C321.ΔA* mutants. Strains were grown at 32^{0}C in 2x YT media in a sterile 96-well polystyrene plates (Costar 3370; Corning Incorporated, Corning, NY, USA) in a Synergy H1 plate reader (BioTek, Winooski, VT, USA). Optical density at 600 nm was measured every 10 min for 12 hr. Error bars represent biological duplicates and technical triplicates.

3.3.2 Reduced protease and nuclease activity enhances sfGFP production in C321.ΔA

To assess the protein production capacity of *C321.ΔA* and its mutants a reporter plasmid and an orthogonal translation system (OTS) plasmid were transformed into all strains. The OTS plasmid expresses all the necessary components to incorporate nsAAs into the reporter protein, sfGFP. We utilized a relatively efficient nsAA system, pAzF, which requires the pEVOL-pAzF plasmid [101]. Testing the inherent ability of the mutants to express protein, all strains were transformed with the pLpp5-sfGFP-wt plasmid, which expresses wild-type sfGFP (sfGFP-wt) off a strong endogenous promoter, Lpp5 [126], and is IPTG-inducible (**Figure 3.3A**).

The *rne*$^-$ mutant was the single mutant with the strongest impact on sfGFP-wt expression, implying that mRNA stability may be the largest limitation for expression of sfGFP-wt under these

conditions. Interestingly, the mutation combination most similar to *BL21(DE3), lon⁻ompT⁻*, expressed sfGFP-wt at levels 44% less than *BL21(DE3).* This discrepancy most likely stems the differences inherent in B-strain and K-strains but was not explored in this work. Furthermore, we observed that the addition of endA⁻ to rne⁻ containing strains added a minor boost in sfGFP-wt expression. The top mutants for expression of sfGFP-wt were *rne⁻*, *endA⁻rne⁻*, and *ompT⁻rne⁻endA⁻,* with the top mutant exceeding *BL21(DE3)* and the parental strain by 2.6-fold.

To test the ability of *C321.ΔA* mutants to express proteins while incorporating nsAAs, sfGFP containing two amber stop codons (sfGFP-2TAG) was expressed with the pAzF OTS (**Figure 3.3B**). In this case, if a nsAA is incorporated into each amber codon, full length sfGFP will be expressed and fluoresce. In the absence of nsAAs, any fluorescence measured will be a result of non-specific incorporation at an amber codon. Under these conditions, BL21(DE3)'s ability to express sfGFP-2TAG is reduced compared to *C321.ΔA* perhaps due to RF-1 being present in BL21(DE3). The *rne⁻* strain and several mutations combined with *rne⁻* showed benefits as in the sfGFP-wt case. The top mutant in this case was *rne⁻endA⁻* with a 2.3-fold improvement as compared to the parental strain. *lon⁻ompT⁻rne⁻endA⁻* was also a top performer. Absolute protein expression was quantified by purifying sfGFP using a Strep-tag-Strep-Tacin column after a 20 hr expression assay in 1 L of 2x YT media (**Supplemental Table A1**).

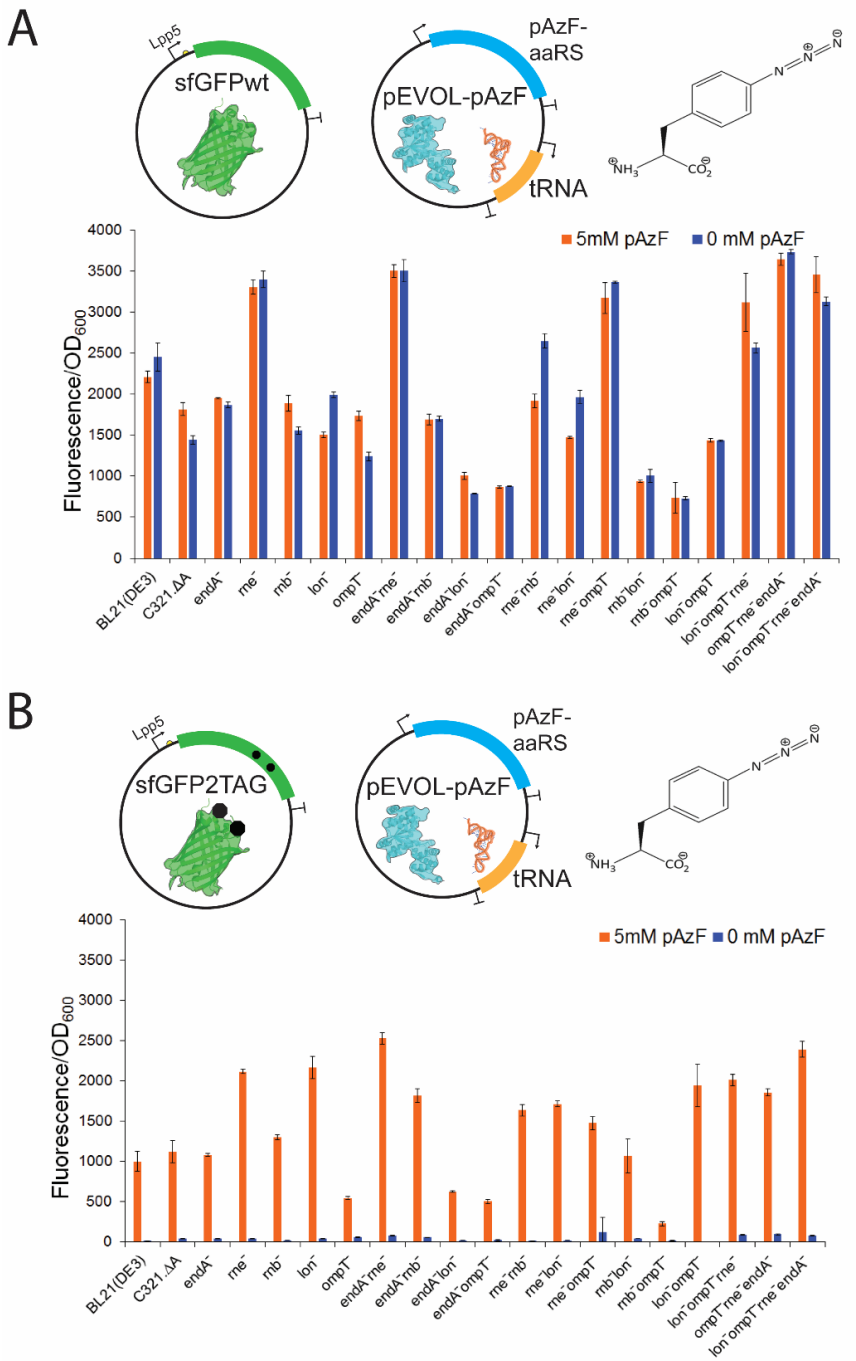


Figure 3.3: Effect of reducing protease and nuclease activity on sfGFP expression in *C321.ΔA*. A) The protein production capability of the modified *C321.ΔA* strains were analyzed by expressing wild-type sfGFP (sfGFP-wt), regulated by a strong endogenous promoter pLpp5, and the pAzF orthogonal translation system expressed on pEVOL-pAzF. For all conditions 1mM IPTG, 0.02% arabinose and 5mM pAzF (orange bars) or 0mM pAzF (blue bars) were added at OD_{600} 0.6-0.8. B) Modified *C321.ΔA* strains were analyzed for the ability to suppress two amber codons in sfGFP at positions 190 and 212 in the presence (orange) or absence (blue) of 5mM pAzF. For all panels error bars represent one standard deviation for biological triplicates and technical triplicates.

In order to test the limits of the system, top sfGFP-wt and sfGFP-2TAG expressing mutants were tested for its ability to incorporate pAzF at 10 amber codons. Because sfGFP cannot incorporate 10 pAzFs without significant loss of fluorescence, we transformed top *C321.ΔA* mutants with a plasmid expressing an elastic-like polymer (ELP) containing 10TAGs with sfGFP-wt fused to its C-terminus and the pEVOL-pAzF plasmid (**Figure 3.4A**). In this case, if there is read-through of all 10TAGs then sfGFP-wt will be expressed and fluorescence will be detected. The advantage of the genomically recoded strain over *BL21(DE3),* which contains RF-1, to incorporate multiple nsAAs is observed here. None of the mutants in this case displayed a significant improvement compared to *C321.ΔA*; however, the system showed a 12-fold improvement over *BL21(DE3).* This suggests that the strain's protein production capability is not the limiting factor under these conditions but rather the ability of the OTS components to incorporate nsAAs.

In order to test a condition where mRNA and protein stability may be a limitation, we expressed sfGFP2TAG and utilized the pyrrolysine (Pyl) OTS system from *Methanosarcina mazei* (*M. mazei*) which was cloned into the pEVOL backbone (pEVOL-MMpyl) [127] (**Figure 3.4B**). The Pyl OTS was chosen because the Pyl synthetase (PylRS) is known to be very difficult to express [128, 129]. We used a pyrrolysine derivative, ProCarb, instead of Pyl due to there being no commercial source for Pyl and its tedious and expensive chemical synthesis [130]. We observed that the *C321.ΔA* mutants had a drastic improvement over *C321.ΔA* and BL21(DE3) for expression of sfGFP containing two ProCarbs. The top mutant, *endA*$^{-}$*rne*$^{-}$*ompT*$^{-}$, showed a 5.6-fold improvement compared to the parental strain. These results demonstrate the advantage of reducing protease and nuclease activity when using known insoluble o-aaRSs, PylRS.

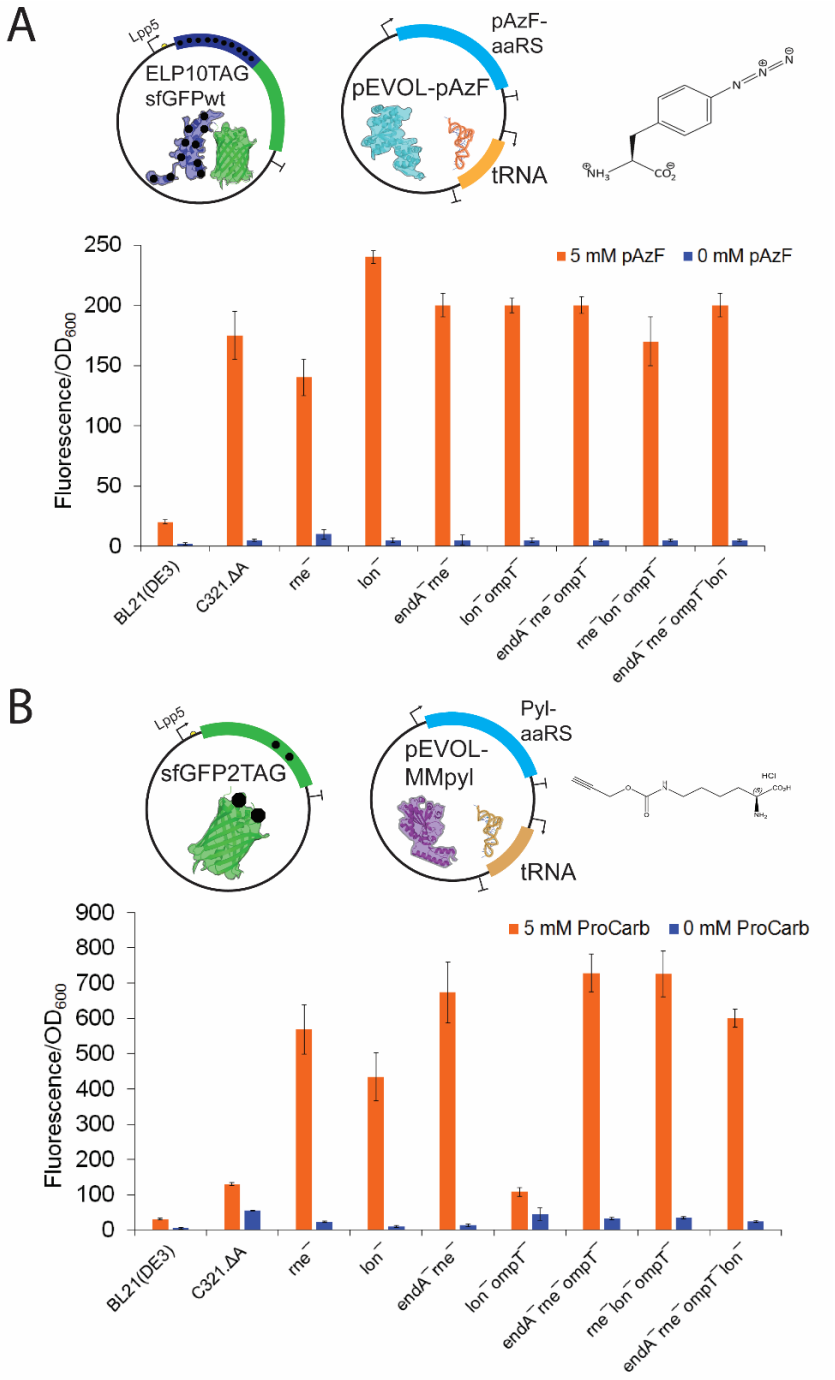


Figure 3.4: Effect of reducing protease and nuclease activity on multiple site incorporation and an additional orthogonal translation system. A) Elastin-like polypeptide (ELP) containing 10 amber codons and fused to sfGFP-wt at its C-terminus was expressed with pEVOL-pAzF in top *C321.ΔA* mutants. For all conditions 1mM IPTG, 0.02% arabinose and 5mM pAzF (orange bars) or 0mM pAzF (blue bars) were added at OD_{600} 0.6-0.8. B) Expression of sfGFP containing two amber codons with the orthogonal translational system for pyrrolysine analog, N_6-(propargyloxycarbonyl)-L-Lysine (ProCarb), was tested in top *C321.ΔA* mutants. For all panels error bars represent one standard deviation for biological triplicates and technical triplicates.

3.3.3 Introduction of T7RNAP to increase the utility of C321.ΔA

Tuning transcription is a powerful tool for efficient recombinant protein production in *E. coli*. Many challenges such as product toxicity, formation of inclusion bodies, and metabolic burden are associated with non-optimal (too high or too low) levels of recombinant protein expression. Tunable expression systems allow for the adjustment of recombinant protein expression using a small molecule inducer and the maximum exploitation of the cell's metabolic capability. Thus, the ability to tune recombinant protein expression is a staple for many protein expression projects. Within this realm, the T7 system within *BL21(DE3)* is the most popular approach for producing proteins due to its high activity, tunability and orthogonality, not expressed by endogenous polymerases. In this system, the DE3 cassette on the genome of *BL21(DE3)* encodes for a phage T7RNAP, induced by the addition of IPTG [113]. The gene of interest is cloned behind a T7 promoter and recognized exclusively by the phage T7RNAP. This allows for the recombinant protein expression to be tunable with the addition of IPTG and orthogonal.

The T7RNAP cassette was incorporated into *C321.ΔA* by amplifying T7RNAP from the BL21(DE3) genome and adding an upstream terminator to transcriptionally isolate the cassette and a CmR gene as a selectable marker (**Supplemental Figure A1**). The T7RNAP cassette also has 45 bp of genomic homology to the genomic insertion site on the 5' and 3' end. Using λ-red mediated homologous recombination [22, 131], the cassette was inserted into the top *C321.ΔA* mutant strains, the CmR marker was removed and the full cassette and insertion site were verified by sequencing (**Figure 3.5**).

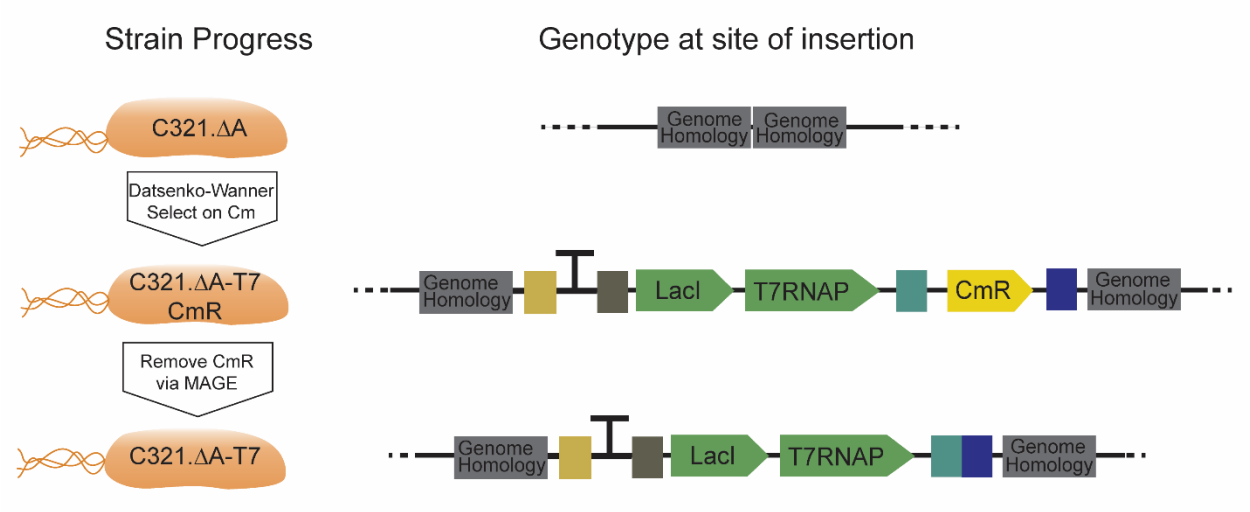


Figure 3.5: Genomic insertion of the T7RNAP cassette. The T7RNAP cassette was inserted into *C321.ΔA* using Datsenko-Wanner. After recovery, transformed cells were plated on LB plates containing 34 µg/mL chloramphenicol, selecting for strains that incorporated the T7RNAP cassette. The selectable marker was then removed through several cycles of multiplex automated genome engineering (MAGE). The resultant strain was termed *C321.ΔA-T7*.

To test functionality of the T7RNAP cassette, a reporter plasmid expressing sfGFP-wt or sfGFP-2TAG, regulated by a T7 promoter, was transformed into the *C321.ΔA-T7* strains along with pEVOL-pAzF. Here, all strains including BL21(DE3) expressed the reporter proteins at levels much lower than sfGFP-wt/2TAG expression regulated by Lpp5 (**Supplemental Figure A2**). Because T7 systems often utilize pET plasmids, the reporter plasmid was switched to a pET28a backbone [132]. With the new plasmid BL21(DE3)'s expression of sfGFP-wt (**Supplemental Figure A3A**) increased to near the same levels as expression of pLpp5-sfGFPwt. However, expression of sfGFP-wt, directed by T7RNAP, remained low in the *C321.ΔA* mutants.

For these reasons, we reconstructed the *C321.ΔA* strains containing T7RNAP by first adding the T7RNAP cassette into *C321.ΔA,* termed *β*, then introduced genomic mutations similar to the BL21(DE3) and BL21 Star (DE3) strains (contains an additional rne131 mutation). The doubling time of the strains in 2x YT media at 32^{0}C were ≤ 18% of the parental strain (**Figure 3.6**). In this case we did find an increase of 2.2-fold and 3.1-fold for sfGFP-wt/pAzF and sfGFP-2TAG/pAzF respectively in top *β* mutants when switching the order of strain construction (**Supplemental Figure A4**).

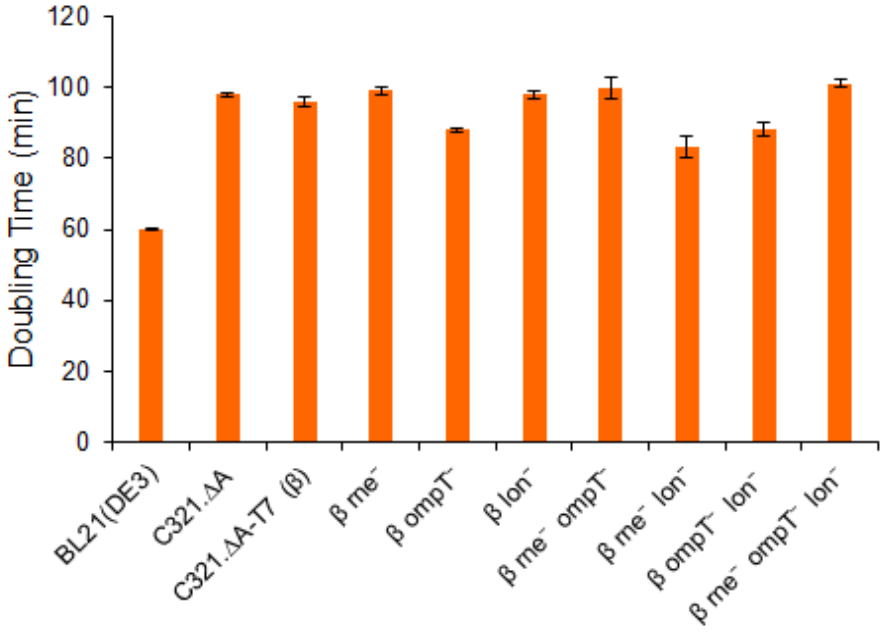


Figure 3.6. Growth rates of *C321.ΔA-T7* and its mutants. Strains were grown at 32^{0}C in 2x YT media in a sterile 96-well polystyrene plates in a Synergy H1 plate reader. Optical density at 600 nm was measured every 10 min for 12 hr. Error bars represent biological duplicates and technical triplicates.

In particular, *β rne⁻ompT⁻lon⁻* showed an improvement in sfGFP-wt of 3.7-fold over the parental strain, the same combination of mutations as BL21 Star (DE3) (**Figure 3.7**). When expressing pET28a-sfGFP2TAG with pEVOL-pAzF, *β rne⁻* showed a 1.9-fold improvement over *β* and 2.3-fold improvement over BL21(DE3). It appears that no matter the RNA polymerase used for expression, rne⁻ is the most beneficial mutation for expression of sfGFP in *C321.ΔA* strains. Lastly, when expressing sfGFP-2TAG with pEVOL-MMpyl, we observed a 17-fold improvement compared to *β*. In this case, we suspect the highest fold improvement was observed due to the poor solubility of PylRS.

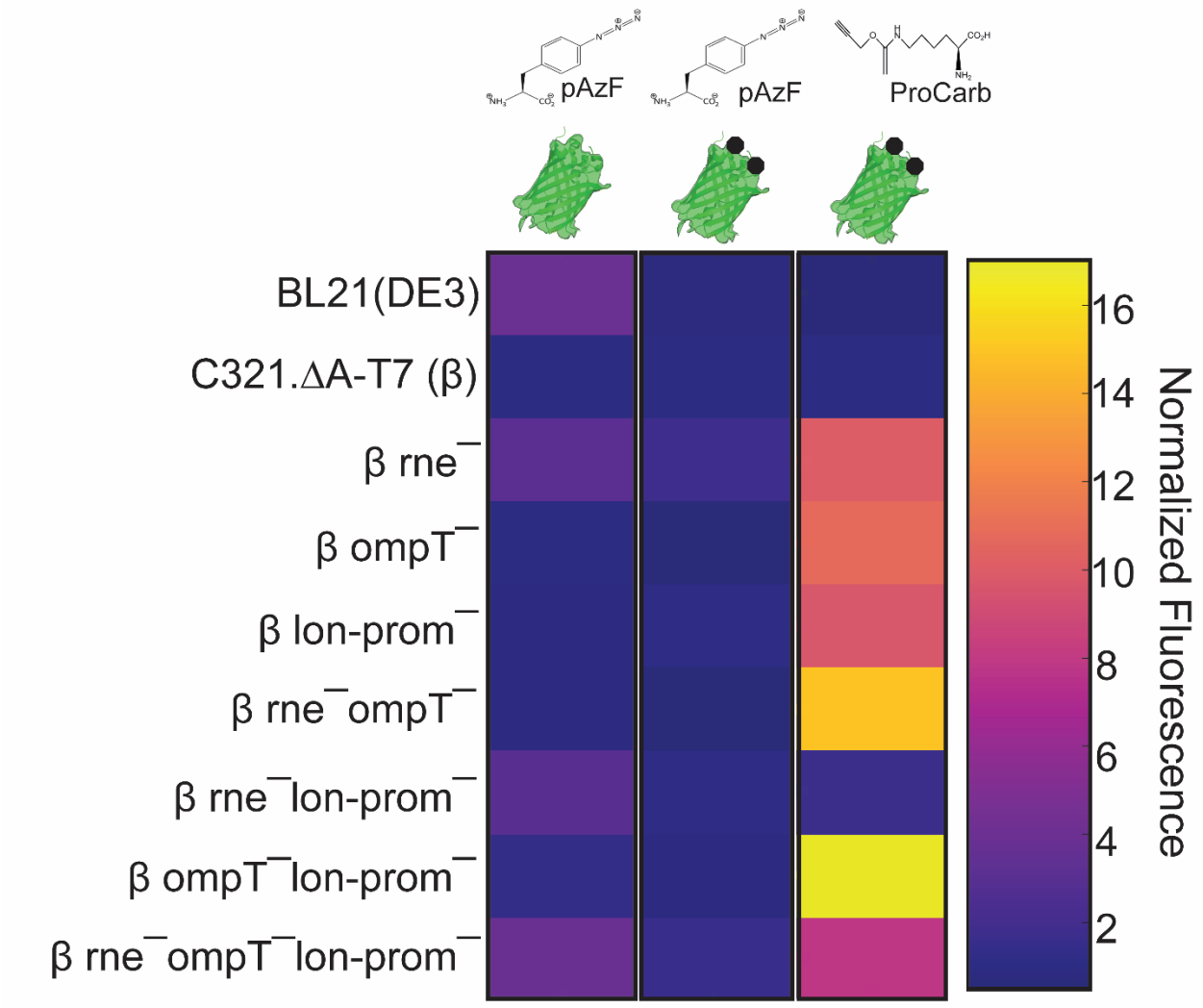


Figure 3.7. **Heat map of the effect of reducing protease and nuclease activity on sfGFP expression using a T7 system in *C321.ΔA*.** Heat map depicts the normalized fluorescence (Fluorescence/OD_{600}) for various reporter proteins relative to *C321.ΔA-T7(β)* within each vertical condition. Column 1: pET28a-sfGFP-wt + pEVOLpAzF + 5 mM pAzF + 1 mM IPTG + 0.02% Arabinose; Column 2: pET28a-sfGFP2TAG + pEVOLpAzF + 5 mM pAzF + 1 mM IPTG + 0.02% Arabinose; Column 3: pET28a-sfGFP2TAG + pEVOL-MMpyl + 5 mM ProCarb + 1 mM IPTG + 0.02% Arabinose. Normalized Fluorescence data is shown in **Supplemental Figures A4 and A5.**

3.4 Conclusions

A major limitation in the nsAA field is producing high quantities of proteins containing multiple, identical nsAAs. Researchers have sought to address this problem by improving aminoacylation rates through protein engineering [26-30, 105], optimizing OTS expression plasmids [101, 133] and removing RF-1 by recoding genomes [42]. This work seeks to contribute to the growing effort towards improving modified protein yields *in vivo*.

We showed that by reducing nuclease and protease activity, production of sfGFP and sfGFP-2TAG containing pAzF or ProCarb can be improved. This approach is especially useful for the synthesis of the difficult to express protein PylRS, where its N-terminal domain is known to be highly insoluble and often aggregates full-length PylRS [128, 129]. Creating a system that is optimized for the expression of PylRS presents the opportunity for further protein evolution studies to improve PylRS aminoacylation rates and its substrate specificity. Studies show that *M. mazei* PylRS' (used in this study) catalytic core reveals a deep hydrophobic pocket for binding of Pyl [134, 135] which allows PylRS to display a remarkably high tolerance towards a variety of substrates [136]. Thus, this work can enable higher yields for modified proteins containing nsAAs compatible with the Pyl OTS system, which has been previously limited by insolubility issues. Additionally, by introducing a T7RNAP cassette into *C321.ΔA* strains with reduced nuclease and protease activity, T7 based expression is productive and will enable wider use of these strains. *C321Δ.A-T7(β)* strains are also superior to BL21(DE3) for production of sfGFP containing multiple nsAAs.

As our ability to increase efficiencies of OTSs improve so will the need for optimized strains to produce proteins containing nsAAs. Additionally, the huge increase in demand for biopharmaceuticals necessitates alternative methods for producing high yielding modified protein. Moreover, the ability to site-specifically incorporate nsAAs into proteins represents a tool of limitless potential for systemic structure/function studies with precision beyond what is possible with conventional site-directed mutagenesis [83]. Often times these studies require nsAAs other than those available through post-translational modification mechanisms within eukaryotic strains, such as biophysical probes [137, 138]

and photoreactive side chains [139, 140]. For these reason, we believe these strains will be of great use to the community and impact the growing nsAA field.

3.5 Materials and Methods

3.5.1 Reagents, Buffers and Plasmids

Chemicals and media were purchased from Sigma Aldrich (St. Louise, MO, USA) unless otherwise designated. Phusion High-Fidelity DNA Polymerase, Taq DNA polymerase with Standard Taq Buffer, T4 DNA ligase, dNTP, Quick-load DNA Ladders, BL21(DE3) and restriction endonuclease were purchased from New England Biolabs (NEB, Ipswich, MA, USA). Multipex PCR Kits used for MASC PCR were purchased from QIAGEN (Hilden, NRW, DE). Plasmids were extracted using Omega E.Z.N.A DNA Isolation Kit (Omega Bio-Tek, Norcross, GA, USA). DNA was column purified or gel extracted using OMEGA HiBind DNA Mini Columns and OMEGA E.Z.N.A Gel Extraction Kit, respectively. Genomic DNA was isolated with Omega E.Z.N.A. Bacterial DNA Kit. All DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). The nsAA pAzF was purchased from P212121, LLC (Ann Arbor, MI, USA) and ProCarb was purchased from BioFine, Inc (Vancouver, BC, CA). SYBR Safe, used in all agarose gels, and DH5α were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Synthetic *E. coli* C321.ΔA (GenBank: CP06698.1) was received as a gift from Farren Isaacs. All oligonucleotides used for cloning are shown in **Supplemental Table A2**. All vectors were cloned using Gibson Assembly [141]. pLpp5 plasmids were derived from the pDTT1 vector. pET vectors were derived from pET28a vectors.

3.5.2 Construction of C321.ΔA mutants

The strains in this study were generated from *C321.ΔA* [42] by disrupting genes of interest using mutagenic oligonucleotides via MAGE [22] (**Supplemental Table A2**). Cultures were grown in LB-Lennox media (10 g/L Trypton, 5 g/L Yeast Extract, and 5 g/L NaCl) at 32^0C and 250 rpm throughout the MAGE cycle steps [22]. Single, double, several triple and quadruple mutations were made to endA, rne, rnb, lon, and ompT, to investigate the effect of reduced nuclease and protease activity on expression of protein containing multiple nsAAs. Multiplex allele-specific colony (MASC) PCR was performed to screen for

gene mutations by using wild-type forward (-wt-f) or mutant forward (-mut-f) primers and reverse primers (-r; **Supplemental Table A2**). Wild-type and mutant forward primers were identical except at the 3'-ends of the oligonucleotide, and the reverse primers were used for detection of both wild-type and mutant alleles. The mutant allele was amplified using the mutant forward and reverse promoter set (-mut-f and –r) which resulted in a band on an electrophoresis gel but not with the wild-type forward and reverse primer set (-wt-f and –r). MASC PCR was performed in 10 µL reactions by using a Multiplex Master Mix at 95^0C for 15 min, with 30 cycles of 95^0C for 30 s, 65^0C for 30 s, and 72^0C for 1 min, and a final extension of 72^0C for 5 min. Selection for lon mutants were performed separately in 10 µL reactions using Taq DNA polymerase with Standard Taq Buffer at 95^0C for 15 min, with 20 cycles of 95^0C for 30 s, 55^0C for 30 s, and 68^0C for 2min, and a final extension of 68^0C for 5 min. Mutant alleles were screened by running PCR products on a 2% agarose gel and confirmed by DNA sequencing by using sequencing primers (**Supplemental Table A2**).

3.5.3 Growth Curves

Overnight cultures of strains were grown in 2x YT (16 g/L Trypton, 10 g/L Yeast Extract, and 5 g/L NaCl) media at 32^0C at 250rpm and were diluted 1:50 in 100 µL of 2x YT media. Diluted cultures (100 µL) were added to 96-well polystyrene plates (Costar 3370; Corning Incorporated, Corning, NY, USA). The OD_{600} was measured at 10 min intervals for 20 hr at 32^0C in orbital shaking mode on a SynergyH1 plate reader (Biotek, Winooski, VT, USA). Growth data for each strain was obtained from three replicate wells and three independent cultures. Doubling time was calculated during exponential growth phase.

3.5.4 Assaying expression of GFP

Strains were freshly transformed with the plasmids of interest. A single colony was inoculated into 5 mL of 2x YT media with 35 µg/mL Kanamycin and 25 µg/mL Chloramphenicol ($Kan_{35}Cm_{25}$) grown overnight at 32^0C, 250 rpm. Overnight cultures were diluted 1:50 into 5 mL of fresh 2x YT media $Kan_{35}Cm_{25}$ in triplicate and grown at 32^0C at 250 rpm. OD_{600} was monitored on a Libra S4 spectrophotometer (Biochrom, Cambridge, UK) until OD_{600} 0.6-0.8 at which point cultures were induced. Inducers consisted of either 5 mM nsAA, 1 mM IPTG, and 0.02% arabinose or 0 mM nsAA, 1 mM IPTG,

and 0.02% arabinose. Cultures were allowed to express for 20-24 hr after induction prior to harvest. To assay fluorescence, overnight cultures were diluted 10-fold in 2x YT media $Kan_{35}Cm_{25}$. The OD_{600} of the 10-fold dilution was measured on a NanoDrop 2000c (Thermo Scientific, Waltham, USA) and multiplied by ten. 100 µL of the 10-fold dilution was added to 96-well polystyrene plates (Costar 3603) in triplicate. Fluorescence of the plates were measured on a Synergy H1 plate reader with a gain of 60. Normalized fluorescence was obtained by dividing fluorescence reading (normalized to 2x YT media $Kan_{35}Cm_{25}$ wells) by OD_{600} read on the NanoDrop 2000c.

3.5.5 Construction of T7RNAP cassette

The T7RNAP cassette was assembled from three pieces: a terminator (TM) piece, a T7RNAP piece, a CmR piece (**Supplemental Table A3**). To transcriptionally isolate from the cassette a 5' terminator was designed upstream the T7RNAP piece. The strong synthetic terminator (L3S2P21) [142] was selected to avoid potential homology with native terminators during genomic insertion. The terminator was order was from IDT as a sense and antisense oligonucleotide (**Supplemental Table A2**). The T7RNAP part was amplified from *BL21(DE3)* genomic DNA. The T7RNAP PCR was performed using Phusion with EDC408 and EDC323 primers, 5 ng genomic DNA per µL of PCR reaction, 3% DMSO at 98^0C for 15 min, with 30 cycles of 98^0C for 30 s, 55^0C for 30 s, and 72^0C for 3 min, and a final extension of 72^0C for 25 min. The CmR piece PCR was performed using Phusion with EDC413 and EDC414 primers and the pAM552C plasmid [143] at 98^0C for 15 min, with 30 cycles of 98^0C for 60 s, 55^0C for 30 s, and 72^0C for 45 s, and a final extension of 72^0C for 25 min. The T7RNAP and CmR PCR reactions each received 1 µL of DpnI per 20 µL of PCR reaction and were incubated at 37^0C for 2 hr. The PCR reactions were column purified and run on a 0.7% agarose gel at 90 V for 45 min. The correct sized band was cut out of the gel and column purified. All three parts were then pool together at equal molar concentrations (75 ng of DNA total) in an overlap PCR reaction using Phusion, 3% DMSO at 98^0C for 10 min, with 15 cycles of 98^0C for 30 s, 55^0C for 30 s, and 72^0C for 4 min, and a final extension of 72^0C for 10 min. The overlap PCR was then diluted 20-fold into a second PCR reaction with EDC410 and EDC414 primers at 98^0C for 3 min, with 24 cycles of 98^0C for 30 s, 55^0C for 30 s, and 72^0C for 4 min, and a final extension of 72^0C for 10 min. PCR reactions were then column purified and run on a 0.7% agarose gel at

90 V for 45 min. The correct sized bands were cut out and column purified. Next, 45 bp of genomic insertion site homology was added to the 5' and 3' end of the assembled T7RNAP cassette using Phusion, 3% DMSO with JGP139 and JGP140 primers at 98^{0}C for 3 min, with 25 cycles of 98^{0}C for 60 s, 65^{0}C for 30 s, and 72^{0}C for 7 min, and a final extension of 72^{0}C for 10 min. PCR reactions were column purified, run on a 0.7% agarose gel at 90 V for 45 min. The correct sized bands were cut out and column purified. The sequence of the fully assembled cassette was confirmed via sequencing.

3.5.6 Datsenko-Wanner of T7RNAP cassette

The T7RNAP cassette was inserted using the λ-red homologous recombination method for PCR products [22, 131]. The *C321.ΔA* strain contains the λ-red recombinase machinery on its genome which enables quick modification of the genome without scars. Briefly, 3 mL of LB-L media was inoculated with overnight culture of the strain of interest at a 1:50 dilution. Cultures were grown at 32^{0}C, 250 rpm until OD_{600} reached 0.7 as read on a Libra S4. The culture was heat shocked at 42^{0}C for 15 min at 100 rpm to activate expression of the λ-red recombinase machinery. Cultures were place on ice for at least 15 min to cool cells down, spinning the culture tube in ice every 3 min. Next, 1 mL of culture was harvested and washed twice with ice-cold sterile deionized water, pelleting cells at 13,000 g at 4^{0}C. The cell pellets were resuspended with 10 ng of the T7RNAP cassette in 100 µL of ice-cold sterile deionized water and electroporated. Cells were then recovered in 1 mL LB-L for at least 3 hr at 32^{0}C, 250 rpm and plated on Cm_{34} plates for 1-3 days at 30^{0}C.

3.5.7 Screening for full T7RNAP cassette insertion

Cells that genomically inserted the CmR portion of the cassette grew on the Cm_{34} plates. To screen for full insertion of the cassette colony PCR was performed. Colonies on the Cm_{34} plate were picked and inoculated into 100 µL LB-L Cm_{25} media in 96-well polystyrene plates (Costar 3370) incubated at 32^{0}C, 250 rpm for at least 3 hr. The cultures were used as the template in colony PCR reactions. To screen for 5' portion of T7RNAP a PCR reaction was performed with MASC PCR reactions using JGP173 and JGP292 primers at 95^{0}C for 15 min, with 30 cycles of 95^{0}C for 30s, 52^{0}C for 30 s, and 72^{0}C for 1 min, and a final extension of 72^{0}C for 10 min. PCR reactions were run on a 2% gel, 110V 45 min. Colony PCR was repeated at a larger scale for the colonies that resulted in a band, reactions were column

purified and submitted for sequencing using JGP173, EDC280 and JGP292 primers. Positive sequence hits were screened for the full T7RNAP region being inserted using Multiplex Master Mix with EDC282 and JGP153 primers at 95^{0}C for 15 min, with 30 cycles of 95^{0}C for 30 s, 53^{0}C for 30 s, and 72^{0}C for 1.5 min, and a final extension of 72^{0}C for 10 min. The PCR reaction were run on a 2% agarose gel. Colony PCR was repeated at a larger scale for the colonies that resulted in a band. The reactions were column purified and submitted for sequencing using EDC282, EDC283, EDC284, EDC285, and JGP153 primers.

3.5.8 Removing antibiotic resistance marker

Clones with full T7RNAP cassette present then underwent MAGE to remove the CmR gene using a mutagenic oligonucleotide, JGP389, with homology on the 5' and 3' end of the CmR gene. After 8 cycles of MAGE, overnight cultures were plated on LB plates at 10^{-6} dilutions in LB-L Cb_{50} media. Colonies were replica-plated onto LB-Cb_{100} and LB $Cb_{100}Cm_{34}$ plates and incubated at 32^{0}C overnight. Colonies that grew on LB-Cb_{100} plates and not LB-$Cb_{100}Cm_{34}$ plates underwent PCR using Multiplex Master Mix with EDC413 and JGP211 primers at 95^{0}C for 15 min, with 30 cycles of 95^{0}C for 30 s, 54^{0}C for 30 s, and 72^{0}C for 1.5 min, and a final extension of 72^{0}C for 10 min. For positive hits colonies the PCR reactions were repeated at a larger scale, column purified and submitted to sequencing with EDC413 and JGP211 primers to confirm the CmR gene was completely removed.

3.5.9 Full-length sfGFP purification and quantification

Strains were freshly transformed with the plasmids of interest. A single colony was inoculated into 5 mL of 2x YT media with $Kan_{35}Cm_{25}$ and grown overnight at 32^{0}C at 250 rpm. Overnight cultures were diluted 1:50 into 40 mL of fresh 2x YT media $Kan_{35}Cm_{25}$ and grown at 32^{0}C, 250 rpm. OD_{600} was monitored on a NanoDrop 2000c until OD_{600} 0.6-0.8 at which point cultures were induced with 5 mM nsAA, 1 mM IPTG, and 0.02% arabinose. Cultures were harvested after 20 hr after induction by pelleting 30 mL of culture at 5,000 g for 10 min at 4^{0}C. The pellet was resuspended in 0.8 mL of 1X phosphate-buffered saline (PBS) buffer for every 1 g of wet cell pellet. Cells were lysed at a frequency of 20 kHz and an amplitude of 50% using a Q125 Sonicator (Qsonica, Newton, CT, USA) with a 3.75 mm diameter probe [25] for 5 cycles of 45 s sonication and 59 s sitting on ice. The input energy (Joules) per cycle averaged to 274. Lysed samples were then centrifuged at 21,000 rpm for 10 min at 4^{0}C. Supernatant was

collected as the soluble fraction. Full-length sfGFP was purified from the soluble fraction by using a C-terminal strep-tag and 0.2 mL gratify-flow Strep-Tactin Sepharose mini-columns (IBA GmbH, Gottingen, DEU). Purified sfGFP was measured using a Quick Start Bradford Kit (BioRad, Hercules, CA, USA) in biological triplicate and technical triplicate.

3.6 Author Contributions

E.D.C and I constructed the strains. E.D.C designed the T7RNAP cassette. O.W., E.D.C. and I constructed, introduced and verified T7RNAP genomic insertion. E.D.C, M.C.J. and I designed experiments. E.D.C. and I performed the experiments. M.C.J. and I wrote the manuscript. All authors discussed the results and commented on the manuscript.

3.7 Acknowledgements

This work was supported by the National Science Foundation (NSF) (MCB-0943393), the Office of Naval Research (N00014-11-1-0363), the Defense Advanced Research Projects Agency Young Faculty Award (DARPA YFA) Program (N66001-11-1-4137), the Army Research Office (W911NF-11-1-0445), the DARPA Living Foundries Program (N66001-12-C-4211), the David and Lucile Packard Foundation (2011-37152), Advanced Research Projects Agency-Energy (ARPA-E) (DE-AR0000435), and the Chicago Biomedical Consortium. J.G.P is supported by the NSF Graduate Research Fellowship (DGE-1144469) and the NIH Initiative for Maximizing Student Development (IMSD) (PAR-17-053). The authors would like to thank Professor Joshua Leonard for his critical review of the manuscript and Professor Peter G. Schultz for providing the pEVOL plasmid.

3.8 Publication Information

This chapter is currently in preparation for submission to a peer-reviewed journal.

CHAPTER 4

4 Design and Optimization of a T7 RNA Polymerase System in a Genomically Recoded *Escherichia Coli*

4.1 Abstract

The ability to incorporation multiple, identical non-standard amino acids has drastically improved with the introduction of a genomically recoded *Escherichia coli* strain, *C321.ΔA*. This strain has been utilized to produce high quantities of modified proteins; however, it is lacking one of the most desired features of protein production strains: a T7 promoter system. T7 promoter systems are highly preferred due to its ability to highly and selectively express recombinant proteins off a plasmid. Here I demonstrated, for the first time, a functional T7 system in *C321.ΔA*. I explored various factors to improve the T7 system including removing negative effectors to protein synthesis (also see Chapter 3.3.1), modifying the T7 inducer system, and probing the genomic context of the T7 RNA polymerase (T7 RNAP) cassette. I also developed a novel method to elucidate the location of a gene at an unknown location on the genome without full genome sequencing. This chapter lays the groundwork for strategies to improve T7-based systems in *C321.ΔA*. As the need for modified proteins increases, so will the need for more versatile strains with the ability to tune expression of recombinant proteins using T7 RNAP.

4.2 Introduction

Expansion of the genetic code to assign additional, non-standard amino acids (nsAAs) is transforming our ability to synthesize modified proteins. Such efforts have enabled the incorporation of (a)biological groups into proteins such as biophysical probes [144] (spin-labeled, fluorescent molecules, and photoactivatable cross linkers), redox active groups [145], and natural post-translational modifications (PTMs) [83, 146-148]. To date more than 100 nsAAs have been co-translationally incorporated into proteins [99].

In order to encode a nsAA, a codon needs to be hijacked. This is often achieved by reassigning the UAG stop codon (least used stop codon in the genome) to the nsAA of interest. Additionally,

expression of an orthogonal translation system (OTS), via the pEVOL plasmid [101], is required. This plasmid encodes an orthogonal aminoacyl-tRNA synthetase (o-aaRS) which aminoacylates an orthogonal tRNA (o-tRNA) with the nsAA. The o-tRNA's anti-codon directs it to the UAG stop codon to complete the recoding process (See Chapter 2.5 for more details).

A huge obstacle to co-translational incorporation of a nsAA is the competition between the aminoacylated o-tRNA and release factor-1 (RF-1). Endogenous RF-1 recognizes UAG codons and subsequently activates hydrolysis and the peptidyl-tRNA to release the peptide chain. RF-1 competes with the nsAA-o-tRNA at the A-site of the ribosome, resulting in a significant amount of truncated product and reduced nsAA incorporation efficiencies. Researchers have recently developed an *Escherichia coli* (*E. coli*) RF-1 deficient strain, termed *C321.*Δ*A*, in which all 321 UAG codons were recoded to the synonymous UAA codon. This completely frees the UAG codon to encode nsAAs and removes the nsAA-o-tRNA/RF-1 competition event [42].

This strain has already been leveraged for exciting applications in biocontainment [149], rapid *in vivo* evolution strategies [105], insertion native PTMs co-translationally, [82] and synthesis of complex function materials [150]. Improvements to this strain will have a large impact on advancing further research in this area. In particular, a widely desired feature of recombinant protein expression strains is the ability to use T7 promoters. The T7 promoter system is extremely popular due to its ability to express recombinant proteins at levels up to 50% of the total cellular protein mass [114, 115]. In this system the gene of interest is cloned behind the T7 promoter which is only recognized by the phage T7 RNAP polymerase (T7RNAP). This highly active polymerase is expressed off the bacterial genome in a prophage (λDE3) under the transcriptional control of a Lac_{UV5} promoter [113]. To control T7RNAP-mediated expression a Lac operator (LacO) is placed downstream the T7 promoter of the gene of interest and the Lac_{UV5} promoter of the T7RNAP gene. LacI, which is expressed off the genome and the expression plasmid, is able to bind to LacO and blocks expression of the downstream gene. Repression by LacI can be removed by the addition of a small molecule like lactose or its non-hydrolyzable analog isopropyl β-$_D$-1-thiogalctopyranoside (IPTG), allowing for inducible expression of recombinant proteins.

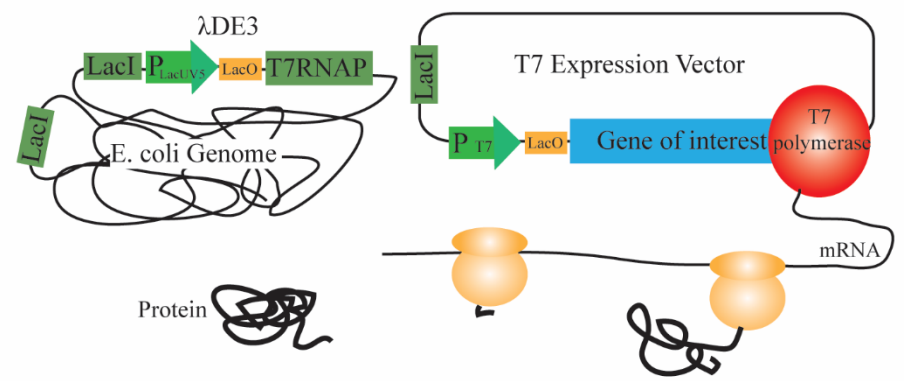


Figure 4.1: Inducible expression of recombinant proteins using the T7 promoter system. The λDE3 cassette on the genome encodes for a T7 RNA Polymerase (T7RNAP), regulated by a Lac_{UV5} promoter and Lac operator (LacO). LacI, which is basally expressed on the λDE3 cassette and the genome, binds to the Lac operator, repressing expression of T7RNAP. The T7RNAP should only be expressed if IPTG, LacI's repressor, is added to the media, releasing LacI from the LacO. This allows for T7RNAP expression, and thus the gene of interest, to be tunable via the addition of IPTG.

Given the advantages of the T7 system for recombinant protein expression we sought to introduce this system into the genomically recoded strain, *C321.ΔA*, to increase its utility. Here I demonstrate, for the first time, a functional T7 system in *C321.ΔA*. I achieved this by exploring numerous variables that were hypothesized to have an effect on the T7 system's productivity and inducibility: (1) negative effectors to protein production, (2) the T7RNAP cassette design, (3) the induction system, and (4) genomic location of the T7RNAP cassette. Through this study it was found that the system suffered from leaky expression when in *C321.ΔA*; however, T7RNAP functions at levels near the state of the art in T7 systems and can be utilized for various applications, as detailed in Chapter 3.

4.3 Results and Discussion

4.3.1 Introducing T7RNAP into C321.ΔA: a proof of concept

A practical challenge in using phage polymerases is the toxic effects they can exhibit [113, 151, 152]. Strains containing λDE3 cassettes are also known to occasionally suffer from unacceptably high levels of expression in the absence of the inducer (e. g. "leakiness"). With these issues in mind, the difficulty of introducing a T7 system into new strains should not be underestimated.

As a proof of concept, we introduced a T7 system into *C321.ΔA* that is most similar to the state of the art in T7 systems, *BL21(DE3).* Although *C321.ΔA*, a K strain, differs from *BL21(DE3)*, a B strain, in several ways [118, 153], we considered *BL21(DE3)* as the best reference for success. Thus, the T7RNAP cassette was amplified off the *BL21(DE3)* genome and stitched together with a 3' antibiotic resistance

marker, KanR, from pKD4 [131]. The 5' and 3' ends of the cassette contained 45 base pairs (bp) of genomic homology to direct the cassette to its genomic insertion site. This T7 cassette is referred to as T7RNAP-KanR (**Supplemental Figure A6**). Using λ-red mediated homologous recombination [22, 131], the cassette was inserted into *C321.ΔA*, the KanR selectable marker was removed and the total cassette sequence and insertion site was verified via sequencing. Four *C321.ΔA-T7RNAP* variants (α, β, γ, and δ) were obtained and tested for changes in growth rate. We observed that the insertion of the T7RNAP cassette did not drastically increase the doubling time of *C321.ΔA-T7RNAP* variants. In fact all strains' doubling times were within 8% of *C321.ΔA* when grown at 32^0C in a plate reader (**Supplemental Figure A7A**).

To test the functionality of T7RNAP in these strains, a reporter plasmid (pT7sfGFP-wt) containing wild-type sfGFP (sfGFP-wt) downstream a T7 promoter was transformed into the *C321.ΔA-T7RNAP* variants. By testing a variety of IPTG concentrations, we observed that *BL21(DE3)* has an ideal inducibility response (**Figure 4.2**). In other words, little fluorescence was detected when no IPTG was present and there was a gradual increase in fluorescence as increasing amounts of IPTG were added until it reached its optimal concentration. On the other hand, all *C321.ΔA-T7RNAP* variants showed a poor inducibility response. A significant amount of basal sfGFP-wt expression was detected when IPTG was not present. This may imply that there is an insufficient intracellular concentration of LacI to bind to LacO and adequately repress T7RNAP expression. This is despite LacI being expressed on the genome and the pT7sfGFP-wt plasmid. Furthermore, the overall yield of sfGFP-wt in *C321.ΔA-T7RNAP* variants is much lower than *BL21(DE3)*. The presence of proteases and nucleases in *C321.ΔA* and not BL21(DE3) could be contributing to the lower yields.

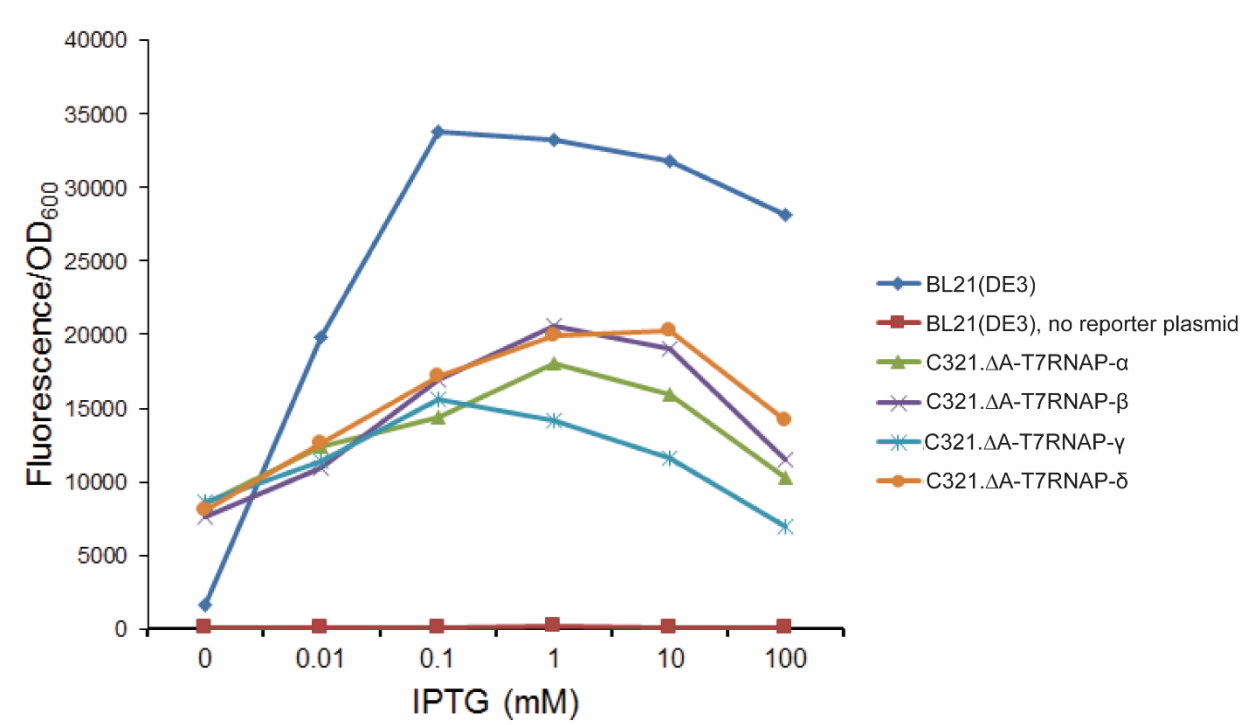


Figure 4.2: Inducibility assay of *C321.ΔA-T7RNAP* variants. Expression of sfGFP-wt was performed in each strain at various IPTG concentrations. All strains were grown in a plate reader at 32ºC until OD 0.7 at which point various amounts of IPTG were added. Each data point represents fluorescence divided by optical density at 600 nm read by a plate reader in a 96-well plate at 12 hr

To address the inducibility issues and low protein yields of the initial *C321.ΔA-T7RNAP* strains, we sought to improve this system by pursuing a variety of strategies. First, to address the lower yields of *C321.ΔA-T7RNAP* strains, we hypothesize that reducing protease and nuclease activity, mirroring similar modifications in *BL21(DE3)*, may help to increase protein yields. Second, we pursued several strategies to reduce basal expression of recombinant proteins. Lastly, we explored the role genomic location has on expression of recombinant proteins regulated by a T7 system.

4.3.2 Reducing protease and nuclease activity in C321.ΔA-T7RNAP to increase protein yield

BL21(DE3) produces high yields of recombinant protein due to its deficiency in the proteases lon, which degrades foreign proteins [154], and ompT, an outer membrane protease [155]. *BL21 Star (DE3),* another widely used *E. coli* protein production strain, is deficient in lon, ompT and RNAse E activity. By truncating the RNAse E gene, rne, at the 131st nucleotide, RNAse E's ability to degrade mRNA is removed while its ribosome RNA processing ability is maintained [110]. This mutation

results in increased stability of recombinant mRNAs and increased protein production. We hypothesize that introducing these same mutations in *C321.ΔA-T7RNAP* strains may increase protein yields. Note here that the strains tested in Section 4.3.2 are the same as those detailed in Section 3.3.3, but were tested under different growth conditions, consistent other experiments within this chapter and as detailed in Section 4.5.7.

To test this hypothesis, we reduced protease and nuclease activity in *C321.ΔA-T7RNAP-β* using multiplex automated genome engineering (MAGE) [121] to introduce mutations into the open reading frame of lon, ompT and rne. Mirroring the mutations in BL21(DE3), the promoter of lon was removed [118]. A point mutation (D103A) was made in ompT which is known to eliminate the proteolytic activity of ompT while maintaining its structure motifs due to ompT's possible chaperone function [119, 120]. The rne gene was truncated at the 131st nucleotide, analogous to *BL21 Star (DE3).* Several combinations of these mutations were made, and the growth rates after introducing these mutations did not differ significantly from the parental strain (**Supplemental Figure A7B**).

The effect of the mutations on protein production was assessed by transforming a plasmid (pT7sfGFP-wt), containing sfGFP-wt and regulated by a T7 promoter, into all *C321.ΔA-T7RNAP-β* (*β*) variants. (**Figure 4.3**). We observed that all mutants expressed sfGFP-wt at levels equal to or greater than the parental strain, *β*, with *β ompT⁻ lon⁻* and *β rne⁻ ompT⁻ lon⁻* being the best performers and closest to the genotype of *BL21 Star (DE3).* High basal expression was still observed in these mutants; however, reduction of protease and nuclease activity would theoretically not affect induction responses. Interestingly, despite *β rne⁻ ompT⁻ lon⁻* containing the same mutations as *BL21(DE3)*, this mutant had decreased sfGFP-wt yields compared to *BL21(DE3).* It is known that *BL21(DE3)* displays higher biomass yields compared to K-12 strains, like *C321.ΔA*, resulting in substantially lower acetate amounts which in turn has a positive effect on recombinant protein production [156, 157]. This factor may ultimately be a limiting factor for *C321.ΔA* expression and is not addressed in this work.

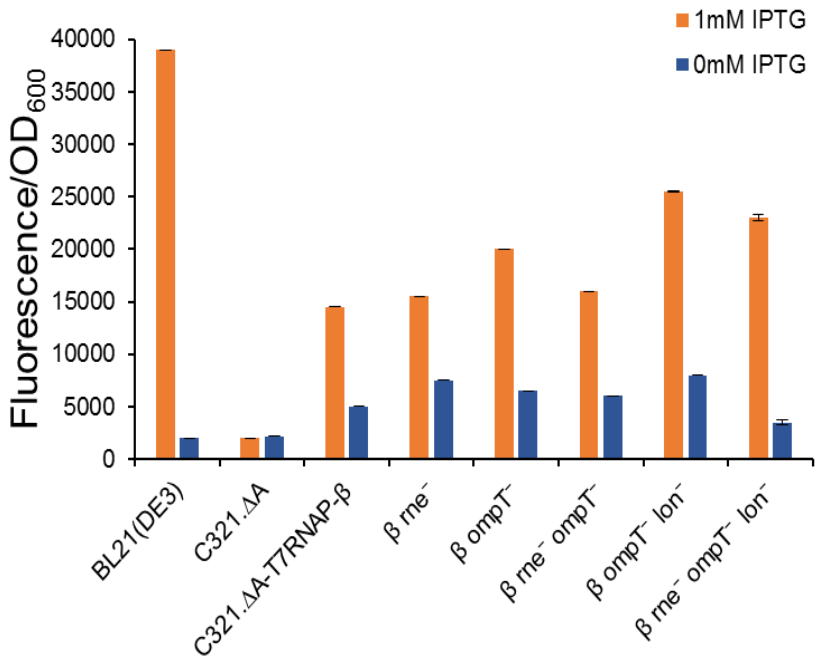


Figure 4.3: Characterizing inducibility of *C321.ΔA-T7RNAP-β* with protease and nuclease mutations. Expression of pT7sfGFP-wt in each strain after 7 hr of induction with (orange bars) or without (blue bars) 1 mM IPTG. Error bars represent biological duplicates and technical duplicates. *β: C321.ΔA-T7RNAP-β*.

Here we showed reduction of nuclease and protease activity can drastically improve sfGFP-wt expression in *C321.ΔA-T7RNAP-β.* In fact, the triple mutant, *β rne⁻ ompT⁻ lon⁻*, showed a 67% improvement compared to the parental strain. This demonstrated that reducing nuclease and protease activity can improve protein production in a genomically recoded strain. In the future a more detailed exploration of positive and negative effectors will be advantageous for further development of *C321.ΔA*. For instance, in Chapter 3 I showed that functional deactivation of DNAse endA, in combination with other nuclease and protease mutations, can increase sfGFP production in the context of *C321.ΔA*. Additionally a study of acetate regulators' effect on protein production in *C321.ΔA* may be advantageous given previous successes in other K-12 strains [156, 157].

4.3.3 Optimization of T7 system to improve inducibility

Basal expression of recombinant proteins in T7 systems is not uncommon. This is because small amounts of basal T7RNAP expression in uninduced cells have large effects on the amount of

recombinant protein expressed due to the high transcriptional efficiency of T7RNAP [152]. We approached this issue in *C321.ΔA* by (i) transcriptionally isolating the T7RNAP cassette, (ii) introducing a T7 lysozyme to reduce basal activity from T7RNAP, and (ii) characterizing a second small molecule inducer system for induction of T7RNAP transcription.

4.3.3.1 Transcriptionally isolating the T7RNAP cassette

Attempting to reduce the basal activity of the β strains, we investigated the insertion site more closely for potential issues. The T7RNAP-KanR cassette was inserted at genomic location (816413-817273) (**Supplemental Figure A8**) to replace a β-lactamase gene (bla) which confers ampicillin resistance. The region around the insertion site was entered into BLAST's nucleotide BLAST program [158] to find potential genomic architectures that might interfere with T7RNAP expression. BLAST uncovered a putative promoter directly upstream the T7RNAP insertion site (**Figure 4.4**). We hypothesized that the upstream promoter may be recruiting endogenous polymerases to localize near the T7RNAP cassette, increasing the chances of a polymerase binding to the Lac_{UV5} promoter and expressing T7RNAP basally.

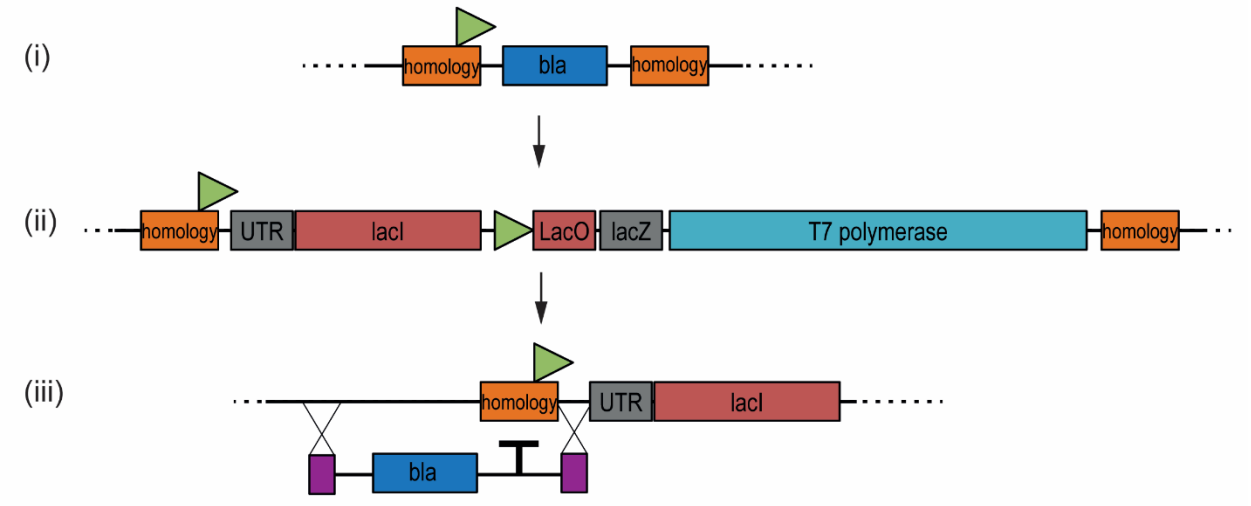


Figure 4.4: Genomic features at the T7RNAP-KanR insertion site. (i) The insertion site of T7RNAP-KanR originally possessed a β-lactamase gene (bla). (ii) The T7RNAP cassette directly replaced the bla gene, retaining the upstream promoter (green arrow). (iii) The bla-T1T2 cassette, containing a selectable marker and a T1T2 terminator, was genomically inserted using Datsenko-Wanner to replace the putative promoter and insert a terminator to transcriptionally isolate the T7RNAP cassette.

To test this hypothesis, λ-red mediated homologous recombination was used to remove the putative promoter and replace it with a PCR product which included an antibiotic resistance marker, bla, and a strong terminator, T1T2 [159]. The terminator was added to prevent potential issues from upstream features, thus transcriptionally isolating the T7RNAP cassette, which has previously been shown to be effective [160]. This strategy was implemented for *β,* resulting in the strain *β-Δbla-T1T2*. The plasmid pT7sfGFP-wt was transformed into *β-Δbla-T1T2* to measure changes in induction response due to the removal of the putative promoter (**Figure 4.5**). The induction response in *β-Δbla-T1T2* did not improve significantly, as compared to *BL21(DE3)* which has an induction ratio (ratio of normalized fluorescence of 1 mM IPTG to 0 mM IPTG) of 17.58; however, *β-Δbla-T1T2* did show a 2-fold improvement in induction ratio and a 2-fold reduction in basal expression, as compared to the parental strain. Although small, basal expression was reduced due to the removal of the putative promoter and the addition of an upstream terminator. Given this improvement, the upstream terminator motif continued to be used for the remainder of this work.

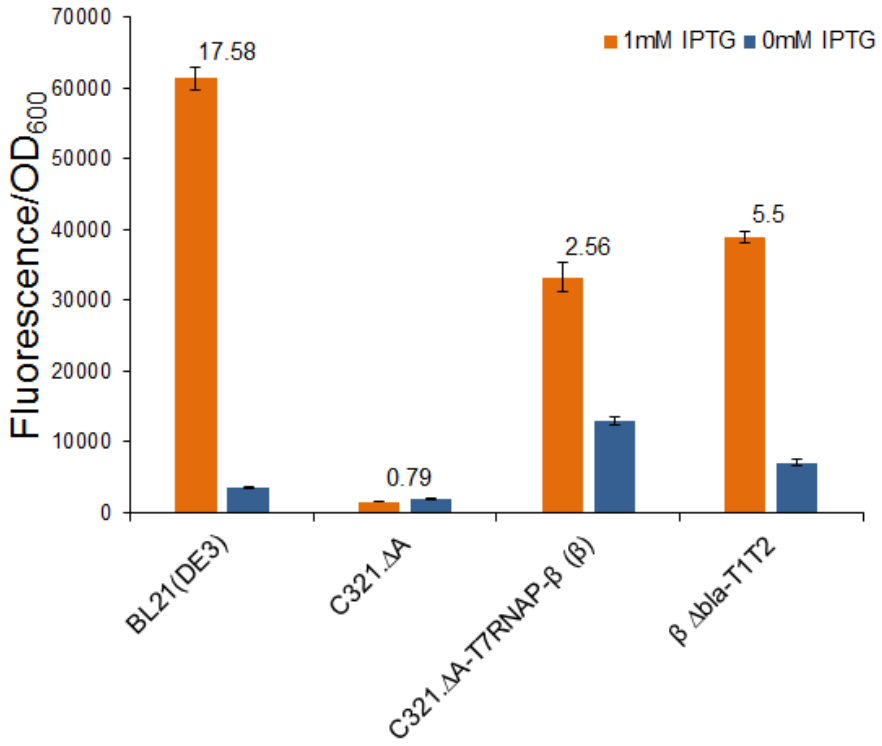


Figure 4.5: Effect of removing putative promoter upstream T7RNAP cassette in *β*. Expression of sfGFP-wt in *C321.ΔA-T7RNAP-β-Δbla-T1T2* and controls were allowed to express for 7 hr after induction with 1 mM (orange bars) or 0 mM (blue bars) IPTG. Numbers above each strain represent the ratio of normalized fluorescence for 1 mM IPTG and 0 mM IPTG conditions. Error bars represent biological duplicates and technical triplicates.

4.3.3.2 Addition of T7 lysozyme to T7 system in C321.ΔA

Previous work to reduce basal expression in T7 systems have found expression of T7 lysozyme, a natural inhibitor of T7RNAP, [161] off a plasmid to be an effective strategy. Promega offers a *BL21(DE3)* strain that contains either the pLysS or pLysE plasmid which expresses lysozyme at different strengths with pLysE being the higher expresser [152]. Because lysozyme also has amidase activity, which can damage *E. coli*'s cell wall, growth defects are often observed in strains containing pLysE [152]. For these reasons we tested pLysS' effectiveness at reducing the basal activity in *β-Δbla-T1T2* (**Supplemental Figure A9**). We observed that pLysS drastically reduced sfGFP-wt expression when present in *β-Δbla-T1T2.* This is most likely do to the toxic effects of lysozyme, reducing the protein production potential of the strain. Subsequently, the pLys plasmids were not used in future studies.

4.3.3.3 Developing an anhydrotetracycline-inducible T7 system

While IPTG-inducible T7 systems are the golden standard, lac promoters (e. g. Lac_{UV5}) are known to be weak in strength and very leaky [162]. Conversely, there are many other inducible systems with tighter regulation such as the araBAD promoter/L-arabinose system [163] and the tetA promoter/anhydrotetracycline (aTc) system [164]. The aTc-inducible system is particularly advantageous due to its independence of metabolic state and *E. coli* strain and low basal expression. For these reasons we sought to develop an aTc-inducible T7 system. For this study slight modifications were made to the original design of the T7RNAP cassette. First, the 5' terminator, T1T2, was replaced with a synthetic terminator L3S2P21 which has a similar termination efficiency as T1T2, [142] but does not contain homology with *C321.ΔA*'s genome. Second, the antibiotic resistance marker was switched to a chloramphenicol resistance gene (CmR, cat). The aTc-inducible T7RNAP cassette consisted of a 5' synthetic terminator, the pTet promoter, the T7RNAP polymerase, and CmR. The 5' and 3' ends of the cassette contained 45 bp of genomic homology. This T7 cassette will be referred to as pTet-T7RNAP.

To test the functionality of the aTc-inducible T7 system, pT7sfGFP-wt-TetR was transformed into the strains containing the pTet-T7RNAP cassette. First, four different ribosome binding site (RBS) strengths [165] were tested in this cassette and found to have no significant impact on sfGFP-wt expression (**Supplemental Figure A10A**). Here we observed very tight regulation when using aTc as the inducer. Little to no expression was observed in the absence of aTc while a very sharp jump in expression occurred after aTc was added, even as low as 10 ng/mL of aTc. However, during these experiments we found that 7 hr after induction the cultures were not as vibrant green as when using the IPTG-inducible system, although the absolute normalized fluorescence numbers were comparable. We measured the OD_{600} using a 1 cm cuvette in a NanoDrop 2000c spectrophotometer and found that the OD_{600} was very low (~1.7) while BL21(DE3) reached an OD of ~6 (**Supplemental Figure A10B and A10C**). We hypothesize that measuring OD_{600} via the plate reader was exaggerating the aTc-inducible system's success. OD_{600} readings from the plate reader were very small (<1) for the aTc-inducible system. Normalizing the fluorescence to OD_{600} from the plate reader caused normalized fluorescence of aTc samples to be comparable to BL21(DE3) although the absolute sfGFP concentrations differed

drastically. Accordingly, we reanalyzed the fluorescence/OD_{600} results with OD_{600} readings read in a 1 cm cuvette. In this case we saw a drastic decrease in the normalized fluorescence for strains containing the pTet-T7RNAP cassette (**Supplemental Figure A10D**). First this suggested that reading OD_{600} with a 1 cm cuvette is required to fairly assess T7 system's effectiveness among each other. This is not trivial because numerous research articles compare normalized fluorescence measured in a plate reader, without quantifying absolute protein concentrations to assess the effectiveness of various genetic circuits in *E. coli*. Second we concluded that the aTc-inducible T7 system was effective at improving the inducibility in a *C321.ΔA* based T7 system, but aTc negatively affects OD_{600} and the protein yield. This system is ideal for protein expression projects where tight control of protein expression is desired, such as expression of toxic modified proteins.

4.3.4 T7RNAP genomic location's role on recombinant protein expression

The genomic location of genes strongly influences expression. Studies conducted in K-12 strains observed that position-dependent gene expression levels can vary by as much as ~300-fold, implying that there are substantial differences in expression potential within bacterial genomes [166]. Thus, exploring a variety of genomic insertion sites for the T7RNAP cassette may have an influence on gene expression levels. Additionally, T7RNAP in *C321.ΔA-T7RNAP-β* and its variants is located on the lagging strand of the genome. This can cause reduced protein expression because for genes on the lagging stand, the DNA and RNA polymerase move in opposite directions, which creates head-on collisions that dramatically reduce the speed of the replication fork in *E. coli* [167]. On the other hand, the majority of genes on the genome are on the leading strand of DNA replication, allowing transcription to occur in the same direction as replication [168]. Because T7RNAP-KanR was inserted downstream the λ-red system, replacing a bla gene, for simplicity, this genomic location may not be optimal for T7RNAP expression. For these reasons we explored various genomic insertion sites for the T7RNAP cassette.

Insertion sites were chosen from previously characterized safe insertion regions (SIRs), regions on the genome that have not been annotated for any coding or regulatory function and have successfully expressed heterologous genes [169]. We targeted the leading strand of SIR.8.9, SIR.8.9b, SIR.12.13, and SIR.32.1, which are all located at various regions on the genome (**Supplemental Figure A8**).

Transcriptional activity characterization of the entire K-12 genome reveals that these locations vary in transcriptional activity which may have an effect on heterologous gene expression [170]. Based on previous work by Jeong et. al. SIR.8.9, SIR.8.9b and β are located in low transcription areas. SIR.12.13 is located in a medium transcription level area, and SIR.32.1 is located at high transcription level area [170].

The assembly of the T7RNAP cassette for this study was performed similarly to T7RNAP-KanR except that the KanR gene was replace with a CmR gene, termed T1T2-T7RNAP-cat. The T1T2-T7RNAP-cat cassettes were inserted to the genome as previously discussed. Genomically inserted T1T2-T7RNAP-cat cassettes were fully sequenced; however, many of the cassettes were not located at the desired insertion site. New primers were designed to assay potential off target insertion sites such as places with homology to internal sites on the cassette (e.g. T1T2 terminators and LacI on the genome), but the cassettes were not located at any of these regions. When tested for expression of sfGFP-wt off a pT7 plasmid, these strains performed quite well (**Supplemental Figure A11**). Thus, we sought to uncover the genomic location of the cassette by developing a method to find a genomically inserted cassette without any prior knowledge of its location (**Figure 4.6**).

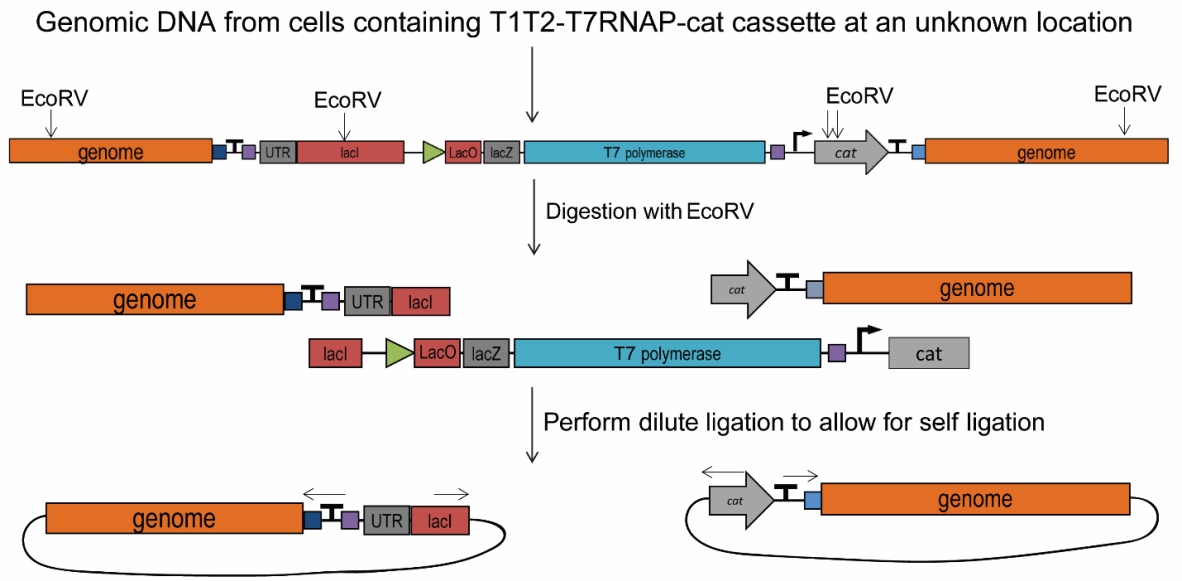


Figure 4.6: Locating the T1T2-T7RNAP-cat cassette at an unknown genomic location. Genomic DNA from T1T2-T7RNAP-cat containing strains was isolated and digested with EcoRV to break up the entire genome into pieces. The digested genome was then ligated at dilute concentrations to favor self-ligation. Internal primers around the EcoRV sites on the T1T2-T7RNAP-cat cassette were designed to amplify outward into the genome. The resulting PCR products were send to sequencing to uncover the genomic insertion site.

For this method genomic DNA was isolated from strains containing the T1T2-T7RNAP-cat cassette and digested with EcoRV. This restriction enzyme was selected because internal EcoRV sites existed on the T7 cassette. Next a dilute ligation was performed to encourage self-ligation. Assuming a fraction of the genomic fragments self-ligated, internal primers on either side of the EcoRV sites on the T1T2-T7RNAP-cat cassette were designed to amplify outward around the circularized genomic fragments. This method yielded clear bands when run on an agarose gel for each reaction and, after sequencing, revealed the genomic insertion sites of the T7 cassettes.

The T1T2-T7RNAP-cat cassette inserted at genomic coordinates 16,654 (16K), 2,513,551 (2.5 M), and 608,541 (608K), only one of which was located at the desired position (**Figure 4.7**). These positions are located in regions of high, medium and low transcriptional activity, respectively [170]. Each location showed an improvement in sfGFP-wt expression over β; however, as sfGFP-wt expression increased so did its basal expression. It would seem that for these genomic locations, sfGFP-wt expression was not affected dramatically. Moreover, there was no correlation of the genomic region's transcriptional activity to sfGFP-wt expression. Due to the difficulty of inserting a ~4.8 kb cassette into the genome, we did not pursue additional genomic insertion locations. However, with the improvement of genomic engineering techniques such as CRISPR-Cas9 [171] a more detailed analysis of this work is possible. More genomic insertion sites would need to be examined to make any conclusions on the genomic insertion location's role on T7-based expression of recombinant proteins.

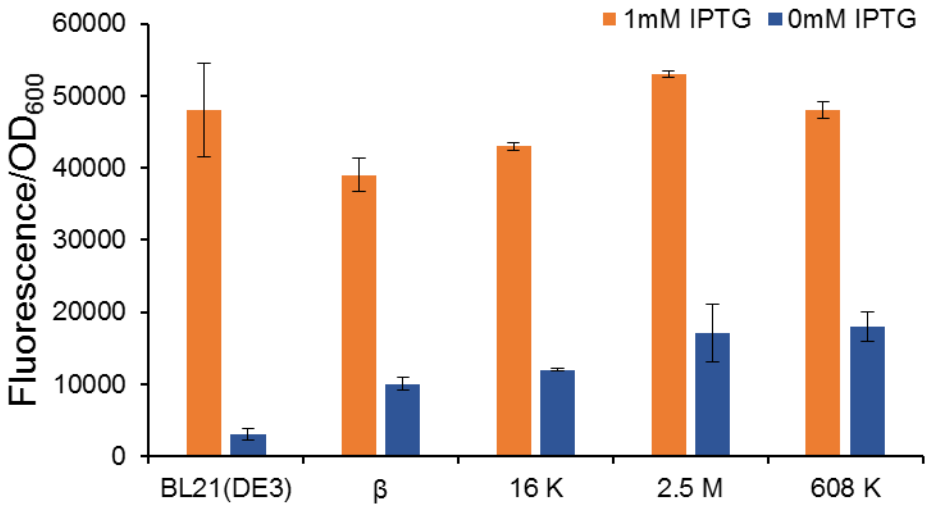


Figure 4.7: Expression of sfGFP-wt with T1T2-T7RNAP-cat cassette at various locations on the genome. Expression of sfGFP-wt 7 hr after induction with or without 1mM IPTG. Here OD_{600} was measured using a plate reader. Error bars represent biological duplicates and technical triplicates. Numbers on the x-axis are the genomic coordinates where the T1T2-T7RNAP-cat cassette is located.

4.4 Conclusions

Here we have demonstrated, for the first time, a functional T7 system in a genomically recoded *E. coli* strain. Through this work we explore various factors that have previously been shown to affect protein production or T7RNAP inducibility. First, we validated that removing negative effectors to protein synthesis can improve sfGFP-wt yields by 67%. Because there are many more proteases, DNAses and RNAses that this study did not explore more negative effectors to protein synthesis could be removed from *C321.ΔA* and tested. Modifications to metabolic pathways could also be studied such as deleting regulators ArcA and IclR which have been shown to reduce acetate production and increase protein yields. [172]

We also introduced a novel aTc-inducible T7RNAP system into *C321.ΔA.* This system was effective at expressing sfGFP-wt and was tightly regulated by the aTc inducer. The aTc-inducible system will be advantageous for protein production projects that require tight control of protein expression, such as toxic modified proteins.

Finally, we explored the role genomic context can play on the expression of recombinant genes off the genome. We found that the addition of terminators to the 5' and 3' end of heterologous cassettes inserted onto the genome do not have negative effects on protein expression and may help to transcriptionally isolate the gene(s) of interest [160]. We tested four genomic insertion sites for expression of the T7RNAP cassette. While, none of these sites had a strong effect on protein production, many more SIRs could be explored to obtain a more comprehensive view on genomic insertion site effects. Through this study we also developed a novel method to find genes inserted onto the genome at an unknown site without having to sequence the entire genome.

In conclusion, we believe that this work is merely a starting point for strategies to improve T7-based system in *C321.ΔA*. As the need for modified proteins increases, so will the need for more versatile nsAA systems with the ability to tune expression of recombinant proteins.

4.5 Materials and Methods

4.5.1 Reagents, Buffers and plasmids

Chemicals and media were purchased from Sigma Aldrich (St. Louise, MO, USA) unless otherwise designated. Phusion High-Fidelity DNA Polymerase, T4 DNA ligase, dNTP, Quick-load DNA Ladders, BL21(DE3) and restriction endonuclease were purchased from New England Biolabs (NEB, Ipswich, MA, USA). Multipex PCR Kits used for MASC PCR were purchased from QIAGEN (Hilden, NRW, DE). Plasmids were extracted using Omega E.Z.N.A DNA Isolation Kit (Omega Bio-Tek, Norcross, GA, USA). DNA was column purified or gel extracted using OMEGA HiBind DNA Mini Columns and OMEGA E.Z.N.A Gel Extraction Kit, respectively. Genomic DNA was isolated with Omega E.Z.N.A. Bacterial DNA Kit. All DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT, Coralville, IA, USA). The nsAA pAzF was purchased from P212121, LLC (Ann Arbor, MI, USA) and ProCarb was purchased from BioFine, Inc (Vancouver, BC, CA). SYBR Safe, used in all agarose gels, and DH5α were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Synthetic *E. coli* C321.ΔA (GenBank: CP06698.1) was received as a gift from Farren Isaacs. The plasmid pLysS was purchased from Promega (Madison, WI, USA). All oligonucleotides used for cloning are shown in **Supplemental Table A4**. All vectors were cloned using Gibson Assembly [141]. The pT7 plasmids were derived from

pET24a vector. pET vectors were derived from pET28a vectors. The pT7sfGFPwt-TetR plasmid was created by amplifying TetR from pGW322 and the backbone of pT7sfGFPwt without the LacI.

4.5.2 Construction of T7RNAP-KanR

The T7RNAP-KanR cassette was assembled from two pieces: a T7RNAP piece and a 5' antibiotic resistance marker, KanR. The T7RNAP piece was synthesized via PCR that was performed using Phusion with EDC326 and EDC323 primers, 0.04% volume saturated BL21(DE3) culture, 3% DMSO at 98^{0}C for 15 min, with 35 cycles of 98^{0}C for 30 s, 60^{0}C for 30 s, and 72^{0}C for 7 min, and a final extension of 72^{0}C for 25 min. The KanR piece was synthesized via PCR that was performed using Phusion with EDC324 and EDC329 using the pKD4 plasmid [131] at 98^{0}C for 5 min, with 30 cycles of 98^{0}C for 60 s, 65^{0}C for 30 s, and 72^{0}C for 4 min, and a final extension of 72^{0}C for 10 min. T7RNAP and KanR PCR reactions each received 1 µL of DpnI per 20 µL of PCR reaction and were incubated at 37^{0}C for 2 hr. PCR reactions were column purified and run on a 0.7% agarose gel at 140 V for 30 min. Correct sized band was cut out of the gel and column purified. T7RNAP and KanR parts were pool together at a 4:1 molar concentration (40 ng of DNA total) in an overlap PCR reaction using Phusion, 3% DMSO at 98^{0}C for 3 min, with 15 cycles of 98^{0}C for 30 s, 60^{0}C for 30 s, and 72^{0}C for 4 min, and a final extension of 72^{0}C for 10 min. The overlap PCR was then diluted 20-fold into a second PCR reaction with EDC326 and EDC329 primers at 98^{0}C for 3 min, with 24 cycles of 98^{0}C for 30s, 60^{0}C for 30 s, and 72^{0}C for 6 min, and a final extension of 72^{0}C for 10 min. These primers added homology to the T7RNAP-KanR cassette targeting the insertion to genomic location (816413-817273), replacing a β-lactamase gene. PCR reactions were column purified and run on a 0.7% agarose gel at 100 V for 45 min. The correct sized bands were cut out and column purified.

4.5.3 Screening for full T7RNAP-KanR cassette insertion

Cells that genomically inserted the antibiotic resistance gene portion of the cassette grew on the antibiotic plates Kan_{50} plates. To screen for full insertion of the cassette colony PCR was performed. Colonies on the antibiotic plate were picked and inoculated into 100 µL LB-Lennox media with Kan_{50} in 96-well polystyrene plates (Costar 3370; Corning Incorporated, Corning, NY, USA), incubated at 32^{0}C, 250 rpm for at least 3 hr. Cultures were used as the template in colony PCR reactions. Five colony PCR

reactions were set up (i, ii, iii, iv, and v) to screen for full insertion of cassette. The PCR reaction was performed with NEB Taq DNA polymerase with Standard Taq Buffer using primers shown in **Supplemental Figure A6** at 95^{0}C for 15 min, with 30 cycles of 95^{0}C for 30 s, 57^{0}C for 30 s, and 68^{0}C for 7 min, and a final extension of 68^{0}C for 15 min. PCR reactions were run on a 2% gel at 110 V for 45 min. Colony PCR was repeated at a larger scale for PCR reaction v (the full cassette) PCR reactions were column purified and submitted for sequencing using EDC258, EDC313, EDC280, EDC315, EDC281, EDC282, EDC283, EDC284, EDC330, EDC285, EDC286, EDC259.

4.5.4 Removing antibiotic resistance marker from T7RNAP-KanR strains

Clones with full T7RNAP-KanR cassette present underwent MAGE [121] to remove the KanR gene using a mutagenic oligo, EDC287, with homology on the 5' and 3' end of the KanR gene. After six cycles of MAGE, overnight cultures were plated on LB plates (10 g/L Trypton, 5 g/L Yeast Extract, 10 g/L NaCl, and 15 g/L Agar) at 10^{-6} dilutions in LB-Lennox media. Colonies were then replica-plated [173] on sterile velvet: (1) LB plates with colonies touched to velvet stamp, (2) sterile Kan_{50} plate touched to velvet stamp, (3) sterile LB plate touched to velvet stamp. All plates were incubated overnight at 32^{0}C. Colonies that grew on LB plates and not Kan_{50} plates were inoculated in LB-L media and grown overnight at 32^{0}C, 250 rpm. Colony PCR was used to verify KanR was removed using NEB Taq DNA polymerase with Standard Taq Buffer with EDC258 and EDC259 at 95^{0}C for 15 min, with 30 cycles of 95^{0}C for 30 s, 57^{0}C for 30 s, and 68^{0}C for 8 min, and a final extension of 68^{0}C for 15 min. PCR reactions column purified and sequence verified.

4.5.5 Growth Curves

Overnight cultures of strains were grown in LB-Lennox (16 g/L Trypton, 10 g/L Yeast Extract, and 5 g/L NaCl) media at 34^{0}C at 250 rpm and were diluted 1:50 in 100 µL of 2x YT media. Diluted cultures (100uL) were added to 96-well polystyrene plates (Costar 3370). The OD_{600} was measured at 10 min intervals for 20 hr at 32^{0}C in orbital shake mode on a SynergyH1 plate reader (Biotek, Winooski, VT, USA). Growth data for each strain was obtained from 3 replicate wells and three independent cultures, unless otherwise noted. Doubling time was calculated during exponential growth phase.

4.5.6 Genomically modifying proteases and a nuclease in C321.ΔA-T7RNAP-β

The strains in this study were generated from *C321.ΔA-T7RNAP* by disrupting genes of interest using mutagenic oligo nucleotides by MAGE [22] (**Supplemental Table A4**). Cultures were grown in LB-Lennox Media (10 g/L Trypton, 5 g/L Yeast Extract, and 5 g/L NaCl) at 32^{0}C and 250 rpm throughout the MAGE cycle steps [22]. Single, double, a triple mutant was made to lon, ompT, rne to investigate the effect of reduced nuclease and protease activity on expression of proteins regulated by the T7RNAP. Multiplex allele-specific colony (MASC) PCR was performed to screen for gene mutations using wild-type forward (-wt-f) or mutant forward (-mut-f) primers and reverse primers (-r; **Supplemental Table A4**). Wild-type and mutant forward primers were identical except at the 3'-ends of the oligonucleotide, and the reverse primers were used for detection of both wild-type and mutant alleles. The mutant allele was amplified when using the mutant forward and reverse promoter set (-mut-f and –r) resulted in a band on an electrophoresis gel but not when the wild-type forward and reverse primer set (-wt-f and –r). MASC PCR was performed in 10 µL reactions by using a multiplex PCR kit (Qiagen) at 95^{0}C for 15 min, with 30 cycles of 95^{0}C for 30 s, 65^{0}C for 30 s, and 72^{0}C for 1 min, and a final extension of 72^{0}C for 5 min. Selection for lon mutants were performed separately in 10 µL reactions using NEB Taq DNA polymerase with Standard Taq Buffer at 95^{0}C for 15 min, with 20 cycles of 95^{0}C for 30 s, 55^{0}C for 30 s, and 68^{0}C for 2min, and a final extension of 68^{0}C for 5 min. Mutant alleles were screened by running PCR products on a 2% agarose gel and confirmation by DNA sequencing by using sequencing primers (**Supplemental Table A4**).

4.5.7 Assaying expression of GFP of C321.ΔA-T7RNAP-β containing genomic modifications

Strains were freshly transformed with pT7sfGFP-wt. A single colony was inoculated into 5 mL of LB-Lennox media with 35 µg/mL Kanamycin (Kan_{35}) and grown overnight at 32^{0}C at 250 rpm. Overnight cultures were diluted 1:50 into 5 mL of fresh LB-Lennox Kan_{35} in duplicate and grown at 32^{0}C at 250 rpm. OD_{600} was monitored on a NanoDrop 2000c until OD_{600} 0.6-0.8 at which point cultures were induced with 1 mM IPTG or 0 mM IPTG. Cultures were allowed to express the proteins of interest for 7 hr after induction. To assay fluorescence, cultures were diluted 10-fold in LB-Lennox Kan_{35} media. 100 µL of the 10-fold dilution was added to 96-well polystyrene plates (Costar 3603) in duplicate. Fluorescence and OD_{600} of each well was measured on a Synergy H1 plate (Biotek, Winooski, VT, USA) with a gain of 60.

The final normalized fluorescence was obtained by dividing normalized fluorescence (to LB-Lennox Kan_{35} media wells) by the OD_{600}.

4.5.8 Replacing putative promoter with T1T2 terminator

The bla-T1T2 cassette, for removing the putative promoter upstream T7RNAP at genomic location 815988, was assembled from two pieces: a bla gene and a T1T2 terminator. The bla gene was amplified using Phusion with EDC398 and EDC399 primers using the pAM552C plasmid [143], 3% DMSO at 98^0C for 3 min, with 25 cycles of 98^0C for 30 s, 55^0C for 30 s, and 72^0C for 45 s, and a final extension of 72^0C for 10 min. The T1T2 terminator was amplified using Phusion with EDC400 and EDC401 using the pDTT1 (derived from pET24a) plasmid at the same conditions listed above detailed. The bla and T1T2 PCR reactions each received 1 µL of DpnI per 20µL of PCR reaction and were incubated at 37^0C for 2 hr. PCR reactions were column purified and run on a 0.7% agarose gel at 140 V for 30 min. The correct sized band was cut out of the gel and column purified. T1T2 and bla parts were pooled together at an equal molar concentration (33 ng of DNA total) in an overlap PCR reaction using Phusion, 3% DMSO at 98^0C for 3 min, with 15 cycles of 98^0C for 30s, 55^0C for 30 s, and 72^0C for 30s, and a final extension of 72^0C for 10 min. The overlap PCR was then diluted 20-fold into a second PCR reaction with EDC398 and EDC401 primers at 98^0C for 3 min, with 24 cycles of 98^0C for 60s, 55^0C for 30 s, and 72^0C for 45 s, and a final extension of 72^0C for 10 min. PCR reactions were column purified and run on a 0.7% agarose gel at 100 V for 45 min. The correct sized bands were cut out and column purified. The bla-T1T2 cassette was inserted into *C321.ΔA-T7RNAP-β* using Datsenko-Wanner, as detailed in Chapter 4.4.2. Colonies that grew on Cb_{50} plates where picked and inoculated into 100 µL of LB-Lennox Cb_{50} media in sterile 96-well plates and grown overnight at 32^0C, 250rpm. MASC PCR was performed using NEB Taq DNA polymerase with Standard Taq Buffer with EDC405 and EDC406 primers, 0.04% volume saturated culture at 95^0C for 15 min, with 30 cycles of 95^0C for 60s, 52^0C for 30 s, and 68^0C for 2 min, and a final extension of 68^0C for 10 min. The bla antibiotic resistance marker was removed with MAGE oligo EDC404 for colonies that resulted in a correct sized band for the previous MASC PCR. The antibiotic resistance marker knockout procedure is equivalent to Chapter 4.4.4 except recovered cells were plated on LB plates. Colonies that grew on LB plates and not Cb_{100} plates, underwent MASC PCR

with EDC405 and EDC406 at a large scale and were submitted to sequencing with the same primers. The resulting strain was termed, *C321.ΔA-T7RNAP-β-Δbla-T1T2.*

4.5.9 Introducing an anhydrotetracycline inducible T7RNAP cassette into C321.ΔA

The pTet-T7RNAP cassette was assembled from four parts: the synthetic terminator L3S2P21 (SynTerm), the pTet promoter, the T7RNAP, and a CmR gene. The SynTerm is made of two complementary single-stranded oligos. The pTet promoter was amplified using Phusion using JGP272 and JGP278/280/282/284, 3% DMSO at 98^{0}C for 3 min, with 25 cycles of 98^{0}C for 30s, 50^{0}C for 30 s, and 72^{0}C for 1 min, and a final extension of 72^{0}C for 10 min. The T7RNAP was amplified using Phusion with JGP279/281/283/285 and EDC323, 3% DMSO at 98^{0}C for 15 min, with 25 cycles of 98^{0}C for 30 s, 52^{0}C for 30 s, and 72^{0}C for 3 min, and a final extension of 72^{0}C for 10 min. The CmR gene was amplified from pAM552C [143] using Phusion with EDC413 and EDC414 primers, 3% DMSO at 98^{0}C for 15 min, with 25 cycles of 98^{0}C for 60 s, 55^{0}C for 30 s, and 72^{0}C for 45 s, and a final extension of 72^{0}C for 25 min. The pTet promoter, the T7RNAP and the CmR PCR reaction each received 1 µL of DpnI per 20 µL of PCR reaction and were incubated at 37^{0}C for 2 hrs. PCR reactions were column purified and run on a 0.7% agarose gel at 140 V for 30 min. The correct sized band was cut out of the gel and column purified. The SynTerm, pTet promoter, T7RNAP and CmR parts were pool together at an equal molar concentration (75 ng of DNA total) in an overlap PCR reaction using Phusion, 3% DMSO at 98^{0}C for 10 min, with 15 cycles of 98^{0}C for 30 s, 50^{0}C for 30 s, and 72^{0}C for 4 min, and a final extension of 72^{0}C for 10 min. The overlap PCR was then diluted 20-fold into a second PCR reaction with EDC410 and EDC414 primers at 98^{0}C for 3 min, with 24 cycles of 98^{0}C for 30s, 55^{0}C for 30 s, and 72^{0}C for 4 min, and a final extension of 72^{0}C for 10 min. PCR reactions were column purified and run on a 0.7% agarose gel at 140 V for 30 min. The correct sized band was cut out of the gel and column purified. The purified cassette was amplified with Phusion, 3% DMSO at 98^{0}C for 3 min, with 25 cycles of 98^{0}C for 30s, 50^{0}C for 30 s, and 72^{0}C for 4 min, and a final extension of 72^{0}C for 10 min. These primers add homology to the pTet-T7RNAP cassette targeting the insertion to genomic location (8.9b) using Datsenko-Wanner. Screening and sequencing were done similarly as detailed above. Supplemental Table A4 for primer details.

4.5.10 Locating genomic insertion of T1T2-T7RNAP-cat cassette

Because the location of cassette was unknown, we developed a strategy to find the insertion site of T7RNAP. First extracted genomic DNA was run on 1% agarose gel to determine that the genomic DNA was intact. Next, the genomic DNA incubated with 1 µL of EcoRV for every 700 ng of DNA, incubated at 37^0C for 1 hr and column purified. Dilute ligation reactions using T7 Ligase were performed to encourage self-ligation with 10 ng of DNA per 10 µL of ligation reaction at 16^0C for 16 hr and 65 ^{0}C for 25 min. Ethanol precipitation was performed to purify DNA by adding 32% by volume ligation reaction, 64% Absolute EtOH, 3% 3mM sodium acetate pH 5.2 and 0.3 µL Saturated Dextran Blue (for visualizing the pellet). The mixture was then incubated at -20^0C for 40 min, spun down at 13,000 rpm at 4^0C for 10 min, and washed with 200 µL of 70% ethanol. The mixture was then spun down at 13,000 rpm for 3 min at room temperature, excess ethanol was removed and pellet was left to dry out overnight and resuspended with 10 µL of deionized water. Amplification of the circularized genomic fragments were amplified using Phusion with JGP218 and EDC406 (5' end) or JGP216 and JGP217 (3' end) with Master Mix with 5 ng/µL of circularized genome at 95^0C for 10 min, with 25 cycles of 95^0C for 30 s, 52^0C for 30 s, and 72^0C for 2.5 min, and a final extension of 72^0C for 5 min. PCR reactions were column purified and run on a 0.7% agarose gel at 140 V for 30 min. The bands were cut out of the gel and column purified. Samples were then sent to sequencing with JGP218 (5' end) and JGP216 (3' end) to evaluate the genome location.

4.6 Author Contributions

I lead, conceived and executed the majority of this study with the guidance and supervision of Dr. Michael Jewett. Oliver Wessier assisted with the construction and genomic insertion of the T1T2-T7RNAP-cat cassette. Erik Carlson designed the original T7RNAP cassette and assisted me in introducing the cassette genomically and introducing the nuclease and protease mutations. I wrote the study summary.

CHAPTER 5

5 Translation system engineering in *Escherichia coli* enhances non-canonical amino acid incorporation into proteins

5.1 Abstract

The ability to site-specifically incorporate non-standard amino acids (nsAAs) into proteins has made possible the study of protein structure and function in fundamentally new ways, as well as the synthesis of unnatural polymers. However, the task of site-specifically incorporating multiple nsAAs into proteins with high purity and yield continues to present a challenge. At the heart of this challenge lies the lower efficiency of engineered orthogonal translation system components compared to their natural counterparts (*e.g.,* translation elements that specifically use a nsAA and do not interact with the cell's natural translation apparatus). Here, we show that evolving and tuning levels of multiple components of the engineered translation system together as a whole enhances nsAA incorporation efficiency. Specifically, we dramatically increase protein yield when incorporating multiple *p*-azido-phenylalanine into proteins by (i) evolving the *Methanococcus jannaschii p*-azido-phenylalanyl-tRNA synthetase anti-codon binding domain, (ii) evolving the elongation factor Tu aminoacyl tRNA binding domain, and (iii) test evolved translation machinery components in fine-tuned expressional vector. Use of the evolved translation machinery in a genomically recoded organism lacking release factor 1 enabled enhanced multi-site nsAA incorporation into proteins. We anticipate that an approach to orthogonal translation system development, like that reported here, will accelerate and expand our ability to site-specifically incorporate multiple types of nsAAs into proteins and biopolymers, advancing new horizons for synthetic and chemical biotechnology.

5.2 Introduction

The site-specific, co-translational incorporation of non-standard amino acids (nsAAs) into proteins expands the range of genetically encoded chemistry in proteins. This expansion can be a powerful tool for both fundamental science and applied technology [174-176]. For example, photo-caged [177], fluorescent

[178], and bio-orthogonal reactive [11, 179, 180] nsAAs have provided new ways to study protein structure and dynamics [176]. In addition, nsAAs that mimic natural post-translational modifications help to elucidate the role of such modifications in previously unattainable ways [41, 83, 146, 175, 181-183]. Further, nsAA incorporation into proteins has opened the way to novel antibody drug conjugates [184, 185] , modified human therapeutics [186] , and protein biomaterials [187], among other applications. With the ability to construct bio-based products beyond the limits of nature, expanding the genetic code has emerged as one of major opportunities in synthetic and chemical biology [100, 176, 188, 189].

Expansion of the genetic code by orthogonal translation systems (OTSs) involves codon re-assignment and engineered translation machinery. The engineered translation machinery is evolved to recognize a nsAA and operate alongside the cell's natural translation apparatus in a parallel and independent fashion. It is orthogonal in the sense that it does not recognize natural amino acids. Most commonly, the amber UAG stop codon is re-assigned to a nsAA in an approach called amber suppression [190]. In this approach pioneered by Schultz and colleagues, an orthogonal tRNA (o-tRNA) anticodon is mutated to allow for decoding of the UAG stop codon [191, 192], and an orthogonal aminoacyl-tRNA synthetase (o-aaRS) is evolved to aminoacylate the nsAA to the o-tRNA. These o-tRNA/o-aaRS pairs are typically derived from phylogenetically distant organisms such as *Methanocaldococcus jannaschii* or *Methanosarcina mazei*, and are not efficiently recognized the aaRS/tRNA pairs of the host, *i.e.*, *Escherichia coli* (*E. coli)* [193, 194]. The o-tRNA/o-aaRS are then evolved so that the o-aaRS charges a nsAA to the o-tRNA.

Once charged, the nsAA-tRNA complex must be properly delivered to the ribosome by elongation factor Tu (EF-Tu). As part of key proofreading steps, the thermodynamic interactions between EF-Tu and aminoacylated tRNAs are finely balanced to bind the canonical amino acids with their cognate tRNAs [195], and not mismatched amino acid/tRNA substrates. Consequently, nsAA-o-tRNA substrates may not bind EF-Tu efficiently because it is a non-native substrate, preventing efficient delivery to the ribosome. EF-Tu may thus require engineering to allow for the efficient incorporation of a given nsAA [41]. Beyond o-aaRSs, o-tRNAs, and EF-Tu, ribosomes must be able to accommodate the nsAA and in some instances ribosome engineering may be necessary. Orthogonal ribosome systems decouple specialized

ribosome function from organism fitness, allowing for potential ribosome engineering to accommodate larger or charged nsAAs [86, 143, 196, 197].

To date, over 150 nsAAs have been co-translationally incorporated into proteins in *E. coli* [100, 198]. While these advances highlight our ever-expanding understanding of the workings of the translational apparatus, they have also revealed areas for improvement. Rigorous analyses of several o-aaRSs reveal that these enzymes, while functional, generally have poor catalytic efficiency as compared to native synthetases [28-30, 199-202]. This has led to limitations in site-specific, multi-site nsAA incorporation into proteins. Moreover, o-aaRSs show polyspecificity, meaning a single OTS can incorporate multiple nsAAs. This is useful for expanding nsAA diversity without the need for further evolution, but hinders incorporation of multiple distinct nsAAs by existing o-aaRSs [203-206]. Native EF-Tu also shows limited capacity for incorporation of bulky or charged nsAAs [41], and could be the target of engineering efforts [207]. Further, the presence of release factor 1 (RF-1) can cause early termination of proteins when using amber suppression technology because it competes for the UAG codon [208, 209]. Additionally, overexpression of OTS system components can be detrimental to cellular growth [210], ultimately affecting protein production and cell viability.

Recent advances have led to major improvements in nsAA incorporation and are beginning to address many of the aforementioned limitations. First, optimized OTS expression systems have been shown to enhance suppression efficiency and allow for two different nsAAs to be incorporated into the same protein [4, 211]. Second, efforts to suppress or delete release factor 1 have removed the competition with the nsAA-o-tRNA species at the UAG codon to increase incorporation efficiencies [2, 208, 212-214]. Third, efforts to further engineer individual OTS components have led to improved nsAA aminoacylation efficiencies and increased overall yields of modified proteins, *e.g.,* o-tRNA [4, 215], o-aaRS [105, 215, 216], EF-Tu [41, 217, 218], ribosome [86, 196, 219, 220]. However, the focus of most engineering efforts remains on evolving OTS components by targeting only individual components, rather than exploring more than one biological part involved in the complex system of protein biosynthesis.

In this work, we sought to explore the effects of evolving both the o-aaRS and EF-Tu for improved *p*-azido-phenylalanine (pAzF) incorporation into proteins in *E. coli*. The foundational principle was that a

systematic engineering approach of evolving multiple components of the orthogonal translation machinery concurrently could provide synergistic opportunities to enhance nsAA incorporation into proteins. Our study involved three steps. First, we evolved the aminoacyl-tRNA synthetase of *Methanococcus jannaschii* to improve pAzF incorporation into proteins. Previous works have already identified *M. jannaschii* o-tRNA/o-aaRS pairs that can be used to site-specifically install pAzF into proteins [11, 105]. However, based on our recent work and that of others [105, 218], we hypothesized that the anticodon recognition domain of the o-aaRS could be further engineered. We explored a total of 10 amino acid residues residing in the anticodon recognition domain of the *M. jannaschii* o-aaRS and demonstrated the ability to isolate variants with increased pAzF incorporation efficiency. Second, we evolved the *E. coli* EF-Tu to improve pAzF incorporation into proteins. We hypothesized that mutations at the amino acid-binding pocket of EF-Tu might yield an enhanced OTS by tuning and optimizing 2 biological parts involved in the complex system of protein biosynthesis. Indeed, our results showed that we could further enhance nsAA incorporation into proteins by combining the beneficial mutants of both the o-aaRS and EF-Tu. Third, we assessed the ability of our engineered OTS to synthesize proteins containing multiple site specifically introduced pAzF residues. Though our engineered system does not benefit from the already high suppression efficiency of a single UAG amber codon, multi-site incorporation of several pAzF residues is significantly improved, an observation that is amplified in a genomically recoded strain lacking RF-1 [2] and consistent with another recent report [221]. In sum, our work demonstrates that a comprehensive engineering approach is for genetic expansion efforts that require highly efficient OTSs. This work has implications for biotechnology, protein engineering, and synthetic biology projects.

5.3 Materials and Methods

5.3.1 Reagents and Buffers

Chemicals and media were purchased from Sigma Aldrich unless designated otherwise. DNA polymerases, T4 polynucleotide kinase, T4 DNA ligase, and restriction endonucleases were purchased from New England Biolabs (NEB). Plasmids were extracted using Omega E.Z.N.A. DNA/RNA Isolation Kits (Omega Bio-Tek). All DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc (IDT). p-Azido-phenylalanine (pAzF) was purchased from P212121, LLC.

5.3.2 Strains and plasmids

DH10-beta competent cells were purchased from New England Biolabs (NEB). All oligonucleotides used for cloning are shown in **Supplemental Table A5**, and all vectors are shown in **Supplemental Figure A15**. The genes of the orthogonal translation system (EF-Tu, pAzFRS, and o-tRNA variants) were cloned into expression vectors using standard methods. The initial source of these genes are as follows: *E. coli* native EF-Tu-coding sequence was amplified from the genome of *E. coli* KC6 strain [222]; p-Azido-phenylalanine aminoacyl tRNA synthetase (pAzFRS) was amplified from the plasmid pEVOL-pAzF [4]; the amber codon suppressor tRNA, *M. jannaschii* $tRNA^{tyr}_{CUA}$, was amplified from the plasmid pEVOL-pAzF [4]. Both EF-Tu and pAzFRS were inserted into plasmid vector pDLppTT1 (kanamycin) flanked by the lpp promoter and the T1T2 terminator. The resulting plasmids were named pDLppEAzRSTT1 and pDLppEFTuTT1 respectively (**Supplemental Figure A11**). Plasmids bearing super folder green fluorescent protein (sfGFP) based reporter genes were constructed by insertion of the sfGFP gene from the plasmid pY71sfGFP [223] with 1 amber codon at position D190 into the plasmid vector pDT7TT2 (Carbenicillin). The resulting plasmid, which harbors a T7 promoter and T1T2 terminator flanking the sfGFP gene, was termed pDT7sfGFP1TAGTT2. From this construct, we used standard cloning procedures to generate the 3-amber (D36, K101, and D190) and 5-amber (D36, K101, E132, D190, and E213) sfGFP reporter plasmids, which were named pDT7sfGFP3TAGTT2 and pDT7sfGFP5TAGTT2 respectively. We also generated a series of constructs with the T7 RNA polymerase promoter replaced by the native promoter PtacI [224] for use in *E. coli* C321.ΔA [2], which lacks a genomic copy of T7 RNA polymerase. These plasmids were named pDPtacIsfGFPXTAGTT2 (X: 1, 3, or 5) (**Supplemental Figure A11**). To create the library of orthogonal translation constructs with different promoter strengths (termed JGPX (X: 1-27), the three OTS component genes were inserted into the JGPX plasmids, with either the T7, Ptac or lpp promoters in all combinations. The pAzFRS, o-tRNA, and EF-Tu were inserted using, NheI and EcoRI, BamHI and EcoRI and BamHI and NheI restriction sites at the 5' and 3' ends, respectively. All plasmid constructs were sequence verified. As for the construction of JGP26BM, we simply replace the parental pAzFRS and wildtype EF-Tu with the evolved mutants respectively in the plasmid JGP26 (**Supplemental Figure A11**).

5.3.3 p-Azido-phenylalanine-tRNA synthetase mutagenesis and selection

Based on the 3-dimensional structure of the *M. jannaschii*-based pAzFRS/tRNA pair (PDB ID: 1J1U), a total of 10 amino acids in the vicinity of the o-tRNA anti-codon loop were mutated to assess the ability to enhance pAzF incorporation efficiency. This was achieved through the use of several libraries as described in the text following a positive-negative selection procedure reported previously in the literature [225, 226]. We found that LB medium caused less false positive colonies than the rich 2xYT during positive-negative coupled selection probably due to the low concentration of 20 canonical amino acids and other small molecules. The selection scheme is shown in **Supplemental Figure A12**. In the positive selection, functional aaRS constructs are tested *in vivo* for their ability to suppress an in frame amber codon in a chloramphenicol acetyl transferase (*cat)* reporter that confers resistance to chloramphenicol. In the negative selection, synthetases that loaded a natural amino acid were selected against using the toxin protein barnase.

In library 1(RsLib1), A233, P258, F261, H283, M285, and R286 were mutated as NNK (N: A, T, G, or C; K: G or T) combinatorially. After construction, the plasmid library was transformed by electroporation into *E. coli* DH10-beta containing the pRepCMD112 plasmid (Supplementary Figure A15) derived from pRepCM3 [227] to produce ~ 10^8 transformants. Winning transformants were identified by their ability to grow on LB solid media supplemented with 25 µg/mL kanamycin and 10 µg/mL tetracycline in the presence of 0.5 mM IPTG, 1mM p-azido-phenylalanine (pAzF), and 10 µg/mL chloramphenicol (first-round selection) or 20 µg/mL chloramphenicol (second-round selection), at 37°C for 36 hours. Plasmids from surviving colonies were extracted and transformed into DH10-beta *E. coli* with pLWJ17B3 (which encodes the barnase gene) for negative selection [226]. The negative selection was carried out as above but in the absence of pAzF. At last, plasmids from surviving colonies were transformed into DH10-beta *E. coli* harboring pDPtacIsfGFP1TAGTT2 plasmid for the final round of screening of fluorescent intensity in 96-well plate in 100 µl of 2xYT media with 25 µg/mL kanamycin and 10 µg/mL tetracycline in the presence of 1 mM IPTG, 1mM p-azido-phenylalanine (pAzF) or 0mM pAzF for negative control samples (**Supplementary Figure A11**).

In library 2 (RsLib2), another 4 amino acids (C231, P232, P284, and K288), which are close to the U35 and A36 of the o-tRNA anticodon loop, were mutated to NNK individually (only one mutation site per DNA molecule). Rather than the antibiotic selection described above, we focused here on the analysis of sfGFP expression by intact cell fluorescence measurements where a sfGFP containing an amber codon will be expressed at a higher level if an orthogonal aaRS/tRNA cognate pair has enhanced activity due to a beneficial mutation. After library construction, the plasmid library was transformed into DH10-beta cell with pDULE-tRNA and pDPtacIsfGFP1TAGTT2 (the sfGFP reporter plasmid) and grown on LB solid media (25 µg/mL kanamycin, 10 µg/mL tetracycline; and 50 µg/mL Carbenicillin) at 37°C for 24 hours. To enable complete library coverage, six hundred single colonies were then picked and grown in 96-well plates in 2xYT liquid media with 25 µg/mL kanamycin, 10 µg/mL tetracycline; and 50 µg/mL Carbenicillin. At an optical density (OD_{600}) = 0.2, 1mM IPTG and 1mM p-azido-phenylalanine (pAzF) were added to the media to induce protein expression allowing for pAzF incorporation into sfGFP containing an amber codon. The fluorescence (excitation/emission = 488/525 nm) of each well was read in Synergy2 (BioTek) with shaking and normalized by optical density with the wavelength of 600 nanometers (Fluorescence per OD_{600}). Each aaRS variant was treated as a separate experiment, and sfGFP levels were compared with the progenitor from which it was evolved.

In library 3 (RsLib3), H283, P284, M285, and R286 were mutated to NNK (N: A, T, G, or C; K: G or T) combinatorially. Selection was performed as described for RsLib1.

5.3.4 EF-Tu mutagenesis and screening

Based on the known 3-dimensional structure of EF-Tu (PDB ID: 1OB2), we created a mutant library in the EF-Tu substrate recognition pocket to assess its ability to enhance pAzF incorporation. This library included a total of 10 amino acids: S65, H66, E215, D216, V217, F218, T228, F261, N273, and V274. As this library was too large to fully mutate to all other amino acids, we pursued an alanine scanning approach whereby we mutated each amino acids residue to alanine in all combinations using the method depicted in **Supplemental Figure A13**. The wildtype EF-Tu-coding sequence was amplified as 5 fragments. Each fragment carried the desired mutations introduced by DNA oligos. Then the 5 fragments, each with overlapping regions, were assembled to full-length EF-Tu sequence by a thermal

cycling procedure: 95°C, 1 min; 20 cycles of 98°C, 20 s; 50°C, 20s; 72°C, 2 min; and a final extension of 72°C, 5 min. The full-length product was cloned into pDLppTT1 plasmids using flanking NdeI and XhoI restriction sites. Twenty samples were sequenced to generate snapshot of the library's diversity (**Supplemental Figure A14**). Next, the EF-Tu library-coding fragment was digested by the restriction endonucleases NheI and BamHI and inserted into pDLppEAzRSTT1 via the same cut sites, resulting in pDLppEAzEFTT1 (**Supplemental Figure A11**). Lastly, the plasmid library was transformed into DH10-beta strain harboring the plasmid pDPtacIsfGFP1TAGTT2 and pDULE-tRNA (**Supplemental Figure A11**). Single colonies were picked from 2xYT solid media plate and inoculated into fresh 2xYT liquid media (25 µg/mL kanamycin, 10 µg/mL tetracycline; and 50 µg/mL Carbenicillin) in a 96-well plate (100 µl media /well). The cells were cultivated to early exponential phase (OD_{600} ~= 0.2), at which point isopropyl β-D-1-thiogalactopyranoside (IPTG) and p-azido-phenylalanine (pAzF) were supplemented with the final concentration of 1 mM each. The cultures were further grown for 8 hours at 30°C with shaking to saturation. Then, the bulk fluorescence (525 nm) of each well was read in Synergy2 (BioTek) and normalized by the cell number measured by the optical density at 600 nm.

5.3.5 Analysis of GFP expression by intact cell fluorescent measurements

Liquid cell cultures of strains harboring plasmid based orthogonal translation systems and GFP reporter plasmids were inoculated from frozen stocks and grown to confluence overnight in 3mL of 2xYT media at 37°C. Cultures were then inoculated at 1:50 dilution in 3mL of 2xYT media with necessary antibiotics, and the cells were allowed to grow at 37°C to an OD_{600} of 0.5–0.7 in a shaking plate incubator at 650 r.p.m. (~3h). Isopropyl β-D-1-thiogalactopyranoside (IPTG) and pAzF were added to a final concentration of 1 mM to induce protein expression for 8 hours. About 5.5 x 10^7 of cells of each sample were collected and washed twice with 200 µL of PBS buffer. The cells were resuspended in 100 µL of PBS buffer in 96-well plate. The fluorescence of sfGFP was measured on a BioTek Synergy 2 spectrophotometric plate reader using excitation and emission wavelengths of 488 and 525 nm, respectively. Fluorescence signals were normalized by the OD_{600} reading.

For *in vivo* fluorescence measurements in the promoter library assays, total 27 plasmids, plasmid JGP1 to 27, were transformed into BL21(DE3) individually. Three mL of LB media was inoculated with a

single colony in a test tube and grown overnight at 37°C. Next, saturated culture was diluted 50-fold into 100 µL of LB media in 96-well plates in triplicate. Cultures and inducers were added individually to each well. Cultures were grown in Synergy (BioTek) plate at 37°C until early mid-exponential phase at which point the cultures were induced with 1 mM IPTG, 0.02% arabinose, and 0 or 10 mM pAzF. Fluorescence and OD_{600} readings were taken at 7.5 hrs after induction for maximum fluorescence/OD_{600} signal.

For the evaluation of solubility of sfGFP harboring p-azido-phenylalanine, 5 OD_{600} x ml of cells of each sample were collected and washed twice with 500 µL of PBS buffer. The cells were resuspended in 500 µL of PBS buffer. The cells were then sonicated with 250-300 joules to extract total protein. Soluble fraction was separated by spinning the lysate at 21, 000 x g for 10 min. The insoluble fraction was prepared by suspending the pellet in 500 µL of PBS buffer. The total protein, soluble fraction, and insoluble fraction were resolved in NuPAGE™ Novex™ 4-12% Bis-Tris Protein Gels (Invitrogen, CA).

5.3.6 Mass Spectrometry

The purified protein was analyzed by nanocapillary LC-MS using a 100 mm × 75 µm ID PLRP-S column in-line with an Orbitrap Elite (Thermo-Fisher, Waltham, MA). All MS methods included the following events: (1) FT scan, m/z 400−2000, 120 000 resolving power and (2) data-dependent MS/MS on the top 2 peaks in each spectrum from scan event 1 using higher-energy collisional dissociation (HCD) with normalized collision energy of 25, isolation width 15 m/z, and detection of ions with resolving power of 60 000. All data were analyzed using QualBrowser, part of the Xcalibur software packaged with the ThermoFisher Orbitrap Elite.

5.4 Results

5.4.1 Evolution of p-azido-phenylalanine–tRNA synthetase.

We hypothesized that mutations at or near the anti-codon recognition domain of the pAzFRS might yield variant aaRSs that more efficiently incorporate pAzF at the amber codon. This hypothesis was based on recent works that show that impaired binding of the pAzFRS to its cognate o-$tRNA_{CUA}$ might reduce efficiency of the system [105], likely because the native *M. jannaschii* TyrRS (*Mj*TyrRS) normally recognizes the GUA anticodon rather than the CUA anticodon to decipher the amber stop codon [12]. The following general approach was used to improve the interaction between the o-tRNA anti-codon loop and

the pAzFRS. First, we used crystal structure-guided information to generate a combinatorial mutation library. Simultaneously, we carried out site-directed mutagenesis to interrogate several individual residues. At last, we sequenced and characterized improved OTS variants obtained via selections.

Crystal structure analysis revealed at least 10 residues in the tRNA-recognition domain of the pAzFRS as targets for modification (**Figure 5.1**). Since a total random combinatorial mutation library of all residues was not feasible (20^10), we created three distinct libraries from the residues identified and sought to evolve the pAzFRS with a plasmid-based orthogonal translation system using a positive and negative selection system previously described [11]. The first library (RsLib1) diversified six residues (A233, P258, F261, H283, M285, and D286) in close proximity to the o-tRNA C34 nucleotide (wildtype G34). After two rounds of positive-negative selection, 300 survived pAzFRS variants were further evaluated quantitatively for their ability to produce superfolder green fluorescent protein (sfGFP) with one amber codon (sfGFP(1UAG)). Direct comparisons to the progenitor enzyme were carried out in a 96-well plate fluorescence-based assay (**Supplemental Figure A12**). Sequencing revealed that 25% of the isolated clones converged to a single variant (**Supplemental Table A6**). The top variant, termed BM1 (P258A, F261P, H283L, M285F, and D286Y) showed approximately a 5-fold increase in normalized fluorescence as compared to the progenitor aaRS and a 2-fold increase as compared to a reported variant harboring D286R single-point mutation, which we term EAzRS in our study [12] (**Figure 5.2A**).

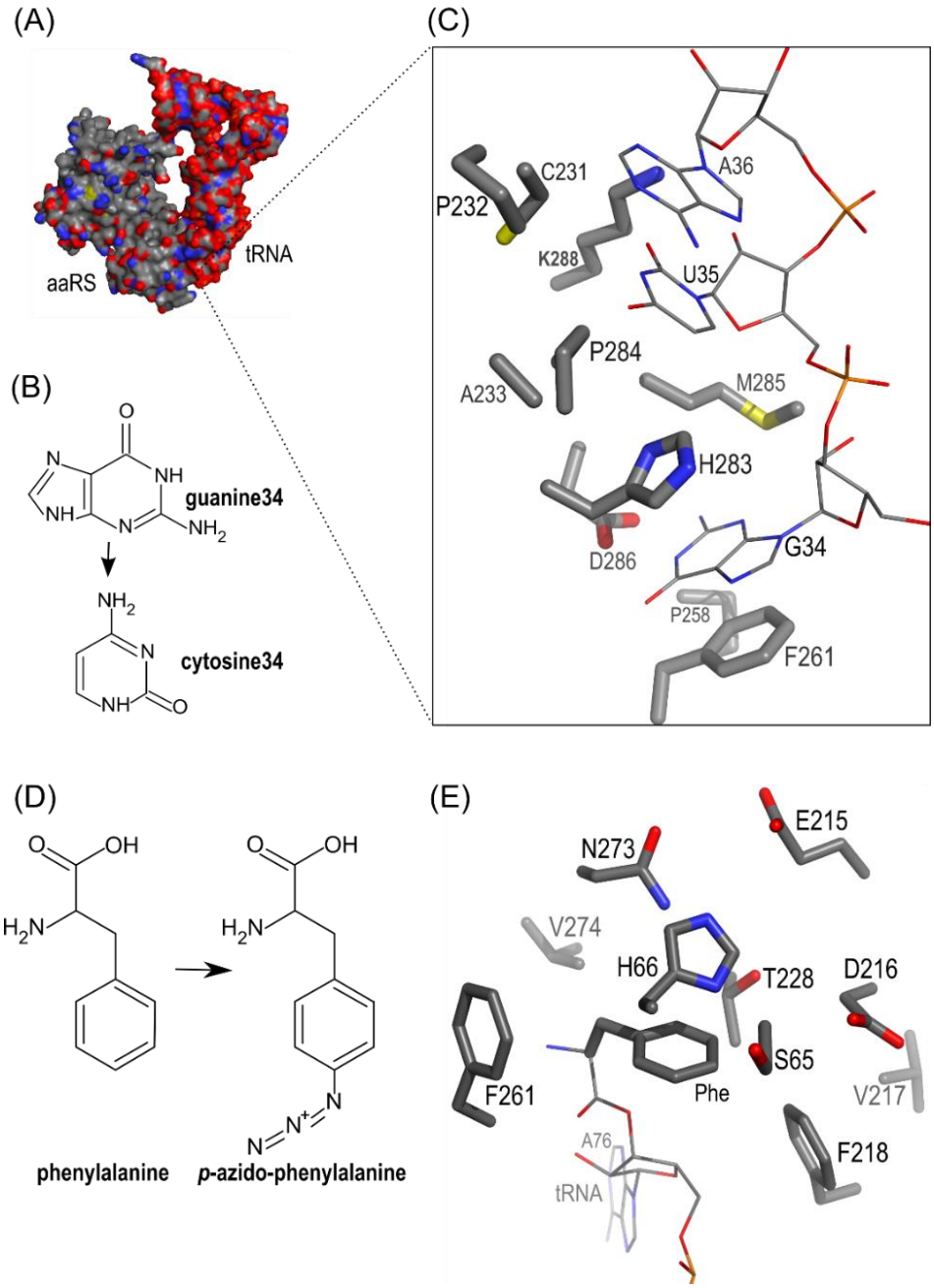


Figure 5.1: Three-dimensional structure information and library design. (A) Crystal structure of the *Methanococcus jannaschii* tyrosyl-tRNA synthetase (left: aaRS) and tyrosine-tRNA^{Tyr} (right: tRNA) complex (PDB: 1J1U); (B) Molecular structure of guanine and cytosine; (C) tRNA-recognition domain of the synthetase and anticodon loop of tRNA. The residues C231, P232, A233, P258, F261, H283, P284, M285, D286, and K288 in tRNA-recognition domain underwent mutagenesis. The nucleotide G34 in tRNA was changed to cytosine for amber codon suppression. (D) Molecular structure of phenylalanine and p-azido-phenylalanine (pAzF); (E) The amino acid-binding pocket of elongation factor Tu and phenylalanyl-tRNA (PDB: 1OB2). Ten amino acids: S65, H66, E215, D216, V217, F219, T228, F261, N273, and V274 were mutated to alanine in this study. Phe: phenylalanine charged in tRNA molecule.

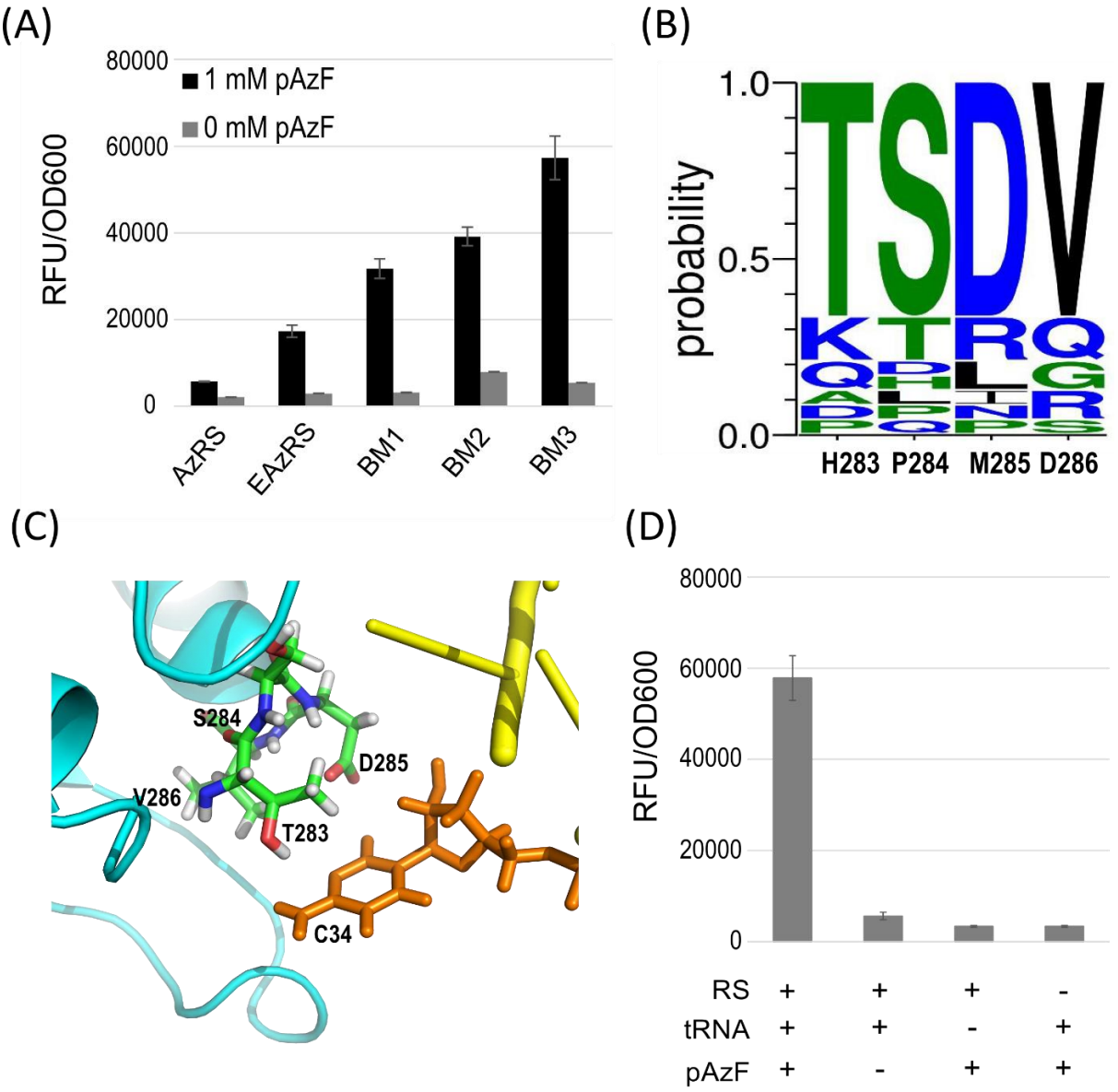


Figure 5.2: The evolution and identification of beneficial variants of *p*-azido-phenylalanine tRNA synthetase (pAzFRS). (A) Representative variants identified during evolution. AzRS: the progenitor pAzFRS first reported by Schultz's group [11]. EAzRS: the pAzFRS harboring a D286R mutation reported by Yokoyama's group [12]. BM1: the top variant obtained from the selection of RsLib1 with the mutations: P258A, F261P, H283L, M285F, and D286Y. BM2: A beneficial variant identified from RsLib2 with a single-site mutation of P284V in addition to the known D286R mutation. BM3: a beneficial variant from the selection of RsLib3 with the mutations: H283T, P284S, M285D, and D286V. BM3 mutant showed approximate 8-fold improvement over the parental pAzFRS and 3-fold over the reported D286R mutant. (B) Sequence convergence after the selection of RsLib3. (C) Highlighted mutated sites of BM3 in 3-D model of pAzFRS. (D) Negative control experiments for BM3 lacking one component for ncAA incorporation. RS: BM3 synthetase; tRNA: o-tRNA; pAzF: p-azido-phenylalanine. Standard deviation accounts for the day-to-day variation in 3 independent experiments and 3 colonies were picked for each samples.

Having identified improvements from modifying the 6 residues closest to the o-tRNA C34 nucleotide, we next sought to explore the impact of another 4 amino acid residues in the anticodon binding domain (C231, P232, P284, and K288). These residues were selected because they are close to the o-tRNA U35 and A36 nucleotides of the anticodon and we hoped they would additionally contribute towards aaRS-$tRNA_{CUA}$ binding optimization. We chose here to use a NNK library to diversify the four positions individually, which still covers all 20 amino acids at each single site. Given the importance of the D286R mutation, shown in our initial selection and in previous works, we chose to incorporate this into our starting pAzFRS gene template. Due to the targeted nature of this library (RsLib2), the 4 target residues were mutated to NNK for a total of 80 variants (20+20+20+20). We picked 600 clones to ensure complete library coverage and screened for improved production of sfGFP(1UAG) using an *in vivo* fluorescence assay in the presence and in absence of pAzF. This screening identified that the mutant BM2 with a P284V mutation resulted in a 2.2-fold improvement in sfGFP(1UAG) synthesis as compared to the EAzRS (**Figure 5.2A**) (**Supplemental Table A7**).

Based on the identification of mutants P284V and D286R in a single secondary structure turn in pAzFRS, we created a third library (RsLib3) focusing only on this turn element directly (H283, P284, M285, and D286). Specifically, we generated a combinatorial mutation library at all 4 positions and performed two rounds of positive-negative selection using the chloramphenicol acetyl transferase/barnase system described above (**Supplemental Figure A12**). Individual colony analysis of twenty-four mutants revealed that 66% or 16 of 24 converged to a single sequence of H283T, P284S, M285D, and D286V (**Figure 5.2B & 5.2C; Supplemental Table A8**). This top variant, termed BM3, with the T283-S284-D285-V286 mutations, was capable of producing ~8-fold more sfGFP(1UAG) than the progenitor enzyme (**Figure & 5.2A**). As a control, we verified that expression of sfGFP(1UAG) with the BM3 encoding plasmid was dependent on both the o-tRNA-bearing plasmid and pAzF (**Figure 5.2D**).

5.4.2 Engineering the amino acid-binding pocket of EF-Tu by combinatorial alanine-scanning method to improve pAzF incorporation.

Once o-tRNAs are charged with a nsAA of interest by an o-aaRS, the charged tRNAs must be shuttled to the ribosome by the EF-Tu. The EF-Tu recognizes the nsAA-tRNA substrate at two regions: (i)

the amino acid binding pocket and tRNA acceptor stem with residues from the β-barrel domain 1 and GTPase domains; and (ii) the T-stem region of the tRNA with residues from the β-barrel domain 2 and the GTPase domain [228]. With the goal of engineering the interaction between pAzF-tRNA$_{CUA}$ and EF-Tu, we sought to identify variants in the amino acid binding pocket of EF-Tu that might favor pAzF-o-tRNA delivery to the ribosome. Based on the 3-dimensional structure of EF-Tu (PDB ID: 1OB2), we identified a total of 10 amino acids around the amino acid-binding pocket of EF-Tu (S65, H66, E215, D216, V217, F218, T228, F261, N273, and V274) to modify (**Supplemental Figure A13**). As with the pAzFRS evolution, there are too many residues to completely diversify and perform a complete screening of the library. Thus, we chose to carry out a more constrained design, by focusing on the construction of an alanine scanning library. By mutating EF-Tu binding pocket residues to the smaller amino acid alanine, the pocket would essentially be enlarged for the nsAA. This library of 1024 members was subjected to screening by intact cell fluorescent measurements and its ability to produce sfGFP(1UAG). In this case, the EAzRS synthetase (the pAzFRS with single D286R mutation) was used along with its cognate o-tRNA partner, because the work here was done in parallel to the synthetase evolution experiments described above. Beneficial EF-Tu mutants were identified by screening for the ability to improve synthesis of sfGFP(1UAG) as compared to wild-type EF-Tu in a 96-well plate fluorescence assay (**Supplemental Figure A12B**). The best mutant, EFBM1, showed approximately a 4-fold improvement as compared to the plasmid-bearing the native EF-Tu (**Figure 5.3A**, the bar “EAzEF1”). Individual colony sequencing revealed that the EFBM1 variant carried 3 alanine mutations at S65A, D216A, and V274A. The S65A and D216A mutations are located near the phenylalanine moiety, so the increased size of the EF-Tu amino acid binding pocket likely allows for improved pAzF accommodation. V274A is located close to the acceptor stem of the tRNA moiety. This mutation most likely alleviates the steric hindrance of the aminoacylated acceptor stem. Finally, we combined the top variants isolated from the aaRS evolution (BM3) and the EF-Tu evolution (EFBM1). The combined system with improved o-tRNA$_{CUA}$ binding and pAzF-tRNA binding resulted in a 6-fold increase in sfGFP(1UAG) production compared to the combination of EAzRS and wildtype EF-Tu (**Figure 5.3A**, the bar “EAzEF2”).

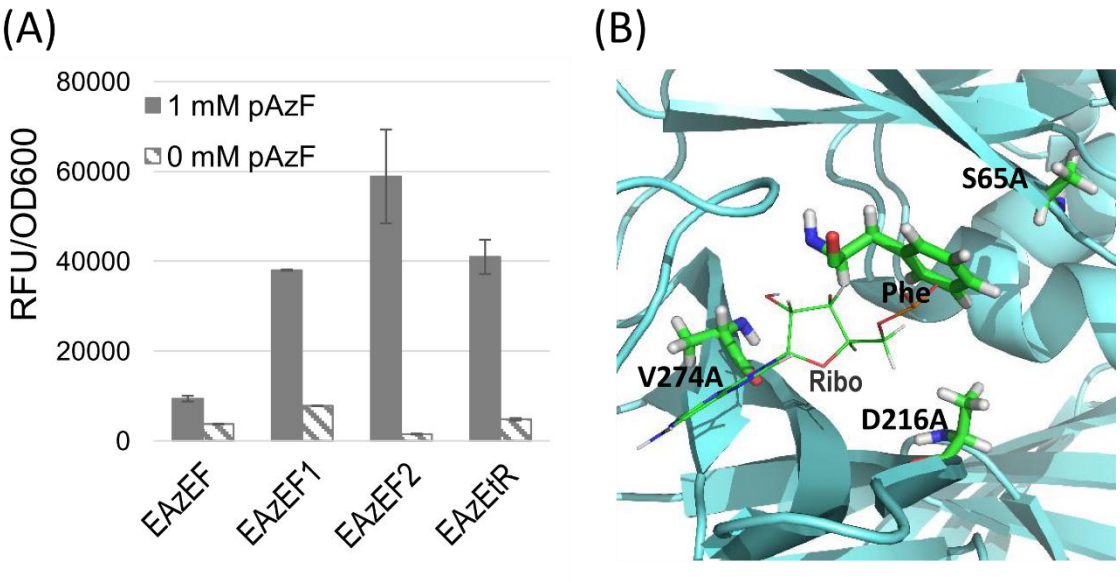


Figure 5.3: The evolution and identification of beneficial EF-Tu mutants for *p*-azido-phenylalanine (pAzF) incorporation. (A) pAzF incorporation efficiency accomplished by wildtype and evolved synthetase and EF-Tu molecules respectively. EAzEF: plasmid-born wildtype EF-Tu and EAzRS (D286R); EAzEF1: plasmid-born beneficial EF-Tu variant EFBM1 and EAzRS (D286R); EAzEF2: plasmid-born evolved EF-Tu EFBM1 and evolved synthetase BM3; EAzEtR: plasmid-born evolved EF-Tu EFBM1 and evolved synthetase BM3 working with a previously optimized orthogonal tRNA [4]. All experiments were performed in the DH10-beta cell with the expression of native EF-Tu from genome. (B) Three dimensional model of EFBM1 with highlighted mutated sites: S65A, D216A, and V274A. Phe: the phenylalanine residue of the aminoacyl-tRNA substrate; Ribo: the ribose ring of the A76 of the aminoacyl-tRNA substrate. Standard deviation accounts for the day-to-day variation in 3 independent experiments and 3 colonies were picked for each samples.

5.4.3 Enhanced nsAA incorporation into proteins by evolved aaRS and EF-Tu

Even though we observed a 6-fold increase in sfGFP(1UAG) production with the combination of evolved pAzFRS and EF-Tu compared to their parental parts, we also noticed that the co-expression pAzFRS and EF-Tu (**Figure 5.3A** sample “EAzEF”) resulted in a lowered sfGFP(1UAG) production compared to the cell expressing only pAzFRS (**Figure 5.2A** sample “EAzRS”). Apparently, co-expression multiple components can negatively influence the cell growth and protein production ability. To better evaluate the evolved translation system components, we have to find a vector to coordinate and optimize the expression of multiple components in an able host. We first sought to place all OTS components on a single plasmid to avoid unnecessary plasmid/antibiotic burden on the cell. Balancing expression levels of orthogonal translation components is essential for cellular fitness and functional activity [229]. To do this, we created a library of plasmids carrying promoters of varying strengths for each component: T7 promoter (high activity) [230], the PtacI promoter (medium activity) [224], and the lpp promoter (low activity) [126].

Promoter activity was validated by placing each promoter in front of a sfGFP gene and assessing normalized fluorescence activity (Fluorescence per OD_{600}). The relative ratios of sfGFP expression were 8:3:1 for the T7, Ptac1, and lpp promoters, respectively, which verified different promoter activity levels (**Supplemental Figure A15**).

We next built 27 individual plasmids, JGP1-27, comprising all possible promoter combinations of the pAzFRS, o-tRNA derived from the pEVOL plasmid [11] and *E. coli*'s endogenous EF-Tu. and, and screened for improved production of sfGFP(1UAG) using an *in vivo* fluorescence assay in the presence pAzF. We observed that the highest protein yield and fastest cellular growth were obtained when using the plasmid JGP26, which has the lpp promoter regulating the o-aaRS and EF-Tu, and a PtacI promoter regulating the o-tRNA (**Supplemental Figure A11 & Supplemental Figure A16**). Using this plasmid and promoter architecture, we then constructed JGP26BM containing our evolved translation system components BM3 and EFBM1. Relative to the more commonly used pEVOL and pUltra [211] system, our approach is unique because we optimized 3 parts (pAzFRS, tRNA, and EF-Tu) in one plasmid.

We next measured enhancements in amber codon suppression and protein yield resulting from our evolved translation system by comparing JGP26, JGP26BM, and pEVOL (**Supplemental Figure A11**) [4]. In our initial test, we co-transformed BL21(DE3) with a plasmid carrying the sfGFP gene with 1, 3, or 5 UAGs, or wild type equivalents and the orthogonal translation system plasmids. The GFP of each sample was measured by the intact cell fluorescence as described in Materials and Methods section. The results indicated a 2-, 5-, and 20-fold increase in incorporation of pAzF into sfGFP(1UAG), sfGFP(3UAG), sfGFP(5UAG), respectively, when using 1mM pAzF (**Figure 5.4A**). We noted that JGP26BM and JGP26 showed the same expression levels of sfGFP(1UAG) under these conditions (**Figure 5.4A**). We then adjusted the concentration of pAzF in the media to 2.5 mM (**Figure 5.4B**) and 5mM (**Figure 5.4C**). Increasing pAzF concentrations increase overall protein production levels and reduced slightly the beneficial effect of our evolved system. However, the evolved components function much more efficiently than progenitor systems and, in all cases, we observed weak expression of proteins with 5 amber codons.

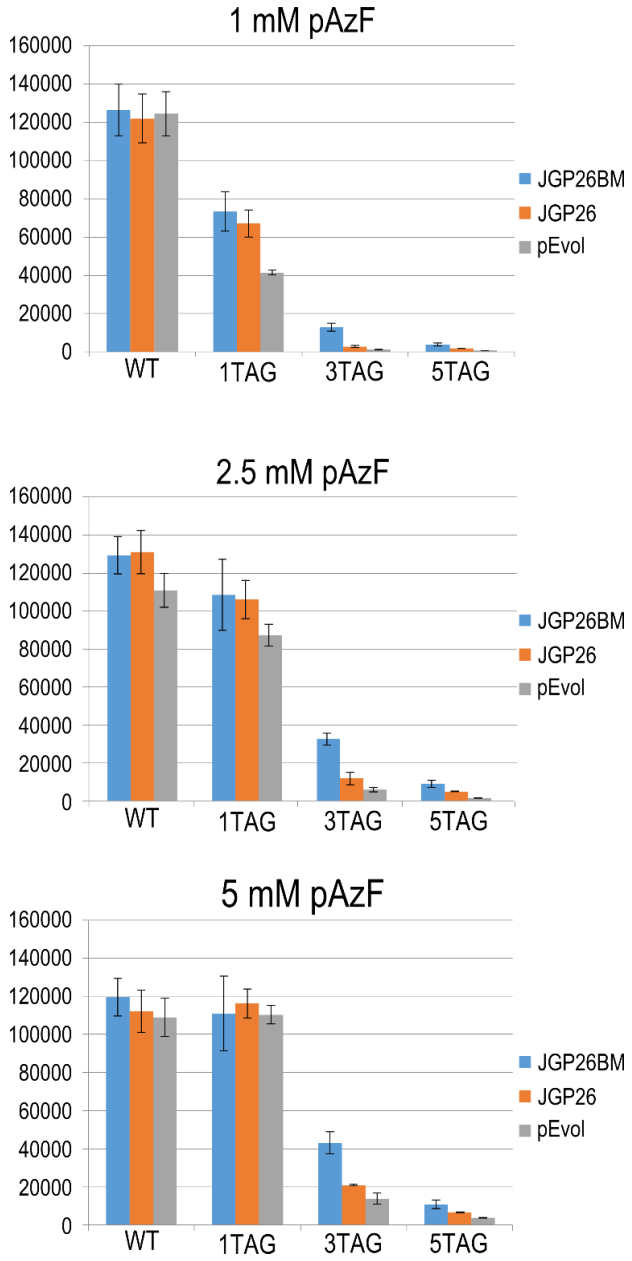


Figure 5.4: Characterization and optimization of pAzF incorporation and protein expression in BL21(DE3) using an evolved synthetase and EF-Tu. The pAzF incorporation efficiency and protein yield of the evolved synthetase and EF-Tu were demonstrated by suppression 1, 3, and 5 amber codons at 1 mM (A), 2.5 mM (B), and 5 mM (C) of pAzF respectively. Wildtype GFP expression was used as control. WT: wildtype GFP; 1TAG: GFP with 1 amber codon; 3TAG: GFP with 3 amber codons; 5TAG: GFP with 5 amber codons. JGP26: a 3-part plasmid containing the synthetase, EF-Tu, and tRNA with the optimized promoter usage: lpp promoter for synthetase and EF-Tu, and PtacI promoter for tRNA; JGP26BM: a 3-part plasmid containing the evolved synthetase (BM3) and EF-Tu (EFBM1) and tRNA with the optimized promoter usage: lpp promoter for BM3 and EFBM1, and PtacI promoter for tRNA. Evol: pEvol-pAzF plasmid reported in previous publications [4]. Standard deviation accounts for the day-to-day variation in 3 independent experiments and 3 colonies were picked for each samples.

5.4.4 Evolved and tuned OTS system enables accurate nsAA insertion at multiple UAG sites

It is known that UAG-suppression competes with termination mediated by release factor 1 (RF-1). While deletion of RF-1 is lethal in normal cells [208, 213, 214, 231, 232], by completely recoding the genome of *E. coli* to replace all 321 occurrences of the UAG codon with the synonymous UAA codon, it is permitting to delete RF-1 and complete reassign the amber codon translation function [1, 2]. Therefore, we next investigated the ability to enable efficient multi-site nsAA incorporation using the genomically recoded *C321ΔA* strain and our engineered translation system components. We co-transformed the plasmid JGP26BM harboring BM3 and EFBM1 with the sfGFP reporter gene constructs containing 1, 3, or 5 UAGs, or the wild type equivalents. We found expression levels with sfGFP constructs with a single UAG amber codon were comparable in wild-type sfGFP, and the absence of RF-1 to be increasingly beneficial for multi-site nsAA incorporation and (**Figure 5.5A**). Specifically, our analysis revealed that the evolved orthogonal translation system could suppress 3-UAG codons at production levels of approximately 80% of wild type sfGFP in the *C321.ΔA* strain. Interestingly, the protein yield of sfGFP with 3 UAGs is slightly higher than that of sfGFP with single UAG. We postulate that the synthesis of nascent peptide of 3-TAG GFP may be slower than single-TAG GFP because the slow kinetics of pAzF incorporation than those of the canonical amino acids. The slow synthesis speed may slightly help to increase the folding efficiency of GFP molecule. Additionally, we observed significantly enhanced production of sfGFP(5UAG) compared the yield we observed in BL21(DE3) cell. The solubility of wildtype sfGFP and its derivatives with 1 pAzF, 3 pAzFs, and 5 pAzFs have the extremely high and similar solubility. Apparently, up to 5 pAzF incorporation does not influence the solubility of the sfGFP molecules. Unfortunately, we could not observe sfGFP (with 1, 3, and 5 amber codons) expression improvement by use of the evolved pAzFRS and EF-Tu compared to the parental (data not shown).

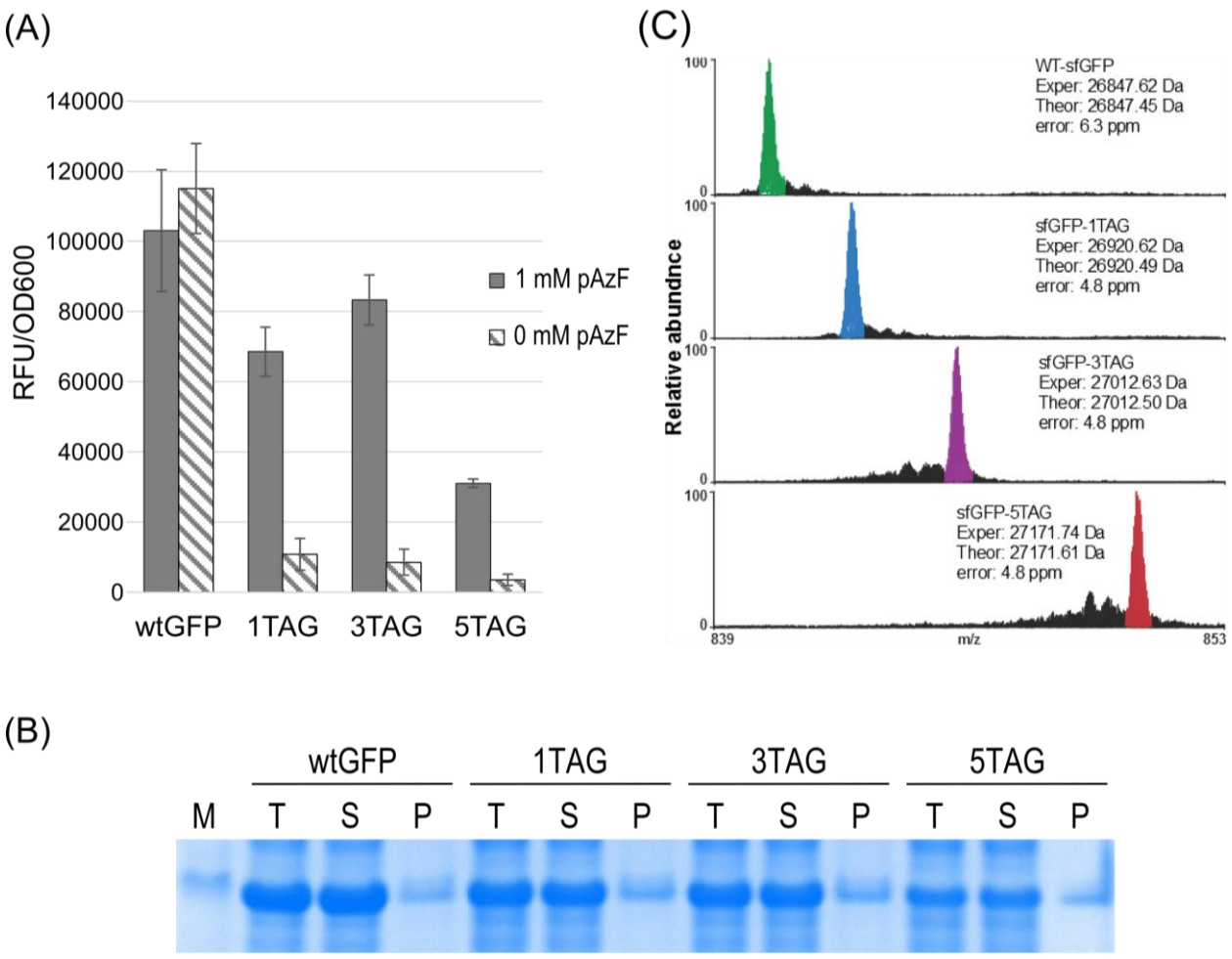


Figure 5.5: Fidelity and efficiency of pAzF incorporation at multiple amber codon at the absence of release factor 1. A) The active protein yield by suppressing 1, 3, and 5 UAGs in reading frame of GFP by the evolved synthetase and EF-Tu in the plasmid JGP26BM. The pAzF was supplied at 1 mM or 0mM. The host is a release factor 1-deficient *E. coli* strain (*C321.ΔA*) with all amber codons in genome replaced with ochre codon [1, 2]. **B)** The solubility of wildtype GFP, GFP with 1 pAzF, 3 pAzFs and 5 pAzFs are high and consistent. T: total protein; S: soluble fraction; P: precipitated fraction. wtGFP: wildtype GFP; 1TAG: GFP with one pAzF; 3TAG: GFP with 3 pAzFs; 5TAG: GFP with 5 pAzFs; M: Molecular weight standard. **C)** Spectrum of wildtype and various modified sfGFP samples, obtained by top-down mass spectrometry and illustrating site-specific incorporation of pAzF at single and multiple sites. Major peaks (color) in each spectrum agree with the theoretical peaks respectively. “Exper” indicates experimentally obtained protein mass, and “Theor” indicates theoretically calculated protein mass. Standard deviation accounts for the day-to-day variation in 3 independent experiments and 2 colonies for each samples.

We next examined the fidelity of multi-site pAzF incorporation into the sfGFP reporter constructs. Specifically, we carried out top-down mass spectrometry (i.e., MS analysis of whole intact proteins) to detect and provide semi-quantitative information for the incorporation of pAzF into sfGFP. **Figure 5.5B**

shows the 32+ charge state of sfGFP and clearly illustrates mass shifts corresponding to the incorporation of one, three, and five pAzF residues. Site-specific incorporation of pAzF, as detected by MS, was greater than 95 % in all samples (**Figure 5.5B**), with less than 3 ppm difference between experimental and theoretical protein masses. In other words, we achieved efficient, high yielding, and pure site-specific pAzF incorporation into sfGFP using our evolved translation system and the *C321ΔA* strain.

5.5 Discussion

Our work joins an ever-growing collection of reports highlighting the ability to repurpose the translation machinery for genetic code expansion. In this study, we demonstrated the importance of pursuing a systems engineering approach when evolving multiple components of orthogonal translation systems and using RF-1 deficient strains to enable high-level and accurate multi-site incorporation of nsAAs into a protein. Our work highlights the importance of including EF-Tu as an engineering target for well-characterized nsAAs that are not bulky and charged, such as pAzF. We thus expect our systems engineering approach of evolving multiple translation system components to facilitate integration of a wide range of nsAA targets.

In terms of the pAzFRS, we used a step-wise library selection and screening procedure to improve aaRS activity by ~8-fold compared to the parental enzyme. The optimization of residues H283-D286 are mainly responsible for the enhancement. In terms of EF-Tu, we used an alanine scanning approach to screen a 1024-member library and observed several mutants having up to 4-fold improvement, each with mutations in different residues. By combining our best mutants and coordinately tuning expression levels, we observed a 2-, 5-, and 20-fold increase in incorporation of pAzF into sfGFP(1UAG), sfGFP(3UAG), sfGFP(5UAG), respectively, as compared to the progenitor system. Notably, the performance of the evolved mutants varied significantly in different context. For instance, when suppressing only one amber codon with a high concentration of pAzF (5mM) in BL21(DE3) strain, the evolved mutants do not showed advantage over their parental ones, while when attempt to suppress 5 amber codons with only 1 mM of pAzF, the evolved mutants showed 20-fold higher efficiency over the parental parts even though the GFP yield is comparatively low. In the strain *C321.ΔA*, due to the lack of

release factor 1, the evolved pAzFRS and EF-Tu can produce abundant GFP by suppressing 5 amber codons, however, we cannot observe the advantage of the evolved mutants over the parental parts again. Apparently, a number of factors are able to influence the performance of the evolved mutants including nsAA concentration, amber codon number, strain, release factor one and *etc*. These factors should be carefully taken into consideration when characterizing beneficial mutants, designing evolution strategies, and measuring the ratio of performance/cost in protein production.

We then used top-down mass spectrometry to confirm multiple instances of nsAA incorporation with near perfect fidelity. Taken together, the engineered translation system reported here demonstrates improved performance for protein yield and accurate protein production as compared with the progenitor enzymes.

In sum, our engineering effort demonstrates new details regarding interactions between proteins involved in translation that might be taken into consideration in future designs for constructing similar engineered translation systems. We believe that only by treating the translation apparatus as a complex system, whereby all biological parts involved in protein biosynthesis are coordinately optimized (*e.g.,* codons, tRNA, aaRS, EF-Tu, and the ribosome together), will we enable more diverse genetic codes and advanced capabilities. Such advances will be important for harnessing a dramatically expanded genetic code for manufacturing novel therapeutics [184, 233, 234] synthesizing genetically-encoded materials [181, 235, 236] advancing medicine [184, 233, 234, 237-241] and elucidating fundamental biological insights (*e.g.,* the histone code [146, 181, 242]).

5.6 Acknowledgments

This work was supported by the Defense Advanced Research Projects Agency (N66001-12-C-4211), the David and Lucile Packard Foundation, the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust, and the Arnold and Mabel Beckman Foundation. This research was carried out in collaboration with the National Resource for translational and Developmental Proteomics under Grant P41 GM108569 from the National Institute of General Medical Sciences. We thank Prof. Peter G. Schultz for providing the pEVOL plasmid.

5.7 Author Contributions

As the second author, I designed and executed the experiments involving the EF-Tu library construction and promoter tuning. I assisted in writing the manuscript.

CHAPTER 6

6 Cell-free synthetic biology: Engineering beyond the cell

6.1 Abstract

Cell-free protein synthesis (CFPS) technologies have enabled inexpensive and rapid recombinant protein expression. Numerous highly active CFPS platforms are now available and have recently been used for synthetic biology applications. In this review, we focus on the ability of CFPS to expand our understanding of biological systems and its applications in the synthetic biology field. First, we outline a variety of CFPS platforms that provide complementary tools for expressing proteins from different organisms. Next, we review the types of proteins, protein complexes and protein modifications that have been achieved using CFPS systems. Finally, we introduce recent work on genetic networks in cell-free systems and its utility for rapidly prototyping *in vivo* networks. Given the flexibility of cell-free systems, CFPS holds promise to be a powerful tool for synthetic biology as well as a protein production technology in years to come.

6.2 Introduction

Cell-free protein synthesis (CFPS) technology was first used over 50 years ago by Nirenberg and Matthaei to decipher the genetic code [243]. In the late 1960s and early 1970s, CFPS was employed to help elucidate the regulatory mechanisms of the *Escherichia coli* lactose [244] and tryptophan [245] operons. Now, in the last two decades, cell-free protein expression platforms have experienced a surge in development to meet the increasing demand for inexpensive and rapid recombinant protein expression technologies, which has resulted in the development of numerous highly active CFPS platforms [24].

This renewed interest in CFPS technology was motivated by the advantages offered by this methodology for the production of recombinant proteins. In particular, the open reaction environment allows for addition or removal of substrates for protein synthesis, as well as precise, on-line reaction monitoring. Additionally, the CFPS reaction environment can be wholly directed toward and optimized for the production of the protein product of interest. In this way, CFPS platforms separate catalyst synthesis

(cell growth) from catalyst utilization (protein synthesis), representing a significant departure from cell-based processes that rely on microscopic cellular “reactors.” CFPS effectively decouples the cell’s objectives (growth & reproduction) from the engineer’s objectives (protein overexpression & simple product purification). Overall, the nature of CFPS technology allows for shortened protein synthesis timelines and increased flexibility for addition or removal of natural or synthetic parts compared to *in vivo* approaches. The versatility of CFPS makes it especially attractive for fundamental discovery and high throughput screening applications.

The ability to prioritize the engineer’s objectives in CFPS has further motivated recent applications of CFPS technology to the exciting and ever-growing field of synthetic biology. For instance, cell-free synthetic biology approaches have enabled development of an *in vitro* prototyping environment for characterization of synthetic parts or genetic networks [246-248]. The open environment and reduced complexity of cell-free systems has also made it possible to develop quantitative models describing cell-free genetic network performance and perform machine learning optimization of cell-free protein synthesis [246, 249]. Additionally, the absence of cell viability constraints has made CFPS an attractive technology for expanding the possible applications of synthetic biology. Recent advances in cell-free synthetic biology include the incorporation of non-natural chemistries into biological polymers [38, 89, 90, 250, 251], *in vitro* assembly of complex biological machines and devices [3], and the development of minimal cells [252-254]. Excitingly, recent work has enabled translation of cell-free technology beyond the lab bench: both to the industrial scale for therapeutic production [255, 256], and to a low-cost, user-friendly format for diagnostic applications [257].

In this review, we focus on the application of CFPS technology to synthetic biology. More detailed reviews on the development of CFPS technology and the types of protein produced in cell-free systems have been published recently [250, 258-261]. Here, we begin by introducing the various CFPS platforms and discuss their technological capabilities. We then outline the types of proteins, protein complexes, and protein modifications that have been achieved using CFPS technologies. Finally, we discuss cutting-edge cell-free synthetic biology applications.

6.3 Multiple cell-free protein synthesis technologies enable production of diverse proteins

The recent technological renaissance has resulted in a variety of highly active CFPS platforms that provide complementary tools for expression of proteins from diverse organisms. Though only *E. coli* and wheat germ extracts have been used in a high-throughput format, all CFPS platforms have the potential to be used for high-throughput screening of DNA libraries and gene products from diverse organisms for biological discovery and synthetic biology applications.

CFPS systems carry out protein synthesis by harnessing the biological catalysts for translation, protein folding, and energy generation from prokaryotic or eukaryotic cells. When combined with a DNA template, amino acids, an RNA polymerase, an ATP regeneration mechanism, salts, and other buffers or environmental stabilizers *(e.g.,* HEPES, spermidine, putrescine), these complex biological catalytic

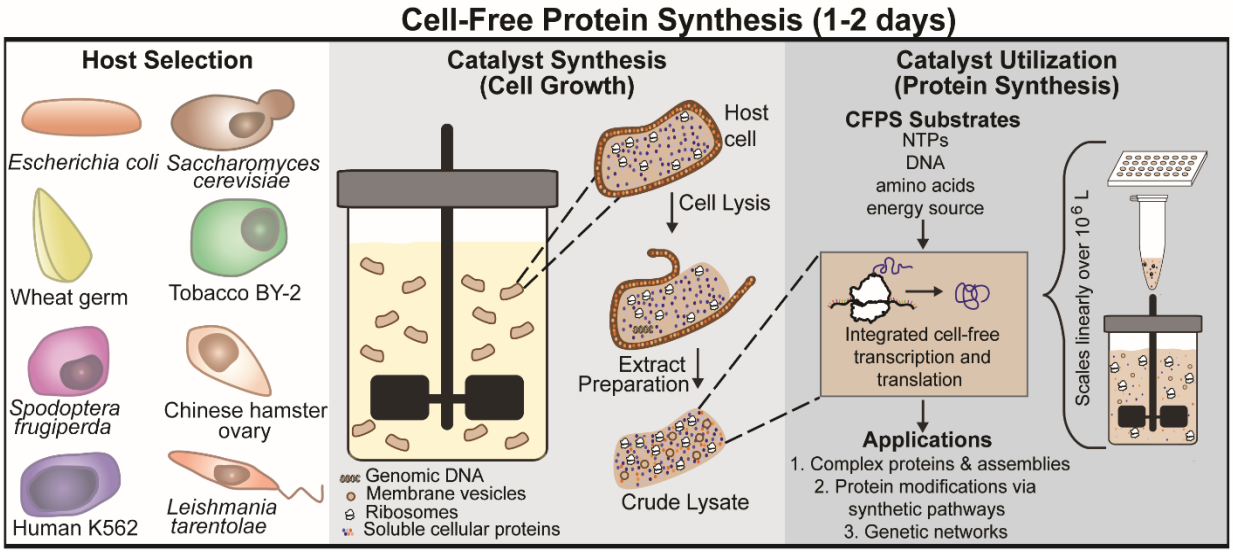


Figure 6.1: CFPS enables rapid production of recombinant proteins. CFPS platforms allow for increased flexibility and shorten process timelines to create a variety of high value recombinant proteins. This technology provides platforms from numerous organisms with varied complexity, to meet the need of the specific target proteins. CFPS also decouples catalyst synthesis and catalyst utilization, traditionally interconnected during *in vivo* recombinant protein production schemes. For CFPS, catalyst synthesis involves cell growth, cell lysis and extract processing which removes genomic DNA to create crude extract. The crude extract may be frozen for future use or used directly for catalyst utilization. This process requires NTPs, DNA, amino acids and an energy source is to produce various proteins with applications in the synthetic biology field.

ensembles carry out sustained protein synthesis *in vitro* [222] (**Figure 6.1**).

There are two categories of CFPS platforms, which represent bottom-up and top-down approaches to *in vitro* protein synthesis. Protein synthesis using purified recombinant elements (PURE) systems build the protein synthesis ensemble from the bottom-up. The PURE approach involves purifying the molecular components required for protein synthesis and subsequently adding them to CFPS reactions [262]. Crude extract systems represent the alternative, top-down approach. Crude cell extract is generated by clarifying whole cell lysate via centrifugation to remove genomic DNA, insoluble biological elements, and unlysed cells. In some platforms, additional measures are taken to degrade endogenous mRNAs [263]. Importantly, the crude extract contains all the biological components required for translation, protein folding, and energy regeneration (*e.g.,* ribosomes, tRNAs, chaperones, metabolic enzymes, elongation factor-Tu (EF-Tu), translation initiation factors, etc.). Crude lysate CFPS platforms have been developed using cells from a number of organisms. Each crude extract CFPS platform has advantages and disadvantages depending on the desired application.

6.3.1 PURE systems

PURE technology was pioneered by Shimizu and colleagues [264]. The PURE approach reconstitutes the transcriptional, translational and energy generation machinery from *E. coli* by purifying recombinantly-expressed histidine (His)-tagged components and adding them to the CFPS reaction mixture. *E. coli* ribosomes are purified using sucrose-density-gradient centrifugation [262].

PURE CFPS systems are commercially available and widely used for both fundamental discovery and synthetic biology applications. Specialized PURE systems have been developed to produce disulfide bond and non-standard amino acid (nsAA) containing proteins [265, 266]. PURE technology has proven particularly useful for the isolated study of recombinant proteins and protein complexes [267-269] due to the simplicity of the system. Furthermore, PURE technology has played a pivotal role in the development of minimal cells [252, 254, 270-272]. However, the expense and cross-reactivity of the His-tag-based component purification process limits its utility in some cases.

6.3.2 Microbial crude extract systems

E. coli and *Saccharomyces cerevisiae* are attractive hosts for CFPS and cell-free synthetic biology because (1) they are easily fermentable and (2) they are model organisms. As model organisms, there exists a wealth of experimental tools and genetic information for both *E. coli* and *S. cerevisiae* available to aid synthetic biology efforts.

The prokaryotic *E. coli* crude extract (ECE) system is one of the most widely adopted platforms for CFPS. This technology is commercially available and is used at the industrial scale [255, 256, 273]. ECE has been widely adopted for two main reasons, (1) its high batch yields, with up to 2.3 g/L of green fluorescent protein (GFP) reported [274], and (2) the fast, scalable, and cost-effective extract preparation process [275]. In addition, simple and rapid methods for extract preparation have been developed [25, 276]. Furthermore, ECE reactions can use nucleoside monophosphates and inexpensive energy substrates, such as glucose or starch to regenerate ATP [261, 277, 278]. ECE has been used to synthesize recombinant human proteins containing post-translational modifications (PTMs) [10] and non-standard amino acids [38, 89, 90, 209, 251, 279, 280]. Furthermore, ECE has been adapted to a high-throughput format for production of antibody variants for drug development and screening studies [281], demonstrating its utility for screening or functional genomic studies. Finally, ECE has been shown to scale linearly over a 10^6 L range in reaction volumes [256], which helped motivated the adoption of this technology for large-scale protein synthesis [255, 273, 282]. While bacterial CFPS are able to produce proteins that would be difficult or impossible in cells, further reductions in cost would enable wider adoption for industrial biomanufacturing.

A second microbial CFPS system uses extract from *S. cerevisiae* (baker's yeast) to catalyze protein synthesis. *S. cerevisiae* cells contain eukaryotic folding machinery, which is beneficial for recombinant human protein production. The first *S. cerevisiae* extract (SCE) system, was pioneered by Iizuka and colleagues over 20 years ago [283]. Recently, the Jewett lab developed and optimized a scalable, low-cost SCE preparation method [284, 285]. Many efforts have been made to improve the efficiency and reduce the cost of SCE CFPS [286-288]. Together, these advances have resulted in a

relatively low-cost eukaryotic CFPS technology with protein yields comparable to other eukaryotic platforms (**Table 1**). However, further improvements in protein synthesis yields will be necessary to encourage adoption of yeast CFPS for synthetic biology applications.

6.3.3 Plant, animal, & insect crude extract systems

Crude extract CFPS technologies derived from higher organisms offer advantages for complex protein production. Importantly, these extracts enable complex co-translational folding mechanisms and the addition of PTMs not currently possible using microbial platforms. However, batch yields from these systems are typically at least an order of magnitude lower, while the system cost is much higher, compared to ECE. Additionally, applications of these technologies to synthetic biology are currently limited. One challenge preventing high yielding eukaryotic CFPS is the difficulty of enabling endogenous translation initiation mechanisms (*e.g.,* 5' capping) *in vitro*.

Two main plant-based crude extract CFPS systems have been developed to date: wheat germ extract (WGE) and tobacco BY-2 cell extract (BYE). The WGE platform, which is derived from wheat seed embryos [289], has achieved both high yields [290] and high-throughput expression [291] of recombinant proteins. Typically, this system is used with continuous substrate replenishment (continuous exchange). For example, Goshima and colleagues used WGE to synthesize approximately 13,000 human proteins in a single study [292]. WGE was also used to synthesize bioactive proteins of diverse species origin, including *Arabidopsis* [293] and the malaria parasite genus *Plasmodium* [294] without codon optimization, in most cases. WGE is the highest-yielding eukaryotic CFPS platform, with yields of up to 1.6 mg/mL GFP in batch [259] and 9.7 mg/mL GFP with continuous exchange reported [290]. Notably, the active lifetime of WGE is remarkably long, up to 60 hours in continuous exchange reactions that replenish substrates [290], because most nucleases and proteases are removed during extract preparation. However, the primary drawback of WGE is the extensive extract preparation process, which takes 4-5 days [289].

The BYE CFPS system was recently developed as an alternative plant-based CFPS platform. Preparation of BYE is greatly simplified compared to WGE and can be completed in just 4-5 hours. Like WGE, BYE has also been used to synthesize proteins from diverse organisms, including a

transmembrane growth factor, a glycosylated *Apergillus niger* enzyme, and a bioactive human antibody [295]. However, unlike WGE, mg/mL yields from BYE have not yet been achieved.

The insect cell extract (ICE) CFPS platform was developed using extract from *Spodoptera frugiperda* (fall army worm) cells and is commercially available. The first ICE system was developed by Ezure and colleagues [296]. A second ICE system, developed by Katzen and Kudlicki, produced extracts containing endoplasmic reticulum (ER) vesicles that retained glycosylation and signal sequence processing activity [297]. ICE has since been used to produce glycoproteins and membrane proteins [297-300].

More recently, CFPS platforms were developed from mammalian cell lines. Mammalian cells are currently the preferred platform for large-scale *in vivo* production of human proteins bearing PTMs. Mikami and colleagues developed a CFPS platform using lysate from HeLa cells, a cervical cancer cell line [301]. In addition, the Kubick lab has described CFPS technologies using extracts from Chinese hamster ovary (CHO) cells and human K562 cells, a myelogenous leukemia cell line [302]. Mammalian crude extract CPFS platforms provide a potentially attractive alternative for the synthesis and high-throughput screening of recombinant human proteins.

6.3.4 Parasitic organism crude extract systems

The *Leishmania tarentolae* extract (LTE) platform, developed by Mureev and colleagues, is the only platform that uses extracts from a parasitic organism. Parasites are attractive candidates for proteomic studies because, in some organisms, >50% of proteins have no identifiable homology or predictable function. However, *in vivo* expression of genes from parasitic or infectious organisms can be challenging due to differences in codon usage and protein synthesis machinery [303]. Parasitic organism crude extract systems are potentially advantageous for generating expressed parasite proteomes. In one instance, a parasitic crude extract CFPS system outperformed ECE in bioactive yields of proteins from *P. falciparium*, a malaria parasite [304].

L. tarantolae, a single-celled lizard parasite, was chosen for CFPS technology development for two main reasons: (1) it is easily fermentable and (2) all endogenous mRNAs share the same leader sequence. The latter feature enables complete repression of background translation via inclusion of an anti-splice-leader oligonucleotide [303]. LTE features a rapid and scalable extract preparation process [305, 306], and is one of the highest yielding eukaryotic systems in batch; yields of up to 300 μg/mL GFP have been reported [303]. Importantly, Mureev and colleagues also developed species-independent translation sequences (SITS) to bypass translational regulation in LTE. The inclusion of SITS allowed cell-free synthesis of GFP in prokaryotic and eukaryotic extracts, including LTE, ECE, WGE, SCE, and ICE, using the same DNA template [303, 305, 306]. LTE has been used to synthesize mammalian, *L. tarentolae*, and *Plasmodium falciparum* proteins.

Despite recent technological advances, there remain challenges facing CFPS technology that need to be addressed. Yields of bioactive human proteins from CFPS remain lower than those achievable *in vivo*. In particular, no eukaryotic CFPS platform is currently capable of producing mg/mL quantities of protein in batch. Increasing eukaryotic CFPS batch yields to 0.5 mg/mL or greater is an important challenge for two reasons. First, these platforms have the potential to achieve higher soluble, bioactive yields of recombinant human proteins than bacterial platforms. Second, eukaryotic CFPS systems could be useful as prototyping tools for genetic circuit characterization or production of recombinant proteins in eukaryotic cells. Additionally, most crude extract CFPS platforms are derived from model organisms, and only the *L. tarantolae* platform is derived from a parasitic organism. Crude extract platforms from non-model organisms could have utility for functional genomics studies. Finally, development and optimization of orthogonal translation initiation strategies, such as SITS, could further increase protein yields and enable expression of proteins from diverse organisms. Such improvements will overcome the need for codon optimization or other template-specific DNA modifications. Addressing these challenges will be important to encourage broad adoption of CFPS technology for fundamental discovery and synthetic biology applications.

6.4 Complex proteins and protein assemblies

6.4.1 Cell-free production of complex proteins

Most cell-free protein synthesis platforms have been applied to the production of complex proteins. CFPS systems offer distinct advantages over *in vivo* protein production for applications that require more precise control over the protein synthesis reaction conditions. Such applications include bispecific antibodies, antibody-drug conjugates (ADCs), vaccines, and membrane proteins.

6.4.1.1 Therapeutics & vaccines

ECE has been used extensively for production of bioactive recombinant therapeutic proteins, including granulocyte macrophage colony-stimulating factor (GM-CSF), antibodies, bispecific antibodies, and antibody-drug conjugates [255, 256, 273, 281, 282, 307]. Further, ECE has been used to produce non-standard amino acid-containing proteins [38, 89, 90, 209, 251, 280], and glycosylated proteins [10]. This was accomplished by modulating the reaction conditions or by adding synthetic parts or enzymatic pathways to the CFPS reaction. Recently, a method for *in vitro* display of antibody fragments was developed in ECE, providing an exciting new technology for high-throughput production and screening of antibody candidates [281].

The eukaryotic WGE, ICE, CHO extract, and K562 extract systems have also been used to produce therapeutic proteins. WGE and ICE have been used to synthesize bioactive antibody fragments, and ICE has been used to produce bioactive tissue-type plasminogen activator [300, 308, 309]. Furthermore, ICE, CHO extract, and K562 extract have been used to synthesize glycosylated erythropoietin [302, 310].

An especially promising application for CFPS systems is the high throughput screening and *in vitro* production of vaccines and virus-like particle (VLP) antigens. The WGE platform was applied to the synthesis of 124 *P. falciparum* genes, and 75% were successfully synthesized without codon optimization [294]. In addition, ECE has been used to synthesize vaccines and VLPs, including a B-cell lymphoma vaccine [311], anti-influenza VLPs [312] and anti-Hepatitis B VLPs [313]. Notably, the VLPs were produced at significantly higher yields than those possible *in vivo*, with yields improved up to 15-fold

[313]. These studies demonstrate the potential of CFPS systems as novel vaccine discovery tools and platforms for large-scale vaccine production. Looking forward, because CFPS systems are abiotic, they are potentially well-suited for portable and on-demand production of vaccines in resource-poor areas. This is an exciting future application area for CFPS technology.

6.4.1.2 Membrane proteins

CFPS systems offer advantages for the production of membrane proteins. Newly synthesized membrane proteins can be stabilized *in vitro* by including membrane mimics (*i.e.*, surfactants, liposomes, nanodiscs) to the cell-free reaction either post-translationally, to solubilize proteins, or co-translationally, to prevent aggregation [299, 314]. The ECE, WGE, ICE, CHO extract, and K562 extract systems have all been used to synthesize membrane proteins including ATP synthase [3], G-protein coupled receptors [315, 316], and epidermal growth factor receptor [310]. The synthesis of membrane proteins is an excellent application for CFPS technology, and has been useful for the structural biology community [317]. Overall, enhanced control over the protein synthesis reaction *in vitro* enables significantly higher yields of soluble, active protein compared to *in vivo* expression.

Taken together, these studies demonstrate the ability of CFPS technology to produce a variety of recombinant proteins with diverse complexity. ECE and WGE in particular demonstrate the potential of CFPS technologies for high-throughput production of many different kinds of proteins of interest for therapeutic development, fundamental discovery, or use as synthetic biology parts.

6.4.2 Macromolecule production and assembly in cell-free systems

Much progress has been made towards *in vivo* synthesis of protein assemblies [318, 319]; however, cellular viability constrains the complexity of the assemblies that can be overproduced. Moreover, cellular complexity may obscure fundamental properties of intricate assembly processes. Recently, CFPS systems have been extended to the production of complex protein assemblies. This advance offers interesting opportunities for studying macromolecule self-assembly and developing synthetic biological devices. CFPS platforms circumvent *in vivo* limitations in three key ways. First, the

cell-free environment accelerates protein engineering efforts by enabling high-throughput experimentation, and simplifying the purification of individual subunits. Second, CFPS platforms are flexible, ranging from crude extract to fully purified transcription-translation systems, which makes it possible to study assembly processes in a controlled reaction environment. Third, CFPS platforms provide an additional level of complexity by coupling encoded protein expression to assembly of protein ensembles, which may elucidate the role of genetic regulation on subunit stoichiometries and assembly mechanisms (**Figure 6.2A**).

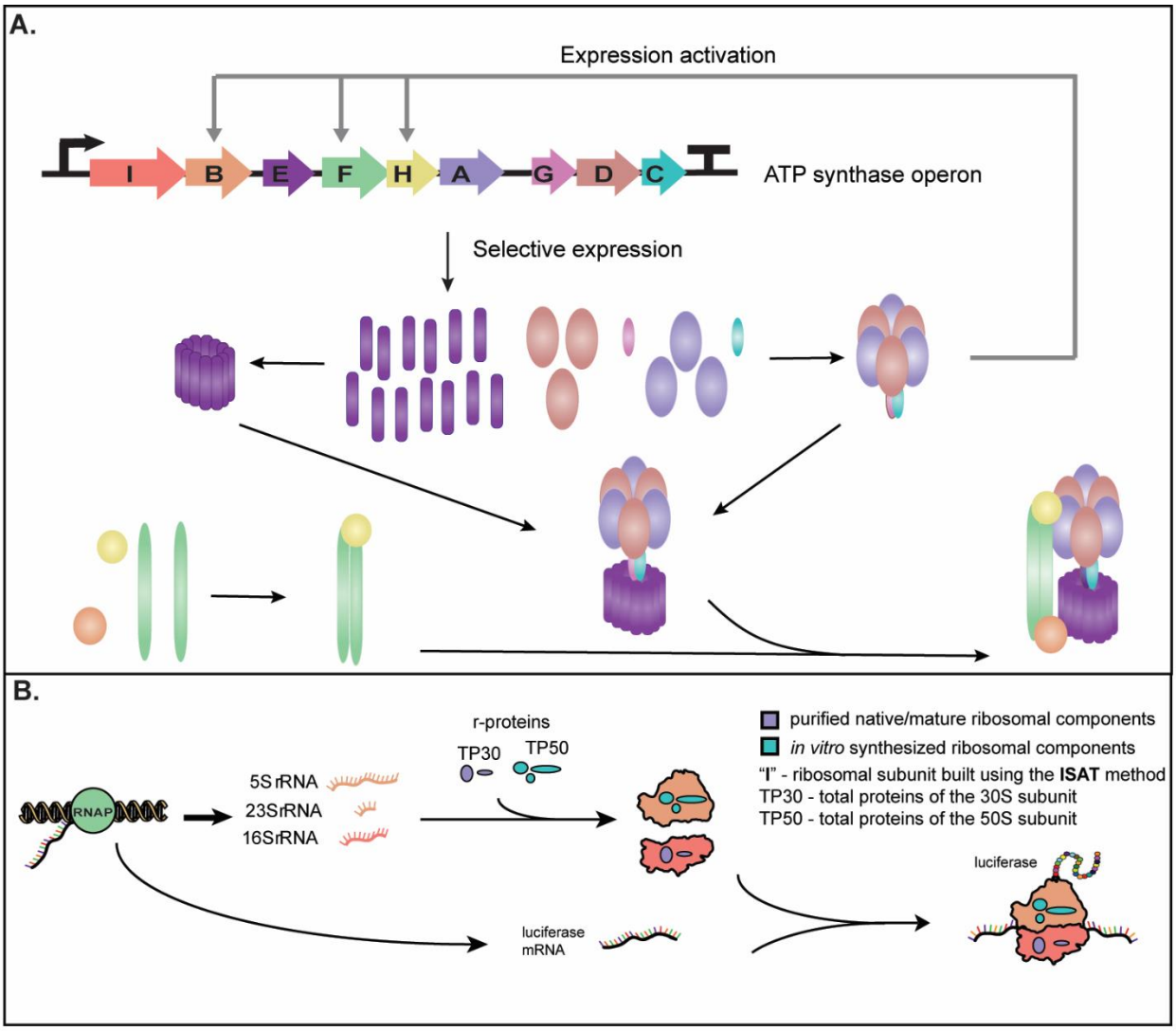


Figure 6.2: Assembly of macromolecules in CFPS reactions. (A) Complete *in vitro* assembly of ATP synthase with hypothesized genetic regulation. The ATP operon is added to a crude *E. coli* extract and transcribed into a single mRNA. Proteins are expressed at various levels determined by operon regulation. Matthies and colleagues suggest assembly of correctly assembled ATP synthase complexes is dependent on specific expression levels on the subunits, correlated to the subunit stoichiometry in the complex [3]. It is hypothesized intermediate assemblies may also activate the expression of other subunits in the operon, allowing for sequential assembly processes [5]. Thus, combined expression and assembly in CFPS systems allows for an additional level of complexity of assembly processes for analysis. **(B) *In vitro* integrated synthesis, assembly, and translation (iSAT) method of constructing synthetic ribosomes enables synthesis of active firefly luciferase in a one-pot reaction**. iSAT enables one-step co-activation of rRNA transcription, assembly of ribosomal subunits and synthesis of active protein by these ribosomes in same compartment. This process begins with T7 RNAP polymerase transcribing rRNA and luciferase mRNA. Ribosomal subunits are reconstituted from mature rRNA and ribosomal components previously purified or synthesized *in vitro*. Newly assembled ribosomes translate mRNA encoding the reporter protein luciferase to assess its activity.

6.4.2.1. Assembly of naturally occurring macromolecules

Early demonstrations of protein assembly in CFPS systems produced macromolecules such as a five sub-unit *E. coli* RNA polymerase [320] and a two sub-unit Hepatitis B core antigen virus-like particle [313]. Even these simple assemblies have many applications. For instance, VLPs have been produced at higher yields than *in vivo*, which can be further applied to the discovery of novel vaccines, and provides straightforward production scale-up processes [313].

As protein production capacity and reaction lifetime increased in CFPS systems, more proteins could be synthesized at once. This allowed for the construction of complex structures, such as the pioneering work of Matthies and colleagues that carried out the synthesis and assembly of *C. thermarum* ATP synthase, a 542-kDa membrane protein consisting of 25 individual proteins arranged into 8 sub-units [3]. Total membrane protein synthesis was possible with detergents in crude *E. coli* extract (**Figure 6.2A**). This demonstration opens the way to the construction of novel *in vitro* energy generating bio-devices that can be rapidly prototyped *in vitro.* Multi-step coordination of protein ensembles has been successfully shown *in vitro*. For instance, Fujiwara and colleagues simulated chromosomal DNA replication in a cell-free system by expressing 13 essential genes [321]. In another study, Shin and colleagues showed complete *in vitro* DNA replication, synthesis and assembly of bacteriophage T7 particles from expression of 60 genes [322]. This is remarkable considering only 35 of the approximately 60 proteins encoded by the T7 genome have known function [323]. Overall, the experiments reaffirm the value of CFPS platforms for interrogating native self-assembly processes and synthesizing self-replicating biological machines in *vitro*.

Beyond protein machines, the ability to assemble macromolecules with RNA and protein has also significantly advanced. For instance, the *E. coli* ribosome, a large protein/RNA assembly, has been integrated *in vitro* using a prokaryotic CFPS system [324] (**Figure 6.2B**). By adding DNA encoding the ribosomal RNA (rRNA) and purified ribosomal proteins to an *E. coli* extract lacking native ribosomes, *in vitro* assembled ribosomes are capable of translation under physiological conditions. Tuning transcription [325] and utilizing a continuous-exchange platform to replenish substrates and remove toxic byproducts

[326] has resulted in reporter protein sfGFP yields of up to 7.5 µmol/L. This new ribosome construction platform, termed iSAT for integrated synthesis, assembly, and translation of ribosomes, makes possible new ways to probe, dissect, and understand ribosome biogenesis. In addition, it contributes meaningfully towards efforts to build minimal cells and construct synthetic ribosomes with novel and useful properties.

6.4.2 2. Assembly and evolution of biological devices (bio-devices)

In order to expand CFPS capabilities of engineering protein assemblies to create bio-devices, cell-free systems can be combined with detection hardware. This is crucial because, in order to follow assembly processes at the nanometer length scale, high molecular structure resolution is needed. Interestingly, Heyman and colleagues developed a device that surface-patterned anti-HA antibodies onto a TEM grid that can trap CFPS synthesized T4 gp18 proteins tagged with HA [327]. This demonstrated that coupling synthesis to assembly in the presence of an anchoring site could enable on-board visualization of protein assembly. Microfluidic affinity assays have also been paired with CFPS platforms, enabling 14,792 on-chip experiments to exhaustively measure protein-protein interactions of 43 *Streptococcus pneumonia* proteins in quadruplicate [328]. This study found several physical interactions between proteins that had been previously unreported. In sum, cell-free systems provide a flexible environment for expanding understanding of native assembly processes and the repertoire of man-made protein assemblies.

6.5 Synthetic pathways in CFPS systems enable site-specific modifications of proteins for novel functionalities

The development of highly active CFPS platforms has enabled the adaptation of CFPS technology towards synthesizing proteins with synthetic modifications and novel functionalities. Synthetic pathways have been added to confer the ability of co-translational and post-translational modification of recombinant proteins. The ability to site-specifically modify proteins could have manifold uses in fundamental and applied science. For example, the co-translational incorporation of acetyl-lysine (AcK) *in vivo* has produced recombinant histone proteins that revealed new understanding of epigenetic regulation

[146]. Site-specific modification will also have utility in the development of therapeutics bearing (a)biological modifications (*i.e.*, phosphorylation, glycosylation, PEGylation, drug conjugates).

Cell-free production of modified proteins complements *in vivo* efforts and offers many advantages. In particular, modulation of substrate concentrations permits direct monitoring of their effects on protein modifications. This is especially important because synthetic protein modification pathways involve multiple proteins that can exhibit cross-talk with the native system. Additionally, an open and abiotic environment allows for the addition of substrates and production of byproducts that cannot be used *in vivo* due to cell membrane permeability or cellular toxicity limitations. CFPS platforms offer greater control over the modification process and expand the types of chemistries that can be added to proteins.

6.5.1 Co-translational incorporation of non-standard amino acids

The co-translational incorporation of non-standard amino acids (nsAAs) into proteins expands the chemistry of life. Such an expansion has enabled the incorporation of (a)biological groups into proteins such as biophysical probes [144] (spin-labeled, fluorescent molecules and photoactivatable cross linkers), redox active groups, or natural post-translational modifications (PTMs). To date, over 100 nsAAs have been co-translationally incorporated into proteins [84]. This has been possible with the introduction of an orthogonal translation system (OTS), which is capable of genetically encoding nsAAs. Traditionally, OTSs consist of an engineered, orthogonal tRNA/aminoacyl tRNA synthetase (o-tRNA/o-aaRS) pair derived from a phylogenetically distant organism, often *Methanocaldococcus jannaschii*. The o-tRNA has a modified anticodon specific to the UAG stop codon (**Figure 6.3**) while the o-aaRS is evolved to bind and aminoacylate the nsAA of interest to the o-tRNA. This process, which was initially pioneered by Schultz and colleagues [190], is called amber suppression as it allows the nsAA to be encoded at the UAG (amber) stop codon. An additional, engineered EF-Tu may be required when the nsAA of interest is too bulky to fit into the native EF-Tu amino acid binding pocket or its charge makes binding inefficient [41]. While it is worth noting that both the WGE[88] and the ICE [329] systems have been used to incorporate

nsAAs, for the purposes of this review we will focus on the extensive work that has been done in the ECE CFPS platform.

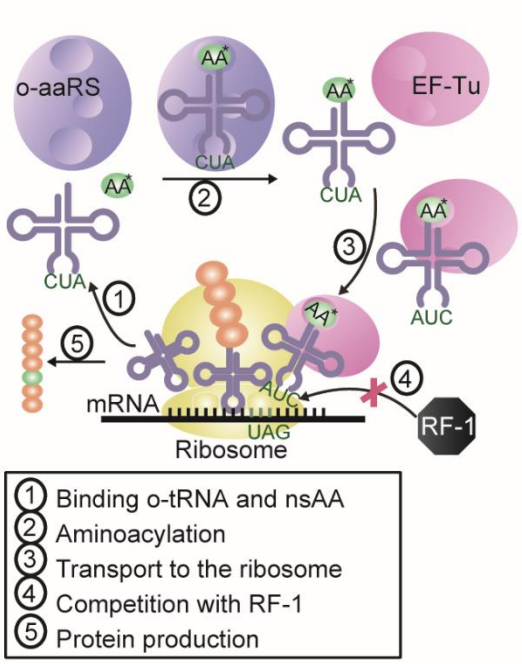


Figure 6.3: Schematic representation of co-translational incorporation of a non-standard amino acid using an orthogonal translation system and amber suppression. The o-aaRS first binds its cognate nsAA and cognate o-tRNA. The o-aaRS then catalyzes the aminoacylation of the o-tRNA. The aminoacyl-tRNA (aa-tRNA) is then released from the o-aaRS and transported to the ribosome by the EF-Tu. The aa-tRNA associates with the A-site of the ribosome and its anticodon binds the complementary triplet codon of the mRNA. The ribosome then ligates the nsAA to the growing peptide chain. When RF-1 outcompetes the aa-tRNA for binding at the UAG amber stop codon, the protein is truncated which results in a decrease of nsAA incorporation efficiency. This problem has been overcome by recode all TAG codons to its synonymous TAA codon, permitting the deletion of RF1 [6-9].

The Swartz lab has made many contributions to the development of CFPS platforms for nsAA incorporation, especially in the area of substrate optimization [38, 251]. Recently, Albayrak and Swartz produced ~1.5 mg/mL of sfGFP protein containing p-azido-phenylalnine (pAzF; azide, “click” chemistry, photocrosslinker) [11] and p-acetylphenylalanine (pAcF; keto, orthogonal reactivity) [330] with a 50-88% amber suppression efficiency [89]. This success was possible by optimizing the concentration of o-tRNA, which was identified as a limiting factor of nsAA incorporation in CFPS reactions [279]. In order to develop a practical, cost-effective method for supplying more o-tRNA to the reaction, the o-tRNA was co-

expressed in the CFPS reaction with the modified protein. The same method was also adopted and validated by Hong and colleagues and proved to reduce cost and setup time of nsAA CFPS reactions [209]. It has also been found that increasing o-aaRS concentrations can improve the production of nsAA containing proteins in CFPS reactions. This not surprising considering that engineered o-aaRS have 500- to 7000-fold less catalytic efficiencies, as compared to their native counterparts due to modifications of the o-aaRS amino acid binding pocket [84]. Increasing concentration of o-aaRS is only possible *in vitro* as o-aaRSs are known to be toxic *in vivo* at medium to high concentrations [250].

The efficiency of nsAA incorporation is greatly limited by nsAA aminoacylation efficiency [84]. Further evolution of o-aaRS/o-tRNA pairs could increase catalytic efficiencies. Using *in vitro* compartmentalized directed evolution strategies [331], CFPS systems can in principle assay larger libraries sizes as compared to *in vivo* evolution methods, which are limited by transformation efficiencies. Additionally, o-aaRS evolved *in vitro* have the potential to reach higher K_m values by limiting the concentration of nsAA. By comparison, intracellular concentration of nsAA are often high due to limited nsAA export and catabolism mechanisms, thus *in vivo* evolution limits the K_m values attainable.

Major advances have also removed a significant obstacle of amber suppression: nsAA-o-tRNA competition with release factor-1 (RF-1) [6-9]. Endogenous RF-1 recognizes amber codons and subsequently activates hydrolysis of peptidyl-tRNA to release the peptide chain. During nsAA incorporation using amber suppression technology, RF-1 competes with nsAA-o-tRNA, resulting in a significant amount of truncated product, reducing nsAA incorporation efficiencies. By recoding TAG codons from seven essential genes to the synonymous TAA ochre codons, the RF-1 gene was deleted without dramatically effecting cellular growth. The RF-1 deletion strain increased o-phosphoserine incorporation by 120-fold *in vivo* [232]. A strain with RF-1 knocked out and 13 essential gene TAG codons recoded to TAA was recently developed as a chassis for CFPS. As a result, a 250% improvement was observed compared to the parent stain with RF-1 present [209]. More recently, researchers developed a RF-1 deletion strain in which all 321 amber codons were recoded to TAA, completely freeing the amber codon to encode nsAAs [6]. While successful insertion of multiple nsAAs is possible without these

modifications [280, 332], a total recoded CFPS chassis has allowed for the incorporation of 5 pAcF residues into sfGFP [90] without additional extract processing. Further development of this chassis can potentially allow tens of nsAAs to be incorporated into a single protein.

Despite the many advances of nsAA incorporation technology, there is room for improvement and development for novel applications. Increasing the ability to incorporate multiple types of nsAAs at multiple locations in recombinant proteins are both important technological challenges for cell-free systems that need to be tackled in the near future. At present, it is not possible to incorporate more than one kind of nsAA into a single protein *in vitro* without significant extract processing [333]. The development of further evolved OTSs for improved nsAA incorporation is another important research direction that could be carried out *in vitro*, where the design space is not constrained by cell viability requirements. In addition, genome-engineering efforts to remove negative effectors of *in vitro* protein synthesis (i.e., nucleases, proteases) will increase yields of nsAA-containing proteins [90]. Beyond extract based systems, it should be noted that purified approaches offer much more freedom of design and control of the translation apparatus for genetic reprogramming [174, 266, 334, 335].

6.5.2 Post-translational incorporation of glycans and metal centers

PTMs can be carried out in CFPS systems by co-activating enzyme pathways and protein synthesis in the cell lysate. The synthetic pathways can be synthesized *in vivo* in the chassis strain prior to extract preparation, *in vitro* via CFPS, or purified and added to the CFPS reaction. These efforts have enabled the *in vitro* production of proteins bearing glycans (sugars) and accessorized with metal centers.

Glycosylation is particularly important for the production of recombinant protein therapeutics, as improper glycosylation can adversely affect the therapeutic activity or circulation half-life of a therapeutic [336]. Glycans are synthesized and attached to proteins by enzymes called glycosyltransferases (GTs) and oligosaccharyltransferases (OSTs), which together comprise the glycosylation machinery of the cell (**Figure 6.4**).

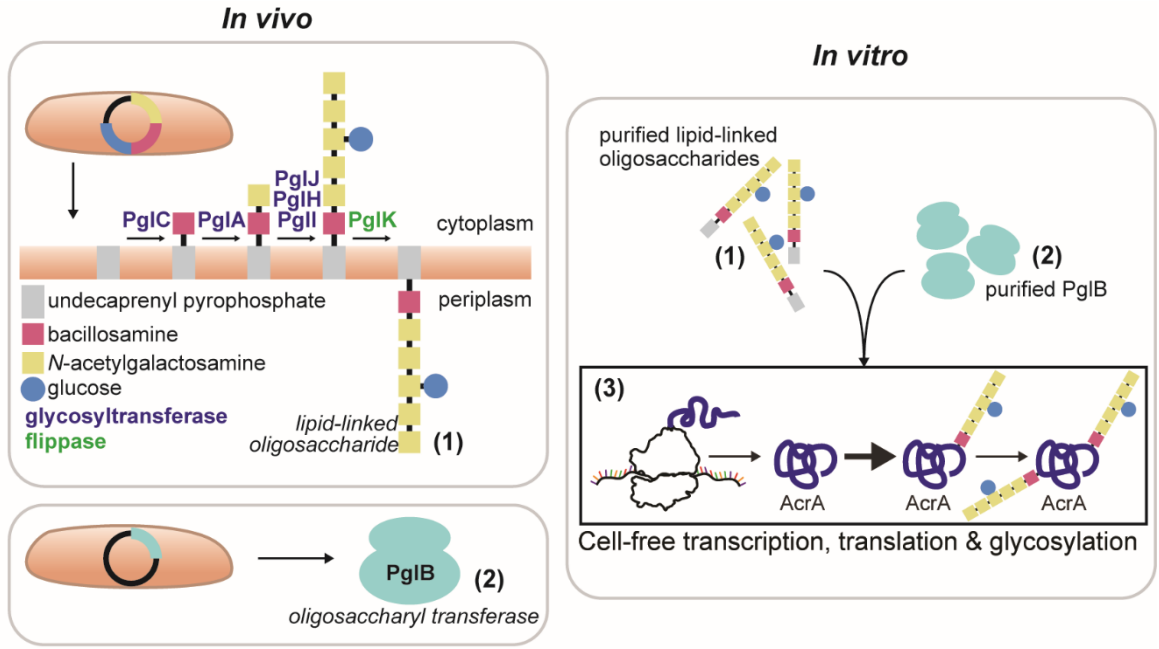


Figure 6.4: Glycoprotein production in cell-free systems [10]**.** The *C. jejuni* pgl locus, which contains glycosyltransferases and a flippase, is expressed *in vivo* in *E. coli*. These enzymes assemble sugar monomers (bacillosamine, *N*-acetylglucosamine, glucose) onto a lipid anchor (undecaprenyl pyrophosphate) in the membrane to form lipid-linked oligosaccharides (LLOs) (top). Similarly, the *C. jejuni* oligosaccharyltransferase (OST), PglB, is expressed *in vivo* in *E. coli* (bottom). LLOs and PglB are purified and added to the *in vitro* glycoprotein synthesis reaction. Addition of purified LLOs and PglB to the CFPS reaction results in the synthesis of fully glycosylated AcrA, a *C. jejuni* glycoprotein.

Glycosylation is possible in some eukaryotic CFPS systems, including ICE, CHO extract, and K562 extract [298, 300, 302, 310]. However, these platforms harness the endogeneous machinery to carry out glycosylation. As a result, this approach restricts the possible glycan structures to those naturally synthesized by the host cells, which may not precisely resemble those found in the native host. In addition, one would need to learn how to recapitulate protein trafficking to achieve some glycosylation patterns. Recent *in vivo* glycoengineering efforts have shown that cellular glycosylation machinery can be engineered to synthesize desired glycan structures [337-340]. More recent efforts are exploiting the open cell-free reaction environment for glycoengineering to enable synthetic glycosylation pathways. In a pioneering study, Guarino and DeLisa demonstrated that glycosylated proteins can be synthesized *in vitro* by adding purified bacterial lipid-linked oligosaccharides and the OST PglB to a CPFS reaction.

Yields of between 50-100 µg/mL of AcrA, a *C. jejuni* glycoprotein, were achieved [10]. This result provides a potential path for the incorporation of synthetic, human glycosylation in cell-free systems through the transfer of synthetic enzyme pathways to bacterial cell-free systems. However, significant development of this technology is still needed.

Metalloproteins are important catalysts in biology [341]. Typically, *in vivo* production of metalloproteins involves apo-protein production and purification, followed by *in vitro* metal loading [342]. CFPS systems potentially offer one-pot synthesis of metal-loaded metalloproteins by addition of free metal atoms and protein chaperones to *in vitro* reaction mixture. Recently, *in vitro* synthesis of the FeFe hydrogenase metalloprotein was described [343, 344]. Three FeS cluster-containing proteins, HydE, HydF, and HydG, are required to assemble the FeFe hydrogenase. This enzyme pathway was reconstituted *in vitro* using *E. coli* lysates containing HydE, HydF, and HydG. When the purified apo-form of the FeFe hydrogenase was added to the reaction, the active hydrogenase was synthesized at yields of 100-200 µg/mL [343]. This study provided evidence that enzymatic pathways can be reconstituted *in vitro*, but the integration of *in vitro* protein synthesis into this system has not yet been demonstrated.

Synthetic PTM mechanisms in cell-free systems have great potential for expanding the types of controllable protein modifications that are possible. This approach offers more flexibility than co-translational methods. Further work in this area will expand opportunities for producing modified proteins with a variety of (a)biological modifications *in vitro*.

6.6 Building up to genetic networks in cell-free systems

In addition to using cell-free synthetic biology for the production of individual proteins or assembly of macromolecules, CFPS systems are now used for the construction of sophisticated genetic networks, often referred to as genetic circuits [345]. Initial efforts to design and implement synthetic genetic networks dominated early *in vivo* synthetic biology efforts [346], serving as a forward engineering approach for studying natural gene regulation and controlling cellular behavior. Much progress has been made towards developing foundational genetic network modules, such as genetic switches [347-349],

logic gates [151, 350, 351], and memory modules [352] as well as engineering cells with practical applications in the areas of bioremediation [353], biosensing [354], biofuel production [355], and therapeutic applications [18]. Despite the aforementioned progress, translational *in vivo* synthetic biology is sometimes limited by cellular viability itself. First, *in vivo* genetic networks are constrained to relatively small networks due to restrictions on cellular metabolic load and limited number of genetic parts. Second, limited high-throughput methods for optimizing regulatory networks slow genetic network design cycles. Third, the interference between host regulation and synthetic networks (cross-talk) results into poor computational modeling, thereby reducing prediction capabilities as the network grows in size. Fortunately, many of these challenges can be addressed by implementing genetic networks *in vitro* using CFPS platforms.

For instance, synthetic gene networks can be studied in isolation as all the endogenous DNA and mRNA are removed during lysate preparation. This eliminates the issue of cross-talk between the host regulation and the synthetic network. In addition, the open nature of cell-free systems allows integration with hardware for high-throughput assays and on-board reaction monitoring. Consequently, automation will drastically reduce current design-build-test cycle time, and will provide new biological modules for *in vivo* characterization. Furthermore, compartmentalizing CFPS reactions within synthetic vesicles will also provide an opportunity to study artificial cells and the fundamental properties of gene regulation in confined environments.

6.6.1 Implementing genetic networks *in vitro*

Simple regulatory elements such as inducible promoters [356], transcriptional switches [357], and multi-stage cascades [356] were the first modules to be implemented *in vitro*. However, because current CFPS systems are optimized for mRNA stability and protein overproduction *in vitro* genetic networks are inadequate when constructing oscillatory circuits [358] or larger networks. Complex genetic circuitry requires resources to be efficiently directed for downstream functionalities of the networks, as well as component half-lives to mimic those found in cells. This is made clear when considering that *in vivo* global mRNA half-life is about 6.8 minutes [359], whereas, *in vitro* mRNA half-life is about 13 minutes [360]. This

discrepancy causes overall limitations of translational machinery for CFPS reactions as mRNA is accumulated in the reaction [361]. To overcome these issues, larger genetic circuits have been implemented *in vitro* by using purified MazF, a sequence-specific endoribonuclease, which has been shown to incrementally inactivate mRNA [360]. In particular, AND gates and negative feedback loops were developed in an *E. coli* cell-free expression system using this system [253]. Karig and colleagues also demonstrated negative feedback loops but without mRNA activation mechanisms, and consequently saw unwanted accumulation of mRNA which ultimately affected the dynamics of the network [362]. CFPS protein half-life has also been found to be much longer than in *in vivo* systems due to reduced proteolytic activity of CFPS system. This last aspect is important for constructing oscillatory circuits and needs to be further investigated.

Additionally, characterization and optimization of regulatory elements for cell-free systems have been explored. Shin and Noireaux have demonstrated the utility of all seven *E. coli* sigma factors [253]. This provides more transcriptional handles to design higher-order genetic networks. Moreover, T7 promoter variants have also been developed in a cell-free context for expanding regulation options and improved switch like activity [362, 363]. In order to limit the number of synthesized proteins, required for a genetic network, alternative transcriptional control mechanisms, such as ligand-sensitive transcriptional controls have been demonstrated [364]. RNA-based genetic circuits are another solution to limit energy loss from protein production and provides the benefit of quicker response times [247, 365].

The open nature of cell-free protein synthesis also allows for *in situ* monitoring of genetic networks. In particular, binary FRET probes have been utilized to monitor mRNA dynamics in real-time [366]. Alternatively, the use of fluorescent RNA aptamers have shown a very simple strategy for monitoring RNA in real-time [367]. High-throughput analysis of biomolecular interactions can also be performed by encapsulating CFPS reactions in nanoliter droplets and assaying interactions via a immunoassay with a temporal resolution of 1 second [368]. HPLC analysis for monitoring metabolite concentrations are more straightforward to use since CFPS reactions can be run directly on an HLPC or MS machine.

Collectively, recent efforts hold promise for employing more complex circuitry and real-world applications [257]. Due to the open environment of cell-free systems and short implementation times, CFPS systems can also be used as a complementary route to prototyping genetic networks in an *in vitro* environment prior to *in vivo* implementation.

6.6.2 Opportunities in prototyping

In vivo synthetic biology applies molecular biology tools to forward-engineer cellular behavior. This is done first through *in silico* design of genetic networks, construction of those networks, followed by testing *in vivo* [369, 370]. However, *in vivo* engineering cycles require a significant amount of time and financial resources [18]. In contrast, *in vitro* biomolecular prototyping promises to improve the overall efficiency of the design-build-test cycle [371] (**Figure 6.5A**). There are several demonstrations of prototyping circuits *in vitro* for *in vivo* implementation, such as negative feedback loops [362], multi-input regulated T7 promoters [363], and riboswitches [372]. However, it is difficult to predict correlations between parameters (*e.g.,* promoters/RBS strengths) found *in vitro* and predicted values *in vivo*. For instance, it was found that characterizing promoter strength *in vitro* must be performed with circularized DNA for there be a correlation to *in vivo* promoter strengths [373]; however, this is limited to low to medium strength, constitutive promoter strengths. Additional work is needed to find unknown parameters involved with the transition from *in vitro* to *in vivo* implementation.

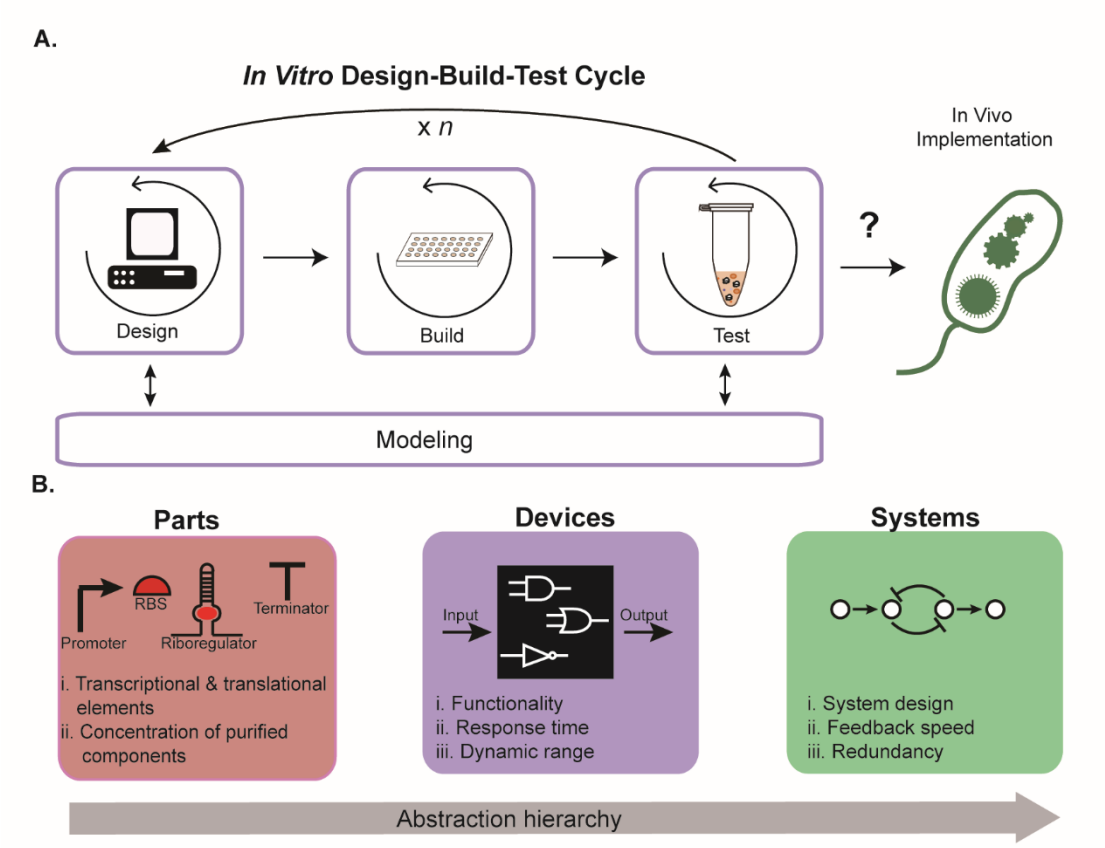


Figure 6.5: *In vitro* prototyping of genetic networks. (A) Overview of *in vitro* prototyping for speeding up *in vivo* design-build-test cycles. *In vitro* prototyping allows for a genetic part or network to be quickly screened for specific characteristics *in vitro* before implementation *in vivo*. First, in the Design stage, the genetic parts or networks are designed, informed by computational models or literature. Next, in the Build stage, several designs are built. In the Test phase, built designs are assayed *in vitro*. If the Test stage does not yield the desired behavior, one reinitiates the cycle *n* number of times until the desired characteristic is achieved. The Test stage can also inform the modeling and allow for better models for the Design stage. Once the desired characteristics are found, top candidates that behave as expected can be implemented *in vivo*, with an increased likeliness of being functional. **(B) Variables for characterization and optimization.** *In vitro* prototyping can occur by characterizing or optimizing various levels on the genetic network "abstraction hierarchy." At the most basic level, transcriptional and translational parts or purified components of a network can be analyzed for functionality. *In vitro* prototyping can also be applied at the device level to assess input-output relationships. Finally, an *in vitro* systems level analysis allows for an isolated study of how multiple genetic devices feed into each other and a preview of the overall network behavior.

Despite these issues, rapid prototyping techniques are compelling and provide a potentially rapid alternative to *in vivo* systems for circuit design and testing. For instance, prototyping will be accelerated through microscale reactions in micro-well plates and microfluidic system. Indeed, these methodologies decrease prototyping time with real-time simultaneous or sequential screening of many system's parameters [374]. Complementary, predictive computational models are being developed to better understand CFPS's limitations in prototyping [375]. As a result, it was confirmed that in CFPS, potential design limitations arise from improperly balanced of energy utilization. Utilizing a continuous exchange set-up can partially address this issue, and has been proven to be a reliable solution for prototyping [356]. Prototyping efforts have even moved beyond a laboratory environment with development of a CFPS system that is both portable and stable because of its ability to reaction on paper [257]. This allows for a low-cost prototyping environment and provides opportunities in using CFPS for diagnostic purposes.

Improvements in CFPS systems have opened new opportunities for cell-free synthetic biology, including studying genetic networks in isolation, characterizing new regulatory elements and speeding up design-build-test cycles for *in vivo* synthetic biology. The size of genetic networks achievable in CFPS systems will only increase as CFPS systems are further developed to meet the specific needs of *in vitro* genetic networks (e.g. downstream resourcing and matching *in vivo* component half-lives). Additionally, further computational modeling cell-free reactions will further improve *in silico* design of in vitro networks and assist in understanding the genetic regulation in isolation. In sum, *in vitro* genetic networks hold much potential for the synthetic biology field as it complements many *in vivo* efforts.

6.7 Conclusion and Future Outlook

Overall, the renewed scientific interest in CFPS in the past two decades has resulted in drastically increased batch yields, active reaction durations, and reaction volumes [24]. The variety of different CFPS technology platforms has further enabled the *in vitro* production of proteins with diverse complexity and species origin. CFPS technology will be invaluable in the near future as an "organism protein factory," to aid in the effort to determine the gene products for the many organisms whose genomes have been sequenced or will be sequenced in the near future due to the rapidly falling cost of DNA sequencing

technology. For example, the Sargasso Seas expedition identified over 1.2 million genes, many of which have unknown function [376].

CFPS technology will also have an increasing role in complementing *in vivo* synthetic biology efforts. Due to its flexibility and high-throughput potential, cell-free systems are perfectly suited to synthesize and assay large libraries, not only for genetic networks as surveyed herein, but also for evolutionary or proteomics studies [377]. This will provide *in vivo* efforts with enriched mutant libraries and accelerated methods for preparing large-scale protein libraries. In sum, CFPS holds tremendous potential to transform our ability to synthesize recombinant proteins and its applications in the synthetic biology field will only grow in the future.

6.8 Acknowledgements

The authors would like to acknowledge Jennifer A. Schoborg, Erik D. Carlson, and Filippo Caschera for assistance in writing this manuscript. Select figures in this text were adapted from Jennifer A. Schoborg and Erik D. Carlson. We gratefully acknowledge the National Science Foundation (MCB-0943393), the Office of Naval Research (N00014-11-1-0363), the DARPA YFA Program (N66001-11-1-4137), the Army Research Office (W911NF- 11-1-044), the NSF Materials Network Grant (DMR - 1108350), the DARPA Living Foundries Program (N66001-12-C-4211), the David and Lucile Packard Foundation (2011-37152), ARPA-E (DE-AR0000435), and the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust for support. JGP is an NSF Graduate Fellow and NIH IMSD Graduate Fellow. JCS is funded, in part, by the Northwestern Biotechnology Training Program supported by NIH (T32GM008449).

6.9 Author Contribution

As co-first authors, Jessica Stark and I, both contributed equally to this book chapter.

6.10 Publication Information

The previous chapter is published with the following citation information: Perez J.G.*, Stark J.C.*, Jewett M.C. Cell-Free Synthetic Biology: Engineering Beyond the Cell. Cold Spring Harb Perspect Biol. 2016 Oct 14. doi: 10.1101/cshperspect.a023853.

CHAPTER 7

7 Summary and Future Directions

7.1 Summary

This work has demonstrated that non-standard amino acid (nsAA) incorporation is a powerful technology capable of a variety of prominent applications. This strategy has the potential to dominate the untapped niche of synthesizing functionally complex proteins. Conventional methods to synthesize modified proteins are limited by length (chemical synthesis) and chemical decoration (eukaryotic post-translational modifications). Expansion of the genetic code is capable of overcoming both of these obstacles by utilizing the synthetic potential and flexibility of the cell's translation apparatus.

In order for this field to reach its full potential several key obstacles must be overcome: (1) underdeveloped strains for expansion of the genetic code and (2) poor nsAA incorporation efficiencies of orthogonal translation systems (OTSs). The majority of this project sought to develop an optimized strain for production of proteins containing nsAAs. This work also addresses the second issue by evolving multiple OTS components in parallel and combining beneficial mutants together to facilitate improved nsAA incorporation efficiencies.

First we hypothesized that removing negative effectors of protein production would increase protein yield. Indeed, we showed a 1.7- to 17-fold improvement in sfGFP-wt and sfGFP2TAG expression by reducing protease and nuclease activity in the genomically recoded strain, *C321∆A*. During this process we learned that genomic modifications allowing BL21(DE3) to have increased protein expression did not translate to similar levels of success in *C321∆A,* a K strain. This study builds upon a body of literature which shows that small differences of genotypes can have a significant, unknown impact on protein expression [118, 153]. This study also revealed that nuclease and protease activity reduction was most advantageous when expressing sfGFP-2TAG and pyrrolysine synthetase (PylRS), a known difficult to synthesize protein due to its insoluble N-terminus.

Next we explored a tunable recombinant protein expression system, the T7RNAP cassette, in *C321.∆A*. This work is the first demonstration of a functional T7 system in a genomically recoded strain.

Our initial system however showed very leaky expression (expression of protein in the absence of the inducer) and low protein yields as compared to BL21(DE3), the gold standard in *E. coli* protein production. We addressed these two issues by introducing a novel aTc-inducible T7 system into *C321.ΔA* and exploring the genomic insertion site's role on protein expression. The aTc-inducible T7 system drastically reduced leaky expression (to undetectable levels); however, addition of aTc reduced overall protein expression levels. Due to the tight induction response of aTc, this system is ideal for projects that require low levels of toxic proteins. Additionally, genomic locations that yielded increased protein expression were identified, but had a negative effect on basal expression (without the inducer). A more comprehensive study is needed to draw any conclusions of genomic location's role on protein expression in a T7 system.

During the process of inserting the T7RNAP cassette at various genomic locations, we encountered issues locating the T7RNAP cassette after its insertion on the genome. In order to find the insertion sites of these cassette we developed a novel method to uncover the location of genes at an unknown genomic location. This method may be advantage for future genomic engineering efforts due to the expense of whole genome sequencing.

Finally, the key obstacle to low nsAA incorporation efficiencies was addressed in Chapter 5. Here we detailed how poor aminoacylation rates of orthogonal synthetases (o-aaRSs) and low binding affinities of the elongation factor Tu (EF-Tu) to the charged orthogonal tRNA (nsAA-o-tRNA) may be responsible for poor nsAA incorporation rates. We sought to address this problem by utilizing a novel protein evolution approach: systematically evolving multiple translational components in parallel (i.e. o-aaRS and EF-Tu) and combining top mutants for synergistic benefits. We also promoter optimized the expression plasmid. Using these strategies, we observed a 2-, 5-, and 20-fold increase in incorporation of p-azido-phenylalanine (pAzF) into sfGFP1TAG, sfGFP3TAG and sfGFP5TAG, respectively, as compared to the original translational components. These efforts highlight the intricacy and interconnectedness of the native translation system. Such a complex system needs to be coordinately optimized (*e.g.,* codons, tRNA, aaRS, EF-Tu, and the ribosome together) to enable efficient expansion of the genetic code.

7.2 Future Directions and Perspective

The work detailed in this dissertation lays the groundwork for further development of genomically recoded strains, exciting applications using the array of strains already developed (detailed in Chapter 3) and novel approaches to protein engineering of OTS components. In my opinion, there are four key areas for continued development and application of this work: (i) synthesizing products that will broaden *C321.ΔA* variants' appeal, (ii) a more detailed study of negative effectors to protein production in *C321.ΔA*, (iii) further exploration of additional genomic insertion sites of the T7RNAP cassette (or other heterologous genes), and (iv) application of systematic protein engineering efforts to other OTSs.

With respect to the first point, we need to synthesize products other than sfGFP to demonstrate the full potential of the *C321.ΔA* mutants with reduced protease and nuclease activity. Synthesizing proteins that are difficult to express in other strains may find success with our strain library. We observed the most dramatic improvement with our mutant *C321.ΔA* strains when expressing sfGFP-2TAG and PylRS, a protein known to be highly insoluble [128, 129]. It is reasonable to believe that other complex, insoluble proteins will have success using our C321.ΔA strain library.

Additionally, utilizing the Pyl OTS in the *C321.ΔA* mutant strains will enable expression of useful proteins containing Pyl and its analogs. Moreover, these strains may prove to be an excellent test bed for further protein evolution studies to improve PylRS aminoacylation rates and its substrate specificity. Studies show that *M. mazei* PylRS' catalytic core contains a deep hydrophobic pocket for binding of Pyl [134, 135] which allows PylRS to display a remarkably high tolerance towards a variety of substrates [136]. Thus, our work could enable higher yields of modified proteins containing nsAAs compatible with the Pyl OTS system, which has been limited by insolubility issues.

Second, in our study we only tested five genomic targets to improve protein production, while there are numerous targets published in the literature that may have additional benefits to the protein production capacity of these strains. For instance, Rna, CsdA and MazF [90] are nucleases that have been shown to reduce recombinant protein yields and were not tested in this work. Modifications to metabolic pathways could also be studied such as deleting regulators ArcA and IclR which have been shown to reduce acetate production and increase protein yields. [172] These studies may also give a

more detailed picture of the functional differences between B and K strains, due to the fact that most of these genomic modifications have been studied in B strains.

Third, a more detailed study of genomic location on heterologous gene expression in *E. coli* would have many benefits, not only for optimizing the T7 system in *C321.ΔA* but also for other projects requiring expression off the genome. DNA replication originates at a fixed positon on the genome (origin of replication, **Supplemental Figure A11**) and rapidly growing bacteria initiate new rounds of replication before the previous rounds have been completed, allowing the copy number of the integrate genes to fluctuate. This can drastically increase or decrease expression of the gene of interest. Transcriptional orientation and local transcriptional activity can also affect expression of recombinant genes. A detailed study of all these components will be advantageous as our ability to incorporate genes on the genome improves. This will also allow us to better understand these factors in native gene expression. The work outlined here provides desirable and undesirable genomic insertion sites (with or without off targets) as starting points for this system. This work has also outlined methods to find genomic insertions if off target insertion does occur.

Finally, utilizing the systematic protein evolution approach outline here towards other OTS systems will give us insight on how generalizable the strategy is. It may also enable improved nsAA incorporation efficiencies over protein evolution strategies of a single OTS component in isolation. This approach could also be applied to OTS components not explored in this study as well as the o-tRNA and the ribosome.

In sum, co-translational incorporation of nsAAs is a growing field that has exciting possibilities in the future. Efforts to improve aminoacylation rates of synthetases is on the rise [99]. Alternative recoding strategies that extend off *C321.ΔA* may provide a route to recoding additional codons. The increasing productivity of cell-free protein synthesis methods will remove many biological constraints at play during *in vivo* work such as limitations on cell membrane permeability and use of toxic nsAAs [378]. Overcoming these technological barriers will enable a technical renaissance, allowing us to produce the next wave of highly functionalized biomaterials and protein therapeutics with broad applications in pharmaceuticals, chemical biology, cell biology and biotechnology.

REFERENCES

1. Isaacs, F.J., et al., *Precise manipulation of chromosomes in vivo enables genome-wide codon replacement.* Science, 2011. **333**(6040): p. 348-53.
2. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* Science, 2013. **342**(6156): p. 357-360.
3. Matthies, D., et al., *Cell-free expression and assembly of ATP synthase.* J Mol Biol, 2011. **413**(3): p. 593-603.
4. Young, T.S., et al., *An enhanced system for unnatural amino acid mutagenesis in E. coli.* Journal of molecular biology, 2010. **395**(2): p. 361-74.
5. Kucharczyk, R., et al., *Mitochondrial ATP synthase disorders: molecular mechanisms and the quest for curative therapeutic approaches.* Biochim Biophys Acta, 2009. **1793**(1): p. 186-99.
6. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* Science, 2013. **342**(6156): p. 357-60.
7. Johnson, D.B., et al., *RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites.* Nat Chem Biol, 2011. **7**(11): p. 779-86.
8. Loscha, K.V., et al., *Multiple-site labeling of proteins with unnatural amino acids.* Angew Chem Int Ed Engl, 2012. **51**(9): p. 2243-6.
9. Ohtake, K., et al., *Efficient decoding of the UAG triplet as a full-fledged sense codon enhances the growth of a prfA-deficient strain of Escherichia coli.* J Bacteriol, 2012. **194**(10): p. 2606-13.
10. Guarino, C. and M.P. DeLisa, *A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins.* Glycobiology, 2012. **22**(5): p. 596-601.
11. Chin, J.W., et al., *Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli.* J Am Chem Soc, 2002. **124**(31): p. 9026-7.
12. Kobayashi, T., et al., *Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion.* Nat Struct Biol, 2003. **10**(6): p. 425-32.
13. Benner, S.A. and A.M. Sismour, *Synthetic biology.* Nat Rev Genet, 2005. **6**(7): p. 533-43.
14. Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age.* Nat Rev Genet, 2010. **11**(5): p. 367-79.
15. Purnick, P.E. and R. Weiss, *The second wave of synthetic biology: from modules to systems.* Nat Rev Mol Cell Biol, 2009. **10**(6): p. 410-22.
16. Endy, D., *Foundations for engineering biology.* Nature, 2005. **438**(7067): p. 449-53.
17. Raimbault, B., J.P. Cointet, and P.B. Joly, *Mapping the Emergence of Synthetic Biology.* PLoS One, 2016. **11**(9): p. e0161522.
18. Ro, D.K., et al., *Production of the antimalarial drug precursor artemisinic acid in engineered yeast.* Nature, 2006. **440**(7086): p. 940-3.
19. Porter, D.L., et al., *Chimeric antigen receptor–modified T cells in chronic lymphoid leukemia.* New England Journal of Medicine, 2011. **365**(8): p. 725-733.
20. Lemon, K.P., et al., *Microbiota-targeted therapies: an ecological perspective.* Science translational medicine, 2012. **4**(137): p. 137rv5-137rv5.
21. Sonnenburg, J.L. and M.A. Fischbach, *Community health care: therapeutic opportunities in the human microbiome.* Science translational medicine, 2011. **3**(78): p. 78ps12-78ps12.
22. Wang, H.H. and G.M. Church, *Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering.* Methods Enzymol, 2011. **498**: p. 409-26.
23. Doudna, J.A. and E. Charpentier, *The new frontier of genome engineering with CRISPR-Cas9.* Science, 2014. **346**(6213).
24. Carlson, E.D., et al., *Cell-free protein synthesis: applications come of age.* Biotechnol Adv, 2012. **30**(5): p. 1185-94.
25. Kwon, Y.C. and M.C. Jewett, *High-throughput preparation methods of crude extract for robust cell-free protein synthesis.* Sci Rep, 2015. **5**: p. 8663.

26. Umehara, T., et al., *N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo.* FEBS Lett, 2012. **586**(6): p. 729-33.
27. Tanrikulu, I.C., et al., *Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo.* Proc Natl Acad Sci U S A, 2009. **106**(36): p. 15285-90.
28. Boniecki, M.T., et al., *CP1-dependent partitioning of pretransfer and posttransfer editing in leucyl-tRNA synthetase.* Proc Natl Acad Sci U S A, 2008. **105**(49): p. 19223-8.
29. Reynolds, N.M., et al., *Cell-specific differences in the requirements for translation quality control.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4063-8.
30. Ling, J. and D. Soll, *Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4028-33.
31. Wang, Q. and L. Wang, *New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast.* Journal of the American Chemical Society, 2008. **130**(19): p. 6066-6067.
32. Liu, W., et al., *Genetic incorporation of unnatural amino acids into proteins in mammalian cells.* Nat Meth, 2007. **4**(3): p. 239-244.
33. Mukai, T., et al., *Genetic encoding of non-natural amino acids in Drosophila melanogaster Schneider 2 cells.* Protein Sci, 2010. **19**(3): p. 440-8.
34. Kondo, K., J. Hodgkin, and R.H. Waterston, *Differential expression of five tRNA(UAGTrp) amber suppressors in Caenorhabditis elegans.* Mol Cell Biol, 1988. **8**(9): p. 3627-35.
35. Yang, H., et al., *A general method for artificial metalloenzyme formation through strain-promoted azide-alkyne cycloaddition.* Chembiochem, 2014. **15**(2): p. 223-7.
36. Cho, H., et al., *Optimized clinical performance of growth hormone with an expanded genetic code.* Proc Natl Acad Sci U S A, 2011. **108**(22): p. 9060-5.
37. Albayrak, C. and J.R. Swartz, *Direct polymerization of proteins.* ACS Synth Biol, 2014. **3**(6): p. 353-62.
38. Bundy, B.C. and J.R. Swartz, *Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation.* Bioconjug Chem, 2010. **21**(2): p. 255-63.
39. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annu Rev Biochem, 2010. **79**: p. 413-44.
40. Currin, A., et al., *Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently.* Chemical Society Reviews, 2015. **44**(5): p. 1172-1239.
41. Park, H.S., et al., *Expanding the genetic code of Escherichia coli with phosphoserine.* Science, 2011. **333**(6046): p. 1151-4.
42. Lajoie, M.J., et al., *Genomically Recoded Organisms Expand Biological Functions.* Science (New York, N.Y.), 2013. **342**(6156): p. 357-360.
43. Opitz, C.A., et al., *Damped elastic recoil of the titin spring in myofibrils of human myocardium.* Proceedings of the National Academy of Sciences, 2003. **100**(22): p. 12688-12693.
44. Walsh, C., *Posttranslational modification of proteins: expanding nature's inventory*. 2006: Roberts and Company Publishers.
45. Prescher, J.A., D.H. Dube, and C.R. Bertozzi, *Chemical remodelling of cell surfaces in living animals.* Nature, 2004. **430**(7002): p. 873-7.
46. Witus, L.S. and M.B. Francis, *Using synthetically modified proteins to make new materials.* Acc Chem Res, 2011. **44**(9): p. 774-83.
47. Schnolzer, M. and S.B. Kent, *Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease.* Science, 1992. **256**(5054): p. 221-5.
48. Fischer, E. and E. Fourneau, *Ueber einige Derivate des Glykocolls.* Berichte der deutschen chemischen Gesellschaft, 1901. **34**(2): p. 2868-2877.
49. Constable, D.J.C., et al., *Key green chemistry research areas-a perspective from pharmaceutical manufacturers.* Green Chemistry, 2007. **9**(5): p. 411-420.
50. Carpino, L.A., et al., *Rapid, continuous solution-phase peptide synthesis: application to peptides of pharmaceutical interest.* Organic process research & development, 2003. **7**(1): p. 28-37.
51. Kent, S.B., *Chemical synthesis of peptides and proteins.* Annual review of biochemistry, 1988. **57**(1): p. 957-989.

52. Albericio, F., *Solid-phase synthesis: a practical guide*. 2000: CRC Press.
53. Stawikowski, M. and G.B. Fields, *Introduction to Peptide Synthesis.* Current protocols in protein science / editorial board, John E. Coligan ... [et al.], 2002. **CHAPTER**: p. Unit-18.1.
54. Thayer, A., *Making peptides at large scale.* Chem Eng News, 2011. **89**(22): p. 81-85.
55. Greenland, W.E., et al., *Solid-phase synthesis of peptide radiopharmaceuticals using Fmoc-N-epsilon-(hynic-Boc)-lysine, a technetium-binding amino acid: application to Tc-99m-labeled salmon calcitonin.* J Med Chem, 2003. **46**(9): p. 1751-7.
56. Zaoral, M., *DDAVP (Desmopressin) and solid phase peptide synthesis.* Biopolymers, 2008. **90**(3): p. 213.
57. Hsiao, T.Y. and S.W. Lee, *Process for the preparation of leuprolide and its pharmaceutically acceptable salts*. 2015, Google Patents.
58. Marx, V., *Roche's fuzeon challenge.* Chem. Eng. News, 2005. **83**(11): p. 16-17.
59. Kramer, A., et al., *Simultaneous synthesis of peptide libraries on single resin and continuous cellulose membrane supports: examples for the identification of protein, metal and DNA binding peptide mixtures.* Pept Res, 1993. **6**(6): p. 314-9.
60. Alberts, B., *Molecular Biology of the Cell with CD*. 2008: Garland.
61. Palomares, L.A., S. Estrada-Moncada, and O.T. Ramírez, *Production of recombinant proteins.* Recombinant Gene Expression: Reviews and Protocols, 2004: p. 15-51.
62. Rader, R.A., *Biopharmaceutical products in the US and European markets*. 2007: BioPlan Associates, Incorporated.
63. Jayapal, K.P., et al., *Recombinant protein therapeutics from CHO cells-20 years and counting.* Chemical Engineering Progress, 2007. **103**(10): p. 40.
64. Kannicht, C. and B. Fuchs, *Post-Translational Modif ications of Proteins*. 2008: Springer.
65. Walsh, G., *Post-translational modifications of protein biopharmaceuticals.* Drug discovery today, 2010. **15**(17): p. 773-780.
66. Bernard, G.R., et al., *Efficacy and safety of recombinant human activated protein C for severe sepsis.* N Engl J Med, 2001. **344**(10): p. 699-709.
67. Muthing, J., et al., *Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24.* Biotechnol Bioeng, 2003. **83**(3): p. 321-34.
68. Yang, M. and M. Butler, *Effects of ammonia and glucosamine on the heterogeneity of erythropoietin glycoforms.* Biotechnol Prog, 2002. **18**(1): p. 129-38.
69. Serrato, J.A., et al., *Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures.* Biotechnol Bioeng, 2004. **88**(2): p. 176-88.
70. Zhang, J., *Mammalian cell culture for biopharmaceutical production*, in *Manual of Industrial Microbiology and Biotechnology, Third Edition*. 2010, American Society of Microbiology. p. 157-178.
71. Macino, G., et al., *Use of the UGA terminator as a tryptophan codon in yeast mitochondria.* Proceedings of the National Academy of Sciences, 1979. **76**(8): p. 3784-3785.
72. Ambrogelly, A., S. Palioura, and D. Soll, *Natural expansion of the genetic code.* Nat Chem Biol, 2007. **3**(1): p. 29-35.
73. Bock, A. and T.C. Stadtman, *Selenocysteine, a highly specific component of certain enzymes, is incorporated by a UGA-directed co-translational mechanism.* Biofactors, 1988. **1**(3): p. 245-50.
74. Gaston, M.A., R. Jiang, and J.A. Krzycki, *Functional context, biosynthesis, and genetic encoding of pyrrolysine.* Curr Opin Microbiol, 2011. **14**(3): p. 342-9.
75. Ruan, B., et al., *Quality control despite mistranslation caused by an ambiguous genetic code.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16502-7.
76. Bezerra, A.R., et al., *Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification.* Proc Natl Acad Sci U S A, 2013. **110**(27): p. 11079-84.
77. Wang, L., et al., *A New Functional Suppressor tRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

78. Wang, L. and P.G. Schultz, *A general approach for the generation of orthogonal tRNAs.* Chem Biol, 2001. **8**(9): p. 883-90.
79. Nguyen, D.P., et al., *Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry.* J Am Chem Soc, 2009. **131**(25): p. 8720-1.
80. Wang, Y.S., et al., *Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant.* ACS Chem Biol, 2013. **8**(2): p. 405-15.
81. Yanagisawa, T., et al., *Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification.* Chem Biol, 2008. **15**(11): p. 1187-97.
82. Pirman, N.L., et al., *A flexible codon in genomically recoded Escherichia coli permits programmable protein phosphorylation.* Nat Commun, 2015. **6**: p. 8130.
83. Oza, J.P., et al., *Robust production of recombinant phosphoproteins using cell-free protein synthesis.* Nat Commun, 2015. **6**: p. 8168.
84. O'Donoghue, P., et al., *Upgrading protein synthesis for synthetic biology.* Nat Chem Biol, 2013. **9**(10): p. 594-598.
85. Contreras-Martinez, L.M., et al., *The ribosomal exit tunnel as a target for optimizing protein expression in Escherichia coli.* Biotechnology Journal, 2012. **7**(3): p. 354-360.
86. Rackham, O. and J.W. Chin, *A network of orthogonal ribosome x mRNA pairs.* Nat Chem Biol, 2005. **1**(3): p. 159-66.
87. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome.* Nature, 2010. **464**(7287): p. 441-444.
88. Sakamoto, K., et al., *Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells.* Nucleic Acids Res, 2002. **30**(21): p. 4692-9.
89. Albayrak, C. and J.R. Swartz, *Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation.* Nucleic Acids Res, 2013. **41**(11): p. 5949-63.
90. Hong, S.H., et al., *Improving cell-free protein synthesis through genome engineering of Escherichia coli lacking release factor 1.* Chembiochem, 2015. **16**(5): p. 844-53.
91. Lodish, H., et al., *Molecular cell biology*. Vol. 3. 1995: Scientific American Books New York.
92. Macino, G., et al., *Use of the UGA terminator as a tryptophan codon in yeast mitochondria.* Proc Natl Acad Sci U S A, 1979. **76**(8): p. 3784-5.
93. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Frontiers in Microbiology, 2014. **5**: p. 172.
94. Beene, D.L., D.A. Dougherty, and H.A. Lester, *Unnatural amino acid mutagenesis in mapping ion channel function.* Curr Opin Neurobiol, 2003. **13**(3): p. 264-70.
95. Cornish, V.W., D. Mendel, and P.G. Schultz, *Probing protein structure and function with an expanded genetic code.* Angewandte Chemie International Edition in English, 1995. **34**(6): p. 621-633.
96. Yang, Y., et al., *Significant Improvement of Oxidase Activity through the Genetic Incorporation of a Redox-active Unnatural Amino Acid.* Chem Sci, 2015. **6**(7): p. 3881-3885.
97. Liu, H., et al., *A method for the generation of glycoprotein mimetics.* J Am Chem Soc, 2003. **125**(7): p. 1702-3.
98. Hammerling, M.J., et al., *Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness.* Nat Chem Biol, 2014. **10**(3): p. 178-80.
99. O'Donoghue, P., et al., *Upgrading protein synthesis for synthetic biology.* Nat Chem Biol, 2013. **9**(10): p. 594-8.
100. Des Soye, B.J., et al., *Repurposing the translation apparatus for synthetic biology.* Curr Opin Chem Biol, 2015. **28**: p. 83-90.
101. Young, T.S., et al., *An enhanced system for unnatural amino acid mutagenesis in E. coli.* J Mol Biol, 2010. **395**(2): p. 361-74.
102. Santoro, S.W., et al., *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity.* Nat Biotechnol, 2002. **20**(10): p. 1044-8.

103. Ryu, Y. and P.G. Schultz, *Efficient incorporation of unnatural amino acids into proteins in Escherichia coli.* Nat Methods, 2006. **3**(4): p. 263-5.
104. Griffiths, A.J., *An introduction to genetic analysis*. 2005: Macmillan.
105. Amiram, M., et al., *Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids.* Nat Biotechnol, 2015. **33**(12): p. 1272-1279.
106. Mandell, D.J., et al., *Biocontainment of genetically modified organisms by synthetic protein design.* Nature, 2015. **518**(7537): p. 55-60.
107. Gopal, G.J. and A. Kumar, *Strategies for the production of recombinant protein in Escherichia coli.* Protein J, 2013. **32**(6): p. 419-25.
108. Makino, T., G. Skretas, and G. Georgiou, *Strain engineering for improved expression of recombinant proteins in bacteria.* Microbial Cell Factories, 2011. **10**: p. 32-32.
109. Borja, G.M., et al., *Engineering Escherichia coli to increase plasmid DNA production in high cell-density cultivations in batch mode.* Microbial Cell Factories, 2012. **11**: p. 132-132.
110. Lopez, P.J., et al., *The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo.* Mol Microbiol, 1999. **33**(1): p. 188-99.
111. Kaplan, R. and D. Apirion, *The involvement of ribonuclease I, ribonuclease II, and polynucleotide phosphorylase in the degradation of stable ribonucleic acid during carbon starvation in Escherichia coli.* J Biol Chem, 1974. **249**(1): p. 149-51.
112. Deutscher, M.P., *E. coli RNases: making sense of alphabet soup.* Cell, 1985. **40**(4): p. 731-2.
113. Studier, F.W. and B.A. Moffatt, *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.* J Mol Biol, 1986. **189**(1): p. 113-30.
114. Baneyx, F., *Recombinant protein expression in Escherichia coli.* Curr Opin Biotechnol, 1999. **10**(5): p. 411-21.
115. Graumann, K. and A. Premstaller, *Manufacturing of recombinant therapeutic proteins in microbial systems.* Biotechnol J, 2006. **1**(2): p. 164-86.
116. Borja, G.M., et al., *Engineering Escherichia coli to increase plasmid DNA production in high cell-density cultivations in batch mode.* Microbial Cell Factories, 2012. **11**(1): p. 132.
117. Kushner, S.R., *mRNA Decay in Escherichia coli Comes of Age.* Journal of Bacteriology, 2002. **184**(17): p. 4658-4665.
118. Studier, F.W., et al., *Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes.* J Mol Biol, 2009. **394**(4): p. 653-80.
119. Spiess, C., A. Beil, and M. Ehrmann, *A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein.* Cell, 1999. **97**(3): p. 339-47.
120. Goerke, A.R. and J.R. Swartz, *Cell-free synthesis of proteins containing unnatural amino acids*. 2008, Google Patents.
121. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution.* Nature, 2009. **460**(7257): p. 894-8.
122. saiSree, L., M. Reddy, and J. Gowrishankar, *IS186 insertion at a hot spot in the lon promoter as a basis for lon protease deficiency of Escherichia coli B: identification of a consensus target sequence for IS186 transposition.* J Bacteriol, 2001. **183**(23): p. 6943-6.
123. Kido, M., et al., *RNase E polypeptides lacking a carboxyl-terminal half suppress a mukB mutation in Escherichia coli.* J Bacteriol, 1996. **178**(13): p. 3917-25.
124. Donovan, W.P. and S.R. Kushner, *Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in Escherichia coli K-12.* Proc Natl Acad Sci U S A, 1986. **83**(1): p. 120-4.
125. Bermardo, G. and C. Cordonnier, *MECHANISM OF DEGRADATION OF DNA BY ENDONUCLEASE I FROM ESCHERICHIA COLI.* J Mol Biol, 1965. **11**: p. 141-3.
126. Inouye, S. and M. Inouye, *Up-promoter mutations in the lpp gene of Escherichia coli.* Nucleic Acids Res, 1985. **13**(9): p. 3101-10.
127. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA.* Science, 2002. **296**(5572): p. 1459-62.

128. Yanagisawa, T., et al., *Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina mazei.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2006. **62**(Pt 10): p. 1031-3.
129. Jiang, R. and J.A. Krzycki, *PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine.* J Biol Chem, 2012. **287**(39): p. 32738-46.
130. Hao, B., et al., *Reactivity and chemical synthesis of L-pyrrolysine- the 22(nd) genetically encoded amino acid.* Chem Biol, 2004. **11**(9): p. 1317-24.
131. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.
132. Richard Lewis, R.P., Jason Leibowitz, Satwant Kaur, Narula Michael, Joseph Ryan *Novel expression systems utilizing bacteriophage T7 promoters and gene sequences*, in *US Patent*. 1985, Merck Sharp & Dohme Corp USA.
133. Chatterjee, A., et al., *A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli.* Biochemistry, 2013. **52**(10): p. 10.1021/bi4000244.
134. Yanagisawa, T., et al., *Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase.* J Mol Biol, 2008. **378**(3): p. 634-52.
135. Kavran, J.M., et al., *Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation.* Proc Natl Acad Sci U S A, 2007. **104**(27): p. 11268-73.
136. Polycarpo, C.R., et al., *Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase.* FEBS Lett, 2006. **580**(28-29): p. 6695-700.
137. Turcatti, G., et al., *Probing the structure and function of the tachykinin neurokinin-2 receptor through biosynthetic incorporation of fluorescent amino acids at specific sites.* J Biol Chem, 1996. **271**(33): p. 19991-8.
138. Arslan, T., et al., *Structurally Modified Firefly Luciferase. Effects of Amino Acid Substitution at Position 286.* Journal of the American Chemical Society, 1997. **119**(45): p. 10877-10887.
139. Pollitt, S.K. and P.G. Schultz, *A Photochemical Switch for Controlling Protein–Protein Interactions.* Angewandte Chemie International Edition, 1998. **37**(15): p. 2104-2107.
140. England, P.M., et al., *Site-specific, photochemical proteolysis applied to ion channels in vivo.* Proc Natl Acad Sci U S A, 1997. **94**(20): p. 11025-30.
141. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
142. Chen, Y.J., et al., *Characterization of 582 natural and synthetic terminators and quantification of their design constraints.* Nat Methods, 2013. **10**(7): p. 659-64.
143. Orelle, C., et al., *Protein synthesis by ribosomes with tethered subunits.* Nature, 2015. **524**(7563): p. 119-24.
144. Cornish, V.W., et al., *Site-specific incorporation of biophysical probes into proteins.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 2910-4.
145. Alfonta, L., et al., *Site-specific incorporation of a redox-active amino acid into proteins.* J Am Chem Soc, 2003. **125**(48): p. 14662-3.
146. Neumann, H., et al., *A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation.* Mol Cell, 2009. **36**(1): p. 153-63.
147. Nguyen, D.P., et al., *Genetically encoding N(epsilon)-methyl-L-lysine in recombinant histones.* J Am Chem Soc, 2009. **131**(40): p. 14194-5.
148. Virdee, S., et al., *Traceless and site-specific ubiquitination of recombinant proteins.* J Am Chem Soc, 2011. **133**(28): p. 10708-11.
149. Rovner, A.J., et al., *Recoded organisms engineered to depend on synthetic amino acids.* Nature, 2015. **518**(7537): p. 89-93.
150. Moatsou, D., et al., *Self-Assembly of Temperature-Responsive Protein-Polymer Bioconjugates.* Bioconjug Chem, 2015. **26**(9): p. 1890-9.
151. Anderson, J.C., C.A. Voigt, and A.P. Arkin, *Environmental signal integration by a modular AND gate.* Mol Syst Biol, 2007. **3**: p. 133.
152. Studier, F.W., *Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system.* J Mol Biol, 1991. **219**(1): p. 37-44.

153. Daegelen, P., et al., *Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3).* J Mol Biol, 2009. **394**(4): p. 634-43.
154. Gottesman, S., *Proteases and their targets in Escherichia coli.* Annu Rev Genet, 1996. **30**: p. 465-506.
155. Grodberg, J. and J.J. Dunn, *ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification.* J Bacteriol, 1988. **170**(3): p. 1245-53.
156. el-Mansi, E.M. and W.H. Holms, *Control of carbon flux to acetate excretion during growth of Escherichia coli in batch and continuous cultures.* J Gen Microbiol, 1989. **135**(11): p. 2875-83.
157. Shiloach, J., et al., *Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by Escherichia coli BL21 (lambdaDE3) and Escherichia coli JM109.* Biotechnol Bioeng, 1996. **49**(4): p. 421-8.
158. Altschul, S.F., et al., *Issues in searching molecular sequence databases.* Nat Genet, 1994. **6**(2): p. 119-29.
159. Orosz, A., I. Boros, and P. Venetianer, *Analysis of the complex transcription termination region of the Escherichia coli rrnB gene.* Eur J Biochem, 1991. **201**(3): p. 653-9.
160. Hoffmann, S.A., S.M. Kruse, and K.M. Arndt, *Long-range transcriptional interference in E. coli used to construct a dual positive selection system for genetic switches.* Nucleic Acids Research, 2016.
161. Zhang, X. and F.W. Studier, *Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme 1.* Journal of Molecular Biology, 1997. **269**(1): p. 10-27.
162. Gronenborn, B., *Overproduction of phage lambda repressor under control of the lac promotor of Escherichia coli.* Mol Gen Genet, 1976. **148**(3): p. 243-50.
163. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* J Bacteriol, 1995. **177**(14): p. 4121-30.
164. Skerra, A., *Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli.* Gene, 1994. **151**(1-2): p. 131-5.
165. Salis, H.M., *The ribosome binding site calculator.* Methods Enzymol, 2011. **498**: p. 19-42.
166. Bryant, J.A., et al., *Chromosome position effects on gene expression in Escherichia coli K-12.* Nucleic Acids Res, 2014. **42**(18): p. 11383-92.
167. French, S., *Consequences of replication fork movement through transcription units in vivo.* Science, 1992. **258**(5086): p. 1362-1365.
168. Rocha, E.P.C., *Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes?* Trends in Microbiology, 2002. **10**(9): p. 393-395.
169. Yu, D., et al., *An efficient recombination system for chromosome engineering in Escherichia coli.* Proc Natl Acad Sci U S A, 2000. **97**(11): p. 5978-83.
170. Jeong, K.S., J. Ahn, and A.B. Khodursky, *Spatial patterns of transcriptional activity in the chromosome of Escherichia coli.* Genome Biology, 2004. **5**(11): p. R86.
171. Jiang, Y., et al., *Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system.* Appl Environ Microbiol, 2015. **81**(7): p. 2506-14.
172. Waegeman, H., et al., *Increasing recombinant protein production in Escherichia coli K12 through metabolic engineering.* N Biotechnol, 2013. **30**(2): p. 255-61.
173. Lederberg, J. and E.M. Lederberg, *Replica plating and indirect selection of bacterial mutants.* Journal of Bacteriology, 1952. **63**(3): p. 399.
174. Forster, A.C., et al., *Programming peptidomimetic syntheses by translating genetic codes designed de novo.* Proc Natl Acad Sci U S A, 2003. **100**(11): p. 6353-7.
175. Davis, L. and J.W. Chin, *Designer proteins: applications of genetic code expansion in cell biology.* Nat Rev Mol Cell Biol, 2012. **13**(3): p. 168-82.
176. Liu, C.C. and P.G. Schultz, *Adding new chemistries to the genetic code.* Annu. Rev. Biochem., 2010. **79**(1): p. 413-444.
177. Wu, N., et al., *A genetically encoded photocaged amino acid.* J Am Chem Soc, 2004. **126**(44): p. 14306-7.
178. Summerer, D., et al., *A genetically encoded fluorescent amino acid.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9785-9.

179. Chin, J.W., et al., *Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli.* Proc Natl Acad Sci U S A, 2002. **99**(17): p. 11020-4.
180. Lang, K., et al., *Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction.* Nat Chem, 2012. **4**(4): p. 298-304.
181. Lee, S., et al., *A facile strategy for selective incorporation of phosphoserine into histones.* Angew Chem Int Ed Engl, 2013. **52**(22): p. 5771-5.
182. Tian, F., et al., *A general approach to site-specific antibody drug conjugates.* Proc Natl Acad Sci U S A, 2014. **111**(5): p. 1766-71.
183. Bröcker, M.J., et al., *Recoding the genetic code with selenocysteine.* Angew. Chem. Int. Ed. Engl., 2014. **53**(1): p. 319-323.
184. Zimmerman, E.S., et al., *Production of site-specific antibody–drug conjugates using optimized non-natural amino acids in a cell-free expression system.* Bioconjug. Chem., 2014. **25**(2): p. 351-361.
185. Axup, J.Y., et al., *Synthesis of site-specific antibody-drug conjugates using unnatural amino acids.* Proc. Natl. Acad. Sci. U. S. A., 2012. **109**(40): p. 16101-16106.
186. Cho, H., et al., *Optimized clinical performance of growth hormone with an expanded genetic code.* Proc. Natl. Acad. Sci. U. S. A., 2011. **108**(22): p. 9060-9065.
187. Albayrak, C. and J.R. Swartz, *Direct polymerization of proteins.* ACS Synth. Biol., 2014: p. on line.
188. Hoesl, M.G. and N. Budisa, *Recent advances in genetic code engineering in Escherichia coli.* Curr. Opin. Biotechnol., 2012. **23**(5): p. 751-757.
189. Quast, R.B., et al., *Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis.* FEBS Lett, 2015. **589**(15): p. 1703-12.
190. Wang, L., J. Xie, and P.G. Schultz, *Expanding the genetic code.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 225-49.
191. Wang, L. and P.G. Schultz, *Expanding the genetic code.* Chem Commun (Camb), 2002(1): p. 1-11.
192. Wang, L., et al., *A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.
193. Ambrogelly, A., et al., *Pyrrolysine is not hardwired for cotranslational insertion at UAG codons.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3141-6.
194. Plass, T., et al., *Genetically encoded copper-free click chemistry.* Angew Chem Int Ed Engl, 2011. **50**(17): p. 3878-81.
195. Schrader, J.M., S.J. Chapman, and O.C. Uhlenbeck, *Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis.* J Mol Biol, 2009. **386**(5): p. 1255-64.
196. Wang, K., et al., *Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion.* Nat Biotechnol, 2007. **25**(7): p. 770-7.
197. Fried, S.D., et al., *Ribosome Subunit Stapling for Orthogonal Translation in E. coli.* Angew Chem Int Ed Engl, 2015. **54**(43): p. 12791-4.
198. Lang, K. and J.W. Chin, *Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins.* Chem Rev, 2014. **114**(9): p. 4764-806.
199. Tanrikulu, I.C., et al., *Discovery of Escherichia coli methionyl-tRNA synthetase mutants for efficient labeling of proteins with azidonorleucine in vivo.* Proc. Natl. Acad. Sci. U. S. A., 2009. **106**(36): p. 15285-15290.
200. Umehara, T., et al., *N-Acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo.* FEBS Lett., 2012. **586**(6): p. 729-733.
201. Nehring, S., N. Budisa, and B. Wiltschi, *Performance analysis of orthogonal pairs designed for an expanded eukaryotic genetic code.* PLoS ONE, 2012. **7**(4): p. e31992.
202. Wiltschi, B., et al., *Expanding the genetic code of Saccharomyces cerevisiae with methionine analogues.* Yeast, 2008. **25**(11): p. 775-86.
203. Miyake-Stoner, S.J., et al., *Generating permissive site-specific unnatural aminoacyl-tRNA synthetases.* Biochemistry, 2010. **49**(8): p. 1667-77.

204. Wang, Y.-S., et al., *Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant.* ACS Chem. Biol., 2012. **8**(2): p. 405-415.
205. Young, D.D., et al., *An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity.* Biochemistry, 2011. **50**(11): p. 1894-1900.
206. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool.* Biochim Biophys Acta, 2014. **1844**(6): p. 1059-70.
207. Wang, J., M. Kwiatkowski, and A.C. Forster, *Kinetics of tRNA(Pyl) -mediated amber suppression in Escherichia coli translation reveals unexpected limiting steps and competing reactions.* Biotechnol Bioeng, 2016. **113**(7): p. 1552-9.
208. Johnson, D.B.F., et al., *RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites.* Nat. Chem. Biol., 2011. **7**(11): p. 779-786.
209. Hong, S.H., et al., *Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation.* ACS Synth Biol, 2014. **3**(6): p. 398-409.
210. Hong, S.H., et al., *Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1.* Chembiochem, 2015.
211. Chatterjee, A., et al., *A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli.* Biochemistry, 2013. **52**(10): p. 1828-1837.
212. Loscha, K.V., et al., *Multiple-site labeling of proteins with unnatural amino acids.* Angew. Chem. Int. Ed. Engl., 2012. **51**(9): p. 2243-2246.
213. Mukai, T., et al., *Codon reassignment in the Escherichia coli genetic code.* Nucleic Acids Res., 2010. **38**(22): p. 8188-8195.
214. Ohtake, K., et al., *Efficient decoding of the UAG triplet as a full-fledged sense codon enhances the growth of a prfA-deficient strain of Escherichia coli.* J. Bacteriol., 2012. **194**(10): p. 2606-2613.
215. Chatterjee, A., H. Xiao, and P.G. Schultz, *Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli.* Proc. Natl. Acad. Sci. U. S. A., 2012. **109**(37): p. 14841-14846.
216. Liu, D.R., et al., *Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo.* Proc. Natl. Acad. Sci. U. S. A., 1997. **94**(19): p. 10092-10097.
217. Doi, Y., et al., *Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system.* J. Am. Chem. Soc., 2007. **129**(46): p. 14458-14462.
218. Wang, N., et al., *Fine-tuning interaction between aminoacyl-tRNA synthetase and tRNA for efficient synthesis of proteins containing unnatural amino acids.* ACS Synth Biol, 2015. **4**(3): p. 207-12.
219. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome.* Nature, 2010. **464**(7287): p. 441-4.
220. Barrett, O.P. and J.W. Chin, *Evolved orthogonal ribosome purification for in vitro characterization.* Nucleic Acids Res, 2010. **38**(8): p. 2682-91.
221. Zheng, Y., et al., *Performance of optimized noncanonical amino acid mutagenesis systems in the absence of release factor 1.* Mol Biosyst, 2016. **12**(6): p. 1746-9.
222. Jewett, M.C., et al., *An integrated cell-free metabolic platform for protein production and synthetic biology.* Mol Syst Biol, 2008. **4**: p. 220.
223. Bundy, B.C. and J.R. Swartz, *Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation.* Bioconjugate chemistry, 2010. **21**(2): p. 255-63.
224. de Boer, H.A., L.J. Comstock, and M. Vasser, *The tac promoter: a functional hybrid derived from the trp and lac promoters.* Proc Natl Acad Sci U S A, 1983. **80**(1): p. 21-5.
225. Wang, L. and P.G. Schultz, *A general approach for the generation of orthogonal tRNAs.* Chemistry & biology, 2001. **8**(9): p. 883-90.
226. Wang, L., et al., *Addition of the keto functional group to the genetic code of Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(1): p. 56-61.

227. Melancon, C.E., 3rd and P.G. Schultz, *One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases.* Bioorganic & medicinal chemistry letters, 2009. **19**(14): p. 3845-7.
228. Nissen, P., et al., *Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog.* Science, 1995. **270**(5241): p. 1464-72.
229. Hong, S.H., Y.C. Kwon, and M.C. Jewett, *Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis.* Frontiers in chemistry, 2014. **2**: p. 34.
230. Ikeda, R.A., C.M. Ligman, and S. Warshamana, *T7 promoter contacts essential for promoter activity in vivo.* Nucleic Acids Res, 1992. **20**(10): p. 2517-24.
231. Johnson, D.B.F., et al., *Release factor one is nonessential in Escherichia coli.* ACS Chem. Biol., 2012. **7**(8): p. 1337-1344.
232. Heinemann, I.U., et al., *Enhanced phosphoserine insertion during Escherichia coli protein synthesis via partial UAG codon reassignment and release factor 1 deletion.* FEBS Lett, 2012. **586**(20): p. 3716-22.
233. Hubbell, J.A., *Drug development: Longer-lived proteins.* Nature, 2010. **467**(7319): p. 1051-2.
234. Wals, K. and H. Ovaa, *Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins.* Front Chem, 2014. **2**: p. 15.
235. Romano, N.H., et al., *Protein-engineered biomaterials: nanoscale mimics of the extracellular matrix.* Biochim Biophys Acta, 2011. **1810**(3): p. 339-49.
236. Zhong, Z., et al., *Synthesis and Catalytic Activity of Amino Acids and Metallopeptides with Catalytically Active Metallocyclic Side Chains.* Organometallics, 2012. **31**(21): p. 7328-7331.
237. Cohen, P., *Protein kinases--the major drug targets of the twenty-first century?* Nat Rev Drug Discov, 2002. **1**(4): p. 309-15.
238. Pray, L.A., *Gleevec: the breakthrough in cancer treatment.* Nature Education, 2008. **1**(1): p. 37.
239. Ramos, J.W., *Cancer research center hotline: PEA-15 phosphoprotein: a potential cancer drug target.* Hawaii Med J, 2005. **64**(3): p. 77-80.
240. Voronkov, M., S.P. Braithwaite, and J.B. Stock, *Phosphoprotein phosphatase 2A: a novel druggable target for Alzheimer's disease.* Future Med Chem, 2011. **3**(7): p. 821-33.
241. Krishnamoorthy, S., et al., *A Novel Phosphopeptide Microarray Based Interactome Map in Breast Cancer Cells Reveals Phosphoprotein-GRB2 Cell Signaling Networks.* PLoS One, 2013. **8**(6): p. e67634.
242. Banerjee, T. and D. Chakravarti, *A peek into the complex realm of histone phosphorylation.* Mol Cell Biol, 2011. **31**(24): p. 4858-73.
243. Nirenberg, M.W. and J.H. Matthaei, *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides.* Proc Natl Acad Sci U S A, 1961. **47**: p. 1588-602.
244. Chambers, D.A. and G. Zubay, *The stimulatory effect of cyclic adenosine 3'5'-monophosphate on DNA-directed synthesis of beta-galactosidase in a cell-free system.* Proc Natl Acad Sci U S A, 1969. **63**(1): p. 118-22.
245. Zalkin, H., C. Yanofsky, and C.L. Squires, *Regulated in vitro synthesis of Escherichia coli tryptophan operon messenger ribonucleic acid and enzymes.* J Biol Chem, 1974. **249**(2): p. 465-75.
246. Siegal-Gaskins, D., et al., *Gene circuit performance characterization and resource usage in a cell-free "breadboard".* ACS Synth Biol, 2014. **3**(6): p. 416-25.
247. Takahashi, M.K., et al., *Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems.* ACS Synth Biol, 2014.
248. Chappell, J., M.K. Takahashi, and J.B. Lucks, *Creating small transcription activating RNAs.* Nat Chem Biol, 2015.
249. Caschera, F., et al., *Coping with complexity: machine learning optimization of cell-free protein synthesis.* Biotechnol Bioeng, 2011. **108**(9): p. 2218-28.
250. Hong, S.H., Y.C. Kwon, and M.C. Jewett, *Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis.* Front Chem, 2014. **2**: p. 34.
251. Goerke, A.R. and J.R. Swartz, *High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids.* Biotechnol Bioeng, 2009. **102**(2): p. 400-16.

252. Caschera, F. and V. Noireaux, *Integration of biological parts toward the synthesis of a minimal cell.* Curr Opin Chem Biol, 2014. **22**: p. 85-91.
253. Shin, J. and V. Noireaux, *An E. coli cell-free expression toolbox: application to synthetic gene circuits and artificial cells.* ACS Synth Biol, 2012. **1**(1): p. 29-41.
254. Stano, P. and P.L. Luisi, *Semi-synthetic minimal cells: origin and recent developments.* Curr Opin Biotechnol, 2013. **24**(4): p. 633-8.
255. Yin, G., et al., *Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system.* MAbs, 2012. **4**(2): p. 217-25.
256. Zawada, J.F., et al., *Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines.* Biotechnol Bioeng, 2011. **108**(7): p. 1570-8.
257. Pardee, K., et al., *Paper-based synthetic gene networks.* Cell, 2014. **159**(4): p. 940-54.
258. Chong, S., *Overview of cell-free protein synthesis: historic landmarks, commercial systems, and expanding applications.* Curr Protoc Mol Biol, 2014. **108**: p. 16.30.1-16.30.11.
259. Harbers, M., *Wheat germ systems for cell-free protein expression.* FEBS Lett, 2014. **588**(17): p. 2762-73.
260. Katzen, F., G. Chang, and W. Kudlicki, *The past, present and future of cell-free protein synthesis.* Trends Biotechnol, 2005. **23**(3): p. 150-6.
261. Lian, Q., H. Cao, and F. Wang, *The cost-efficiency realization in the Escherichia coli-based cell-free protein synthesis systems.* Appl Biochem Biotechnol, 2014. **174**(7): p. 2351-67.
262. Shimizu, Y., et al., *The PURE system for protein production.* Methods Mol Biol, 2014. **1118**: p. 275-84.
263. Pratt, J., *Coupled transcription-translation in prokaryotic cell-free systems.* Transcription and translation: a practical approach, 1984: p. 179-209.
264. Shimizu, Y., et al., *Cell-free translation reconstituted with purified components.* Nat Biotechnol, 2001. **19**(8): p. 751-5.
265. Shimizu, Y., T. Kanamori, and T. Ueda, *Protein synthesis by pure translation systems.* Methods, 2005. **36**(3): p. 299-304.
266. Murakami, H., et al., *A highly flexible tRNA acylation method for non-natural polypeptide synthesis.* Nat Methods, 2006. **3**(5): p. 357-9.
267. Kuruma, Y., T. Suzuki, and T. Ueda, *Production of multi-subunit complexes on liposome through an E. coli cell-free expression system.* Methods Mol Biol, 2010. **607**: p. 161-71.
268. Kuruma, Y., et al., *Functional analysis of membranous Fo-a subunit of F1Fo-ATP synthase by in vitro protein synthesis.* Biochem J, 2012. **442**(3): p. 631-8.
269. Matsubayashi, H., Y. Kuruma, and T. Ueda, *In vitro synthesis of the E. coli Sec translocon from DNA.* Angew Chem Int Ed Engl, 2014. **53**(29): p. 7535-8.
270. Matsubayashi, H. and T. Ueda, *Purified cell-free systems as standard parts for synthetic biology.* Curr Opin Chem Biol, 2014. **22**: p. 158-62.
271. Nishimura, K., et al., *Cell-free protein synthesis inside giant unilamellar vesicles analyzed by flow cytometry.* Langmuir, 2012. **28**(22): p. 8426-32.
272. Nishimura, K., et al., *Stochasticity in Gene Expression in a Cell-Sized Compartment.* ACS Synth Biol, 2014.
273. Groff, D., et al., *Engineering toward a bacterial "endoplasmic reticulum" for the rapid expression of immunoglobulin proteins.* MAbs, 2014. **6**(3): p. 671-8.
274. Caschera, F. and V. Noireaux, *Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system.* Biochimie, 2014. **99**: p. 162-8.
275. Swartz, J., *Developing cell-free biology for industrial applications.* J Ind Microbiol Biotechnol, 2006. **33**(7): p. 476-85.
276. Shrestha, P., T.M. Holland, and B.C. Bundy, *Streamlined extract preparation for Escherichia coli-based cell-free protein synthesis by sonication or bead vortex mixing.* Biotechniques, 2012. **53**(3): p. 163-74.
277. Calhoun, K.A. and J.R. Swartz, *An economical method for cell-free protein synthesis using glucose and nucleoside monophosphates.* Biotechnol Prog, 2005. **21**(4): p. 1146-53.

278. Calhoun, K.A. and J.R. Swartz, *Energizing cell-free protein synthesis with glucose metabolism.* Biotechnol Bioeng, 2005. **90**(5): p. 606-13.
279. Albayrak, C. and J.R. Swartz, *Using E. coli-based cell-free protein synthesis to evaluate the kinetic performance of an orthogonal tRNA and aminoacyl-tRNA synthetase pair.* Biochem Biophys Res Commun, 2013. **431**(2): p. 291-5.
280. Ozawa, K., et al., *High-yield cell-free protein synthesis for site-specific incorporation of unnatural amino acids at two sites.* Biochem Biophys Res Commun, 2012. **418**(4): p. 652-6.
281. Stafford, R.L., et al., *In vitro Fab display: a cell-free system for IgG discovery.* Protein Eng Des Sel, 2014. **27**(4): p. 97-109.
282. Zimmerman, E.S., et al., *Production of site-specific antibody-drug conjugates using optimized non-natural amino acids in a cell-free expression system.* Bioconjug Chem, 2014. **25**(2): p. 351-61.
283. Iizuka, N., et al., *Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae.* Mol Cell Biol, 1994. **14**(11): p. 7322-30.
284. Hodgman, C.E. and M.C. Jewett, *Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis.* Biotechnol Bioeng, 2013. **110**(10): p. 2643-54.
285. Choudhury, A., et al., *Evaluating fermentation effects on cell growth and crude extract metabolic activity for improved yeast cell-free protein synthesis.* Biochemical Engineering Journal, 2014. **91**: p. 140-148.
286. Gan, R. and M.C. Jewett, *A combined cell-free transcription-translation system from Saccharomyces cerevisiae for rapid and robust protein synthe.* Biotechnol J, 2014. **9**(5): p. 641-51.
287. Hodgman, C.E. and M.C. Jewett, *Characterizing IGR IRES-mediated translation initiation for use in yeast cell-free protein synthesis.* N Biotechnol, 2014. **31**(5): p. 499-505.
288. Schoborg, J.A., et al., *Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis.* Biotechnol J, 2014. **9**(5): p. 630-40.
289. Madin, K., et al., *A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes.* Proc Natl Acad Sci U S A, 2000. **97**(2): p. 559-64.
290. Endo, Y. and T. Sawasaki, *Cell-free expression systems for eukaryotic protein production.* Curr Opin Biotechnol, 2006. **17**(4): p. 373-80.
291. Sawasaki, T., et al., *A cell-free protein synthesis system for high-throughput proteomics.* Proc Natl Acad Sci U S A, 2002. **99**(23): p. 14652-7.
292. Goshima, N., et al., *Human protein factory for converting the transcriptome into an in vitro-expressed proteome.* Nat Methods, 2008. **5**(12): p. 1011-7.
293. Nozawa, A., et al., *Construction of a protein library of Arabidopsis transcription factors using a wheat cell-free protein production system and its application for DNA binding analysis.* Biosci Biotechnol Biochem, 2009. **73**(7): p. 1661-4.
294. Tsuboi, T., et al., *Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates.* Infect Immun, 2008. **76**(4): p. 1702-8.
295. Buntru, M., et al., *A versatile coupled cell-free transcription-translation system based on tobacco BY-2 cell lysates.* Biotechnol Bioeng, 2014.
296. Ezure, T., et al., *Cell-free protein synthesis system prepared from insect cells by freeze-thawing.* Biotechnol Prog, 2006. **22**(6): p. 1570-7.
297. Katzen, F. and W. Kudlicki, *Efficient generation of insect-based cell-free translation extracts active in glycosylation and signal sequence processing.* J Biotechnol, 2006. **125**(2): p. 194-7.
298. Kubick, S., et al., *In vitro synthesis of posttranslationally modified membrane proteins.* Current Topics in Membranes, 2009. **63**(2): p. 25-49.
299. Sachse, R., et al., *Membrane protein synthesis in cell-free systems: from bio-mimetic systems to bio-membranes.* FEBS Lett, 2014. **588**(17): p. 2774-81.
300. Stech, M., et al., *A continuous-exchange cell-free protein synthesis system based on extracts from cultured insect cells.* PLoS One, 2014. **9**(5): p. e96635.
301. Mikami, S., et al., *An efficient mammalian cell-free translation system supplemented with translation factors.* Protein Expr Purif, 2006. **46**(2): p. 348-57.

302. Brodel, A.K., D.A. Wustenhagen, and S. Kubick, *Cell-free protein synthesis systems derived from cultured Mammalian cells.* Methods Mol Biol, 2015. **1261**: p. 129-40.
303. Mureev, S., et al., *Species-independent translational leaders facilitate cell-free expression.* Nat Biotechnol, 2009. **27**(8): p. 747-52.
304. Kovtun, O., et al., *Towards the construction of expressed proteomes using a Leishmania tarentolae based cell-free expression system.* PLoS One, 2010. **5**(12): p. e14388.
305. Johnston, W.A. and K. Alexandrov, *Production of eukaryotic cell-free lysate from Leishmania tarentolae.* Methods Mol Biol, 2014. **1118**: p. 1-15.
306. Kovtun, O., et al., *Leishmania cell-free protein expression system.* Methods, 2011. **55**(1): p. 58-64.
307. Xu, Y., et al. *Production of bispecific antibodies in "Knobs-into-Holes" using a cell-free expression system*. in *mAbs*. 2014. Taylor & Francis.
308. Kawasaki, T., et al., *Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ.* Eur J Biochem, 2003. **270**(23): p. 4780-6.
309. Stech, M., et al., *Production of functional antibody fragments in a vesicle-based eukaryotic cell-free translation system.* J Biotechnol, 2012. **164**(2): p. 220-31.
310. Brodel, A.K., et al., *IRES-mediated translation of membrane proteins and glycoproteins in eukaryotic cell-free systems.* PLoS One, 2013. **8**(12): p. e82234.
311. Ng, P.P., et al., *A vaccine directed to B cells and produced by cell-free protein synthesis generates potent antilymphoma immunity.* Proc Natl Acad Sci U S A, 2012. **109**(36): p. 14526-31.
312. Lu, Y., J.P. Welsh, and J.R. Swartz, *Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines.* Proc Natl Acad Sci U S A, 2014. **111**(1): p. 125-30.
313. Bundy, B.C., M.J. Franciszkowicz, and J.R. Swartz, *Escherichia coli-based cell-free synthesis of virus-like particles.* Biotechnol Bioeng, 2008. **100**(1): p. 28-37.
314. Junge, F., et al., *Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins.* N Biotechnol, 2011. **28**(3): p. 262-71.
315. Wang, X., et al., *Peptide surfactants for cell-free production of functional G protein-coupled receptors.* Proc Natl Acad Sci U S A, 2011. **108**(22): p. 9049-54.
316. Kaiser, L., et al., *Efficient cell-free production of olfactory receptors: detergent optimization, structure, and ligand binding analyses.* Proc Natl Acad Sci U S A, 2008. **105**(41): p. 15726-31.
317. Bernhard, F. and Y. Tozawa, *Cell-free expression--making a mark.* Curr Opin Struct Biol, 2013. **23**(3): p. 374-80.
318. Selkrig, J., et al., *Assembly of β-barrel proteins into bacterial outer membranes.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2014. **1843**(8): p. 1542-1550.
319. Du, D., H.W. van Veen, and B.F. Luisi, *Assembly and operation of bacterial tripartite multidrug efflux pumps.* Trends in Microbiology, (0).
320. Asahara, H. and S. Chong, *In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme.* Nucleic Acids Res, 2010. **38**(13): p. e141.
321. Fujiwara, K., T. Katayama, and S.M. Nomura, *Cooperative working of bacterial chromosome replication proteins generated by a reconstituted protein expression system.* Nucleic Acids Res, 2013. **41**(14): p. 7176-83.
322. Shin, J., P. Jardine, and V. Noireaux, *Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction.* ACS Synth Biol, 2012. **1**(9): p. 408-13.
323. Chan, L.Y., S. Kosuri, and D. Endy, *Refactoring bacteriophage T7.* Mol Syst Biol, 2005. **1**: p. 2005.0018.
324. Jewett, M.C., et al., *In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation.* Mol Syst Biol, 2013. **9**: p. 678.
325. Fritz, B.R. and M.C. Jewett, *The impact of transcriptional tuning on in vitro integrated rRNA transcription and ribosome construction.* Nucleic Acids Res, 2014. **42**(10): p. 6774-85.
326. Liu, Y., et al., *Characterizing and Alleviating Substrate Limitations for Improved in vitro Ribosome Construction.* ACS Synth Biol, 2014.

327. Heyman, Y., et al., *Cell-free protein synthesis and assembly on a biochip.* Nat Nanotechnol, 2012. **7**(6): p. 374-8.
328. Gerber, D., S.J. Maerkl, and S.R. Quake, *An in vitro microfluidic approach to generating protein-interaction networks.* Nat Methods, 2009. **6**(1): p. 71-4.
329. Ezure, T., et al., *A cell-free translocation system using extracts of cultured insect cells to yield functional membrane proteins.* PLoS One, 2014. **9**(12): p. e112874.
330. Lemke, E.A., *Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities.* Methods Mol Biol, 2011. **751**: p. 3-15.
331. Nishikawa, T., et al., *Directed Evolution of Proteins through In Vitro Protein Synthesis in Liposomes.* J Nucleic Acids, 2012. **2012**: p. 923214.
332. Loscha, K.V., et al., *Multiple-Site Labeling of Proteins with Unnatural Amino Acids.* Angewandte Chemie International Edition, 2012. **51**(9): p. 2243-2246.
333. Oh, S.-J., et al., *Translational incorporation of multiple unnatural amino acids in a cell-free protein synthesis system.* Biotechnology and Bioprocess Engineering, 2014. **19**(3): p. 426-432.
334. Goto, Y., T. Katoh, and H. Suga, *Flexizymes for genetic code reprogramming.* Nat Protoc, 2011. **6**(6): p. 779-90.
335. Passioura, T. and H. Suga, *Flexizyme-mediated genetic reprogramming as a tool for noncanonical peptide synthesis and drug discovery.* Chemistry, 2013. **19**(21): p. 6530-6.
336. Li, H. and M. d'Anjou, *Pharmacological significance of glycosylation in therapeutic proteins.* Curr Opin Biotechnol, 2009. **20**(6): p. 678-84.
337. Fisher, A.C., et al., *Production of secretory and extracellular N-linked glycoproteins in Escherichia coli.* Appl Environ Microbiol, 2011. **77**(3): p. 871-81.
338. Meuris, L., et al., *GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins.* Nat Biotechnol, 2014. **32**(5): p. 485-9.
339. Valderrama-Rincon, J.D., et al., *An engineered eukaryotic protein glycosylation pathway in Escherichia coli.* Nat Chem Biol, 2012. **8**(5): p. 434-6.
340. Hamilton, S.R., et al., *Humanization of yeast to produce complex terminally sialylated glycoproteins.* Science, 2006. **313**(5792): p. 1441-3.
341. Kunamneni, A., et al., *Laccases and their applications: a patent review.* Recent Pat Biotechnol, 2008. **2**(1): p. 10-24.
342. Lawton, T.J. and A.C. Rosenzweig, *Detection and characterization of a multicopper oxidase from Nitrosomonas europaea.* Methods Enzymol, 2011. **496**: p. 423-33.
343. Kuchenreuther, J.M., S.A. Shiigi, and J.R. Swartz, *Cell-free synthesis of the H-cluster: a model for the in vitro assembly of metalloprotein metal centers.* Methods Mol Biol, 2014. **1122**: p. 49-72.
344. Boyer, M.E., C.W. Wang, and J.R. Swartz, *Simultaneous expression and maturation of the iron-sulfur protein ferredoxin in a cell-free system.* Biotechnol Bioeng, 2006. **94**(1): p. 128-38.
345. Brophy, J.A.N. and C.A. Voigt, *Principles of genetic circuit design.* Nat Meth, 2014. **11**(5): p. 508-520.
346. Gardner, T.S., C.R. Cantor, and J.J. Collins, *Construction of a genetic toggle switch in Escherichia coli.* Nature, 2000. **403**(6767): p. 339-342.
347. Ham, T.S., et al., *Design and construction of a double inversion recombination switch for heritable sequential genetic memory.* PLoS One, 2008. **3**(7): p. e2815.
348. Ham, T.S., et al., *A tightly regulated inducible expression system utilizing the fim inversion recombination switch.* Biotechnol Bioeng, 2006. **94**(1): p. 1-4.
349. Bayer, T.S. and C.D. Smolke, *Programmable ligand-controlled riboregulators of eukaryotic gene expression.* Nat Biotechnol, 2005. **23**(3): p. 337-43.
350. Win, M.N. and C.D. Smolke, *Higher-order cellular information processing with synthetic RNA devices.* Science, 2008. **322**(5900): p. 456-60.
351. Rackham, O. and J.W. Chin, *Cellular logic with orthogonal ribosomes.* J Am Chem Soc, 2005. **127**(50): p. 17584-5.
352. Friedland, A.E., et al., *Synthetic gene networks that count.* Science, 2009. **324**(5931): p. 1199-202.

353. Gilbert, E.S., A.W. Walker, and J.D. Keasling, *A constructed microbial consortium for biodegradation of the organophosphorus insecticide parathion.* Appl Microbiol Biotechnol, 2003. **61**(1): p. 77-81.
354. Rajendran, M. and A.D. Ellington, *Selection of fluorescent aptamer beacons that light up in the presence of zinc.* Anal Bioanal Chem, 2008. **390**(4): p. 1067-75.
355. Inokuma, K., et al., *Improvement of isopropanol production by metabolically engineered Escherichia coli using gas stripping.* J Biosci Bioeng, 2010. **110**(6): p. 696-701.
356. Noireaux, V., R. Bar-Ziv, and A. Libchaber, *Principles of cell-free genetic circuit assembly.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12672-7.
357. Kim, J., K.S. White, and E. Winfree, *Construction of an in vitro bistable circuit from synthetic transcriptional switches.* Mol Syst Biol, 2006. **2**: p. 68.
358. Elowitz, M.B. and S. Leibler, *A synthetic oscillatory network of transcriptional regulators.* Nature, 2000. **403**(6767): p. 335-8.
359. Selinger, D.W., et al., *Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation.* Genome Research, 2003. **13**(2): p. 216-223.
360. Shin, J. and V. Noireaux, *Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system.* J Biol Eng, 2010. **4**: p. 9.
361. Selinger, D.W., et al., *Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation.* Genome Res, 2003. **13**(2): p. 216-23.
362. Karig, D.K., et al., *Expression optimization and synthetic gene networks in cell-free systems.* Nucleic Acids Res, 2012. **40**(8): p. 3763-74.
363. Iyer, S., et al., *Multi-input regulation and logic with T7 promoters in cells and cell-free systems.* PLoS One, 2013. **8**(10): p. e78442.
364. Iyer, S. and M.J. Doktycz, *Thrombin-mediated transcriptional regulation using DNA aptamers in DNA-based cell-free protein synthesis.* ACS Synth Biol, 2014. **3**(6): p. 340-6.
365. Chappell, J., M.K. Takahashi, and J.B. Lucks, *Creating small transcription activating RNAs.* Nat Chem Biol, 2015. **11**(3): p. 214-20.
366. Niederholtmeyer, H., L. Xu, and S.J. Maerkl, *Real-time mRNA measurement during an in vitro transcription and translation reaction using binary probes.* ACS Synth Biol, 2013. **2**(8): p. 411-7.
367. Chizzolini, F., et al., *Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength.* ACS Synth Biol, 2014. **3**(6): p. 363-71.
368. Rendl, M., T. Brandstetter, and J. Ruhe, *Time-resolved analysis of biological reactions based on heterogeneous assays in liquid plugs of nanoliter volume.* Anal Chem, 2013. **85**(20): p. 9469-77.
369. Litcofsky, K.D., et al., *Iterative plug-and-play methodology for constructing and modifying synthetic gene networks.* Nat Methods, 2012. **9**(11): p. 1077-80.
370. Sleight, S.C. and H.M. Sauro, *Randomized BioBrick assembly: a novel DNA assembly method for randomizing and optimizing genetic circuits and metabolic pathways.* ACS Synth Biol, 2013. **2**(9): p. 506-18.
371. Sun, Z.Z., et al., *Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system.* ACS Synth Biol, 2014. **3**(6): p. 387-97.
372. Martini, L. and S.S. Mansy, *Measuring riboswitch activity in vitro and in artificial cells with purified transcription-translation machinery.* Methods Mol Biol, 2014. **1111**: p. 153-64.
373. Chappell, J., K. Jensen, and P.S. Freemont, *Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology.* Nucleic Acids Res, 2013. **41**(5): p. 3471-81.
374. Lee, K.H., et al., *Real-time monitoring of cell-free protein synthesis on a surface plasmon resonance chip.* Anal Biochem, 2007. **366**(2): p. 170-4.
375. Karzbrun, E., et al., *Coarse-grained dynamics of protein synthesis in a cell-free system.* Phys Rev Lett, 2011. **106**(4): p. 048104.
376. Venter, J.C., et al., *Environmental genome shotgun sequencing of the Sargasso Sea.* science, 2004. **304**(5667): p. 66-74.
377. Takemori, N., et al., *High-throughput synthesis of stable isotope-labeled transmembrane proteins for targeted transmembrane proteomics using a wheat germ cell-free protein synthesis system.* Mol Biosyst, 2015. **11**(2): p. 361-5.

378. Quast, R.B., et al., *Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis.* FEBS Letters, 2015. **589**(15): p. 1703-1712.

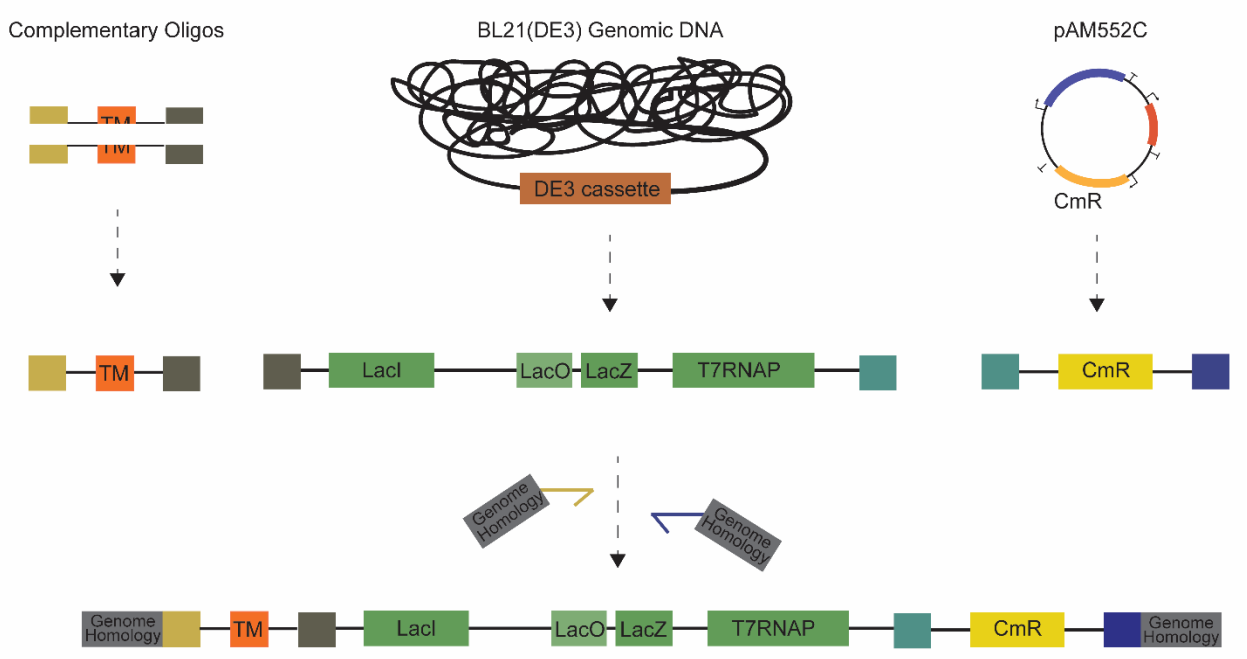
APPENDIX

A. Supplemental Information

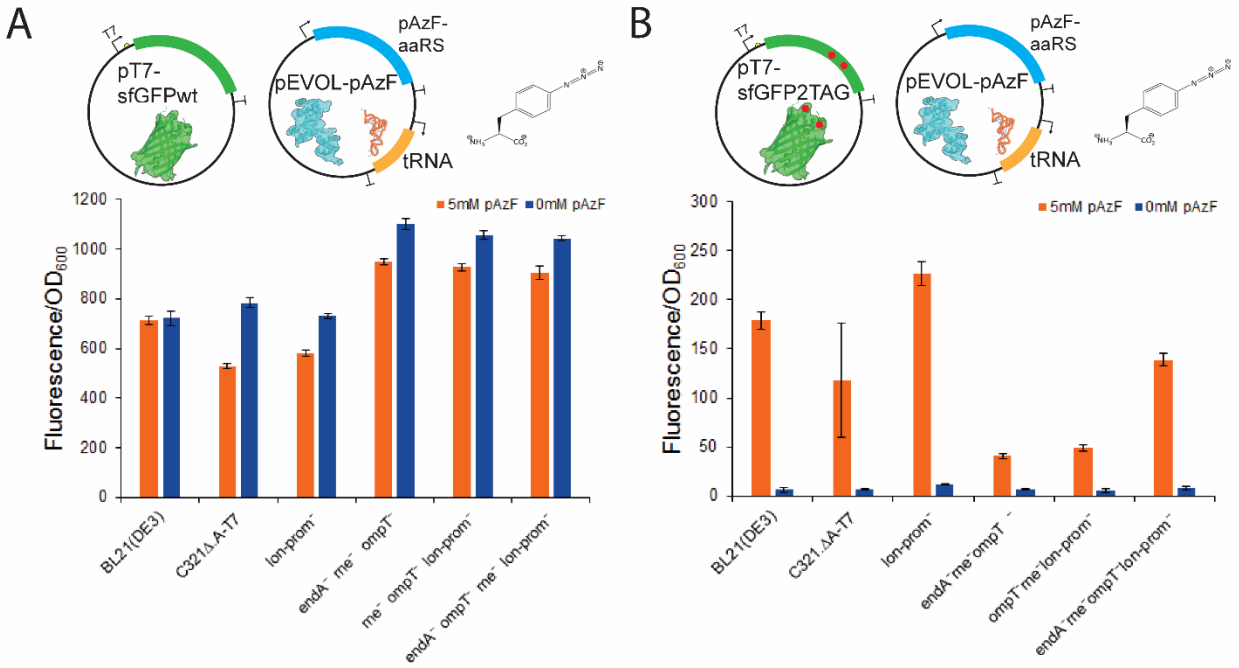
A.1 Supplemental Information to Chapter 3

Supplemental Table A1: Quantification of *in vivo* protein concentrations of top mutant strains.

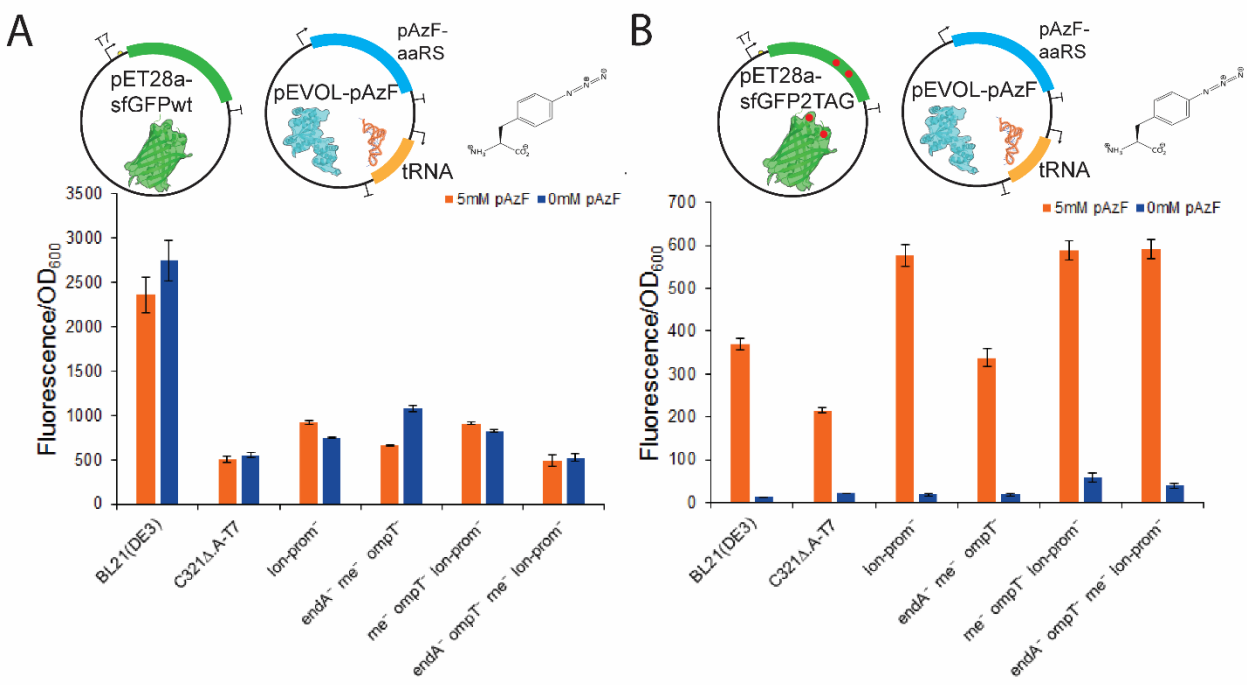
Strain	Reporter Plasmid	OTS Plasmid	µg/mL
C321.ΔA	pLpp5-sfGFP-wt	pEVOLpAzF	389.4 ± 72
endA⁻rne⁻ompT⁻	pLpp5-sfGFP-wt	pEVOLpAzF	757.4 ± 137
C321.ΔA	pLpp5-sfGFP-2TAG	pEVOLpAzF	18 ± 0.5
endA⁻rne⁻	pLpp5-sfGFP-2TAG	pEVOLpAzF	67 ± 3.8



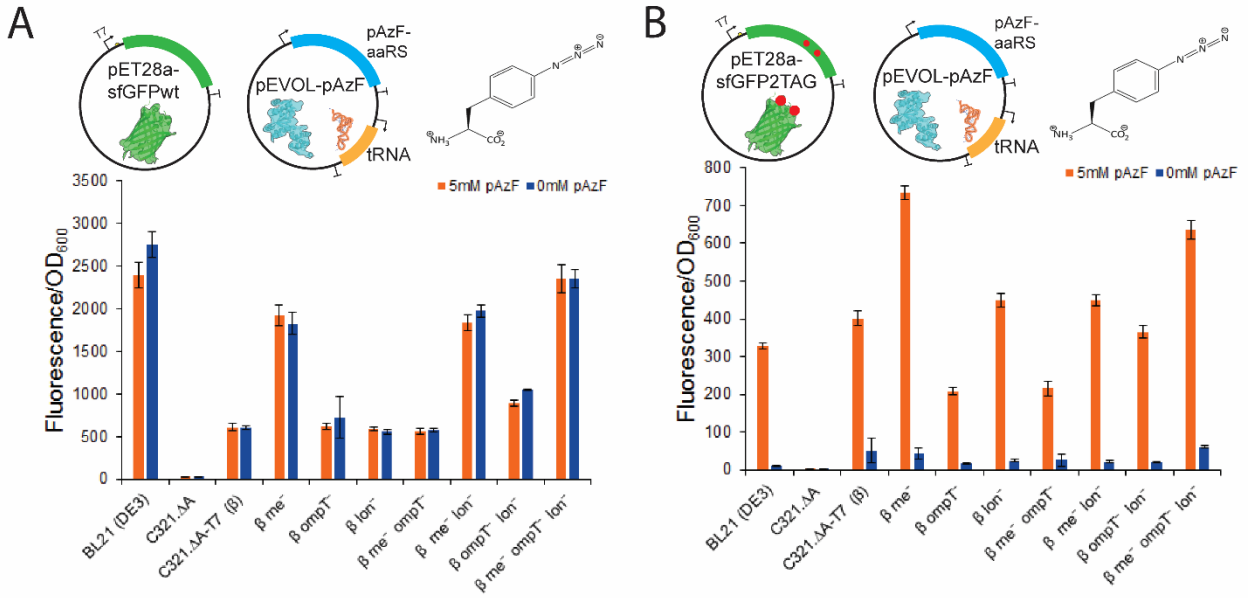
Supplemental Figure A1: Construction of T7RNAP cassette. The T7RNAP cassette consists of three parts: 5' Terminator (TM), T7 RNA Polymerase (T7RNAP) and antibiotic resistance marker CmR. These parts were amplified from various sources and were stitched together via overlap PCR. During the stitching process homology to the genomic insertion site was added to the 5' and 3' end of the cassette. See Chapter 3.5 for more details.



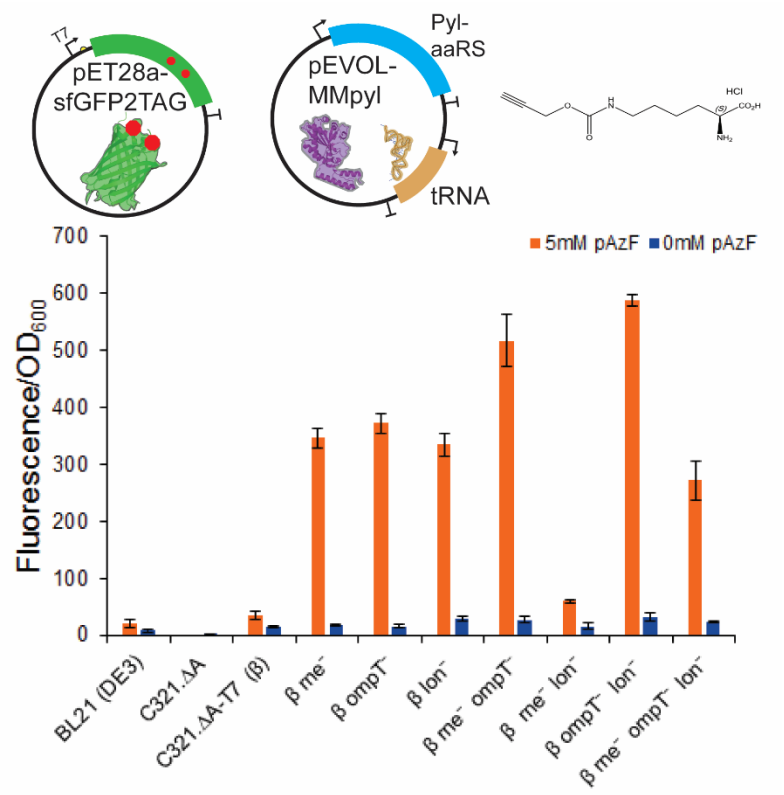
Supplemental Figure A2: Expression of sfGFP in top *C321.ΔA* mutants containing genomically expressed T7RNAP cassette. A) T7RNAP function was tested in the top *C321.ΔA* mutants by expressing wild-type sfGFP (sfGFP-wt) with the pAzF orthogonal translation system expressed on pEVOL-pAzF. For all conditions 1 mM IPTG, 0.02% arabinose and 5 mM pAzF (orange bars) or 0mM pAzF (blue bars) was added at OD_{600} 0.6-0.8. **B)** Modified *C321.ΔA-T7* strains were analyzed for the ability to suppress two amber codons at positions 190 and 212 in the presence (orange) or absence (blue) of 5 mM pAzF. For all panels error bars represent one standard deviation for biological triplicates and technical triplicates.



Supplemental Figure A3: Expression of sfGFP in *C321.ΔA-T7* strains using a pET reporter plasmid. A) SfGFP-wt and the pAzF orthogonal translation system was expressed using a pET plasmid and pEVOL-pAzF, respectively. For all conditions 1 mM IPTG, 0.02% arabinose and 5 mM pAzF (orange bars) or 0 mM pAzF (blue bars) were added at OD_{600} 0.6-0.8. **B)** Expression of sfGFP with two amber codons at positions 190 and 212 in the presence (orange) or absence (blue) of 5 mM pAzF was tested with pEVOL-pAzF. For all panels error bars represent one standard deviation for biological triplicates and technical triplicates.



Supplemental Figure A4: Expression of sfGFP in *C321.ΔA-T7* strains produced by first inserting T7RNAP, followed by the introduction of protein and nuclease mutations. A) Expression of sfGFP-wt was performed using pET28a-sfGFP-wt and pEVOLpAzF. For all conditions 1 mM IPTG, 0.02% arabinose and 5 mM pAzF (orange bars) or 0 mM pAzF (blue bars) were added at OD_{600} 0.6-0.8. **B)** Expression of sfGFP with two amber codons at positions 190 and 212 in the presence (orange) or absence (blue) of 5 mM pAzF was tested with pEVOL-pAzF. For all panels error bars represent one standard deviation for biological triplicates and technical triplicates.



Supplemental Figure A5: Expression of sfGFP-2TAG utilizing the pyrrolysine orthogonal translation system in β strains. Expression of sfGFP with two amber codons at positions 190 and 212 in the presence (orange) or absence (blue) of 5 mM ProCarb was tested with pEVOL-MMpyl. Error bars represent one standard deviation for biological triplicates and technical triplicates.

Supplemental Table A2: Oligonucleotides used in this study.

MAGE Primers	Nucleotide Sequence
JGP290_endA-MAGE	C*G*G*T*AAAAGTCCACGCTGACGCGCCCGGTACGTTTTATTGCTAACTGAAAAATTAACTGGCAGGG CAAAAAAGGCGTTGTTGATCTGCA
JGP287_rne131-MAGE	C*T*G*T*TGAGCCGCTTCTTCGGCGCACTGAAAGCGCTGTTCAGCTAACTGAGAAGAAACCAAACCGA CCGAGCAACCAGCACCGAAAGCAGA
JGP286_rnb-MAGE	A*T*T*T*TGTCACCATCGACAGTGCCAGCACAGAAGATATGGATTAACTGACTTTTCGCTAAGGCGTTG CCGGATGACAAACTTCAGCTGAT
JGP368_lonprom-MAGE	C*A*G*A*TGACACGACTGTGCTTCACGCCATCTATTAACATGTACGTCAGATAGAGGAAAAATTAAAG GGGAGATAAAATCCCCCCTTTTTGG
ompT-MAGE	T*G*G*A*CAACTCTCGGCAGCCGAGGTGGCAATATGGTCGCGCAGGACTGGATGGATTCCAGTAACC CCGGAACCTGGACGGATGAAAGTAGA

MASC Sequencing Primers	Nucleotide Sequence
JGP362_endA-wt-f	CCCGGTACGTTTTATTGCGGATGT
JGP363_endA-mut-f	CCCGGTACGTTTTATTGCTAACTGA
JGP364_end-r	GCTGGCGCTGGTAATTTCGGCGTCA
end-seq-f	ATGTACCGTTATTTGTCTATTGCTGC
JGP346_rne-wt-f	CACTGAAAGCGCTGTTCAGCGGTGGT
JGP347_rne-mut-f	CACTGAAAGCGCTGTTCAGCTAACTGA
JGP348_rne-r	GTGCGACTACCGCTTCTTCGGCTAC
JGP349_rne-seq	CAGATGGAAACCCCGCACTACCACG
JGP359_rnb-wt-f	CCAGCACAGAAGATATGGATGACGCC
JGP360_rnb-mut-f	CCAGCACAGAAGATATGGATTAACTGA
JGP361_rnb-r	TCACTTTCAGGCTGCCAGTCACCGG
rnb-seq	CTGAAAGGCGATCGTTCTTTCTATG
JGP370_lon-wt-f	CTTTAATTTTTCCTCTATTCTCG
JGP371_lon-f	GATTGCAGTACGCACCAGC
JGP369_lon-seq	CTGCGTTCCATCGTAGAAGC
JGP365_ompT-wt-f	CAGCCGAGGTGGCAATATGGTCGAT
JGP366_ompT-mut-f	CAGCCGAGGTGGCAATATGGTCGCG
JGP367_ompT-r	GAGTTCAAAATCTTCATAACGATAAC
JGP345_ompT-seq-f	CTGACAACCCCTATTGCGATCAGCTC

T7RNAP cassette construction	Nucleotide Sequence
JGP221_L3S2P21_sense	GAT TTT CAG CCT GAT ACA GCT CGG TAC CAA ATT CCA GAA AAG AGG CCT CCC GAA AGG GGG GCC TTT TTT CGT TTT GGT CCC CTT TTT GCG TTT CTA CA
JGP221_L3S2P21_antisense	TGTAGAAACGCAAAAAGGGGACCAAAACGAAAAAAGGCCCCCCTTTCGGGAGGCCTCTTTTCTGGAA TTTGGTACCGAGCTGTATCAGGCTGAAAATC
EDC408_DE3-0-f	GCTTTTTTATACTAAGTTGGCATTATAAAAAAG
EDC323_iT7396-r	GACTTACGGCTGACGAATACCTG
EDC413_cat-f	CAGGTATTCGTCAGCCGTAAGTCAACTCTTCCTGTCGTC
EDC414_cat-1-r	GAGATTATCAAAAAGGATCTTCACC

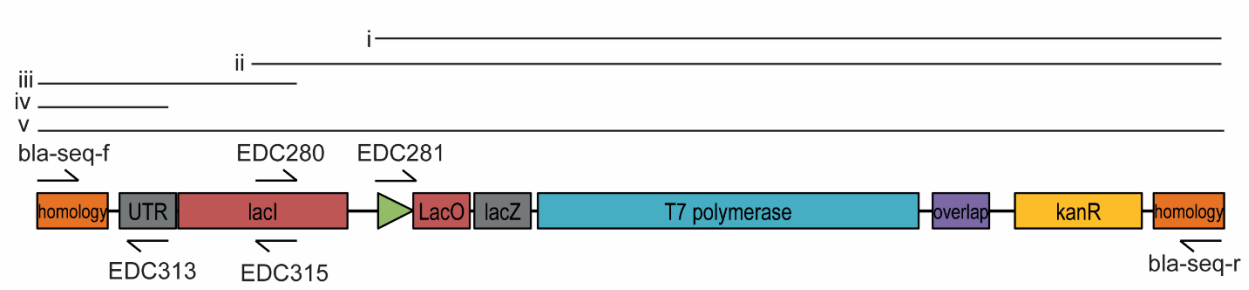
JGP146_Loc9.8b_T1T2DE3cat_f	G*A*TTGCAGGAAAACTCGGTTAACGGAGTGATCGAGTTAACATTGTGATTTTCAGCCTGATACAG
JGP147_Loc9.8b_T1T2DE3cat_r	*A*CACAGCAACCATGTAAATCGGAGTTGAAACCAATATTTAACTTAGAGATTATCAAAAAGGATCTTCACC
JGP173_Hom8.9bseq_f3	GAACTTAATGCAACACACCA
JGP292_T7RNAPr_seq	CCGCGCTTAGCTTTCACT
EDC280_T7seq1	CCATCAACAGTATTATTTTCTCC
EDC282_T7seq3	GGTGAAGCACGCTTCC
JGP153_Hom8.9bseq_r	CGTGTGCGTGTCCATAGA
JGP389_T7cassCmR_ko_MAGE	G*T*C*T*GCAGAAAAAAAGCCCGCTTATTAGGCGGGCTAGGAGTGAGTGTGTGTTTAAGTTCCACACATTATTCGAGCCGGATGAGTAATTGT
EDC413_cat-f	CAGGTATTCGTCAGCCGTAAGTCAACTCTTCCTGTCGTC
JGP211_Hom8.9bscreen_r1	GCACCATGAAATCCTGCA

Supplemental Table A3: T7RNAP cassette parts

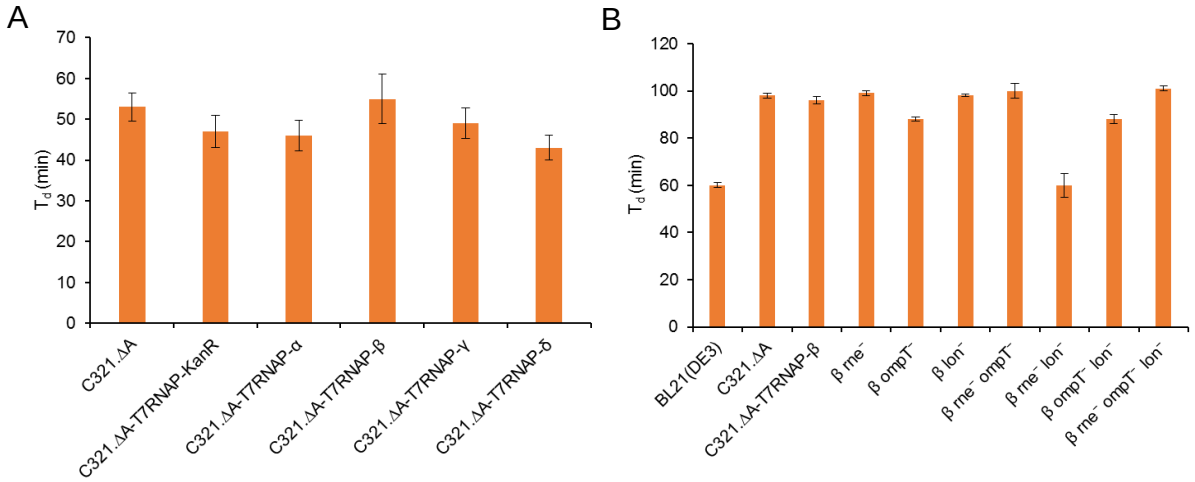
T7RNAP parts	DNA Sequence
TM piece	CGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACA
T7RNAP piece	CCTTTTTGCGTTTCTACAGCTTTTTTATACTAAGTTGGCATTATAAAAAAGCATTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTGATTTCAATTTTGTCCCACTCCCTGCCTCTGTCATCACGATACTGTGATGCCATGGTGTCCGACTTATGCCCGAGAAGATGTTGAGCAAACTTATCGCTTATCTGCTTCTCATAGAGTCTTGCAGACAAACTGCGCAACTCGTGAAAGGTAGGCGGATCCAGATCCCGGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTCGGGATCTGATCCGGATTTACTAACTGGAAGAGGCACTAAATGAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAACTGGCTGCTATCCCGTTCAACACTCTGGCTGACCATTACGGTGAGCGTTTAGCTCGCGAACAGTTGGCCCTTGAGCATGAGTCTTACGAGATGGGTGAAGCACGCTTCCGCAAGATGTTTGAGCGTCAACTT

	AAAGCTGGTGAGGTTGCGGATAACGCTGCCGCCAAGCCTCTCATCACTACCCTACTCCCTAAGAT GATTGCACGCATCAACGACTGGTTTGAGGAAGTGAAAGCTAAGCGCGGCAAGCGCCCGACAGCC TTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCCGTAGCGTACATCACCATTAAGACCACTCTGGCT TGCCTAACCAGTGCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTCGGGCCATTGA GGACGAGGCTCGCTTCGGTCGTATCCGTGACCTTGAAGCTAAGCACTTCAAGAAAAACGTTGAGG AACAACTCAACAAGCGCGTAGGGCACGTCTACAAGAAAGCATTTATGCAAGTTGTCGAGGCTGAC ATGCTCTCTAAGGGTCTACTCGGTGGCGAGGCGTGGTCTTCGTGGCATAAGGAAGACTCTATTCA TGTAGGAGTACGCTGCATCGAGATGCTCATTGAGTCAACCGGAATGGTTAGCTTACACCGCCAAA ATGCTGGCGTAGTAGGTCAAGACTCTGAGACTATCGAACTCGCACCTGAATACGCTGAGGCTATC GCAACCCGTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCTTGCGTAGTTCCTCCTAA GCCGTGGACTGGCATTACTGGTGGTGGCTATTGGGCTAACGGTCGTCGTCCTCTGGCGCTGGTG CGTACTCACAGTAAGAAAGCACTGATGCGCTACGAAGACGTTTACATGCCTGAGGTGTACAAAGC GATTAACATTGCGCAAAACACCGCATGGAAAATCAACAAGAAAGTCCTAGCGGTCGCCAACGTAAT CACCAAGTGGAAGCATTGTCCGGTCGAGGACATCCCTGCGATTGAGCGTGAAGAACTCCCGATGA AACCGGAAGACATCGACATGAATCCTGAGGCTCTCACCGCGTGGAAACGTGCTGCCGCTGCTGTG TACCGCAAGGACAAGGCTCGCAAGTCTCGCCGTATCAGCCTTGAGTTCATGCTTGAGCAAGCCAA TAAGTTTGCTAACCATAAGGCCATCTGGTTCCCTTACAACATGGACTGGCGCGGTCGTGTTTACGC TGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAAGGACTGCTTACGCTGGCGAAAGGTA AACCAATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACGGTGCAAACTGTGCGGGTGTCGAT AAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAAAACCACGAGAACATCATGGCTTGCGC TAAGTCTCCACTGGAGAACACTTGGTGGGCTGAGCAAGATTCTCCGTTCTGCTTCCTTGCGTTCTG CTTTGAGTACGCTGGGGTACAGCACCACGGCCTGAGCTATAACTGCTCCCTTCCGCTGGCGTTTG ACGGGTCTTGCTCTGGCATCCAGCACTTCTCCGCGATGCTCCGAGATGAGGTAGGTGGTCGCGC GGTTAACTTGCTTCCTAGTGAAACCGTTCAGGACATCTACGGGATTGTTGCTAAGAAAGTCAACGA GATTCTACAAGCAGACGCAATCAATGGGACCGATAACGAAGTAGTTACCGTGACCGATGAGAACA CTGGTGAAATCTCTGAGAAAGTCAAGCTGGGCACTAAGGCACTGGCTGGTCAATGGCTGGCTTAC GGTGTTACTCGCAGTGTGACTAAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAGTTCGG CTTCCGTCAACAAGTGCTGGAAGATACCATTCAGCCAGCTATTGATTCCGGCAAGGGTCTGATGTT CACTCAGCCGAATCAGGCTGCTGGATACATGGCTAAGCTGATTTGGGAATCTGTGAGCGTGACGG TGGTAGCTGCGGTTGAAGCAATGAACTGGCTTAAGTCTGCTGCTAAGCTGCTGGCTGCTGAGGTC AAAGATAAGAAGACTGGAGAGATTCTTCGCAAGCGTTGCGCTGTGCATTGGGTAACTCCTGATGG TTTCCCTGTGTGGCAGGAATACAAGAAGCCTATTCAGACGCGCTTGAACCTGATGTTCCTCGGTCA GTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGATTGATGCACACAAACAGGAGTC TGGTATCGCTCCTAACTTTGTACACAGCCAAGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGG CACACGAGAAGTACGGAATCGAATCTTTTGCACTGATTCACGACTCCTTCGGTACCATTCCGGCTG ACGCTGCGAACCTGTTCAAAGCAGTGCGCGAAACTATGGTTGACACATATGAGTCTTGTGATGTAC TGGCTGATTTCTACGACCAGTTCGCTGACCAGTTGCACGAGTCTCAATTGGACAAAATGCCAGCAC TTCCGGCTAAAGGTAACTTGAACCTCCGTGACATCTTAGAGTCGGACTTCGCGTTCGCGTAACGC CAAATCAATACGACTCCGGATCCCCTTCGAAGGAAAGACCTGATGCTTTTCGTGCGCGCATAAAAT ACCTTGATACTGTGCCGGATGAAAGCGGTTCGCGACGAGTAGATGCAATTATGGTTTCTCCGCCA AGAATCTCTTTGCATTTATCAAGTGTTTCCTTCATTGATATTCCGAGAGCATCAATATGCAATGCTGT TGGGATGGCAATTTTTACGCCTGTTTTGCTTTGCTCGACATAAAGATATCCATCTACGATATCAGAC CACTTCATTTCGCATAAATCACCAACTCGTTGCCCGGTAACAACAGCCAGTTCCATTGCAAGTCTG AGCCAACATGGTGATGATTCTGCTGCTTGATAAATTTTCAGGTATTCGTCAGCCGTAAGTC
CmR Piece	CAGGTATTCGTCAGCCGTAAGTCAACTCTTCCTGTCGTCATATCTACAAGCCGGCGCGCCAAATTG ACAATTACTCATCCGGCTCGAATAATGTGTGGAACTTAAACACACACAGGAGGAAAACATATGGAA AAAAAAATCACCGGCTACACCACCGTTGACATCTCTCAGTGGCACCGTAAAGAACACTTTGAAGCG TTCCAGTCTGTCGCGCAGTGTACCTACAACCAGACCGTTCAGCTAGACATCACCGCGTTCCTGAAA ACCGTTAAAAAAAACAAACACAAATTCTACCCGGCGTTCATTCACATCCTGGCGCGTCTGATGAAC GCGCACCCGGAATTTCGTATGGCGATGAAAGACGGTGAACTGGTTATCTGGGACTCTGTTCACCC GTGCTACACCGTTTTCCACGAACAGACCGAAACCTTCTCTTCTCTGTGGTCTGAATACCACGACGA CTTCCGTCAGTTCCTGCACATCTACTCTCAGGACGTTGCGTGCTACGGTGAAAACCTGGCGTACTT CCCGAAAGGTTTCATCGAAAACATGTTCTTCGTTTCTGCGAACCCGTGGGTTTCTTTCACCTCTTTC GACCTGAACGTGGCGAACATGGACAACTTCTTCGCGCCGGTTTTCACTATGGGTAAATACTACACC CAGGGTGACAAAGTTCTGATGCCGCTGGCGATCCAGGTTCACCACGCGGTTTGCGACGGTTTCCA CGTTGGTCGTATGCTGAACGAACTCCAGCAGTATTGCGACGAATGGCAGGGTGGTGCGTAAACTC ACTCCTAGCCCGCCTAATAAGCGGGCTTTTTTTCTGCAGACCAAGTTTACTCATATATACTTTAGAT TGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTC

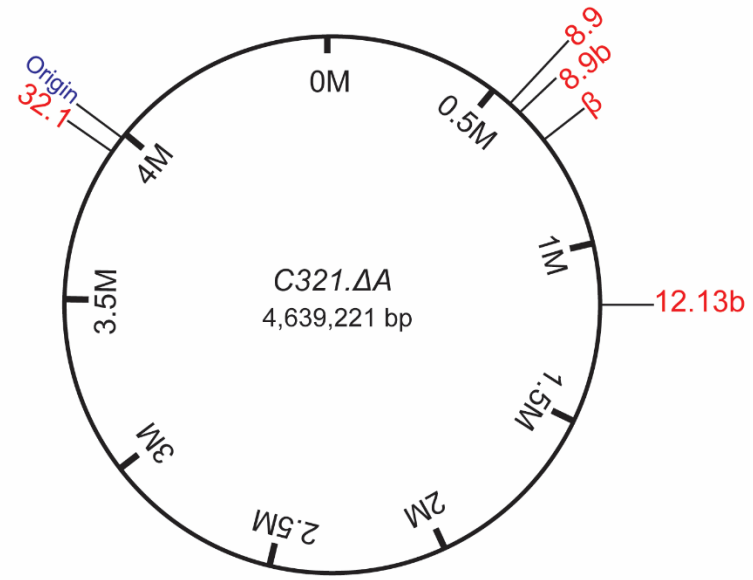
A.2 Supplemental Information to Chapter 4



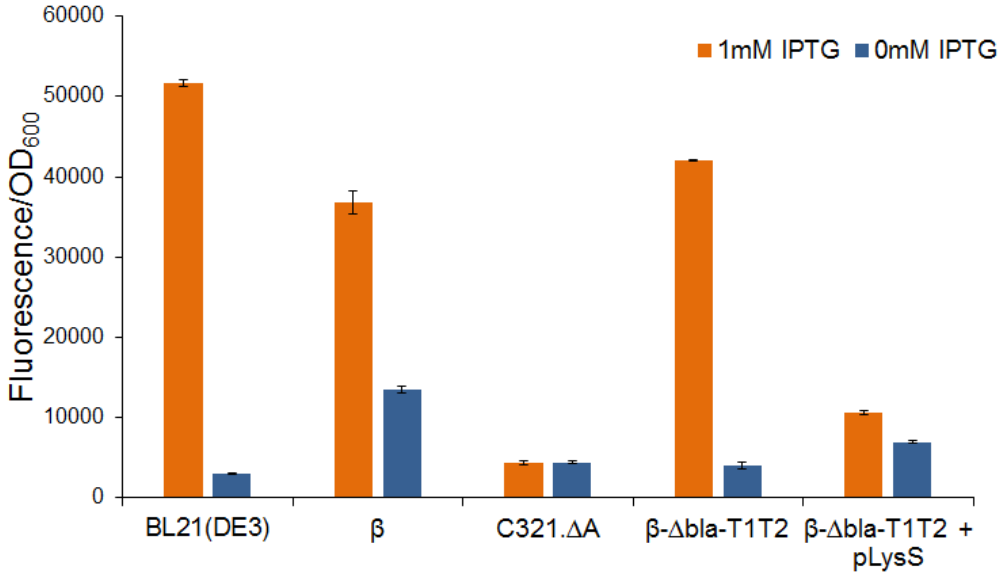
Supplemental Figure A6: Diagram of the T7RNAP-KanR cassette with screening primers. The T7RNAP-KanR cassette is made of a T7RNAP portion of the λDE3 cassette in *BL21(DE3)* and an antibiotic resistance marker, KanR. The 5' and 3' ends of the T7RNAP-KanR cassette contain 45 base pairs of homology that will direct the cassette to its insertion site on *C321.ΔA* via Datsenko-Wanner. Primers shown above were used to synthesize the amplicons (i-v) used for confirming full insertion of the T7RNAP-KanR cassette.



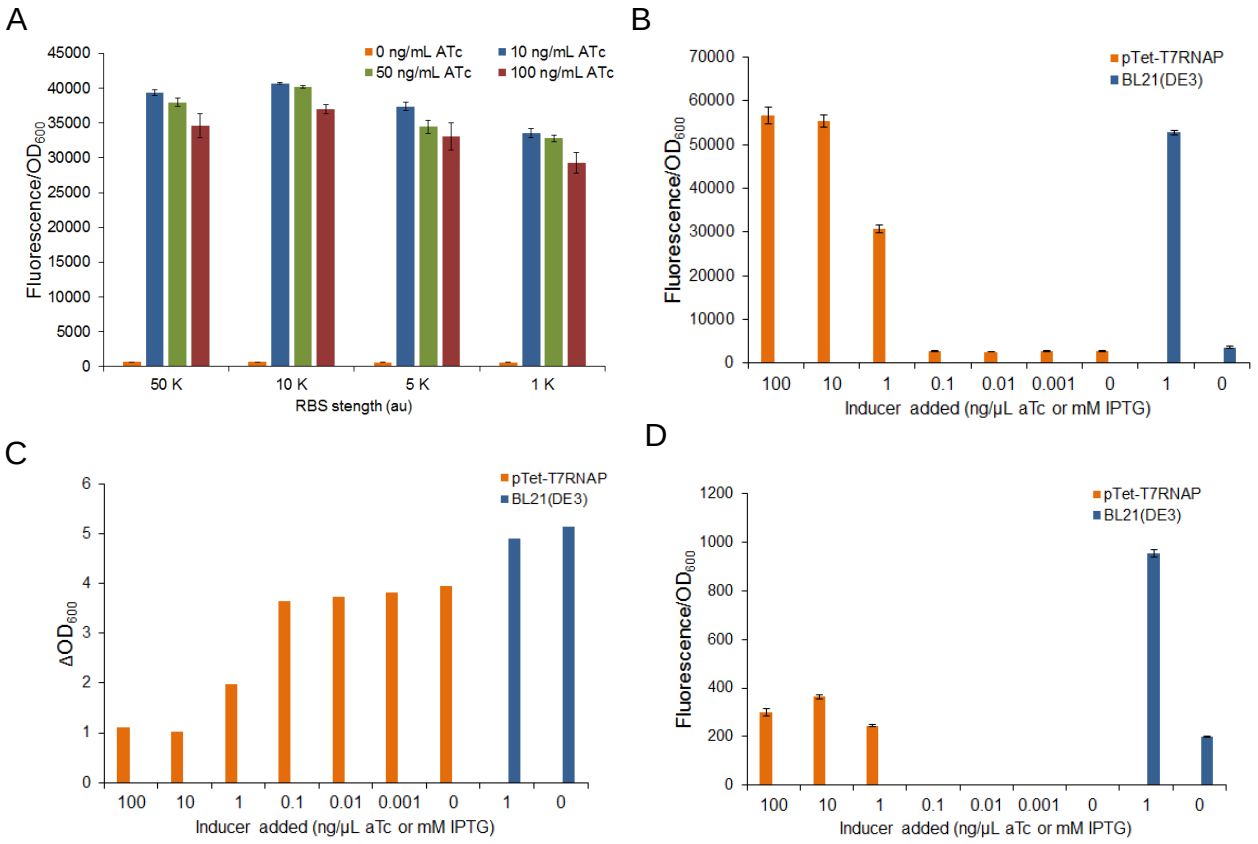
Supplemental Figure A7: Growth rates of *C321.ΔA-T7RNAP* variants. A) Doubling times for strains containing the T7RNAP-KanR cassette. *C321.ΔA-T7RNAP-KanR* represents strains that successfully inserted the full T7RNAP-KanR cassette via Datsenko-Wanner and removed KanR via MAGE (e.g. *α, β, γ, δ*). B) Doubling times for β mutants with reduced protease and nuclease activity.



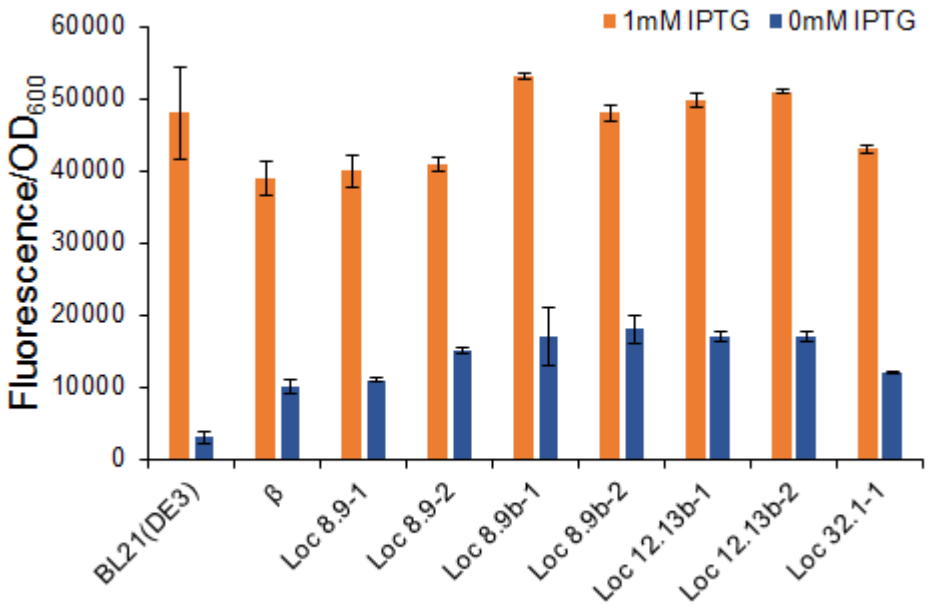
Supplemental Figure A8: Map of *C321.ΔA* genome with T7RNAP insertion sites annotated. Dark tick marks represent half of a mega (M) base pair section. Insertion sites of T7RNAP are labeled in red. The origin of replication is labeled in blue.



Supplemental Figure A9: Addition of T7 lysozyme to T7 system in *β-Δ-T1T2*. All strains were transformed with the pT7sfGFP-wt plasmid. Fluorescence was measured 7 hr after induction with 1 mM (orange bars) or 0 mM (blue bars) IPTG. Error bars represent biological duplicates and technical triplicates. β: *C321.ΔA-T7RNAP-β*



Supplemental Figure A10: Characterization of aTc-inducible T7 system in *C321.ΔA*. A) Expression of pT7sfGFP-wt-TetR 7 hrs after induction with 0 ng/mL (orange), 10 ng/mL (blue), 50 ng/mL (green), 100 ng/mL (red) aTc. Four different RBS strengths were tested upstream the T7RNAP gene. The au unit is based off of the Salis ribosome binding calculator. B) Expression of pT7sfGFP-wt-TetR 7 hrs after induction with titrating amounts of aTc inducer. Here OD_{600} was measured on a plate reader. C) Change in OD_{600} during expression of pT7sfGFP-wt-TetR 7hrs after induction with titrating amounts of aTc inducer. Here OD_{600} was measured using a 1 cm cuvette. D) Expression of pT7sfGFP-sfGFP-wt-TetR after normalizing to OD_{600} measured in 1 cm cuvette. Error bars for all panels represent biological duplicate and technical triplicate.



Supplemental Figure A11: Expression of sfGFP-wt with T7RNAP inserted at various genomic locations in *C321.ΔA*. The T1T2-T7RNAP-cat cassette was inserted at various locations in *C321.ΔA* (designed to insert at Loc 8.9, 8.9b, 12.13b and 32.1). Expression of pT7sfGFP-wt was performed in each strain with or without 1 mM IPTG. Here OD_{600} was measured using a plate reader. Error bars for all panels represent biological duplicate and technical triplicate.

Supplemental Table A4: Oligonucleotides used in this study.

Construction C321.ΔA-T7RNAKanR	**Sequence**
EDC326_iiT7486-f	A*C*ATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTGATCAGAAGGACGTTGATCG
EDC323_iT7396-r	GACTTACGGCTGACGAATACCTG
EDC324_iKan310OL-f	CAGGTATTCGTCAGCCGTAAGTCGAAGCGGAACACGTAGAAAG
EDC329_iiKan-r	T*T*TGCCGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAAGGTGGAATCGAAATCTCGTGATG

Screening for C321.ΔA-T7RNAKanR	**Sequence**
EDC258	ATGAGTATTCAACATTTCCGTGTC
EDC313	CATACTCTGCGACATCG
EDC280	CCATCAACAGTATTATTTTCTCC
EDC315	GGAGAAAATAATACTGTTGATGG
EDC281	CCGGCTCGTATAATGTGTG
EDC259	TTTGCCGACTACCTTGGTG
EDC282	GGTGAAGCACGCTTCC

EDC283	CTGGTGGTGGCTATTGG
EDC284	CTTGCGTTCTGCTTTGAG
EDC285	TGTGGCAGGAATACAAGAAG
EDC313	CATACTCTGCGACATCG
EDC315	GGAGAAAATAATACTGTTGATGG
EDC330	GGAAGATACCATTCAGCCAGC
EDC286	GCAGGTAGCTTGCAGTG

Removing KanR from T7RNAP-KanR cassette	**Sequence**
EDC287_KanR.K.O.	G*G*T*T*GGGCGTCGCTTGGTCGGTCATTTCGAACCCCAGAGTCCCGCCATGCGAAACGATCCTCATCCTGTCTCTTGATCAGATCTTGATCC

MAGE oligos	
JGP368_lonprom-MAGE	C*A*G*A*TGACACGACTGTGCTTCACGCCATCTATTAACATGTACGTCAGATAGAGGAAAAATTAAAGGGGAGATAAAATCCCCCCTTTTTGG
ompT-MAGE	T*G*G*A*CAACTCTCGGCAGCCGAGGTGGCAATATGGTCGCGCAGGACTGGATGGATTCCAGTAACCCCGGAACCTGGACGGATGAAAGTAGA
JGP287_rne131-MAGE	C*T*G*T*TGAGCCGCTTCTTCGGCGCACTGAAAGCGCTGTTCAGCTAACTGAGAAGAAACCAAACCGACCGAGCAACCAGCACCGAAAGCAGA

MASC and MASC Sequencing Primers	**Sequence**
JGP346_rne-wt-f	CACTGAAAGCGCTGTTCAGCGGTGGT
JGP347_rne-mut-f	CACTGAAAGCGCTGTTCAGCTAACTGA
JGP348_rne-r	GTGCGACTACCGCTTCTTCGGCTAC
JGP349_rne-seq	CAGATGGAAACCCCGCACTACCACG
JGP370_lon-wt-f	CTTTAATTTTTCCTCTATTCTCG
JGP371_lon-mut-f	GATTGCAGTACGCACCAGC
JGP369_lon-seq-f	CTGCGTTCCATCGTAGAAGC
JGP365_ompT-wt-f	CAGCCGAGGTGGCAATATGGTCGAT
JGP366_ompT-mut-f	CAGCCGAGGTGGCAATATGGTCGCG
JGP367_ompT-r	GAGTTCAAAATCTTCATAACGATAAC
JGP345_ompT-seq-f	CTGACAACCCCTATTGCGATCAGCTC

Replacing putative promoter with T1T2 terminator	**Sequence**
EDC398_bla-H-f	A*T*GGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACAGTGGGATTTTGGTCATGAG
EDC399_bla-r	AACTCTTCCTGTCGTC
EDC400_t1t2-f	GACGACAGGAAGAGTTGATTTTCAGCCTGATACAG
EDC401_t1t2-r	A*G*GCTTCAACGGATTCATTTTTCTATTTCATAGCCCGGAGCAACCTGTAGAAACGCAAAAAGG
EDC402_bla-t1t2-f	ATGGAGCAAAAGTACATTTAGG

EDC403_bla-t1t2-r	AGGCTTCAACGGATTCATTTTTC
EDC404_bla.K.O.	A*G*A*C*CGCTTCTGCGTTCTGATTTAATCTGTATCAGGCTGAAAATCCTGCAGACCAA GTTTACTCATATATACTTTAGATTGATTTAAAAC
EDC405_seq-f	CACCATACTCACTTTTGC
EDC406_seq-r	CAGTATCGTGATGACAGAG

SynTM-T7RNAP-CmR	**Sequence**
JGP221_L3S2P21_sense	GAT TTT CAG CCT GAT ACA GCT CGG TAC CAA ATT CCA GAA AAG AGG CCT CCC GAA AGG GGG GCC TTT TTT CGT TTT GGT CCC CTT TTT GCG TTT CTA CA
JGP221_L3S2P21_antisense	TGTAGAAACGCAAAAAGGGGACCAAAACGAAAAAAGGCCCCCCTTTCGGGAGGCCTC TTTTCTGGAATTTGGTACCGAGCTGTATCAGGCTGAAAATC
EDC408_DE3-0-f	GCTTTTTTATACTAAGTTGGCATTATAAAAAAG
EDC323_iT7396-r	GACTTACGGCTGACGAATACCTG
EDC413_cat-f	CAGGTATTCGTCAGCCGTAAGTCAACTCTTCCTGTCGTC
EDC414_cat-1-r	GAGATTATCAAAAAGGATCTTCACC
JGP139_Loc9.8_T1T2DE3cat_f	C*G*CGAGGCATCGTCTTAACGAGGCACCGAGGCGTCGCATTCTTCAGATTTTCAGCC TGATACAG
JGP140_Loc9.8_T1T2DE3cat_r	T*G*CGGGGATTACTCCCATAAGCGCTAACTTAAGGGTTGAACCATCGAGATTATCAAA AAGGATCTTCACC
JGP173_Hom8.9bseq_f3	GAACTTAATGCAACACACCA
JGP292_T7RNAPr_seq	CCGCGCTTAGCTTTCACT
EDC280_T7seq1	CCATCAACAGTATTATTTTCTCC
EDC282_T7seq3	GGTGAAGCACGCTTCC
JGP153_Hom8.9bseq_r	CGTGTGCGTGTCCATAGA
JGP389_T7cassCmR_ko_MAGE	G*T*C*T*GCAGAAAAAAAGCCCGCTTATTAGGCGGGCTAGGAGTGAGTGTGTGTTTAA GTTCCACACATTATTCGAGCCGGATGAGTAATTGT
EDC413_cat-f	CAGGTATTCGTCAGCCGTAAGTCAACTCTTCCTGTCGTC
JGP211_Hom8.9bscreen_r1	GCACCATGAAATCCTGCA

A.3 Supplemental Information to Chapter 5

Supplemental Table A5: Oligonucleotides used in this study.

Primers	Sequence
EFTu-f	GGTGGTCATATGTCTAAAGAAAAGTTTGAACG
EFTu-r	GGTGGTCTCGAGTTAGCTCAGAACTTTTGCTACAA
EvolAzRS-f	GGTGGTCATATGGACGAGTTCGAAATGAT
EvolAzRS-r	GGTGGTCTCGAGTTACAGACGTTTGCGAATTGG
pDAK-f	CGGCCGCACTCGAGCACCA
pDAK-r	GGTGGTCATATGATATCTCCTTCTTAAAGTT
DAKPnative-f	GGAATTGTGAGCGGATAACA
DAKPtacI-r	TCCACACATTATACGAGCCGATGATTAATTGTCAACAGCTCCGATCCTCTACG CCGGAC
DAKlpp-r	ACAAGTATTACACAAAGTTTTTTATGTTGAGAATATTTTTTTGATCGATCCTCTA CGCCGGAC
Sf1TAG-f	AGATATACATATGAGCAAAGGTG
Sf1TAG-r	GGTGGTCTCGAGTTATTTTTCGAACTGCGGATG
Sf3TAG-f	TAGGATGATGGCAAATATAAAACGCC
Sf3TAG-r	AAAGCTAATCGTACGTTCCTG
G77-f	GCGGCGACCTGACCGTCAACAGCTACGAGGAGCTGGAAAGCTTGTTTAAG
G77-r	CGGTCAGGTCGCCGCCMNNTTTCTCMNNGCGCTTAATCGT
G78-f	CTGAAAAACGCCGTGGCG
G78-r	GCCACGGCGTTTTTCAGMNNMNNCGGMNNCAGTTCTTTGTTCTTAAACAAGC TTTCCAGC
PLV-f	GCTAGCGGCACCTGTCCTACGAGTTG
PLV-r	GAATTCGTATCAGCTCACTCAAAGGCG
G166-f	AGAAAGCGTACTGCCCGNNKGGCGTGGTTGAGGGTA
G151-f	NNKCCGGCAGGCGTGGTTGA
G151-r	GTACGCTTTCTTGATCTTCG
G152-f	NNKGCAGGCGTGGTTGAGG
G152-r	GCAGTACGCTTTCTTGATCT
G154-f	NNKATGCGCCTGAAAAACGCC
G154-r	ATGCAGTTCTTTGTTCTTAAAC
G158-f	NNKAACGCCGTGGCGGAAGA
G158-r	CAGGCGCATCGGATGCAG
G220-r	MNNMNNMNNMNNCAGTTCTTTGTTCTTAAACAAGC
G79-f	CACCATCAACACTGCG CACGTTGAATACGA
G80-f	CACCATCAACACTTCTGCG GTTGAATACGA
G123-f	CACCATCAACACTGCG GCG GTTGAATACGA
G123-1-f	CACCATCAACACTTCTCACG
G123-r	GATCGGCAGCAGGAACGGCT
G81-f	TCCTGCTGCCGATCGCG GACGTATTCTCCATCTCCGG
G125-f	TCCTGCTGCCGATCGAAGCG GTATTCTCCATCTCCGG
G126-f	TCCTGCTGCCGATCGAAGACGCG TTCTCCATCTCCGG
G127-f	TCCTGCTGCCGATCGCG GCG GTATTCTCCATCTCCGG
G128-f	TCCTGCTGCCGATCGAAGCG GCG TTCTCCATCTCCGG
G129-f	TCCTGCTGCCGATCGCG GACGCG TTCTCCATCTCCGG
G130-f	TCCTGCTGCCGATCGCG GCG GCG TTCTCCATCTCCGG
G131-f	TCCTGCTGCCGATCGAAGAC
G231-f	TCCTGCTGCCGATCGCG GACGTAGCG TCCATCTCCGGTCG
G232-f	TCCTGCTGCCGATCGAAGCG GTAGCG TCCATCTCCGGTCG

G233-f	TCCTGCTGCCGATCGAAGACGCG GCG TCCATCTCCGGTCG
G234-f	TCCTGCTGCCGATCGCG GCG GTAGCG TCCATCTCCGGTCG
G235-f	TCCTGCTGCCGATCGAAGCG GCG GCG TCCATCTCCGGTCG
G236-f	TCCTGCTGCCGATCGCG GACGCG GCG TCCATCTCCGGTCG
G237-f	TCCTGCTGCCGATCGCG GCG GCG GCG TCCATCTCCGGTCG
G238-f	TCCTGCTGCCGATCGAA GAC GTA GCG TCCATCTCCGGTCG
G131-r	CTCACCAGCACGGCCTTC
G132-r	AGTGTTGATGGTGATACCAC
G82-f	TGCTGCCGATCGAAGCG GTATTCTCCATCTCCGG
G83-f	TGCCGATCGAAGACGCG TTCTCCATCTCCGG
G84-f	GTGGTACCGTTGTTGCG GGTCGTGTAGAACG
G85-f	CTGGCGTTGAAATGGCG CGCAAACTGCTGGA
G86-f	GCCGTGCTGGTGAGGCG GTAGGTGTTCTGCTGCGT
G87-f	CGTGCTGGTGAGAACGCG GGTGTTCTGCTGCGT
G124-f	GCCGTGCTGGTGAGGCG GCG GGTGTTCTGCTGCGT
G124-1-f	GCCGTGCTGGTGAGAAC
N=A,T,G,C M=A,C K=G,T	

Supplemental Table A12: Amino acid sequences of 24 mutants from the o-aaRS library 1.

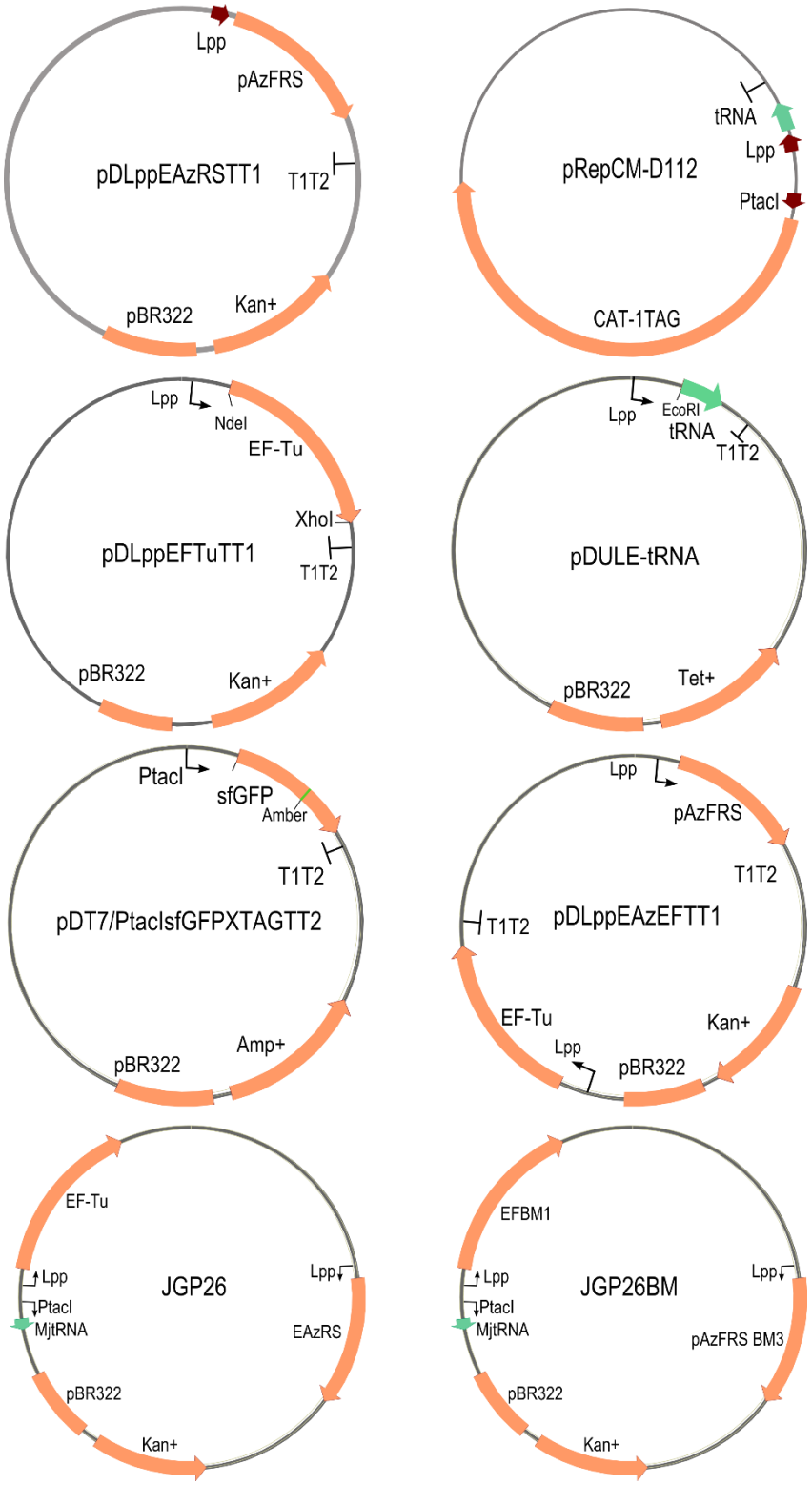
Sample	A233	P258	F261	H283	M285	D286	Variant
1	A	A	P	L	F	Y	BM1
2	A	A	P	L	F	Y	
3	A	A	P	L	F	Y	
4	A	A	P	L	F	Y	
5	A	A	P	L	F	Y	
6	A	A	P	L	F	Y	
7	L	F	G	N	H	F	BM1-1
8	L	F	G	N	H	F	
9	L	F	G	N	H	F	
10	S	P	F	L	W	E	BM1-2
11	S	P	F	L	W	E	
12	N	T	F	I	W	E	BM1-3
13	T	S	H	T	S	Y	BM1-4
14	C	H	W	T	L	T	BM1-5
15	L	H	T	G	V	S	BM1-6
16	I	P	L	G	S	R	BM1-7
17	I	T	S	L	K	R	BM1-8
18	V	N	S	P	V	T	BM1-9
19	A	N	G	N	H	F	BM1-10
20	V	P	N	G	Y	T	BM1-11
21	H	S	T	M	M	F	BM1-12
22	A	Y	T	Y	A	L	BM1-13
23	I	P	F	I	T	V	BM1-14
24	A	P	F	H	M	R	BM1-15

Supplemental Table A7: Amino acid sequences of 5 selected mutants from the o-aaRS library 2.

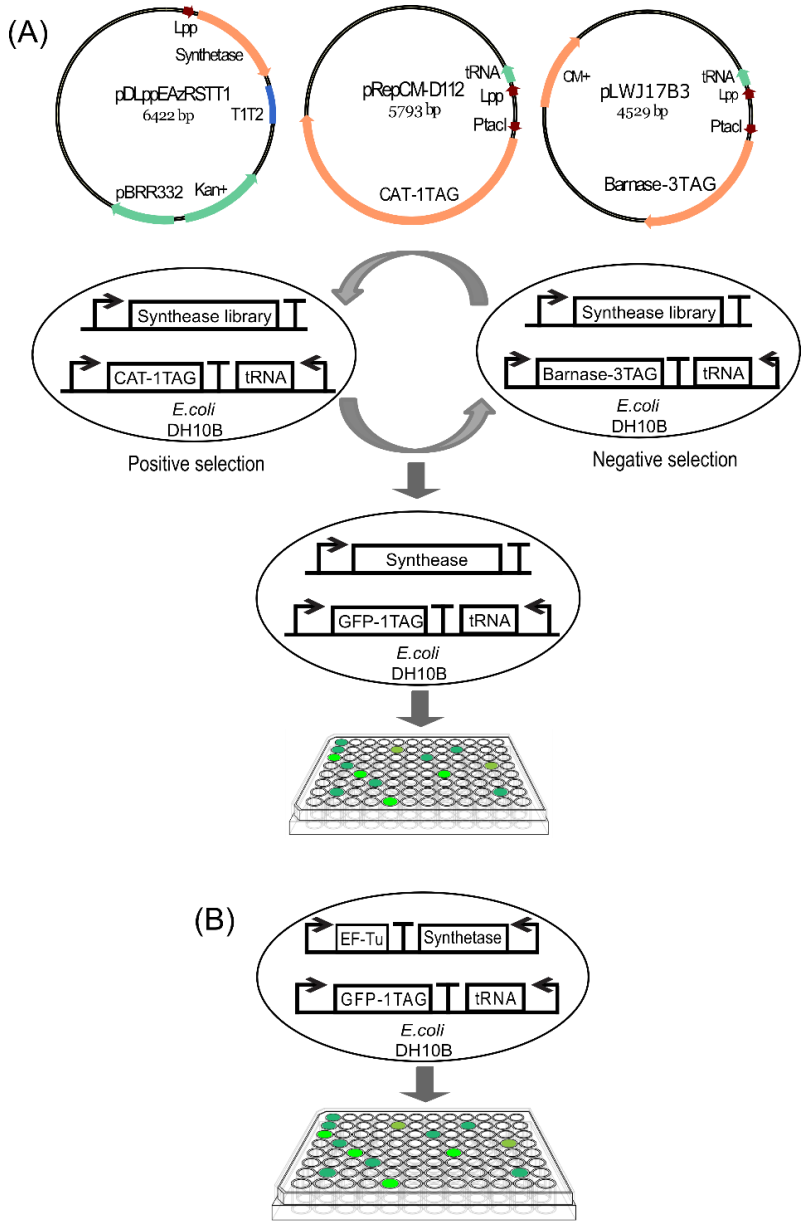
Sample	C231	P232	P284	K288	
1	C	P	V	K	BM2
2	C	P	V	K	
3	C	P	D	K	BM2-1
4	C	P	T	K	BM2-2
5	R	P	P	K	BM2-3

Supplemental Table A8: Amino acid sequences of 24 selected mutants from the o-aaRS library 3.

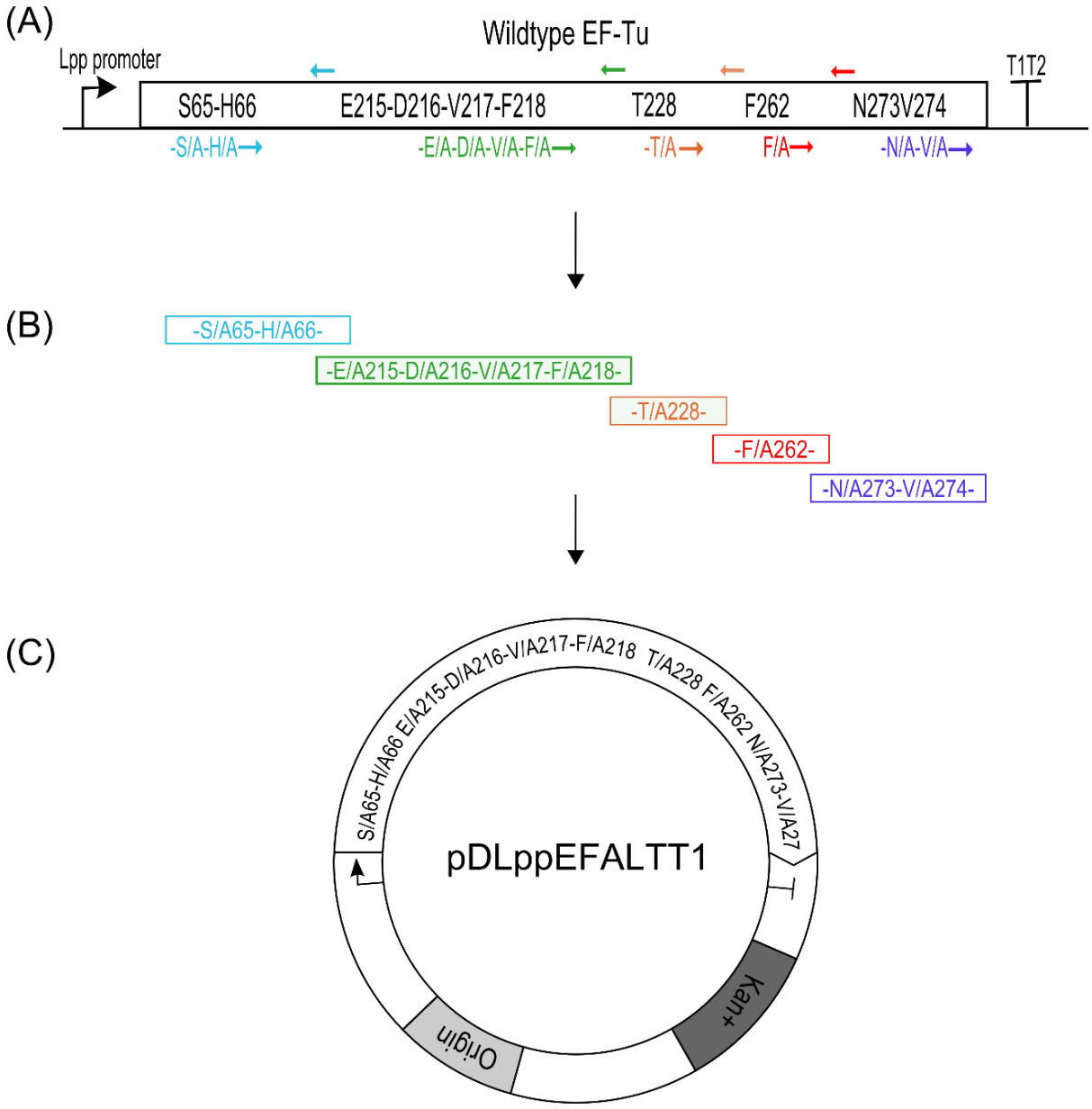
Sample	H283	P284	M285	R286	Variant
1	T	S	D	V	BM3
2	T	S	D	V	
3	T	S	D	V	
4	T	S	D	V	
5	T	S	D	V	
6	T	S	D	V	
7	T	S	D	V	
8	T	S	D	V	
9	T	S	D	V	
10	T	S	D	V	
11	T	S	D	V	
12	T	S	D	V	
13	T	S	D	V	
14	T	S	D	V	
15	T	S	D	V	
16	T	S	D	V	
17	K	T	R	Q	BM3-1
18	K	T	R	Q	
19	K	T	R	Q	
20	Q	L	N	G	BM3-2
21	P	Q	P	R	BM3-3
22	Q	H	I	S	BM3-4
23	A	D	L	R	BM3-5
24	D	P	L	G	BM3-6



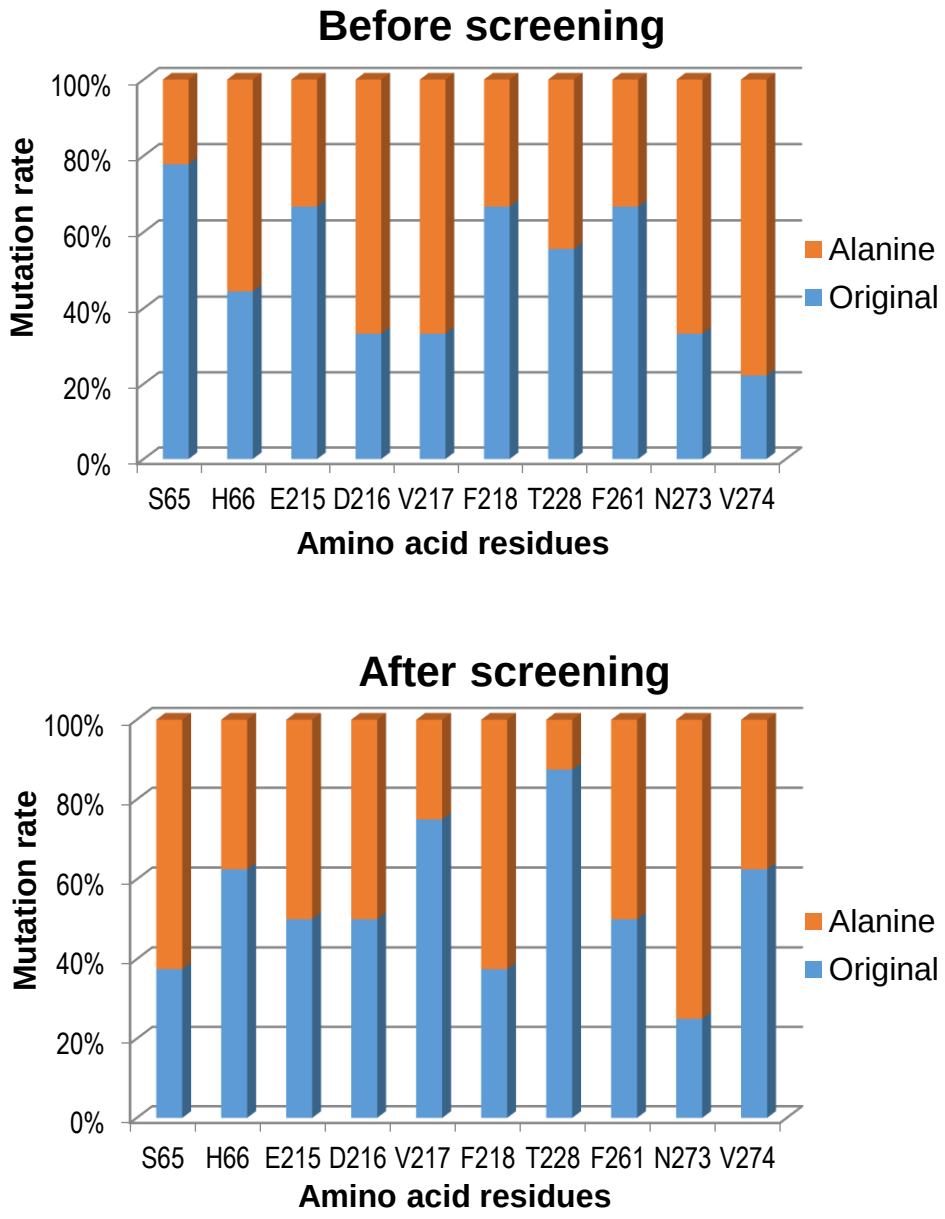
Supplemental Figure A15: Maps of plasmids used in this study. The construction methods were described in the Materials and Methods section.



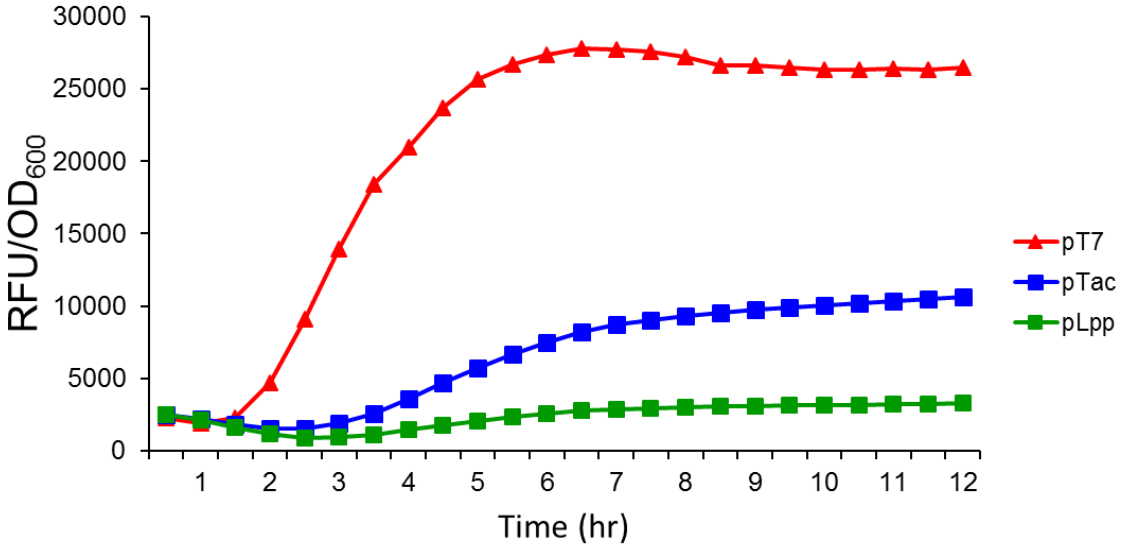
Supplemental Figure A12: Library selection and screening. (A) Selection of p-azido-phenylalanine aminoacyl tRNA synthetases (aaRS) from a library of variant sequences. The plasmid library was transformed by electroporation into *E. coli* DH10β containing pRepCMD112 plasmid for positive selection. In positive selection, transformants were grown on LB solid media (25 μg/ml kanamycin, 10 μg/ml tetracycline, 0.5 mM IPTG) at 37°C for 36 hours against 10 μg/ml (the first round) and 20 μg/ml (the second round) chloramphenicol respectively. The plasmid library was isolated from surviving colonies and transformed into DH10β *E. coli* with pLWJ17B3 for negative selection, based on the toxin barnase. Selection was performed for a total of 4 rounds (positive-negative-positive-negative) according to previous publications with slight modifications. Following the positive and negative selections, plasmids from surviving colonies were transformed into DH10β *E. coli* harboring pDPtacIsfGFP1TAGTT2 plasmid for the final round of screening of fluorescent intensity in 96-well plate. (B) Screening of EF-Tu library. Beneficial EF-Tu mutants were identified using normalized fluorescence in a 96-well plate assay.



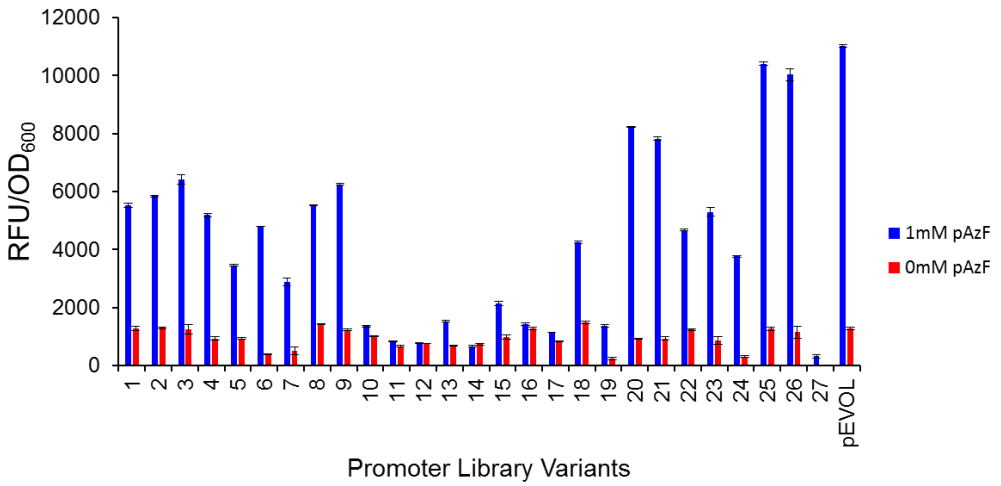
Supplemental Figure A13: Cartoon schematic demonstrating the construction of the EF-Tu alanine-scanning library. (A) Wild type EF-Tu was used as template. The full-length template was amplified in 5 fragments by DNA primers that contain desired mutations and overlapping sequences to adjacent fragments at both ends. (B) The 5 fragments were subject to overlap extension PCR to restore the full-length EF-Tu-coding sequence with desired mutations. (C) Finally, the EF-Tu library was inserted into an expression vector and then transformed into DH10β cell for screening.



Supplemental Figure A14: The mutation rate of the 10 altered residues of the EF-Tu amino acid binding pocket before and after screening. (A) Twenty mutants were randomly picked and sequenced from the library before screening. The frequency of alanine and the original residue was counted to validate the expected mutation efficiency. (B) After screening, eight mutants having >1.5-fold enhancement in their ability to incorporate pAzF into proteins were sequenced. The frequency of alanine and original residue at each site was counted again to observe the preference of mutation



Supplemental Figure A15: Promoter characterization. Normalized fluorescence of promoter candidates. A high (T7), medium (PtacI) and low (lpp) promoter was added upstream of sfGFP-coding gene to construct the reporter plasmids. The plasmids were transformed into BL21(DE3) and analyzed for their ability to drive the expression of sfGFP in the presence of with 1 mM IPTG. Readings were taken from a cultivation time course up to 12 hrs after induction for fluorescence/OD_{600} signal at 37C.



	AzRS	EF-Tu	o-tRNA
1	T7	T7	T7
2	T7	T7	TacI
3	T7	T7	Lpp
4	T7	TacI	T7
5	T7	TacI	TacI
6	T7	TacI	Lpp
7	T7	Lpp	T7
8	T7	Lpp	TacI
9	T7	Lpp	Lpp
10	TacI	T7	T7
11	TacI	T7	TacI
12	TacI	T7	Lpp
13	TacI	TacI	T7
14	TacI	TacI	TacI
15	TacI	TacI	Lpp
16	TacI	Lpp	T7
17	TacI	Lpp	TacI
18	TacI	Lpp	Lpp
19	Lpp	T7	T7
20	Lpp	T7	TacI
21	Lpp	T7	Lpp
22	Lpp	TacI	T7
23	Lpp	TacI	TacI
24	Lpp	TacI	Lpp
25	Lpp	Lpp	T7
26	Lpp	Lpp	TacI
27	Lpp	Lpp	Lpp

Supplemental Figure A16: Promoter optimization of orthogonal translation system components. Normalized fluorescence of promoter library variants. A high (T7; dark green), medium (PtacI; medium green) or low (lpp; light green) promoter was added upstream of AzFRS, EF-Tu, and o-tRNA in JGP plasmid to create a library of 27 members. The plasmid library was transformed into BL21(DE3) and analyzed for their ability to suppress an 1TAG in sfGFP in the presence of with 1 mM IPTG, and 10 mM pAzF. Readings were taken at 7.5 hrs after induction for maximum fluorescence/OD_{600} signal. Error bars, SD (n=3).