NORTHWESTERN UNIVERSITY

Cell-free Platforms for Synthesis of Non-standard Polypeptides in vitro

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biological Sciences

By

Benjamin James Des Soye

EVANSTON, ILLINOIS

December 2018

ABSTRACT

Cell-free Platforms for Synthesis of Non-standard Polypeptides in vitro

Benjamin James Des Soye

Proteins represent a critical class of biomolecules, universally employed by all living organisms to fulfill essential structural, functional, and enzymatic roles necessary to support life. In nature, these polymers are composed generally of twenty natural amino acid (AA) building blocks, which can be modified with covalent adducts known as post-translational modifications (PTMs) to effect changes in functionality and behavior. Complete understanding of the biological role played by a given protein necessitates an understanding of its various PTMs and their influence on its structure and function. However, our ability to identify and characterize specific patterns of PTMs is limited by our ability to produce pure, homogeneous samples of proteins featuring a specific set of modifications. In this thesis, I sought to develop platforms for enabling preparative scale synthesis of proteins featuring user-definable PTM patterns to facilitate downstream fundamental discovery. The key idea was to genetically encode PTMs and then directly incorporate modified amino acids (as non-canonical amino acids, ncAAs) into proteins by repurposing the amber stop codon as a coding channel, a technique known as amber suppression. Unfortunately, up to now these approaches have been conventionally limited by competition from release factor 1 (RF1) and the fact that the orthogonal translation systems (OTSs) used to incorporate ncAAs are toxic to cells. To address these limitations, I hypothesized that applying cell-free protein synthesis (CFPS) systems derived from cells lacking RF1 would enable the use of OTSs without associated toxicity effects while simultaneously eliminating release factor

competition. I further hypothesized that the elimination of putative negative effectors of CFPS or the use of highly-active translational components would yield highly-productive CFPS systems, enabling preparative scale synthesis of proteins featuring specific PTMs for downstream characterization and thus allowing me to meet my goal.

I first contributed to the development of a CFPS system from a partially-recoded strain of *Escherichia coli* deficient RF1. As hypothesized, stabilization of template DNA and mRNA via removal of nucleases in the system increased productivity fourfold, and ncAA incorporation was improved significantly in the absence of RF1. Next, I embarked on a similar effort to generate a CFPS system from a fully-recoded RF1-deficient strain in which all native instances of the amber stop codon were removed. In further agreement with my initial hypotheses, the functional inactivation of putative negative effectors of CFPS yielded a 4.5-fold increase in platform productivity and the absence of RF1 activity in the system facilitated insertion of 40 identical ncAAs into proteins with ≥98% fidelity. In my next aim, I hypothesized that imbuing our best recoded RF1-deficient source strain with the ability to synthesize the viral T7 RNA polymerase would expand the capabilities of our recoded CFPS platform to yield a one-pot system, simplifying use of the platform and facilitating its adoption. After genomically incorporating a construct encoding the polymerase into the strain and subsequently installing mutations protecting the enzyme from proteolysis, the platform was capable of robust ncAA incorporation independent of purified polymerase supplementation, agreeing with my hypothesis. Next, I set out to pioneer a novel CFPS system derived from the fast-growing non-model bacterium *Vibrio natriegens* based on the hypothesis that its lysates would be enriched with highly-active translational components. Elucidation of optimal cell growth, lysis, and reaction conditions culminated in a highly-

productive CFPS platform comparable to the state-of-the-art in agreement with this hypothesis. I also found that the system was uniquely capable of synthesizing short peptides, which have historically been difficult to produce recombinantly. Finally, I pursued the synthesis of proteins featuring specific patterns of serine phosphorylation, a widely studied PTM. In this effort, I reasoned that incorporation of *o*-phosphoserine (Sep) in CFPS could be improved by utilizing improved Sep-specific translation components in an engineered strain background optimized for Sep incorporation. In an illustrative example of what I was trying to accomplish, I applied our systems to the synthesis of the glycolytic enzyme triosephosphate isomerase bearing specific serine phosphorylations to investigate their effects on the enzyme. Taken together, my work will facilitate efforts to interrogate the effects of specific PTMs on protein structure and function, further increasing our understanding of the chemistry of life.

ACKNOWLEDGMENTS

I would first like to thank my amazing wife Caitlin for her unfaltering emotional and moral support through the years. She celebrated with me when things went well, and propped me up when things weren't working out. During our daily recounting of our workdays, she even managed to muster some pretty authentic interest and enthusiasm for what I was working on! Completion of a PhD is never easy, but she made it easier. Thanks for everything, girl – your support meant everything to me.

Next, I wish to extend my thanks to my excellent advisor, Professor Michael Jewett, for his guidance and mentorship over the past six years. His passion and enthusiasm for science and research scholarship are as infectious today as they were when I first sat in his synthetic biology course back in 2012, and these values inspire everyone in his orbit to pursue excellence. I would also like to thank my thesis committee members – chair Professor Neil Kelleher, Professor Karl Scheidt, and Professor Sadie Wignall – for their collective wisdom, feedback, and support. By challenging me in my qualifying exam and committee meetings, they forced me out of my scientific comfort zone to take into account considerations that I had missed. I have no doubt that I am a better scientist because of their leadership.

I must also thank the many Jewett lab members who influenced and shaped my grad school experience. Intellectual and material contributions to specific projects are listed at the end their respective chapters, but here I would like to highlight my thanks to individuals for their less tangible contributions to my development as a scientist. I would like to thank my undergraduate Sam Davidson for being a second set of hands for experiments, a sounding board for ideas, and a

fun conversational partner to distract from the monotony of pipetting dozens of cell-free reactions. He was a welcome presence for over three years, and I surely would not have accomplished as much as I did without his help. I thank Seok Hoon Hong, Yong Chan Kwon, Rey Martin, and Javin Oza for teaching me just about everything I know about working with bacteria and running cell-free reactions. I thank the members of Office 4 (NOT "The Bro Office", that was a branding fail), past and present, for their excellent company: Weston Kightlinger, Mike Hammerling, Andrew Hunt, Jennifer Kay, Yi Liu, Oliver Weisser, Mark Anderson, and Arnaz Ranji. I thank my PCR sensei Erik Carlson for answering all of my questions and teaching me how to assemble and manipulate DNA. I thank the past and present members of Team ncAA not already mentioned for their constant support and excellent input: Rui Gan, Jessica Perez, Jasmine Hershewe, Dubby Wiseman, and Charlotte Abrahamson. Finally, I extend my thanks to all other Jewett lab members, past and present, for their help and comraderie during our time together. You all answered my questions and gave sage advice, and for that I am extremely grateful.

I would next like to thank my family. I would like to thank my parents, Eric and Michelle, who have unconditionally supported all of my pursuits for nearly 30 years. They've always believed in my capabilities, which has aided my belief in myself. I would also like to thank my brother, Gabe, and my sister, Bethany, for supporting and (very occasionally) visiting me. They also do a pretty good job taking care of Mom and Dad! Distance may separate us now that I live in Evanston, but I still feel their love and support every single day.

Next, I want to thank the many friends I've made during my time at Northwestern. The weekend board games, Magic: the Gathering, cookouts, and general hangs injected a ton of fun

into my life. Thanks to Shaunna, Kaleigh, Steve, Lauren, Brooks, Dave, Ruby, Luke, Chris, Andrew, Laura, Steph, Warren, and Scott.

Many thanks to those who helped edit this thesis: Weston Kightlinger, Andrew Hunt, Luis Schachner, and Michael Jewett.

Lastly, I want to thank Dr. Thomas Helm for connecting me to my first research experience during my undergraduate studies. It was a small thing at the time, but it ultimately sparked my enthusiasm for research and in doing so altered the entire trajectory of my academic career. It is because of this act that I now find myself composing a PhD thesis.

LIST OF ABBREVIATIONS

λHR……..….λ-Red mediated homologous recombination

AA……amino acid

aaRS……..…aminoacyl-tRNA synthetase

AMP…….…antimicrobial peptide

CAGE……...…..…conjugative assembly genome engineering

CAT…….chloramphenicol acetyltransferase

CFPS……cell-free protein synthesis

CHO……..Chinese hamster ovary

CRMAGE……..….CRISPR/Cas9 enhanced MAGE

DHFR……...dihydrofolate reductase

EF-Tu……....elongation factor Tu

EF-Sep…….…orthogonal Sep-specific elongation factor

ELP……...elastin-like polypeptide

GRO……..genomically recoded organism

IP……..immunoprecipitation

IPTG……...isopropyl-D-thiogalactopyranoside

LC……..liquid chromatography

m/z…….mass-to-charge ratio

MAGE……..multiplex advanced genome engineering

MASC-PCR……..multiplex allele-specific colony polymerase chain reaction

mGM-CSF……..…..murine granulocyte-macrophage colony-stimulating factor

MPC…….multi-proteoform complex

MS……....mass spectrometry

MW……molecular weight

mRFP……...……red fluorescent protein

ncAA……non-canonical amino acid

nTDMS……..…native top-down mass spectrometry

NTP……...…nucleotide triphosphate

o-tRNA……..orthogonal amber suppressor tRNA

OTS……..…orthogonal translation system

pAcF……...……*p*-acetylphenylalanine

pAcFRS……..orthogonal pAcF tRNA synthetase

PANOx-SP……PEP, amino acids, NAD, oxalic acid, spermidine, putrescine

pAzF……...……*p*-azidophenylalanine

pAzFRS……..orthogonal pAzF tRNA synthetase

PCR……..polymerase chain reaction

PEP…….....phosphoenol pyruvate

POI……..……protein of interest

pPaF……..……..*p*-propargyloxyphenylalanine

pPaFRS……..……orthogonal pPaF synthetase

PTM…….post-translation modification

RF1……...release factor 1

S12…….. *E. coli* crude cell lysate clarified by centrifugation at 12,000 x *g*

Sep……...*o*-phosphoserine

SepRS……....orthogonal Sep tRNA synthetase

sfGFP……..…superfolder green fluorescent protein

sfGFP-2amb/UAG….……sfGFP containing 2 amber codons at positions N212 and T216

sfGFP-5amb…sfGFP containing 5 amber codons at positions D36, K101, E132, D190, and E213

sfGFP-S2X……...sfGFP containing 1 amber codon at position S2

sfGFP-T216X……...…sfGFP containing 1 amber codon at position T216

T7RNAP……...T7 RNA polymerase

TPI…….……..triosephosphate isomerase

tRNASep…….orthogonal Sep tRNA

TX-TL……..…transcription-translation

Vnat.....*Vibrio natriegens*

Dedicated to the loving memory of my grandmother, Denise Gendron Bacardi.

I wish you could have stayed with us long enough to have seen what I was studying. I would have loved to have shown everything to you, and I know you would have been fascinated by it.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES AND SEQUENCES

CHAPTER 1

1 Introduction: Overview and Review of Literature

1.1 Dissertation Overview

1.1.1 Motivations and objectives

Proteins play centrals roles in most of the essential structural, functional, and enzymatic processes necessary to support life. These templated polymers are composed generally of twenty natural amino acid (AA) building blocks, which can be combined in a near-infinite number of combinations to generate an impressive level of structural and functional diversity. This complexity is compounded by the existence of post-translational modifications (PTMs), covalent adducts that are attached to specific amino acid sidechains of proteins in specific patterns to effect changes in protein functionality and behavior. **Complete understanding of the biological role played by a given protein necessitates an understanding of its various PTMs and their influence on its structure and function**.

Understanding the role of specific PTMs remains difficult. Detection of naturally-occurring patterns of PTMs is a particular challenge – many PTMs are temporally transient, and persist only as long as needed to enact a specific function[1-4]. In order to detect such modifications, cells must be interrogated at the correct time under very specific conditions. Additionally, many enzymes are present only at very low levels in cells such that variants featuring specific PTM patterns fall well below the detection limit for even modern detection assays[4, 5]. Even when a protein's PTMs are known, complete characterization of the modifications can be difficult to perform. In particular, crystal structure determination and *in vitro* experimental

characterization both require homogeneous, pure samples of a protein bearing only a specified set of PTMs (a specific proteoform) which is often prohibitively difficult (or even impossible) to obtain using conventional means [6-11].

Direct genetic encoding of PTMs into the gene sequence of a protein is one method by which homogeneous proteoform samples can be prepared. In this context, a given natural amino acid covalently conjugated to a PTM of interest is collectively considered a type of non-canonical amino acid (ncAA). Pioneering work by Peter Schultz and others has yielded a method enabling the cotranslational, site-specific installation of ncAAs into proteins in response to the amber stop codon (UAG), a process known as amber suppression[12-18]. Using this technique, a single type of PTM can be directly incorporated into nascent proteins site-specifically by encoding amber codons into the product gene at positions where the PTM is desired, a relatively simple task with the ever-decreasing cost of *de novo* DNA synthesis and modern DNA assembly techniques.

Transformative though it has been, amber suppression is not without its limitations. In cells, ncAA incorporation can be limited by poor bioavailability. Large and/or charged ncAAs in particular can have difficulty crossing cell membranes, limiting their availability for amber suppression[6, 19, 20]. Amber suppression in cells can also be limited by cytotoxicity of the engineered translation components required for incorporation of ncAAs[21, 22]. Finally, amber suppression has historically been limited by the activities of release factor 1 (RF1), a protein which is natively responsible for terminating translation at amber codons. RF1 interference at amber codons during amber suppression frequently leads to premature product truncation, fundamentally limiting the number of ncAAs that can realistically be incorporated into a single protein[21, 23, 24] .

Recently, efforts were made to address this limitation by pursuing the generation of *E. coli* strains in which the *prfA* gene (encoding RF1) is removed, thereby eliminating RF1 interference in amber suppression. Initial attempts to simply knock out this key translational regulator were stymied by cell inviability. Indeed, 321 genes in *E. coli* require the activity of RF1 in order to properly terminate translation[21], making a *prfA* knockout lethal in a wild type background. It wasn't until after the arrival of next-generation genome engineering techniques (*e.g.* multiplex advanced genome engineering (MAGE[25]) and conjugative assembly genome engineering (CAGE[26])) that a complete *prfA* knockout could be pursued in earnest, enabled by the use of these editing technologies to modify (recode) the stop signals of key genes such that they no longer required RF1 activity for proper termination. Partially recoded strains such as *rEc.E7* (in which 7 essential genes are recoded[24]) and *rEc.E13* (in which 7 essential and 6 highly-expressed genes are recoded[21, 24]) remain viable with a *prfA* deletion (yielding *rEc.E7.ΔA* and *rEc.E13.ΔA*, respectively), though both strains experience significant growth defects upon removal of the release factor[24]. Recoding efforts were ultimately completed in 2013, when Lajoie *et al*. reported the development of strain *C321* in which all genes dependent on RF1 for termination are recoded, permitting the removal of *prfA* (*C321.ΔA*) with minimal effects on cell health[21]. These RF1-deficient strains have already demonstrated improved capacity for ncAA incorporation *in vivo*[21, 27].

In this thesis, I was primarily interested in advancing our ability to characterize and understand PTMs by developing a new generation of tools for improved preparative scale synthesis of specific proteoforms using amber suppression. Since the impact of a new technology correlates with the rate of its adoption, I was also interested in increasing the accessibility of the tool. I

hypothesized that applying cell-free protein synthesis (CFPS) systems derived from cells lacking RF1 would enable the use of OTSs without associated toxicity effects while simultaneously eliminating release factor competition. I further hypothesized that the elimination of putative negative effectors of CFPS or the use of highly-active translational components would yield highly-productive CFPS systems, enabling preparative scale synthesis of proteins featuring specific PTMs for downstream characterization and thus allowing me to meet my goal. Through functional inactivation of negative effectors of CFPS, protein production in lysates derived from both *rEc.E13.ΔA* and *C321.ΔA* was improved ~four-fold, and lysates from both strains demonstrated superior capabilities for ncAA incorporation, in agreement with my hypotheses. *C321.ΔA*-derived lysates in particular supported incorporation of a record 40 ncAAs into a single peptide. After optimizing recoded strains for use in CFPS, I was next interested in expanding the capabilities of the platform to both enable scale-up to large reaction volumes as well simplify the use of the system to facilitate its adoption by new laboratories. I hypothesized that imbuing our best recoded RF1-deficient source strain with the ability to synthesize the viral T7 RNA polymerase would expand the capabilities of our recoded CFPS platform to yield a "one-pot" platform. To test this, I genomically incorporated a synthetic DNA construct into our best *C321.ΔA* derivative and installed mutations to protect the polymerase from proteolytic degradation. As hypothesized, the resulting system is self-contained and highly productive using only the biological components present in the extract. Next, I pursued the development of a novel CFPS platform based on extracts derived from an up-and-coming molecular biological workhorse, the bacterium *Vibrio natriegens*. I reasoned that preparative scale synthesis of ncAA-containing proteins could benefit from *V. natriegens*' supercharged suite of translation machinery[28, 29]. By

optimizing cell growth, lysis, and reaction parameters, I established a highly-productive CFPS platform based on *V. natriegens* extracts. The platform demonstrates utility in the synthesis of short peptide products (another non-standard target that is difficult to synthesize recombinantly[30, 31]) and is capable of supporting ncAA incorporation in CFPS. My *V. natriegens* CFPS system is remarkably easy to prepare and use, which I believe will increase the accessibility of CFPS systems. Finally, equipped with a highly-productive CFPS system optimized for ncAA incorporation, I wanted to begin synthesizing and interrogating proteoforms featuring specific PTMs. Specifically, I was interested in generating proteoforms featuring different patterns of serine-linked phosphorylation by directly incorporating *o*-phosphoserine (Sep) into targets. I hypothesized that Sep incorporation in CFPS could be improved by using by applying improved Sep translational components in a Sep-specific engineered lysate. This effort is ongoing. What I hope to achieve by the end of this work is i) a set of tools enabling the synthesis of specific proteoforms to accelerate efforts to understand and characterize patterns of PTMs, and ii) strategies for facilitating the expansion of these tools into new fields for compelling new applications and discoveries.

1.1.2 Dissertation outline

The rest of Chapter 1 will detail the review of literature, spanning from an introduction to protein biology and modification, to a discussion of methods of ncAA incorporation and cell-free biology. This will be used a means of motivating my work to endeavor to develop novel platforms for designer proteoform synthesis to enable fundamental discovery.

Chapters 2-4 will describe efforts to improve ncAA incorporation in cell-free systems using lysates derived from recoded strains of *E. coli*. When I entered this area of study, the state-of-the-

field was limited to only a small number (1-3) of ncAA incorporations into a single peptide, limiting the scope of potential proteoforms that could be generated. Around that time, efforts to improve ncAA incorporation culminated with the generation of viable *E. coli* strains lacking RF1. Chapter 2 focuses on the development of a CFPS platform from one such strain, the partially-recoded *rEc.E13.ΔA*. I will describe genomic modifications made to the strain which collectively increased the productivity of its lysates fourfold, and demonstrate that this productivity increase is caused by stabilization of nucleic acid substrates in the system. I will go on to highlight the improved capacity for ncAA incorporation in the system due to its lacking RF1 activity. Limitations in the partially-recoded lysate and the contemporaneous advent of the fully-recoded strain *C321.ΔA* led us to pursue a similar effort in *C321.ΔA*, as described in Chapter 3. Here I will describe genomic modifications made to the strain which increased the productivity of its lysates > fourfold, and present data supporting hypotheses regarding the effects of those modifications. I will go on to demonstrate that the complete recoding in the strain permits incorporation of up to a staggering 40 ncAAs into a single peptide with near-perfect fidelity. Finally, Chapter 4 will describe the creation of a one-pot platform for *in vitro* ncAA incorporation. Here I will present a series of synthetic genomic inserts designed to imbue a *C321.ΔA* derivative with the ability to natively synthesize the viral T7 RNA polymerase that is commonly used to catalyze transcription in cell-free systems. I go on to generate a series of T7 RNA polymerase mutants and describe a set of protective mutations that prevent OmpT-mediated proteolysis of the enzyme during lysate preparation. The final platform described operates independently of supplementation with purified polymerase, making it suitable for scale-up to industrial reaction volumes and facilitating expansion of the system into new fields for compelling synthetic biology applications.

In Chapter 5 I describe the establishment of a novel cell-free protein synthesis platform based on lysates derived from the fast-growing bacterium *Vibrio natriegens*. I report the stepwise series of process optimizations that ultimately yield a CFPS platform comparable to state-of-the-art systems composed of *E. coli* lysates. I go on to demonstrate potential applications of the platform, which include a high capacity for synthesis of short peptide products and the ability to support ncAA incorporation. Overall, the platform described is facile to grow, prepare, and use, and is perfect for early attempts at the use of CFPS. We anticipate that it will help grow the field.

Chapter 6 will recount my efforts to develop an improved CFPS system for incorporation of phosphorylated ncAAs, namely *o*-phosphoserine (Sep). I will describe the recombinant parts necessary to incorporate Sep, and present my efforts to encode improved parts into a single plasmid for in-cell expression. I will also present my parallel effort to develop a Sep-optimized chassis strain. Finally, I will highlight one application for this technology, presenting results of an exploration into the role phosphorylation plays in regulating the glycolytic enzyme triosephosphate isomerase.

In Chapter 7, I summarize my work and look towards future directions for both ncAA incorporation in CFPS as well as the use of these systems for enabling studies of PTMs.

1.2 Review of Literature

1.2.1 Post-translational Modification of Proteins

Proteins are involved in practically every aspect of life, serving roles ranging from structural to catalytic. These templated polymers are synthesized by the ribosome, a massive biomolecular machine that decodes the information contained in messenger RNA (mRNA) by

translating 61 triplet nucleotide codons into 20 canonical amino acids which are covalently linked at the peptidyl transferase center to form linear chains referred to as “primary structures” – a process known as translation. Nearly all proteins require folding of the amino acid chain, mediated by noncovalent hydrogen bond formation, into higher order secondary (e.g. beta sheet and alpha helix) and tertiary (e.g. beta barrel) structures in order to become active. The final structure of a protein is critically important. Indeed, investigations in the relationship between protein structure and function have demonstrated that these two characteristics are so closely related such that one implies the other.

After emerging from the ribosome and folding into its final conformation, a protein may not yet be fully functional. The newly-formed polypeptide chain must undergo additional chemical processing steps to fully activate. These steps are most often mediated by a tightly regulated suite of enzymes which catalyze the chemical modifications necessary to promote activation of a target protein. Because they occur after translation has completed, such modifications are collectively known as “post-translational modifications”, or PTMs, and can be divided broadly into two classes. In the first class, an immature protein is proteolytically cleaved to remove some of the polypeptide fragments. In the second class, enzymatic processing of a nascent protein results in the chemical modification of amino acid sidechains that modify protein function. This thesis will focus primarily on this second class of PTM, though for the sake of completeness a brief introduction to both classes of PTM (including the post-translation cleavage of proteins) will be presented in the next section.

1.2.1.1 Post-translational Processing by Proteolysis

The modification of proteins through removal of fragments from the nascent polypeptide is a ubiquitous phenomenon. This process is mediated by protease enzymes, which catalyze targeted cleavage of peptide bonds in the target via hydrolysis resulting in the irreversible formation of new N- and C-termini and separation of the resulting peptides. Such proteolytic processing events can result in protein activation, inactivation, altered function, or changes in cellular localization, and regulate a large number of cellular processes including DNA replication, cell proliferation, and cell death[32, 33].

There are many illustrative examples of protein function being altered by protease activity. Perhaps the best-known example is the regular removal of the N-terminal (leader) methionine residue in eukaryotic and prokaryotic proteins during their maturation. Catalyzed by a class of proteases known as methionine aminopeptidases, this PTM stabilizes proteins by essentially replacing the bulky leader methionine with a nonbulky N-terminal residue[34]. In another well-established example, in many vertebrates the digestive protease trypsin is initially synthesized in an inactive form (trypsinogen) to delay its activity until it has localized to the small intestine. Once there, the enzyme enteropeptidase cleaves a six-amino-acid peptide from the N-terminus of trypsinogen to yield an active trypsin protein[35]. The vertebrate metabolic hormone insulin is similarly initially synthesized as proinsulin, an immature form of the protein with limited signaling activity. Excision of a 31-amino acid peptide from proinsulin by the coordinated activities of the proteases PC1, PC2, and PC3 is necessary to generate the fully active hormone[36]. Finally, many eukaryotic protein hormones are synthesized in a single large amino acid chain that must be cleaved by proteases to liberate individual hormone moieties. For example, in humans the

polypeptide pro-opimelanacortin is cleaved in a tissue-specific manner to yield several important peptide hormones including the melanotropins and β-endorphin[37].

While these modifications are certainly extremely important and worth mentioning here, this thesis will not focus on polypeptide cleavage PTMs. Rather, the focus will be on the class of PTMs in which chemical side-chains are modified which is introduced in the next section.

1.2.1.2 Post-translational Modification by Side Chain Alteration

Proteins are commonly regulated by the covalent modification of specific amino acid sidechains to enact changes in function **(Figure 1.1)**. These processes are mediated by a diverse set of modification enzymes, which generally specialize in the addition or removal of a particular functional group at specific residue side chains resulting in conformational changes in the target protein leading to a change in function. Unlike proteolytic processing, this type of PTM is often reversible such that modifications can be enacted only at specific times to generate needed functions. In other words, a single protein molecule can be combinatorially modified with several different functional groups at different positions, each contributing to the overall structure of the protein such that the final functional result is influenced by the entire pattern of modifications – then, once that particular function is no longer needed, modifications can be removed to revert the protein to a different functional state. A single protein may be differentially regulated in this way in different contexts, such as different cellular compartments or different tissue types, generating a massive level of regulatory and functional complexity. To get a sense of the impact that the use of these modifications can have, one need only consider the bacterium *Escherichia coli* and a human. Relatively speaking, the human does not have that many more protein-coding genes than *E. coli* (~22,000 genes[38] and 4,288 genes[39], respectively) but the structural diversity permitted

by PTM addition expands the repertoire of functionality accessible to the human genes, contributing to the tremendous amount of difference between the two species[40]. Importantly, regulation of protein function via PTMs enables cell to respond to stimuli much faster than by enacting changes in gene expression via transcription and translation. Understandably, understanding these kinds of modifications – the modification enzymes involved, the underlying regulatory networks, the ways in which different modifications work together to activate a given function – remains a challenge of significant interest to the research community.

Here I will introduce a handful of the best-characterized functional groups. This review will not be exhaustive, and a number of larger (albeit extremely important) modifications such as glycosylation[41] and ubiquitination[42] will not be considered. Instead, my focus will be on the smaller adducts that could conceivably be introduced co-translationally into proteins using techniques introduced later in this chapter.

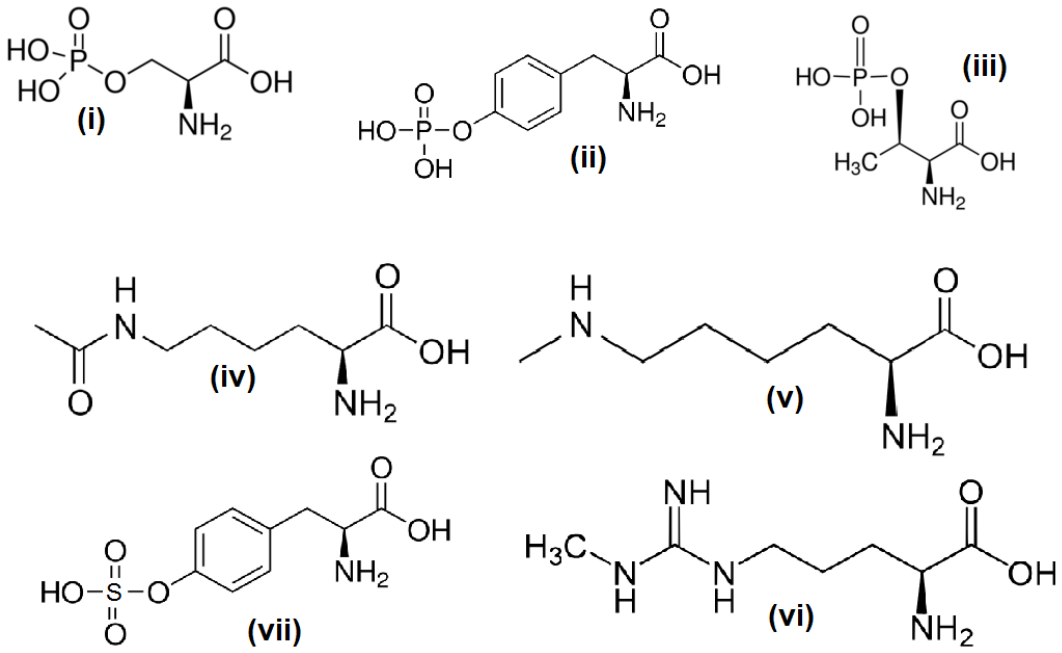


Figure 1.1. Chemical structures of amino acids bearing the post-translational modifications discussed in this chapter. (i) phosphoserine; **(ii)** phosphotyrosine; **(iii)** phosphothreonine; **(iv)** acetyllysine; **(v)** mono-methyllysine. Note that the terminal nitrogen can bond to up to two more methyl groups; **(vi)** methylarginine; **(vii)** sulfotyrosine.

1.2.1.2.1 Phosphorylation

Phosphorylation is the reversible attachment of a phosphate group (R-**O-PO$_3^{2-}$**) to a protein **(Figure 1.1.i-iii)**. Usually observed at the side chains of serine, threonine, and tyrosine, this modification is regulated by the opposing activities of kinase enzymes (which catalyze the transfer of a γ-phosphate from adenosine triphosphate to the amino acid sidechain) and phosphatase enzymes (which facilitate removal of phosphate groups via hydrolysis)[8]. Phosphorylation is one of the most important polypeptide chain modifications in nature and plays a pivotal role in protein folding, function, stability, and localization[8]. This ubiquitous modification is present in all domains of life: approximately one-third of all eukaryotic proteins are predicted to be phosphorylated[43] and phosphorylated proteins have been found in archaea[44] and bacteria[45].

Protein phosphorylation is critical for the development, growth, function and survival of all organisms and is involved in the regulation of a myriad of biological processes, including cell cycle control, receptor-mediated signal transduction, differentiation, proliferation, and metabolism[10, 46-50]. Moreover, perturbation of phosphorylation patterns and breakdown of associated control systems is often a hallmark of disease states such as cancer, inflammation, Alzheimer's disease, and congenital disorders[51-54].

A vast number of phosphorylation events have been described. Many of the most well-known examples center on the use of so-called kinase cascades to amplify biological signals. In such a system, receipt of a particular stimulus causes activation of a kinase, which we'll call kinase A. Kinase A, now active, then goes on to activate kinase B via phosphorylation. Kinase B goes on to activate kinase C, which goes to activate kinase D, *etc.* Ultimately, a response is triggered by kinase-mediated phosphorylation of the regulatory molecules that initiate the required physiological process. Because a single kinase enzyme can phosphorylate many targets, at each step of the cascade the number of active agents increases. In this way, even a small stimulus can be amplified by many orders of magnitude to enact a significant response[55, 56].

Cells frequently employ this strategy to enact sweeping changes in gene expression in response to stimuli. In these cases, the final targets of the cascade are transcription factors which upon activation bind to genomic DNA to enact new genetic algorithms. The most characterized examples here are the mitogen-activated protein kinase (MAPK) cascades. In mammals, four MAPK cascades have been described, all built around three core kinases (MAPKKK, MAPKK, and MAPK)[55]. These cascades are involved in cellular responses to numerous stimuli, triggering a reaction by the eventual phosphorylation of hundreds of cellular substrates[55].

There are also many examples of individual protein functions being regulated by phosphorylation. The so-called “guardian of the genome”[57], p53 is a tumor suppressor protein that plays a critical role in the regulation of cell proliferation[58, 59]. Activation of this protein in response to cell damage or perturbations in cellular physiology can lead to cell cycle arrest or even apoptotic cell death[59], a response designed to protect the multicellular organism at the expense of the cell. Twenty-four different phosphorylation sites have been identified in p53[60, 61], several of which have been implicated in directing facets of the protein’s stability and activity. For instance, several studies have identified N-terminal phosphorylation of human p53 at serine 15 and serine 20 as an early step in activation by promoting stability of p53 and allowing it to build up in the damaged cell[62-64]. Serine 392 phosphorylation increases after UV-induced DNA damage, and several studies have concluded that this modification stimulates the DNA binding activity of the protein[65-68]. Still, the exact mechanisms underlying p53 regulation are incompletely understood[63]. Indeed, at least one conflicting report has concluded that phosphorylation of key serine residues is not actually required for activity[69]. A second example is glycogen synthase, a key enzyme involved in the conversion of glucose into glycogen. When glycogen synthesis isn’t needed (*i.e.* when environmental glucose levels are low) this enzyme is phosphorylated by the aptly-named glycogen synthase kinase 3 (GSK3) at serine residues 640, 644, 648, and 652, which collectively deactivate it[70, 71]. Upon an increase in glucose levels, these phosphorylations are removed by protein phosphatase 1 reactivating glycogen synthase and permitting glycogen synthesis[72].

Phosphorylation is also used to modulate the ability of a modified protein to noncovalently interact or associate with other proteins. This is best exemplified in DNA-wrapped histone

proteins, where the addition/removal of phosphate groups regulates associations with a variety of transcription factors and histone modification enzymes. A well-established phenomenon in this space is the hyper-phosphorylation of histone protein H2A(X) at serine 139 at damaged sections of genomic DNA[73-75]. This modification is primarily catalyzed by the kinase ataxia-telangiectasia mutated (ATM), which interacts with a number of other substrates in a larger overall response to genotoxic stress[76]. The addition of the phosphate group to this histone facilitates repair of the DNA lesion by promoting association with a suite of protein involved in chromatin accessibility and break repair[77].

1.2.1.2.2 Methylation

Methylation is the addition of up to 3 methyl groups (R-**CH_3**) to a protein, canonically at the side chains of lysine and arginine residues **(Figure 1.1.v-vi)**. Methylation is catalyzed by methyltransferases, which transfer a CH_3 group from *S*-adenosylmethionine to the recipient side chain. Like phosphorylation, this action is reversible, with the removal of methyl groups performed by demethylases.

Though it was first characterized in 1959[78], study of protein methylation didn't really take off until recent decades[79] when advances in molecular biological techniques enabled investigations into the physiological impact of this PTM. Now, a wealth of information has been gleaned about this PTM, largely in the realm of histone modification and chromatin biology. Methylation is believed to influence chromatin and gene expression by acting as a sort of "handle" for recruiting various proteins to the area to enact a specific response[79]. For instance, methylation of lysine 4 in histone 3 is highly conserved and is known to facilitate transcription by recruiting chromatin remodeling factors CHD1 and BPTF which open the chromatin to facilitate

RNA polymerase binding[79-82]. A more complex methylation event has been characterized at lysine 9 of histone 3 (H3K9), where di- and tri- methylation is linked with the establishment of closed, transcriptionally-inactive chromatin by recruitment of heterochromatin protein 1[83-86]; conversely, mono methylation at H3K9 is associated with the promoters of highly-transcribed genes[83]. Many other histone methylation events have been catalogued, and their rich variety helped lead to the now widely-accepted notion of a "histone code" describing the net effects that various combinations of histone PTMs have on chromatin structure, DNA accessibility, and gene regulation[87].

Non-histone examples of protein methylation have also been described. Most of this work has centered around a familiar protein, p53. Methylation of four lysine residues and three arginine residues dictate some p53 behavior, both as an activator and a repressor[88-92]. As in histones, reversibility of methylation has been demonstrated with p53: lysine-specific demethylase 1 has been implicated in the demethylation of lysine 370, repressing p53 function in the DNA damage response[93]. A second example in eukaryotes is methylation of arginine residues in RNA-binding proteins, where the modification has been observed regulating ribonucleoprotein assembly and localization[94, 95], splicing[96], and mRNA stability[97].

1.2.1.2.3 Acetylation

Acetylation is the addition of an acetyl group (R-**O-CH$_3$**) to a protein **(Figure 1.1.iv)**. Unlike the modifications discussed above, with regards to location this modification can take on two slightly different forms. In the first form, an acetyl group is covalently transferred from acetyl-coenzyme-A (Ac-CoA) to the α-amino group at the N-terminus of a nascent protein by ribosomally-associated N$^{\alpha}$ acetyltransferase complexes. This modification has been observed in

all kingdoms of life, though it is especially common in higher eukaryotes where it present on nearly all proteins[98, 99]. A universal role for N-terminal acetylation has not been established, though in some specific cases it has been observed regulating protein-protein interactions[100-102], localization[103, 104], and degradation[105, 106]. The second form of protein acetylation is analogous to the modifications discussed above, and describes the covalent addition of acetyl groups to amino acid sidechains. It is this second form of acetylation that this thesis is most concerned with. Canonically, this occurs at lysine residues, and is mediated by the activity of acetyltransferase enzymes which transfer the acetyl group from Ac-CoA to the ε-amino group of the target lysine. This modification is reversible, with deacetylase enzymes catalyzing acetyl group removal via a hydrolysis mechanism[107].

Like methylation, acetylation is most thoroughly characterized in the context of histone biology[107]. For instance, acetylation of histones has a defined role in influencing their nuclear localization. An example is in histone 4, where transient acetylation occurs immediately after synthesis at histones 5 and 12 resulting in chaperone recruitment and nucleosome assembly. After nucleosome assembly, these acetylations are removed, and the histones can be acetylated again in a chromatin-specific manner[108-110]. Additionally, acetylation causes the lysine side chain to lose its positive charge which, in histones, disrupts the protein's ability to form salt bridges with the negatively charged phosphate backbone of DNA[111-113]. This loosens the surrounding chromatin and increases accessibility of RNA polymerases in the region, leading to the generally-accepted notion that stable histone acetylations lead to transcriptional activation in the surrounding chromatin. This has been directly observed for acetylation of histone 4 at lysine 16 (H4K16),

H3K56, H3K64, and H3K122[113, 114]. By an analogous mechanism, H4K91 acetylation loosens chromatin by disrupting histone-histone interactions to destabilizes nucleosomes[115].

Acetylation has also been characterized in non-histone proteins. Once again, p53 exemplifies the use of this PTM in a regulatory role. In this case, p53 can be activated by acetylation at K120 and K164 leading to activation of p53-responsive genes and cell cycle arrest[116, 117]. Conversely, the activities of deacetylases HDAC1 and SIRT1 result in deacetylation of p53, thereby deactivating transcriptional activity. In a second exemplary case, the transcription factor FOXO1 can become reversibly acetylated at multiple sites, regulating its localization in the nucleus and altering its genetic targets[118].

1.2.1.2.4 Sulfation

Sulfation is the covalent addition of a sulfate group (R-**O-SO$_3^-$**) to a protein **(Figure 1.1.vii)**. Though first reported in the literature over 50 years ago[119], this chemical modification has received relatively little attention and has only recently been accepted as a significant PTM. To date, sulfation has been observed predominantly in higher eukaryotes at the side chains of tyrosine residues, where sulfotransferase enzymes catalyze the transfer of the sulfate group from phosphoadenosylphosphosulfate to the side chains of the target tyrosine[3]. Despite being a relative newcomer to the PTM stage, sulfation has been implicated in a number of key cellular processes, including hemostasis[120, 121], visual functions[122-124], viral entry into cells[125], and ligand/receptor binding and recognition[126].

Most of the investigations into sulfation conducted to date have concluded that the PTM primarily influences protein activity by regulating protein-protein interactions, with the sulfate group acting as an essential recognition element to enable binding[126]. One example occurs in

higher vertebrates during the formation of a thrombus, or platelet aggregate, at the site of blood vessel injury. While the complete mechanism underlying this phenomenon has yet to be fully unraveled, it is known that the von Willebrand factor (vWf) and the platelet membrane protein GP Ibα are involved in the early steps of the process[127, 128]. A key event in this process is the binding of platelets to vWf, an association which is heavily dependent on the sulfation of three residues (Y276, Y278, and Y279) within GP Ibα[120, 129]. Sulfation has also recently been identified as a key regulator of chemokine signaling in humans. For example, the chemokine receptor CCR5 was shown to be tyrosine sulfated, and incubation of CCR5-expressing cells in culture with a sulfation inhibitor (sodium chlorate) decreased binding of MIP-1α and MIP-1β – two natural chemokine ligands for CCR5[125]. Sulfation of tyrosine 14 in CCR5 seems to be particularly indispensable for binding MIP-1α[126].

Unlike the PTMs discussed earlier in this chapter, sulfation so far does not appear to be nearly as transient or dynamic. Sulfotransferases appear to be constitutively active wherever they are found, and the majority of sulfated proteins identified to date are either secreted or displayed on exterior membrane-bound proteins, which may preclude removal of the sulfate adduct[126]. Sulfatase enzymes (*i.e.* enzymes that catalyze removal of a sulfate group via hydrolysis) exist, but none are known definitively to facilitate removal of the sulfate group from tyrosine[126]. Sulfates, once added to proteins, appear able to remain there indefinitely.

In closing this section, it is important to again stress that any given protein will usually receive many different kinds of PTMs at many different sites concurrently. The structure and overall function of a modified protein is attributable to the net effects of the entire pattern of

modifications, making it very difficult to conclusively attribute a specific function or behavior to any one specific modification event.

1.2.1.3 Methods for Identifying and Characterizing PTMs

As stressed a number of times already in this chapter, acquiring a complete understanding of a protein of interest (POI) requires a complete deciphering of all PTMs present on the POI, including an understanding of the regulatory timing of those modifications and the ways in which the PTMs collectively alter the structure of POI to enact some change in function. These studies are challenging, as PTMs are not genetically encoded and so cannot be identified or described via DNA sequencing. Accordingly, researchers have spent decades developing techniques for identifying PTM patterns on proteins of interest. In the next section I will provide a brief introduction to some of these techniques. For the purposes of this thesis I will very broadly divide these approaches into two groups – i) antibody-based approaches and ii) mass-spectrometry-based approaches. I will finish with a discussion of methods for functionally characterizing specific PTMs.

1.2.1.3.1 Antibody-based Approaches for PTM Identification

Antibodies are Y-shaped protective proteins that play a critical role in the immune responses of many higher eukaryotes[130-132]. In this context, antibodies bind to moieties – known as “antigens” – on invading pathogens, thus targeting the invader for attack by other parts of the immune system. Outside of an immune response, antibodies retain the ability to recognize and bind to their specific targets, a feature that has been exploited by researchers for decades in the development of countless analytical assays. The techniques introduced in this section will be presented through the lens of PTM identification, but all of them have been modified many times

for use on different kinds of protein targets. Antibody use is by no means limited to investigations into PTMs.

1.2.1.3.1.1 Immunoprecipitation

Immunoprecipitation (IP) is the process by which a POI is selectively removed from a complex mixture based on its ability to be bound by antibodies that are typically immobilized on some solid substrate such as beads or resin **(Figure 1.2.A)**[133]. Proteins which do not bind are washed away, and by the end of the process only the POI remains bound to the antibody array. The isolated POI can then be separated from the antibodies for downstream analysis[133].

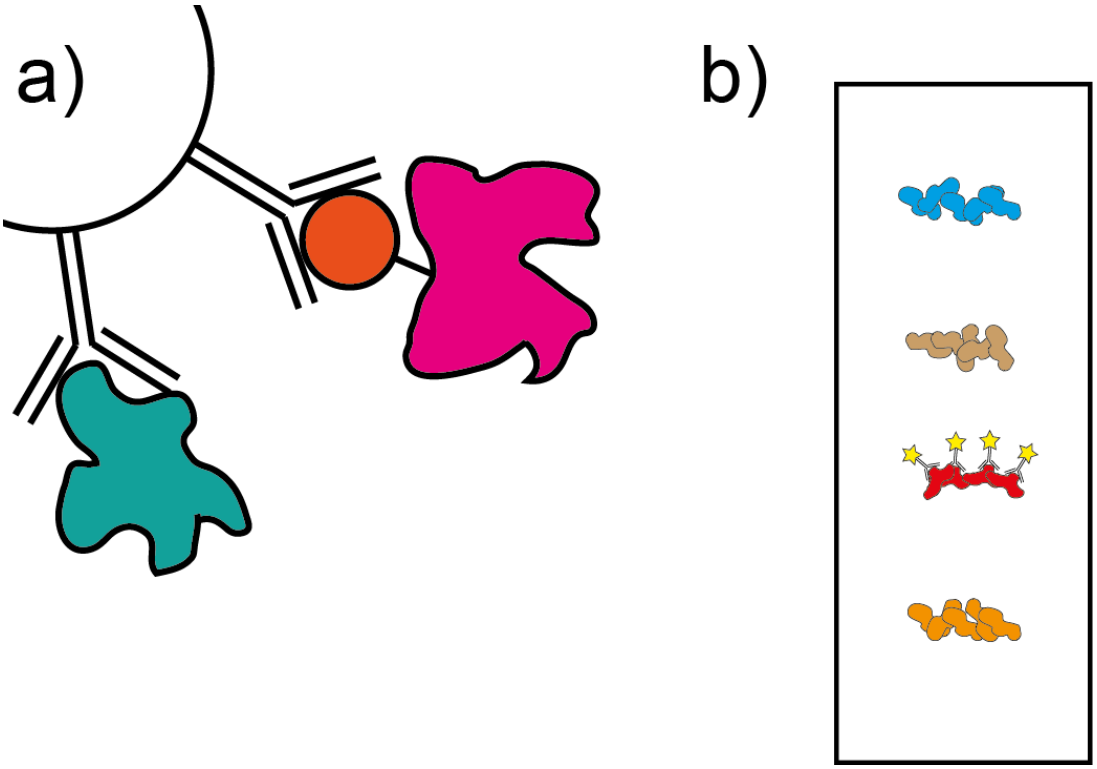


Figure 1.2. Antibody approaches to PTM identification. A) Immunoprecipitation. Immobilized antibodies bind to a protein (left) or PTM (right) of interest to selectively retain them for further study. **B)** Western blot. The protein of interest (red) is visualized via the binding of an antibody fused to a visualization domain. Other proteins in the sample are pictured here for clarity, but in practice would not be visible absent antibody binding.

In PTM identification, most of what can be accomplished by IP depends on the antigen and specificity of the antibody(ies) being used. Antibodies have been generated that bind to all instances of a particular POI **(Figure.1.2.A, left)** or all instances of a specific kind of PTM (*e.g.* phosphoserine[134], trimethyllysine[135]; **Figure 1.2.A, right**) but due to their generality, in both of these cases the acquisition of meaningful information regarding protein-PTM associations requires additional downstream analysis. In a handful of very rare cases, antibodies have been developed that specifically only bind a specific PTM attached to a given protein[136, 137], but in general, IP is most commonly performed as a precursor to a more in-depth analysis that explicitly establishes a POI-PTM association[133].

Chromatin immunoprecipitation (ChIP) is a specific type of IP that is widely used to identify where specific histone PTMs are located throughout the genome. This approach uses antibodies that bind instances of specific PTMs on histones of interest. By chemically crosslinking nucleosomes to bound DNA, histone proteins bound by a particular antibody are covalently connected to the DNA bound by the encompassing nucleosome complex. By coupling this technique with next-generation high-speed sequencing (ChIP-Seq) a tremendous amount of genomic histone PTM occupancy data has been generated, establishing the prevailing view that histone modifications demarcate functional elements in eukaryotic genomes[83, 138-143].

1.2.1.3.1.2 Western Blotting

Western blotting is the process by which POIs are visualized by the binding of an antibody **(Figure 1.2.B)**[144]. In this case, the antibody is usually fused to a visualization domain, commonly an enzyme that can generate a colored or luminescent product (*e.g.* horse radish peroxidase) or a fluorescent marker (*e.g.* Alexa dyes) such that the presence of a target of interest

can be inferred by antibody binding and the subsequent emergence of a colored or luminescent substrate and/or fluorescent signal. In performing a western blot, proteins in a sample are typically separated via SDS-PAGE electrophoresis and transferred to a membrane, where they can be exposed to antibody/visualization domain fusions. Proteins featuring the antigen for the antibody become bound and can later be visualized by virtue of the associated visualization moiety[144].

A western blotting step is one possible analytic follow-up to an IP, serving to establish a connection between a PTM and a POI. This step differs in the antibody used, depending on the sort of antibody used in the initial IP step, such that the western blot complements the IP in order to establish a PTM-POI relationship. If the IP step utilized an antibody that binds a particular POI, then the western blot utilizes antibodies that bind specific kinds of PTMs[145]. In this way, any antibodies in the western blot step that enable visualization of the POI indicate that their target PTM is present somewhere in the POI. Conversely, if the IP step used an antibody that binds a particular kind of PTM, then the western blot step uses antibodies that bind various POIs[146]. In this case, any antibodies in the western blot step that enable visualization of the sample indicate that their target POI is associated with the PTM initially isolated during the IP step. In either case, the combination of IP and western blot serve to associate specific kinds of PTMs with protein(s) of interest.

Western blots can also be used independently of IP to establish the presence of a given PTM on a POI. In this case, two blots can be performed in parallel with one blot using an antibody against a PTM and the other using an antibody against a POI. After visualization, overlap of the bands on each blot establish a link between the POI and PTM being investigated[147].

1.2.1.3.1.3 Limitations to Antibody-based Approaches for Identifying PTMs

The use of antibodies to discern targets of post-translational modification has been foundational in advancing the field and is still regularly used. Still, this approach suffers from some significant drawbacks. The most obvious limitation is that, in all cases, verified antibodies have to actually exist against your various target(s) of interest[5, 148]. The creation of antibodies against novel targets is non-trivial, and a project can be stopped right in its tracks if any of the targets to be interrogated lack extant antibodies to bind them. Additionally, antibodies aren't always perfect, and errant binding of off-targets is known to occur[5, 149, 150]. As a result, false positive results are always a concern using these approaches. Conversely, for IP in particular proper antibody-target binding is dictated by the structures of the two proteins – in specific cases, antibody binding to a target can be limited by structural changes in the target brought on by PTMs and other effects, leading to false negative results[150]. Finally, the most significant drawback to these approaches is that they, at best, establish that a given PTM is present on a POI, and are usually incapable of producing positional data revealing the specific residues at which the PTMs occur.

1.2.1.3.2 Mass spectrometry approaches for PTM identification

For more than 60 years, mass spectrometry (MS) has enabled the separation and subsequent detection of molecules ionized in the gas phase[151, 152]. The name "mass spectrometry" is actually a bit of a misnomer, as the process actually determines the mass-to-charge (m/z) ratio of analytes rather than their mass. The fundamental output of a MS analysis is the mass spectrum, which is a plot of ion abundance versus m/z presented in terms of Daltons (Da) per unit charge[151].

At its most basic level, mass spectrometry can be divided into 3 processes. First is ionization, which generates ions in the gas phase from a sample of interest. This is followed by mass analysis, during which the newly-formed gaseous ions are transferred into the vacuum of a mass analyzer and separated based on their m/z values. Last is detection, which in some instruments occurs when particular ions are able to exit the mass analyzer and collide with the detector[151]. More modern instruments perform mass analysis and detection simultaneously, generating mass spectra based on the circular movement of trapped ions[153]. Collectively, these processes constitute a mass analysis cycle. These mass analysis cycles can be completed 10s of times per second, allowing for real-time monitoring and recording of the ions entering the analyzer.

Early ionization processes (*e.g.* electron ionization) were "hard", and would frequently result in the fragmentation of the molecule(s) being ionized[151]. This limited the use of MS to relatively low-molecular-weight analytes, which precluded analysis of larger biological species. However, the development of "soft" ionization techniques in the 1980s[154, 155] made possible the ionization (and, by extension, MS analysis) of large, biologically-relevant molecules. In the decades since, an enormous amount of effort has gone into developing MS-based approaches for the analysis of proteins. In particular, practically every investigation performed today to analyze and identify PTMs includes rigorous MS analysis of the protein samples[4, 5, 156, 157]. Such efforts have been aided tremendously by advances in computational power.

The prevailing approaches to MS-based proteomics (namely, bottom-up and top-down; **Figure 1.3**) will be introduced in this section. A key concept common to both of these approaches is that ions can be subjected to multiple rounds of MS in tandem (MS/MS)[158, 159]. Very simply put, MS/MS entails the isolation of precursor protein/peptide ions with particular m/z values in a

mass analyzer which are then fragmented in such a way that cleavage products generally break randomly along peptide bonds. The cleavage products are then introduced into a second mass analyzer, which generates a mass spectrum describing the abundance and m/z values of ions generated during fragmentation. The MS/MS spectra generated (commonly referred to as the progenitor ion's "mass fingerprint") can then be computationally or manually compared to a series of theoretical spectra to identify the precursor ion. MS/MS can be done iteratively (*e.g.* MS^3, MS^4) with each subsequent mass analysis performed on ions generated by fragmenting precursor ions from the preceding mass analysis. This approach is the foundation on which practically all MS-based proteomics is built[4, 158, 160].

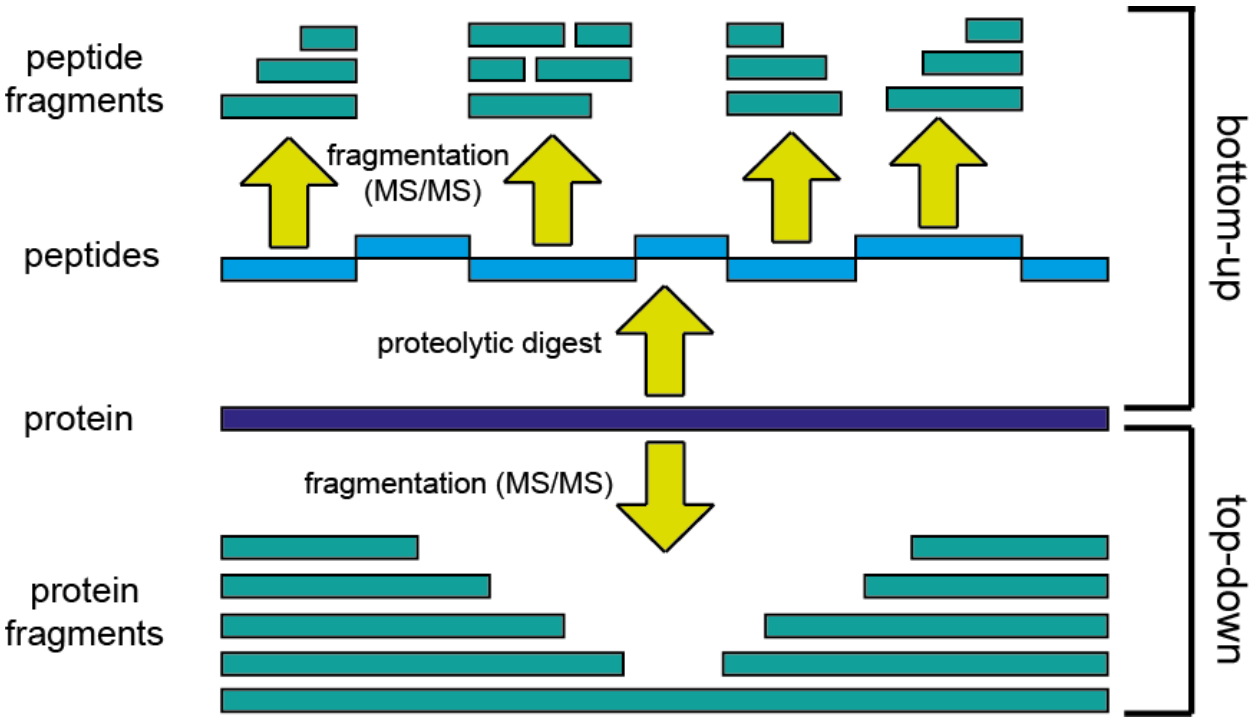


Figure 1.3. Simplified cartoon schematic depicting mass spectrometry-based proteomics approaches. Bottom-up mass spectrometry is the analysis of proteolyzed peptides and their fragment ions. Top-down mass spectrometry is the analysis of intact proteins and their fragment ions.

1.2.1.3.2.1 PTM Identification Using Bottom-up Proteomics

Bottom-up proteomics refers to the identification of a protein/proteoform in a sample based on MS data generated using peptides released from the protein via proteolysis **(Figure 1.3, top)**. At its most basic, this technique can be used to investigate a single target. In this case, the protein of interest is typically separated from most of the other proteins in the biological sample (usually by electrophoresis or affinity purification). The most commonly used protease for this is trypsin due to its high specificity for its canonical cleavage site after lysine and arginine residues[161]. In some cases tryptic peptides are insufficient for analysis (being either too long or too short); in such cases, other proteases such as chymotrypsin, AspN, Lys-N, or Glu-C can be used[162]. Following proteolysis, a second round of enrichment separates peptides bearing a specific PTM from the rest of the peptides, commonly via some of the antibody-based approaches discussed above[4]. Finally, the peptide mix is typically fractionated on a high-pressure chromatography column before ionization (usually via electrospray) and entry into the instrument[158]. As the m/z values for intact peptides are rarely sufficient for identification, tandem MS/MS is almost always applied, and protein inference is achieved by assigning peptide sequences to proteins. For PTM identification, the theoretical spectra used to achieve peptide identification can be made to include the mass of a PTM of interest at residues that could feature that PTM – in this way, site-specific PTM localization is possible[4, 158-160]. Advances in methodology and computational processing power have expanded the scale and throughput of this approach to enable the simultaneous analysis of many proteins in a sample, an approach referred to as “shotgun bottom-up proteomics”[159, 163].

Despite being one of the earliest techniques developed for MS-based explorations of PTMs, bottom-up remains the gold standard technology in this space. An exemplary case, Liu *et al* recently identified 40 di- and trimethylated peptides in a lysate derived from a human cell line using a shotgun approach[164]. By a similar approach, Svinkina *et al* employed a novel antibody against acetylation for enrichment and subsequently identified 10,000 acetylated peptides corresponding to ~3,000 proteins in human Jurkat cells using shotgun proteomics[165]. In another example, Zappacosta *et al* conducted a high-throughput bottom-up analysis of proteins from rat liver samples and identified 16,000 unique phosphorylation sites from a single experiment[166].

While bottom-up MS approaches remain widely used to this day, there are many key limitations to these methods that must be carefully considered when designing experiments and interpreting the data obtained. The first of these is complications in protein inference due to peptide redundancy. Because the peptides analyzed in bottom-up approaches are relatively small, it is possible for a peptide detected to be theoretically derived from multiple isoforms of a protein or even from multiple *different* proteins. In such a case, it becomes extremely difficult to make a protein inference based on the detection of that peptide[167, 168]. A second limitation is that even a very successful bottom-up experiment can provide incomplete information about the PTMs on a protein due to low sequence coverage – the modifications on any peptides that aren't detected during the experiment remain unknown[157, 167, 168]. Finally, these approaches are unable to establish connectivity between PTMs detected at distal parts of a protein, as detection of two modified peptides does not guarantee that those peptides were initially derived from the exact same precursor protein molecule[157, 167, 168].

1.2.1.3.2.2 PTM Identification Using Top-down Proteomics

As opposed to the bottom-up approach, which features peptide ions as the fundamental analyte, top-down proteomics ionizes *undigested, intact protein molecules* and directs them into the mass spectrometer for analysis **(Figure 1.3, bottom)**[157, 163, 167, 169]. Because protein molecules are analyzed intact, the presence of covalent adducts (such as PTMs) can be inferred from characteristic shifts in m/z values from the initial mass spectra generated using this approach, simplifying the association of PTMs with protein(s) of interest. Tandem MS can be employed as in the bottom-up approach, with the fundamental difference being that in top-down the precursor ions are intact proteins as opposed to peptide fragments. Similar to bottom-up MS/MS, mass analysis of the cleavage products in a top-down experiment generates "mass fingerprint" spectra which can be compared to a library of theoretical spectra to identify proteins and site-specifically locate PTMs[157, 167].

By analyzing proteins in their fully-intact forms, top-down methods alleviate many of the limitations to the bottom-up approach described above. Mass analysis and fragmentation of entire protein molecules provides 100% sequence coverage, ensuring that all stable modifications to the protein are detected. In the same way, connectivity between distal PTMs is readily established by integration of data from both precursor ion and fragment ion spectra. Finally, in the absence of potentially-redundant peptides the "protein inference problem" is all but eliminated.

Adoption of high-throughput, "shotgun" techniques based on top-down proteomics has led to the rapid, unequivocal identification of protein PTM patterns. An early application of this technology focused on mouse cardiac myosin binding protein C, identifying all of the phosphorylation sites in both the truncated and full-length forms of the protein and characterizing sequential phosphorylation events[170]. In a more recent study, Anderson *et al* leveraged a high-

throughput top-down approach to identify 684 unique proteins featuring thousands of different combinations of PTMs from a human cancer cell line[171]. One particularly intriguing application of this strategy, so-called native proteomics, ionizes noncovalently associated protein complexes for infusion into the mass spectrometer[172]. This approach could eventually be used to establish functional roles for specific PTMs with regards to protein-protein interactions.

Despite the many key advantages enjoyed by top-down proteomics, research groups have been slow to adopt the method for several reasons. First, it is technically challenging[157]. Advanced instruments with high mass-resolution are required[157, 173], and the large precursor ions analyzed are capable of obtaining many different charge states and generate many different fragment ions which both contribute to the generation of extremely complex spectra[157, 173]. Second, whole proteins are generally more difficult to handle than peptides. In particular, larger proteins often have solubility issues under general liquid chromatography (LC)-MS conditions, and almost all membrane proteins require the use of MS-incompatible detergents to be solubilized[174]. Third, the limit of detection of mass spectrometers for proteins is much lower than for peptides, increasing the amount of sample needed and accordingly reducing experimental throughput[173, 174]. Finally, protein tertiary structure tends to become more difficult to disrupt as MW increases which limits the fragmentation efficiency of larger protein ions, limiting the applicability of top-down approaches to smaller proteins with MW < ~50 kDa[173, 175]. Key advances in instrumentation, experimental design, and data interpretation are expected to address many of these concerns and expand the functionality and accessibility of this approach[157].

1.2.1.3.2.3 General challenges in MS-based identification of PTMs

In addition to the approach-specific limitations noted in the preceding sections, MS-based approaches for the detection and identification of PTMs encounter several universal challenges that must be regularly overcome when employing these techniques. One challenge is grounded in the observation that many PTMs, in particular the more transient regulatory modifications, are often present at substoichiometric levels in samples[4, 158, 176]. As mass spectrometer ion detection is biased towards the most abundant ions in a sample, especially rare or short-lived proteoforms can go undetected[177]. A second challenge is that PTMs and other covalent adducts are occasionally labile in the gas phase[4, 176, 178]. In these cases, the departure of a modifying chemical group during ionization stymies efforts to observe it in the mass analyzer. This is a particular challenge when using MS to investigate phosphorylation events, as phosphoserine and phosphothreonine residues are highly labile[179].

1.2.1.3.3 Methods for Characterizing PTMs

Establishing the existence of a particular PTM on a protein of interest is an important first step in elucidating the complete functional role that the protein plays. However, simply demonstrating that a protein can obtain a specific modification is far from the whole story. Indeed, complete characterization next requires a determination of the change in protein function that results from the addition of the PTM, ideally paired with an investigation into the structural basis for the observed functional change as well as identification of the enzymes responsible for the addition and/or removal of the chemical group. In the next section, I will provide a brief introduction to the most common experimental approaches that have been developed for functionally investigating proteins bearing PTMs. For the purposes of this thesis I will very broadly divide these approaches into two categories – i) *in vivo* approaches and ii) *in vitro* approaches.

1.2.1.3.3.1 In vivo *Approaches for Characterizing PTMs*

At a high level, the ultimate goal of determining the role(s) played by PTMs in modulating protein function is to obtain a more thorough understanding of the global workings of the cell. In this context, it is logical to go directly into cells to study the effects of PTMs of interest. Accordingly, many of the earliest investigations into PTM function were conducted *in vivo*, and these approaches still enjoy widespread use today.

One strategy seeks to assign function to patterns of post-translational modifications by associating them with a cellular response to a particular stimulus, such as heat shock or induction with a chemical signal[180]. Very simply stated, this strategy entails exposure of one population of cells to a particular stimulus while leaving a second control population unstimulated, followed by proteomic analysis to determine how PTM occupancy differs between the two samples. For example, Kim *et al* investigated patterns of tyrosine phosphorylation in response to heat shock in mouse RIF-1 cells. They obtained protein samples from both heat shocked and thermotolerant cells, separated them using electrophoresis, and probed each sample with an anti-phosphotyrosine antibody. The authors observed 93 proteins that showed significant changes in phosphorylation between the two conditions, and were able to identify 81 of them using mass spectrometry[181]. More recently, Melo-Braga *et al* used a higher-throughput MS-based approach to compare PTM occupancy from samples derived from grapevine plants (*Vitis vinifera*) with and without infection with *Lobesia botrana* pathogen. The authors identified 110 phosphorylation and 20 acetylation sites differing in occupancy between the two samples, suggesting regulatory roles for these two modifications in the organism's response to infection[182].

In establishing correlations between PTM occupancy and a cellular condition, assays such as those described above certainly help shine some light on system-wide roles for those modifications. However, these strategies are insufficient for obtaining information associating changes in protein structure/function with PTMs at specific locations. For resolving effects of site-specific modifications, researchers have generally turned to mutational screens, employing genome editing techniques to install point mutations at specific amino acid positions in protein(s) of interest. Using this approach, it is relatively simple to obtain information about the effects of a specific PTM by eliminating it from the protein – the amino acid at the position involved can be changed to one that cannot receive the modification, thereby preventing its addition to the protein[183, 184]. Converse investigations are more difficult, as site-specific addition of modifications is practically impossible to perform *in vivo*. Still, approaches have been developed that simulate the presence of a PTM by encoding a structural mimetic at the amino acid position of interest. For instance, aspartic and glutamic acid residues have been used to mimic phosphorylated amino acids[183], glutamine has been used to mimic acetylated amino acids[183, 184], and methionine/leucine have been used to mimic methylated amino acids[185]. Liu *et al* applied these approaches to investigate the effects of PTMs on the structural protein tubulin in a HeLa cell model[186]. Mutation of lysine 394 and threonine 257 in the protein prevented assembly into microtubules, suggesting roles for acetylation and phosphorylation at those positions, respectively, in promoting assembly[186]. In a separate study, Kashiwagi *et al* assessed the role of phosphorylation at serine 75 in mouse Src kinase. By generating two mouse lines in which the serine of interest was mutated to either aspartic acid (a phosphorylation mimic) or alanine (thereby eliminating the phosphorylation site) the authors were able to observe phenotypic effects of that

phosphorylation. They observed that the phosho-mimicking mutant, but not the non-phosphorylatable mutant, was associated with increased retinal ganglion cell loss in aging mice, implicating phosphorylation of serine 75 in regulating retinal ganglion cell survival in those animals[187].

As useful as these mutational approaches have been, they are not without limitations. Perhaps the most impactful limitation is that while site-specific PTMs are associated with an observed functional change, the underlying mechanisms (*i.e.* structural changes in the protein elicited by a PTM) cannot be ascertained using these assays[180]. Additionally, the PTM mimetics discussed above differ both structurally and electrostatically from the modified amino acids that they are standing in for, and do not always faithfully replicate the effects of their associated PTM. Indeed, phosphorylation mimetics in particular often fail to correctly imitate the presence of a phosphate group due to differences in charge density[136, 188]. Lastly, the modification state of positions examined in this way are absolute (*i.e.* always "modified" or incapable of being modified) which does not at all accurately mirror the normally transient nature of most PTMs[189].

1.2.1.3.3.2 **In vitro *Approaches for Characterizing PTMs***

Approaches for the isolation and subsequent study of a protein of interest bearing specific PTM(s) greatly facilitate the complete characterization of that proteoform. By these approaches, a proteoform of interest is isolated *in vitro* allowing for the direct interrogation of such parameters as efficiency, substrate specificity and binding affinity, and even structure. In this way, in-depth characterizations of PTMs (including functional and mechanistic insights) are achievable.

In a seminal example of the power of *in vitro* analysis of PTMs, Russo *et al* purified cyclin-dependent kinase (CDK) and phosphorylated it via incubation with so-called CDK-activating

kinase to obtain a homogeneous sample of phosphorylated CDK, from which they were able to obtain a 2.6 Å crystal structure. From the structure, they were able to describe conformational shifts in the kinase upon phosphorylation that promote association with a key binding partner, providing a mechanism by which the phosphorylation event modulates CDK activity[190]. Another exemplary study by Gu and Roeder demonstrated that p53 is acetylated near the C-terminus by the acetyltransferase p300, and that the acetylation event is sufficient to activate the DNA-binding activity of the protein[191].

These *in vitro* analyses complement investigations *in vivo* very well, but because of several key limitations are generally unable to stand entirely on their own. Because most PTMs exist only briefly during the lifetime of a protein at substoichiometric levels, it is difficult-to-impossible to purify most proteoforms of interest from cells[158, 176, 178]. As a result, proteins bearing specific patterns of PTMs must be directly synthesized via some other means. The simplest way to obtain highly pure samples of a specific proteoform is to express the unmodified protein and modify it *in vitro*; however, this approach requires knowledge of the modification enzyme(s) responsible for performing those modifications and the ability to obtain those enzymes in a pure, active configuration, which is not always possible[176]. Even when the enzyme is known and obtainable, it can be difficult to direct it towards the modification of only specific sites of interest rather than simply modifying every canonical side chain in the sample. Additionally, the *in vitro* reaction environment differs greatly from the tightly regulated environment that the enzyme usually operates in *in vivo* such that errant modification of unnatural targets may occur[192]. Indeed, the differences between a minimalist reaction environment *in vitro* and the regulated, crowded chemical milieu *in vivo* may cause dramatic differences in protein behavior such that any findings

in vitro must be supported with evidence derived *in vivo* to derive conclusions with high confidence. This is particularly true for kinase enzymes, which are often more promiscuous *in vitro* than *in vivo* leading to the generation of many false positive results when trying to assess substrate specificity[193]. Many past efforts to circumvent these synthesis limitations have sought to use solid phase peptide synthesis to directly incorporate modified amino acids into peptides to use as substrates for screens, or to chemically ligate modified peptides together to generate a specific proteoform[176, 194-196] – however, these approaches remain limited by the fact that peptide fragments do not always faithfully resemble into the correctly folded, intact protein substrates that they are mimicking[7]. Furthermore, ligation of peptides into full proteins imposes some amino acid sequence requirements that make it unsuitable for the synthesis of many targets of interest[176, 195].

1.2.2 Genetic Code Expansion

Proteins represent a critical class of biomolecules, universally employed by *all* living organisms to fulfill essential structural, functional, and enzymatic roles necessary to support life. In nature, these polymers are composed generally of twenty natural amino acid (AA) building blocks, which can be combined in a near-infinite number of combinations to generate an impressive level of structural and functional diversity **(Figure 1.4.A)**. However, many interesting chemistries cannot be accessed using only these natural building blocks; accordingly, for some time there has been an interest in the incorporation of non-canonical amino acids (ncAAs) featuring novel functional sidegroups to expand the repertoire of protein functions.

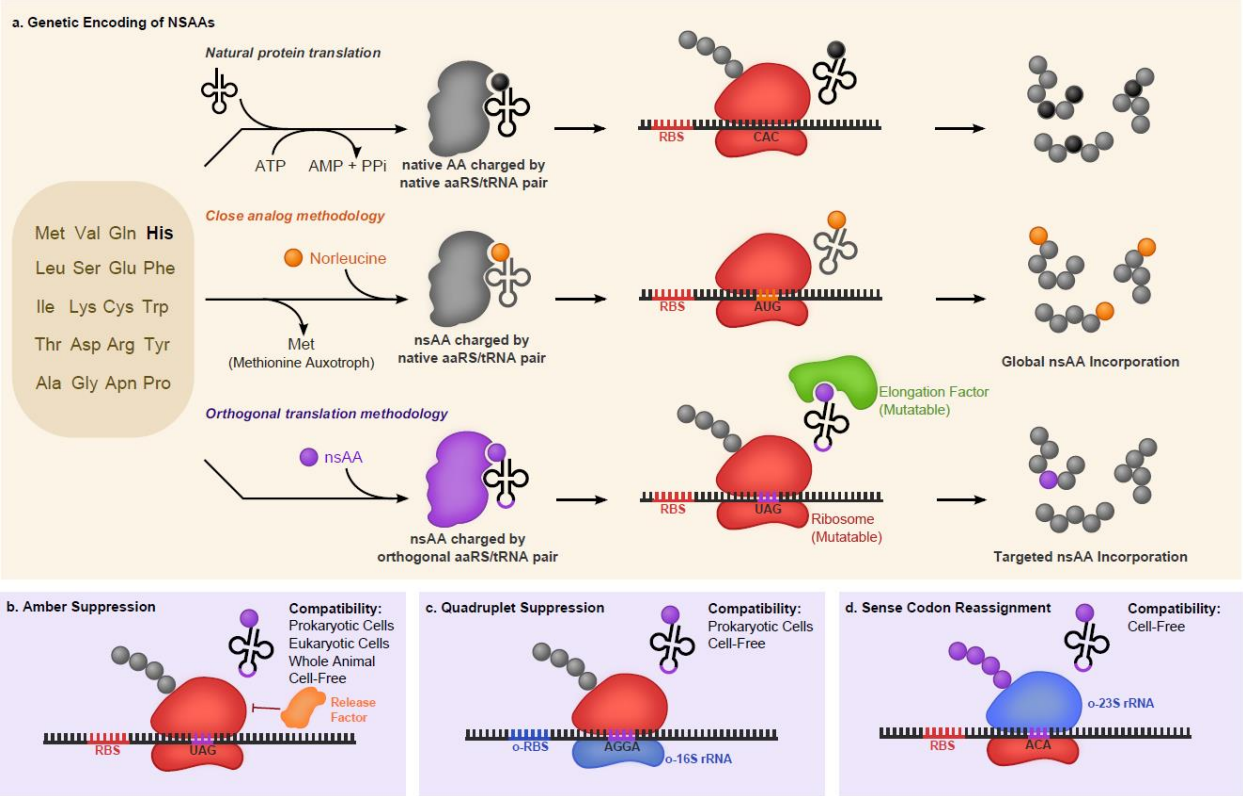


Figure 1.4. Methods for Genetic Code Expansion a. Two general paradigms exist for the genetic incorporation of nonstandard amino acids into proteins contrasted with the natural process of encoding the canonical amino acids. The close analog methodology complements a natural amino acid auxotrophy with a close nonstandard analog, enabling global protein labeling by native translational machinery. The orthogonal translation methodology introduces orthogonal translational machinery engineered to charge an orthogonal tRNA with a nonstandard amino acid, enabling site-specific targeted genetic incorporation. Certain ncAAs may require additional mutations in the elongation factor or the ribosome. **b.** For targeted genetic incorporation, amber suppression is the most widely used technique. Competition with release factors limits efficiency, and methods are discussed to overcome this. **c.** Quadruplet suppression can be performed with appreciable efficiency with the use of an engineered orthogonal 16S ribosomal subunit[197]. **d.** Sense codons can be completely reassigned by using an orthogonal 23S ribosomal subunit, engineered to accept a synthetic set of tRNAs[198].

Broadly speaking, ncAAs can be divided into two classes. *Synthetic ncAAs* are chemically synthesized and can bear little resemblance to their naturally occurring counterparts. *Posttranslational modifications* (PTMs) are modified derivatives of canonical amino acids. In recent years, two distinct approaches for the incorporation of ncAAs into proteins have emerged **(Figure 1.4.A)**. One such approach is *global suppression.* This method uses auxotrophic strains

that are incapable of synthesizing a particular AA. When grown in the presence of a ncAA that bears close structural resemblance to the "missing" AA, the organism's native translational machinery incorporates the ncAA instead[199, 200]. An alternative approach uses Orthogonal Translation Systems (OTSs) to genetically encode an ncAA of interest site-specifically by reassignment of codons, typically the amber stop codon (TAG) in a strategy known as *amber suppression*[201].

To date, >150 ncAAs have been incorporated by OTSs into peptides[202] for a wide range of applications including the introduction of bioorthogonal handles for protein tagging[203, 204], alteration of protein stability[205, 206], monitoring of protein localization, and genetic encoding of PTMs[6, 7, 207, 208]. As a result of these impressive efforts and the transformative potential to construct bio-based products beyond natural limits, expanding the genetic code has emerged as one of several major defining opportunities and points of synergy in chemical and synthetic biology.

The following section focuses on recent developments in repurposing the translation system for novel functions, with a focus on codon reassignment. I will first examine development of the molecular machinery at the heart of genetic code expansion. Next, I will discuss ncAA incorporation in several contexts, including whole-genome recoding in prokaryotic and eukaryotic systems *in vivo*. I will end with a discussion of current challenges in the field and provide commentary on future opportunities.

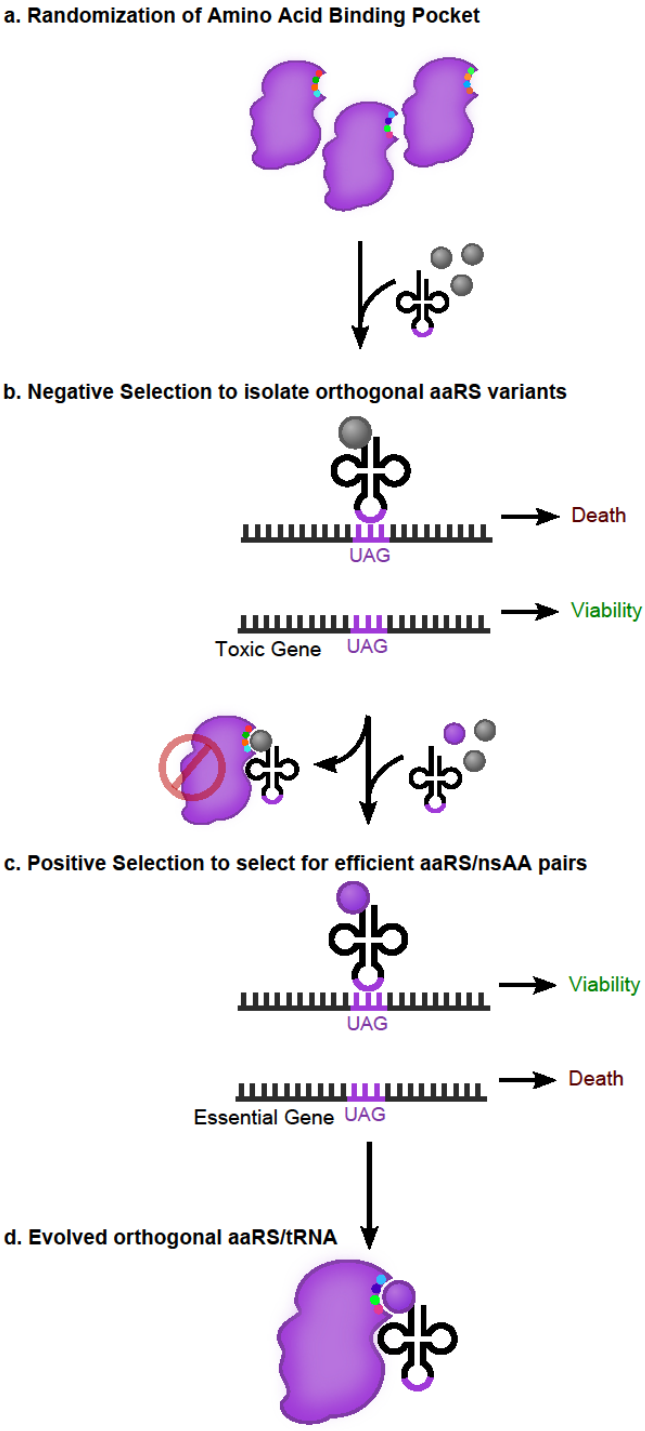

a. Randomization of Amino Acid Binding Pocket
b. Negative Selection to isolate orthogonal aaRS variants
Death
UAG
Viability
Toxic Gene
UAG
c. Positive Selection to select for efficient aaRS/nsAA pairs
Viability
UAG
Death
Essential Gene
UAG
d. Evolved orthogonal aaRS/tRNA

Figure 1.5 (previous page). General methodology for engineering orthogonal translation systems for novel NCAAs. a. Scaffold orthogonal aaRS/tRNA pairs are selected from distant organisms. Residues in the amino acid binding pocket of the aaRS are randomized. **b.** A negative selection is performed to exclude aaRS variants capable of charging natural amino acids. In the absence of the ncAA, clones are selected by the inability to suppress a toxic gene. **c.** A positive selection is performed to retrieve aaRS variants capable of charging the ncAA, selecting for the ability to suppress a selectable marker gene, such as an antibiotic resistance marker. **d.** The selected orthogonal translation system consists of the orthogonal aaRS, tRNA, and ncAA.

1.2.2.1 Genetic Code Expansion Using Orthogonal Translation Systems

Amber suppression seeks to "hijack" the amber translational stop codon (TAG), recoding it into a sense codon corresponding to a ncAA of interest. Generally, this is accomplished using a *suppressor tRNA* that has been mutated to decode the amber codon and an *aminoacyl-tRNA synthetase* (aaRS) that has been mutated to accept the ncAA of interest and covalently load it onto the suppressor tRNA. These components are typically sourced from distant archeal species to ensure orthogonality to host translation machinery, undergoing directed evolution to improve their compatibility with a new ncAA and enable its site-specific incorporation into proteins.

Directed evolution is the most widely-used approach for the generation of novel OTS components[209-211] **(Figure 1.5)**. These efforts start with the selection of a scaffold aaRS/tRNA pair. To date, several aaRS/tRNA pairs have been used in the creation of new OTSs. The *Methanocaldococcus jannaschii* TyrRS/$tRNA^{Tyr}$ pair is arguably the most common pair used, but is generally limited to aromatic amino acids and is not orthogonal in eukaryotes[202, 211, 212]. The PylRS/$tRNA^{Pyl}$ pair from *Methanosarcina* species (*M. mazei, M. barkeri*) has shown compatibility with eukaryotic systems[84, 212], and is an especially attractive starting point for evolution as the native PylRS natively demonstrates polysubstrate specificity[213] and $tRNA^{Pyl}$ natively decodes the amber codon[214]. Other starting components have included the *o*-phosphoserine (Sep)RS from *Methanococcus maripaludis*[6] and the TrpRS/$tRNA^{Trp}$ pair from

Saccharomyces cerevisiae[215]. The TyrRS/tRNA^{Tyr} and LeuRS/tRNA^{Leu} pairs from *Escherichia coli* have also been used in OTS development for use in higher organisms[216, 217]. After selecting a scaffold pair, crystal structural data is commonly used to identify specific residues in the aaRS that interact with the amino acid and the tRNA[209, 218]. These residues are randomized to create a library of mutant aaRS variants *in vitro* and subsequently transformed *in vivo*. Finally, alternating rounds of selection identify mutant variants that are both functional and orthogonal to native machinery[209, 211]. Negative selections eliminate variants that can charge the *o*-tRNA with native amino acids based on the synthesis of a toxic gene (*e.g.*, barnase) in the absence of the cognate ncAA. Positive selections in the presence of the ncAA isolate variants that can charge the ncAA of interest onto the *o*-tRNA based on the suppression-dependent synthesis of a selectable marker (*e.g.*, antibiotic resistance gene) or reporter (*e.g.*, GFP). By subjecting "winners" to alternating rounds of these screens, a ncAA-specific aaRS/tRNA pair can be identified.

While most new OTSs have focused on the generation of new ncAA-aaRS-tRNA pairs, some efforts have expanded OTSs to include additional components. In an exemplary study, incorporation of Sep was *only possible* after the development of a Sep-specific elongation factor[6]. More recently, researchers were able to increase incorporation efficiency of selenocysteine (Sec) to >90% by engineering an elongation factor optimized for Sec[219]. NcAA incompatibility with the ribosome may also be an impediment to incorporation[220]. Platforms for ribosomal engineering and evolution will be integral to the elucidation and optimization of ribosome/ncAA interactions. Two recent approaches permit construction of modified ribosomes by decoupling organism fitness from ribosome function. In the first, researchers engineered a 30S subunit that is made orthogonal to natural components by a mutant 16S rRNA; this orthogonal

subunit can be mutated to evolve novel function without impairing host viability[197, 221]. In a parallel approach, a system for the *in vitro* assembly of functional ribosomes has been developed[222-224].

For the purposes of this thesis, it is important to note that functional OTSs have already been developed for phosphoserine[6, 225], phosphothreonine[226], phosphotyrosine[227], acetyllysine[207], methyllysine[226], and sulfotyrosine[228], enabling the genetically-encoded site-specific installation of these important post-translational modifications.

1.2.2.2 Prokaryotic Strain Engineering Tailored for Genetic Code Expansion

Historically, ncAA incorporation via natural codon suppression has been significantly limited by native translational components that have evolved essential function to faithfully decode all codons within an open reading frame. In the case of amber suppression, *release factors* are a class of proteins responsible for facilitating the termination of translation in response to ribosomal stalling at a stop codon. Competition between loaded suppressor tRNAs and release factor proteins at amber codons meant to encode ncAAs severely hinders successful suppression, with release factor activity resulting in the premature truncation of most of the protein product[23]. Resultant yields of the target ncAA-containing protein under these conditions are very low, especially for proteins containing multiple instances of the ncAA[24]. Conversely, the presence of OTSs on high copy plasmids in the presence of high concentrations of ncAA *in vivo* can also drive the incorporation of ncAAs at >300 amber codons that terminate native genes, resulting in cellular toxicity[21].

Consequently, much effort has gone into the elimination of release factor activity to improve ncAA incorporation. Initial attempts to outright delete the essential gene *prfA* that encodes

release factor 1 (RF1) were stymied by cell inviability. Several early approaches, including release factor engineering[229] and supplementation *in trans* with partially-recoded versions of the amber-dependent essential genes[230] permitted subsequent removal of *prfA* from the organism. More recent efforts have recoded the genome of *E. coli*. In such efforts, researchers edited the amber-dependent essential genes to terminate instead with the synonymous ochre codon (TAA) and deleted *prfA* from the organism. Incorporation of Sep was greatly enhanced in one such RF1-deficient recoded strain[231]. More recently, a recoded RF1-deficient strain was used in the preparation of cell lysates to improve the incorporation of the ncAA p-propargyloxy-L-phenylalanine (pPaF) *in vitro*[24, 232].

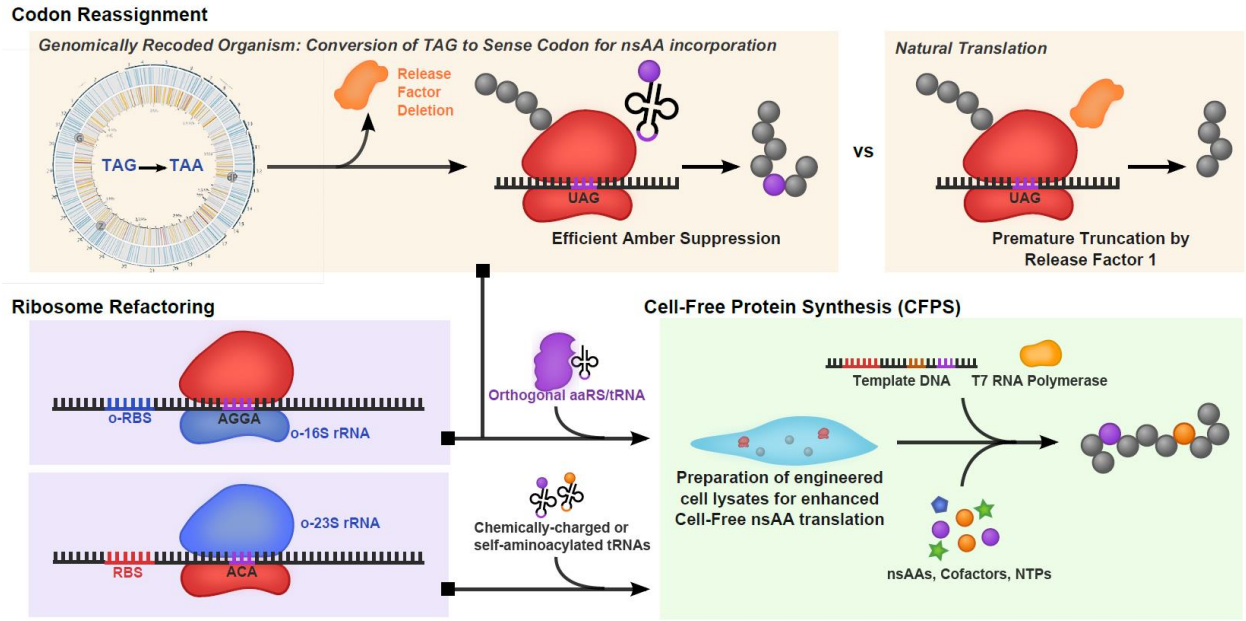


Figure 1.6. Genomically recoded organisms (GROs), Strain Engineering and *in vitro* ncAA translation. A genomically recoded organism reassigns the TAG stop codon to an open sense coding channel for enhanced ncAA amber suppression. Ribosomal engineering promotes other classes of codon suppression. These engineered strains can be used *in vivo* and, via cell lysate, to develop *in vitro* translation systems with open coding channels for the production of ncAA-containing proteins.

These strain engineering efforts culminated with the recently-reported completion of the first completely genomically recoded organism (GRO), an *E. coli*-derived strain that lacks all amber codons and RF1[21, 26]. In this study, the authors systematically reassigned *all 321* native instances of the amber (TAG) codon to the ochre (TAA) codon and deleted the *prfA* gene **(Figure 1.6)**. In the resulting strain (C321Δ*A*) the amber codon is orthogonal and unrecognized by the remaining translational machinery, freeing it for use as a dedicated codon for ncAA incorporation. This strain has demonstrated improved properties for incorporation of ncAAs[21] and has more recently been engineered to depend on ncAAs as a 21st synthetic biochemical building block[233, 234].

Beyond amber suppression, some effort has been made to access additional codons for ncAA assignment, such as the "ochre" (TAA) [235] and "opal" (TGA) stop codons[236]. In a parallel approach, researchers instead sought to access non-natural *quadruplet* codons to encode ncAAs[237, 238]. Combined with amber suppression, quadruplet suppression has enabled the incorporation of multiple distinct ncAAs into a single peptide[239-241].

1.2.2.3 Genetic Code Expansion in Eukaryotic Systems

Amber suppression has been adapted to eukaryotic systems as a tool for the interrogation of cellular biology. Engineered orthogonal aaRS/tRNA pairs have been used to genetically encode ncAAs that modulate posttranslational modifications by photocaging lysine[242] and serine[243] residues, promote chromatin condensation with crosslinking residues[244], introduce chemical handles for fluorescent protein labeling and live cell imaging[245], among many other applications. Most often, OTS components are scaffolded on *M. mazei* PylRS/tRNA^{Pyl}, *E. coli*

TyrRS/tRNA^{Tyr} or *E. coli* LeuRS/tRNA^{Leu} pairs, engineered in *E. coli* or *S. cerevisiae*[216], and then ported to mammalian vectors.

In addition to unicellular organisms and tissue culture, amber suppression with engineered PylRS/tRNA^{Pyl} derivatives has been demonstrated in the nematode *Caenorhabditis elegans*[246]. Similar work has been successful in the fly *Drosophila melanogaster*, where ncAAs could be incorporated both site-specifically and tissue-specifically into proteins[247].

As with prokaryotic systems, competition with release factors remains a major barrier for efficient amber suppression in eukaryotes. Though orthogonal ribosomes or genome-wide codon reassignment have not yet been demonstrated in these systems, efforts are being made in this direction. Recently, a *de novo* synthesis of *S. cerevisiae* chromosome III included TAG->TAA stop codon reassignments[248]. As this effort is extended to the remaining chromosomes, the resulting synthetic strain will lack native TAG codons, motivating the need to engineer the specificity of the single eukaryotic release factor to exclude amber recognition.

An additional challenge is native quality control machinery. Nonsense mediated decay (NMD) surveys transcripts for nonsense codons excessively distal from the 3' end, triggering degradation in response to their presence[249]. Addressing this limitation, the pathway has been knocked out in *C. elegans* to boost amber suppression efficiency[246].

A final consideration in these systems is the expression of the orthogonal tRNA, as eukaryotic RNA polymerase III is typically recruited to intragenic A and B-Box promoter sequences. One solution has been to identify orthogonal aaRS/tRNA pairs containing A and B-like elements. For instance, E. coli's TyrRS/tRNA^{Tyr} is orthogonal in *S. cerevisiae* and contains its consensus A and B-box[216]. However, OTSs scaffolded on PylRS/tRNA^{Pyl}, lack such intragenic

sequences. Instead these tRNAs have been successfully expressed by co-opting dicistronic tRNA scaffolds such as $tRNA^{Arg}_{UCU}$[250], or by using the extragenic U6 RNA polymerase III promoter[251].

1.2.2.4 Current Challenges and Future Outlook

Looking forward, three major challenges define the trajectory of this field: the development of more efficient OTSs, accession of more open coding channels to enable multi-site incorporation of multiple ncAAs, and the generation of OTSs engineered for more exotic ncAAs.

Engineered OTSs suffer poor enzymatic efficiencies relative to native translational machinery[220]. Overall, this represents a problem of insufficient aaRS evolution, as typically only 6-8 residues proximal to the amino acid binding pocket are mutagenized. Further diversification is limited by library size constraints – the randomization of 6 residues results in a library of nearly 10^8 members, quickly saturating standard screening techniques. The use of computational protein modeling may enable more rational engineering, guided by *in silico* binding predictions[252]. Further, during evolution OTSs are screened for the ability to suppress just a few codons in selectable markers, whereas native translational machinery faces a load of thousands of codons. Increasing the load on OTSs may promote the selection of more catalytically active enzymes. Amber suppression of native essential genes has proven to be an effective strategy to tie strain viability to ncAA incorporation[233, 234], offering a potential route to drive protein evolution under more stringent selective conditions.

Multi-site incorporation of a single ncAA has remained elusive due to the use of codons that have their cognate translation components (*e.g.* RF1) out-competing the incorporation of the ncAA, as well as the impaired activity of OTSs. These limitations are compounded in efforts to

achieve multi-site incorporation of *multiple distinct* ncAAs. This pursuit will require advances in our ability to suppress multiple codons simultaneously, and in the development of mutually-orthogonal OTSs. *In vivo*, the field is currently limited to suppressing two distinct codons simultaneously. The use of nonstandard nucleotide bases[253] may enable crossing this barrier in the near future. Alternatively, radical recoding strategies that extend off the RF1-deficient GRO may provide a route to more codons. Along these lines, progress is being made towards a synthetic *E. coli* strain utilizing only 57 codons via the removal of redundant sense codons from the genome[228]. Even with more codons, new advances in orthogonal OTSs are needed. OTSs have been selected against interactions with *native* translational machinery, but not necessarily against other OTSs. This raises the potential for OTS cross-reactions when expressed simultaneously. Work is being done to develop orthogonal tRNA acceptor stems[254] and additional aaRS/tRNA scaffolds to mitigate this cross-orthogonality.

Finally, certain biological constraints limit the scope of ncAA diversity. These include limitations on cell membrane permeability and steric incompatibility with the ribosome and other translational components such as elongation factors. Further engineering these components with cell-free systems or synthetic ribosomes is necessary for increasing the available ncAA chemical space.

Overcoming some of these technological barriers will better enable the creative potential of nonstandard protein development in producing the next wave of highly functionalized biomaterials with broad applications in medicine, materials science and biotechnology.

1.2.3 Cell-free Protein Synthesis Systems

Cell-free protein synthesis (CFPS) systems have revolutionized our ability to understand, harness, and expand the capabilities of biological systems. Since their early use in the deciphering of the genetic code[255], CFPS platforms have improved steadily such that they now represent a *bona fide* alternative to *in vivo* overexpression for simple and efficient bulk synthesis of protein products. Indeed, for the past decade a technical renaissance in the field has transformed CFPS in the interest of leveraging the unique advantages offered by these approaches for applications including biomolecular breadboarding[256-259], expression of toxic products[260-263], production of complex protein products that are poorly soluble *in vivo*[264-267], manufacture of glycoproteins[147, 268-270], detection of disease[30, 271, 272], on demand biomanufacturing[30, 265, 273-276], and education[277, 278].

In this next section I will provide a brief introduction to CFPS systems, including a summary of CFPS, a discussion of cell-free approaches as contrasted with protein expression in cells, an overview of CFPS systems developed to date, and a discussion of recent trends in CFPS platform development. I will finish with a discussion of specific applications of CFPS, with a focus on the use of these platforms towards the synthesis of proteins featuring non-canonical amino acids (ncAAs).

1.2.3.1 Cell-free Protein Synthesis: A Primer

To synthesize protein products of interest, CFPS systems harness cellular machinery responsible for catalyzing protein synthesis and energy regeneration from purified components or crude lysates derived from chassis organism cells. Such systems contain the necessary elements for translation and protein folding (*e.g.* ribosomes, aminoacyl-tRNA synthetases, translation initiation and elongation factors, ribosome release factors, chaperones, foldases, *etc.*) and energy

regeneration (*e.g.* enzymes involved in glycolysis, oxidative phosphorylation) such that upon being combined with raw materials (amino acids, nucleotide triphosphates, energy, enzyme cofactors, buffer, salts, DNA template) these biological processes can be coordinately activated in a test tube in the absence of intact cells. Protein synthesis can then proceed until one or more substrates is depleted or byproduct accumulation reaches inhibitory levels[279-281].

These systems enjoy several distinct advantages over *in vivo* methods for recombinant protein production[279-283]. First, cellular compartmentalization is eliminated, permitting direct access to the reaction volume. This simplifies the eventual purification of the protein product of interest and enables experimenters to directly sample and analyze reaction contents. Additionally, the lack of physical barriers permits the direct supplementation of key/limiting substrates to the reaction volume to prolong or enhance synthesis. Second, the absence of intact, living cells circumvents many of the common problems encountered during recombinant expression *in vivo*. For instance, CFPS enables the robust synthesis of toxic products that otherwise could not be produced in living cells. Furthermore, cellular programming is removed via the removal of host organism genomic DNA during crude lysate preparation such that CFPS reactions cannot enact responses to stresses imposed by the synthesis demands of the experimenter in a nonproductive or otherwise deleterious manner. The removal of genomic DNA also ensures that the only DNA present in CFPS reactions is the template (encoding a product of interest) supplied by the experimenter – in this way, 100% of CFPS reaction volume is directed towards synthesis of a desired product. Third, cellular proteins in CFPS reactions are diluted up to 20-fold relative to cellular concentrations. This dilute reaction environment helps to facilitate the folding of complex multi-domain eukaryotic products which are difficult to solubly express in cells. Finally, CFPS

reactions do not require the use of plasmid DNA to template products of interest. Alternative templates that are significantly faster to synthesize and mutagenize, such as PCR products, can be used instead, accelerating efforts to screen multiple different products in parallel.

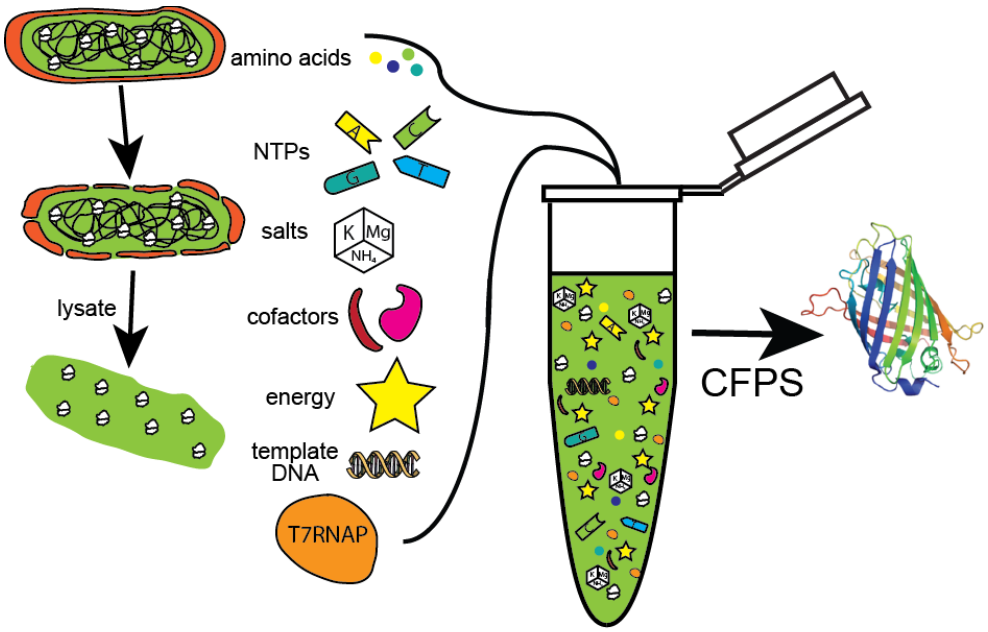


Figure 1.7. Simplified schematic of the production and utilization of crude lysates from cells to catalyze cell-free protein synthesis (CFPS). Reactions are supplemented with enzymatic cofactors, energy, and other substrates required for protein synthesis as well as plasmid DNA template directing the system towards the production of a product of interest. Once all substrates are present, protein synthesis activates to synthesize the specified product. T7RNAP: T7 RNA polymerase; NTPs: nucleotide triphosphates.

In the interest of leveraging these advantages for protein synthesis in many organismal backgrounds, CFPS systems from crude cellular lysates have been developed from an ever-increasing number of different chassis organisms. In the following sections I will highlight some of the more widely-used CFPS platforms that have been developed, comparing and contrasting their relative advantages and disadvantages.

1.2.3.2 CFPS in *Escherichia coli*

CFPS platforms built on extracts derived from the prokaryote *Escherichia coli* are by far the most popular[280]. The particularly widespread adoption of these platforms is due to several factors. First, *E. coli* is perhaps the most heavily studied and well-characterized model organism of all time. Accordingly, protocols for everything from growing and handling *E. coli* to modifying the *E. coli* genome are well-established. Second, *E. coli* is easily fermented at large scale using cheap growth media and easily lysed using high-pressure homogenizers or sonication, simplifying extract preparation[280, 283]. Third, *E. coli* CFPS systems generally have the highest protein yields with titers up to 2.3 g/L reported in the literature[284]. Fourth, *E. coli* CFPS can be very inexpensive due to the ability to use cheap substrates such as glucose as the primary energy source in these reactions[283, 285]. Lastly, due in large part to the preceding features *E. coli* cell-free systems are commercially available and have been demonstrated at the industrial scale, greatly facilitating expansion of their use[280].

Trailblazing investigations from Swartz and others have been instrumental in advancing the capabilities of *E. coli* CFPS to where it currently stands. Based on the early observation that the amino acids alanine, aspartic acid, and asparagine were being synthesized in *E. coli* CFPS (suggesting that central metabolism was active), Kim and Swartz developed the PANOx cell-free system using phosphoenol pyruvate as the energy source and overcoming amino acid depletion by increasing the concentration of the 20 canonical amino acids to 2 mM in the reaction environment. Collectively these changes increased the yields of *E. coli* CFPS 4-fold to ~700 μg/mL of chloramphenicol acetyltransferase (CAT)[286]. Jewett *et al* further improved the system by replacing the crowding agent polyethylene glycol (PEG) with spermidine and putrescine, which

stabilized nucleic acids in the reaction volume in addition to crowding cellular components. In the same work, the reaction formulation was altered such that cation concentrations in the reaction mix more closely align with what is experienced in the cell. The resulting system saw CAT yields increase to over 700 μg/mL, establishing the guiding principle in the development of CFPS systems that cytoplasmic mimicry enables highly productive systems[287]. More recent efforts by Swartz, Bundy, Noireaux, and others have sought to improve yields by optimization of lysate preparation parameters and stabilization of reaction substrates via genomic knockout of genes encoding negative effectors of CFPS[284, 288-291]. As a result of these and other important works, *E. coli* CFPS reaction durations now exceed 20 hours, with yields regularly surpassing 2 g/L. Reaction scale has also steadily increased, culminating recently with the report of a 100 L reaction[264].

Though powerful, CFPS platforms derived from *E. coli* struggle with several aspects of protein synthesis, limiting their applicability towards the synthesis of some products. First, the rate of amino acid addition to nascent polypeptides in the *E. coli* ribosome is significantly faster than in eukaryotic ribosomes[292]. Consequently, even in the relatively dilute environment of a CFPS reaction, *E. coli* translation machinery struggles to solubly synthesize some complex proteins of eukaryotic origin[280]. Second, despite several advances, *E. coli* CFPS still has a limited ability to correctly fold proteins featuring multiple disulfide bonds and integral membrane components[280]. Third, *E. coli* systems have a very limited capacity for the synthesis of products featuring post-translational modifications (PTMs)[280, 282]. These collective limitations have fueled interest in the development of eukaryotic CFPS platforms.

1.2.3.3 CFPS Using Eukaryotic Systems

In contrast to *E. coli* CFPS platforms, eukaryotic systems are often capable of synthesizing complex eukaryotic proteins with higher solubility. Existing eukaryotic CFPS systems are commonly derived from wheat germ, yeast, insect cells, and Chinese hamster ovary cells. I will briefly introduce each of these systems in the following sections.

1.2.3.3.1 CFPS in Wheat Germ Extract

CFPS systems using wheat germ extract (WGE) have been the most productive eukaryotic platforms developed to date[280]. These systems are derived from isolated wheat seed embryos, and typically produce between several hundred μg to mg of protein per mL of reaction volume[293]. Importantly, the WGE platform shows advantages over *E. coli* for the soluble synthesis of complex eukaryotic products[294]. For instance, Goshima *et al* were able to apply the WGE system to the synthesis of over 13,000 human proteins in a single study[295]. Impressively, Endo and Sawasaki reported the synthesis of nearly 10 g/L of green fluorescent protein (GFP) using WGE CFPS in semicontinuous mode[296]. Furthermore, the platform is able to achieve several PTMs on products[279]. However, WGE CFPS suffers from a handful of limitations. Extract preparation is complex and yields only a very small volume of extract from cells[280]. Additionally, this organism is severely lacking in a set of tools for genetic modification[280].

1.2.3.3.2 CFPS in *Saccharomyces cerevisiae*

The development of CFPS systems from yeast (*S. cerevisiae*) seeks to get the best of both worlds with regards to prokaryotic vs. eukaryotic platforms. On the one hand, this chassis organism is very "bacteria-like" in that it is unicellular and is easily fermentable in liquid growth media[280]. Indeed, in terms of growth and handling of cells, yeast is more similar to bacteria than it is to other eukaryotic cell cultures. On the other hand, it is a eukaryote, and so theoretically could benefit

from such features as the slower eukaryotic ribosome and capacity for installing PTMs[280, 282]. Additionally, *S. cerevisiae* is a highly-studied model organism – accordingly, the knowledge base for the organism is broad and includes a fully-sequenced genome and tools for facile genetic engineering[279].

Early works by Iizuka applied yeast extracts to the investigation of translation mechanisms in the organism[297]. This work was later followed by Wang *et al* who presented a method for lysing yeast cells which starts by removing the other membrane of the cell wall using lyticase to yield a protoplast, which is then lysed via passage through a 25-guage needle. The resulting lysates synthesized ~70 ng/uL protein[298]. More recent efforts by Hodgman and Jewett have led to a new, optimized method which includes the use of high-pressure homogenization for cell lysis, combined transcription/translation without mRNA capping, and technically simple extract preparation methods to yield a platform capable of synthesizing 7.69 ± 0.53 μg/mL active luciferase, two orders of magnitude higher than previously reported yeast platforms[299]. In a followup study, Choudhury *et al* optimized fermentation parameters to improve the yields of yeast CFPS to 8.86 ± 0.28 μg/mL[300], after which Schoborg *et al* improved the productivity of the system further by using a semi-continuous reaction format to simultaneously feed limiting substrates (creatine phosphate, nucleotide triphosphates, *etc.*) while removing toxic byproducts (inorganic phosphate), doubling the yield of the system to 17.0 ± 3.8 μg/mL[301].

Ultimately, yeast CFPS is limited by low yields[279, 280]. Additionally, some organelles feature a high abundance of degradative enzymes such as ATPases and phosphatases which may impair reaction productivity by consuming critical substrates in a non-productive manner[280].

1.2.3.3.3 CFPS in Insect Cells

Due to a superior capacity for synthesizing products featuring post-translational modifications, CFPS from insect cells (specifically, from the Fall Army worm *Spodoptera frugiperda*) is one of the fastest-growing CFPS platforms[280, 302]. Typically yielding ~ 45 µg/mL active luciferase[279], these systems have been successfully used to achieve PTMs including acetylation[303]. However, these platforms are hampered by two significant disadvantages that have been the primary limiters of the technology[280]. First, cell cultivation is time consuming and expensive. Second, as is the case with several of the non-model organisms presented here, there is a very limited set of tools for genetic modification of insect cells.

1.2.3.3.4 CFPS in Chinese Hamster Ovary Cells

Chinese hamster ovary (CHO) cells are widely used as industrial workhorses for the expression of human recombinant proteins. For the development of CFPS systems based on CHO cell lysates, Kubick and colleagues developed a lysate preparation method using mechanical lysis and mild treatment of the extract that retains microsomal vesicles of the endoplasmic reticulum within the extract[304]. Such lysates have been applied to the synthesis of 30-50 µg/mL of the protein of interest, including membrane proteins which are difficult to synthesize using other platforms[304-306]. A recent report by Martin *et al* optimized a commercial CHO CFPS system, increasing the yield of superfolder GFP to > 700 µg/mL and demonstrating robust synthesis (>100 µg/mL) of biologically active, intact monoclonal antibodies[307]. Still, use of these platforms is limited by relatively low batch yields and the requirement of a relatively higher percent volume of extract in fully-assembled CFPS reactions[282].

1.2.3.4 Trends in CFPS

Several ongoing trends can be observed in the development of the platforms discussed above **(Figure 1.8)**. First, yields continue to increase for CFPS **(Figure 1.8.A-B)**. These improvements are in part a result of improved solubility in *E. coli* platforms. One approach that has proved beneficial in this case is the use of fusion partners to increase solubility of aggregation-prone proteins[308, 309]. *E. coli* CFPS productivity has also been improved by leveraging genetic engineering tools to functionally inactivate "negative effector" genes (*i.e.* genes whose products destabilize critical CFPS substrates) in the source strains[288, 289]. Additional improvements have resulted from the use of novel energy regeneration systems. For instance, Caschera and Noireaux recently reported the synthesis of 2.3 mg/mL GFP using maltose as the primary energy molecule[284]. Eukaryotic CFPS system productivities have increased as well, generally via efforts to reduce background translation[279].

Second, there has been an effort to reduce the cost of CFPS. This has been pursued on one front by switching from expensive high-energy phosphate bond energy sources such as phosphoenol pyruvate (PEP) or creatine phosphate to cheaper substrates. These efforts have seen great success in *E. coli* CFPS systems, which have been shown to successfully utilize glucose, starch, and maltose as primary sources of energy[284, 285]. Unfortunately, eukaryotic systems have so far not shown the ability to use non-phosphorylated energy substrates[279]. On a second front, cost reduction has been achieved by the development of lysis procedures that do not require expensive homogenization equipment, typically turning to inexpensive methods such as sonication to rupture cells[290, 291].

Third, in recent years efforts to synthesize more complex products have intensified. In particular, innovation using *E. coli* platforms by Jewett and others has enabled limited synthesis of membrane proteins as well as products featuring disulfide bonds or glycosylation events[268, 269, 307, 310]. These improvements highlight the versatility afforded to CFPS systems due to their open nature, as all of these efforts were enabled by the ability to directly add critical components to the reaction environment.

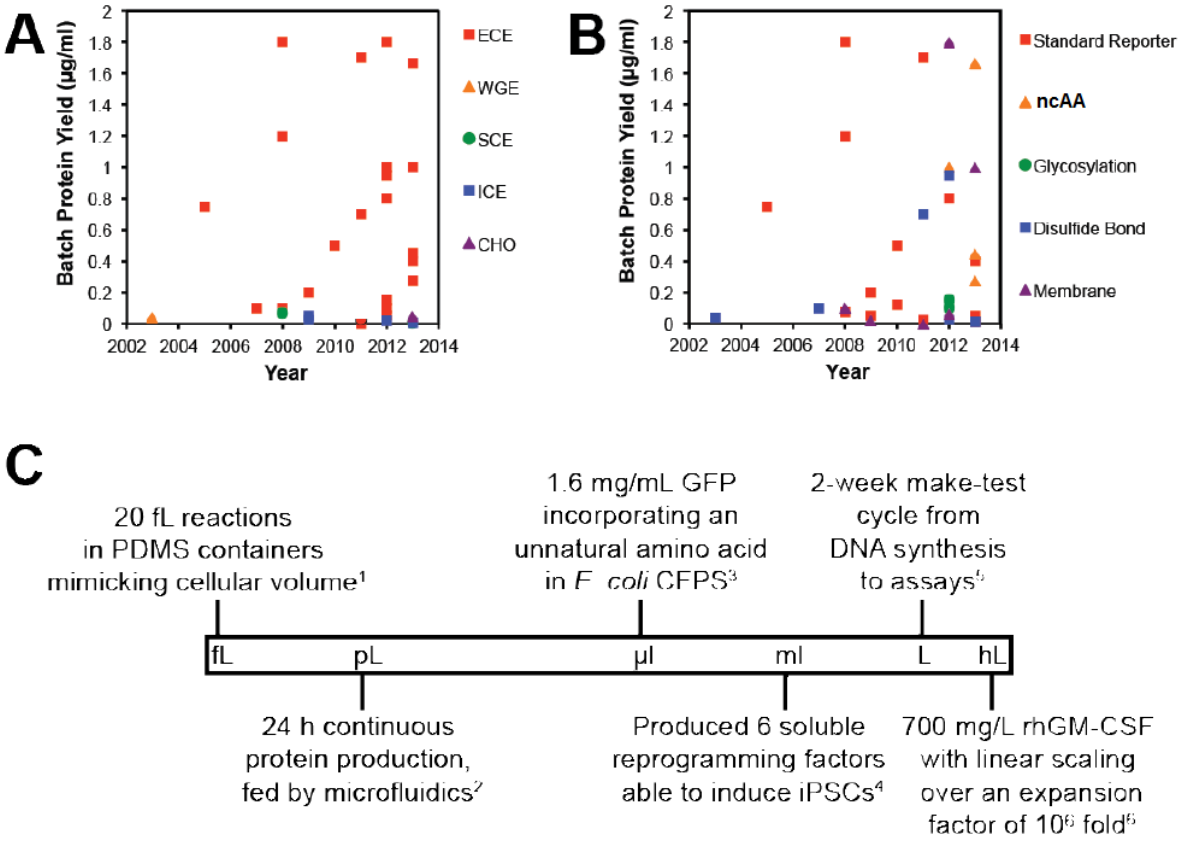


Figure 1.8. Recent trends in CFPS. Batch protein yields from the systems summarized in this chapter are arranged by platform **(A)** and target type **(B)**. Also shown are the volume scales at which CFPS has been successful **(C)**. Abbreviations: ECE – *E. coli* extract, WGE – wheat germ extract, SCE – *S. cerevisiae* extract, ICE – insect cell extract, CHO – Chinese hamster ovary cell extract, PDMS - polydimethylsiloxane, GFP – green fluorescent protein, iPSCs – induced pluripotent stem cells, rhGM-CSF – recombinant human granulocyte macrophage colony-stimulating factor. Figure adapted from [279], Figure 3.

Fourth, the use of CFPS has been expanded by adaptation of the approach to novel applications. For example, several recent studies have leveraged CFPS and improved automation to rapidly screen and characterize biological targets *in vitro*[311, 312]. Other works by Bundy and others have used lyophilized CFPS systems to enable on-demand product synthesis in the absence of a cold chain[30, 274, 275, 313]. In addition, pioneering efforts by Noireaux, Collins, and others have applied CFPS to the generation of genetic circuits *in vitro*[257, 258, 314-317].

Lastly, CFPS reaction size is now able to span 17 orders of magnitude. On the macro side, Zawada *et al* demonstrated that *E. coli* CFPS scales linearly from reaction volumes of 250 µL to 100 L (an expansion factor of 10^6) with yields around 700 µg/mL[264]. On the micro side, recent efforts have sought to miniaturize CFPS reactions for high-throughput applications with reaction volumes as small as fL (*i.e.* 10^{-15} L) described in the literature[318, 319].

1.2.3.5 Non-canonical Amino Acid Incorporation in CFPS

CFPS lends itself particularly well to the use of orthogonal translation systems (OTSs) for co-translational incorporation of non-canonical amino acids (ncAAs) **(Figure 1.6)**. The elimination of physical boundaries facilitates efforts to incorporate bulky and/or charged ncAAs that typically exhibit poor membrane permeability that reduces their bioavailability *in vivo* (*e.g.* p-propargyloxy-L-phenylalanine)[19, 23]. Moreover, the open nature of CFPS coupled with the elimination of cell viability concerns permits the experimenter to directly supplement OTS components (which have known toxicity effects *in vivo*) to reactions at high concentrations typically needed to overcome the poor efficiencies of these enzymes[23]. Finally, CFPS-based ncAA-incorporation efforts uniquely benefit from the use of self-aminoacylating tRNAs featuring

flexizyme ribozymes to catalyze ncAA insertion without the need for laborious ncAA-aaRS-tRNA cognate pair development[320].

To achieve incorporation of the non-canonical moiety in cell-free reactions, typically one or more OTS components are expressed in the extract source strain such that those components are enriched in the resulting lysate[23]. Early methods developed by Swartz and colleagues express only the orthogonal (o)-tRNA during cell growth, with the remaining OTS components (aminoacyl-tRNA synthetase and ncAA) supplied directly to reactions to unite and activate the complete OTS. In an early example, Goerke and Swartz demonstrated synthesis of up to 660 µg/mL of protein containing the ncAA p-azido-L-phenylalanine (pAzF) using crude extracts enriched with o-tRNA[321]. In an adaptation of this approach, the o-tRNA is co-expressed in the CFPS reaction in the form of a transzyme construct featuring self-cleaving hammerhead ribozymes that remove themselves from nascent transcripts to liberate active o-tRNA into the reaction[322]. First validated by Albayrak and Swartz, this approach enabled synthesis of up to 1.7 mg/mL of protein containing a ncAA[322], and was subsequently leveraged by Hong *et al* to improve ncAA incorporation at multiple sites in a single protein[24]. Finally, despite concerns with orthogonal aminoacyl-tRNA synthetase cellular toxicity, several groups have achieved successful ncAA incorporation in CFPS using lysates derived from cells expressing both o-tRNA and synthetase, typically sourced from a pEVOL plasmid[14, 136, 323]. An important exemplary case from Oza *et al* showcased the synthesis of MEK1 kinase featuring user-defined serine phosphorylation events via direct incorporation of phosphoserine[136].

Beyond OTS-mediated incorporation of ncAA species via amber suppression, CFPS approaches also hold great promise for large-scale genetic code reassignment. For instance,

translation systems reconstituted from purified components[324] or crude cell extracts depleted of native tRNAs[325] can be selectively supplied with purified tRNAs to essentially create a custom genetic code with select sense codons left "blank" for ncAA reassignment[326, 327]. Alternatively, in a recent effort a 50S ribosomal subunit was developed with mutations at the peptidyl transferase center that abolish its ability to use several native tRNAs for translation[198]. The use of this ribosome together with tRNAs containing compensatory point mutations enabled small-scale reprogramming of the genetic code, representing a proof of concept of a potential approach for genetic code rewriting and expansion.

As *in vivo*, efforts to co-translationally incorporate ncAAs into proteins in CFPS are limited by competition with release factor 1 (RF1)[23, 24]. Historically, a number of approaches have leveraged the unique features of CFPS to circumvent this competition and increase the efficiency of ncAA incorporation. By one approach, CFPS is performed using a suite of purified translation components defined and assembled by an experimenter. In this case, RF1 activity can be fully avoided by simply withholding the protein from the reaction mix. However, this approach is extremely expensive and unsuitable for large scale production of ncAA-containing products. In crude lysate systems, the barrier-less nature of CFPS has allowed for RF1 to be selectively removed from extracts by augmenting it with an affinity tag (*e.g.* chitin-binding domain[328] or His-tag[329]) and removing it from the system via affinity chromatography. The conditional inactivation of RF1 in lysates has been achieved by installing mutations to make the protein a better substrate for the periplasmic protease OmpT, which RF1 only can interact with during lysate preparation[330]. Finally, generation of lysates from recoded RF1-deficient strains is a promising

future avenue for the application of these technologies towards the production of proteins bearing ncAAs[21, 24].

1.3 This Work: Expanding Capabilities for Studying PTMs

My thesis seeks to build upon recent advances in the use of cell-free systems for site-specific ncAA incorporation to generate platforms for preparative scale synthesis of proteins featuring user-definable sets of PTMs. This was achieved primarily by identifying biological processes limiting the productivity of lysates derived from RF1-deficient strains and genetically modifying those strains in order to eliminate or circumvent those processes. First, I contributed to an effort that sought to improve the protein yield from a system derived from a partially-recoded RF1-deficient *E. coli* strain that had previously been described in the literature. As described in Chapter 2, our work led to dramatic increases in the synthesis capabilities of the platform and informed a set of universal negative effector processes that could be targets for future engineering efforts. Next, based on the observation that errant truncation was still present in the system, I worked on an effort that similarly sought to improve product yield via genome engineering from a fully-recoded RF1-deficient strain, as summarized in Chapter 3. In the interest of facilitating expansion of the technology into new laboratories, I next set out to reduce both the cost and complexity of the system by transforming it into a one-pot platform as illustrated in Chapter 4. Because preparative protein synthesis benefits in general from highly active translation machinery, I next set out to pioneer a CFPS platform based on lysates derived from the fast-growing bacterium *Vibrio natriegens*. This project is presented in Chapter 5. Finally, I set out to improve our capacity for synthesizing homogeneously phosphorylated proteins for fundamental studies into the roles that phosphorylation plays in regulating protein structure/function. My efforts here are

documented in Chapter 6. Complete understanding of the roles played by PTMs has long been opposed in many cases by the inability to obtain large, homogeneous samples of proteins with specific modifications. This thesis addresses this perennial limitation, presenting a synthesis route for designer proteoforms to facilitate fundamental studies into the biology of PTMs.

1.4 Publication Information

Sections of this chapter were published with the following citation information:

Des Soye B.J.*, Patel J.R.*, Isaacs F.J., and Jewett M.C. Repurposing the translation apparatus for synthetic biology. *Curr. Opin. Chem. Biol*. 2015 Jul 15;16: 83-90. doi: 10.1016/j.cbpa.2015.06.008.

CHAPTER 2

2 Improving Cell-free Protein Synthesis Through Genome Engineering of *Escherichia coli* Lacking Release Factor 1

2.1 Abstract

Site-specific incorporation of non-standard amino acids (ncAAs) into proteins opens the way to novel biological insights and applications in biotechnology. Here, we describe the development of a high yielding cell-free protein synthesis (CFPS) platform for ncAA incorporation from crude extracts of genomically recoded *Escherichia coli* lacking release factor 1. We used genome engineering to construct synthetic organisms that, upon cell lysis, lead to improved extract performance. We targeted five potential negative effectors to be disabled: the nuclease genes *rna*, *rnb*, *csdA*, *mazF*, and *endA*. Using our most productive extract from strain MCJ.559 (*csdA*$^{-}$ *endA*$^{-}$), we synthesized 550±40 μg/mL of modified superfolder green fluorescent protein containing *p*-acetyl-L-phenylalanine. This yield was increased to ~1300 μg/mLwhen using a semicontinuous method. Our work has implications for using whole genome editing for CFPS strain development, expanding the chemistry of biological systems, and cell-free synthetic biology.

2.2 Introduction

Crude extract-based cell-free protein synthesis (CFPS) has emerged as a powerful technology platform to study, exploit, and expand the capabilities of biological systems[280, 281,

331]. In recent years, for example, *Escherichia coli*-based CFPS platforms have been applied to the clinical manufacture of therapeutics at the 100 L scale[264], biomolecular breadboarding[316, 317], and site-specific incorporation of non-canonical amino acids (ncAAs) into proteins[24, 322, 328, 332]. Site-specific incorporation of ncAAs into proteins has opened new opportunities for the production and study of biopolymers with chemical properties, structures, and functions that are impossible to create from only the 20 canonical amino acids. Some illustrative examples of these pioneering efforts include the synthesis of antibody-drug conjugates[333, 334], the direct polymerization of protein-based materials[335], and structural assessment of enzyme inhibitors for drug discovery[336].

Amber suppression is the most common approach for opening coding channels to incorporate ncAAs. In this approach, an orthogonal tRNA is reprogrammed to suppress an in frame amber stop codon. This requires multiple biological parts. These include: 1) ncAA-charged o-tRNA substrates that can decode the amber codon (typically produced by an orthogonal aminoacyl-tRNA synthetase (o-aaRS) that is only able to charge a ncAA to its cognate o-tRNA, which is not aminoacylated by the cell's endogenous aaRSs), 2) proper delivery of ncAA-charged o-tRNA substrates by elongation factor Tu (EF-Tu) to the ribosome, and 3) compatible ribosomes[337-339]. Many seminal works by Schultz and others have established and driven the field forward[7, 13, 16, 17, 235, 247, 340-345]. Unfortunately, the technology has been limited by release factor 1 (RF1) competition, which leads to poor expression yields, mainly due to the high level of truncated product[229], and inefficient incorporation of multiple identical ncAAs[21, 24]. Many elegant efforts have focused on alleviating this limitation *in vitro*. For example, researchers have 1) omitted RF1 using reconstituted systems[21, 229-231, 346, 347], 2) silenced RF1 with antibodies

or aptamers[348, 349], or 3) removed tagged RF1 from crude lysates[328, 350]. Recently, *in vivo* efforts have made significant advances in the production of RF1-deficient strains[21] that could be used as a chassis to make crude extracts for CFPS[23]. Indeed, we recently showed the ability to use a genomically recoded RF1-deficient *E. coli* strain (*rEc.E13.ΔprfA*) to improve production of modified full length soluble superfolder green fluorescent protein (sfGFP) containing *p*-propargyloxy-L-phenylalanine (pPaF)[24]. We observed a shift from 20% full-length product (with RF1) to 80% full-length product (without RF1).

The goal of this work was to improve CFPS yields from *rEc.E13.ΔprfA* crude extracts. Such an advance will facilitate new technological applications and provide opportunities to take advantage of cost benefits, yield improvements, and freedom of design for ncAA incorporation as compared to *in vivo* methods[23]. Because *rEc.E13.ΔprfA*[21] was not previously optimized for CFPS[24], we exploited multiplex automated genome engineering (MAGE)[351] to improve extract performance. The key idea was to functionally inactivate negative effectors in the host strain such that they would not be present in the lysate. Previously, deletion of genes for stabilizing DNA template[289], amino acid supply[288], and protein degradation[352] has improved CFPS systems from other source strains.

2.3 Materials and Methods

2.3.1 Strains and Plasmids

The bacterial strains and plasmids used in this study are listed in **Table 2.1**. Spectinomycin (20 μg/mL) was used for culturing strains, kanamycin (50 μg/mL) was used for maintaining pY71-

based plasmids, and tetracycline (20 μg/mL) was used for maintaining the pDULE-o-tRNA plasmid.

2.3.2 Strain Construction and Verification

The strains in this study were generated from *rEc.E13*.Δ*prfA*[21] by disrupting genes of interest with mutagenic oligonucleotides by MAGE[25] **(Table 2.2)**. Cultures were grown in lysogeny broth (LB)-Lennox medium (10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl)[351] at 32°C at 250 rpm throughout MAGE cycling steps. MAGE oligonucleotides were designed to introduce an internal stop codon and frameshift of ~1/4 into the target gene sequence, thereby causing early translational termination as previously reported[25]. Single, double, triple, and quadruple disruptions of *csdA*, *rnb*, *mazF* and *endA*, including singly *rna* disruption, were generated to investigate the effect of their inactivation on CFPS **(Table 2.1)**. Multiplex allele-specific colony PCR was performed to verify gene disruptions[351] by using wild-type forward (-wt-F) or mutant forward (-mut-F) primers and reverse primers (-R; **Table 2.2**). Wild-type and mutant forward primers were identical except at the 3'-ends of the oligonucleotides, and the reverse primers were used for detection of both wild-type and mutant alleles. The mutant allele could be amplified by using the mutant forward and reverse primer set (-mut-F and -R) but not the wild-type forward and reverse primer set (-wt-F and -R). MASC PCR was performed in 20 μL reactions by using a multiplex PCR kit (Qiagen) at 95°C for 15 min, with 30 cycles of 94°C for 30 s, 65°C for 30 s, and 72°C for 1 min, and a final extension of 72°C 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 **(Table 2.2)**.

2.3.3 Growth Rate Assessment

Overnight cultures of engineered strains grown in Luria-Bertani (LB)[353] at 250 rpm at 34°C were diluted to an OD_{600} of 0.05 in 2xYTPG media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L K_2HPO_4, 3 g/L KH_2PO_4, and 18 g/L glucose; adjusted pH to 7.2 with KOH). Diluted cultures (100 µL) were added to 96-well polystyrene plates (Costar 3370; Corning Incorporated, Corning, NY). The OD_{600} was measured at 15 min intervals for 15 h at 34°C in fast shaking mode on a Synergy2 plate reader (Biotek, Winooski, VT). Growth data of each strain was obtained from six replicate wells with two independent cultures. Doubling time was calculated during the early exponential growth phase (OD_{600} = 0.05-0.2).

2.3.4 Cell Extract Preparation

Cells were grown to an OD_{600} of 4.0 in 2xYTPG medium (2 L) in Tunair® shake flasks (1 L culture in each 2.5 L flask) at 34°C and 220 rpm for rapid prototyping of engineered strains. In order to maintain pH ~ 7, KOH (1 mL, 1 N) was added at OD_{600} = 2.0. For the MCJ.559 strain harboring pDULE-o-tRNA, the best CFPS performer, cells were grown in 2xYTPG medium (10 L) in a BIOSTAT C-plus fermenter (Sartorious AG, Göttingen, Germany) to an OD_{600} of 3.0 at 34°C. Cells were pelleted by centrifuging for 15 min at 5000 × *g* and 4°C, washed with cold S30 buffer (3x10 mM tris-acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, 1 mM dithiothreitol)[354], and stored at -80°C. To make cell extract, cell pellets were thawed and suspended in S30 buffer (0.8-1 mL per gram of cells) and lysed in an EmulsiFlex-C3 homogenizer (Avestin, Ottawa, Canada) with a single pass at a pressure of ~138-172 MPa. A chilled syringe was used to inject resuspended cells and collect lysed cells for the small volume of cell suspension

prepared from shake flasks, and a chilled hopper was used for the cells harvested from fermentation. Cell debris and insoluble components were removed by two rounds of centrifugation for 30 min at 30,000 × *g* and 4°C. The supernatant was incubated for 80 min at 120 rpm at 37°C in an empty run-off reaction to optimize the extract activity and then centrifuged for 15 min at 15,000 × *g* at 4°C. The supernatant was flash-frozen in liquid nitrogen and stored at -80°C until use. The total protein concentration of the extracts was 40-50 mg/mL, as measured by Quick-Start Bradford protein assay kits (Bio-Rad, Hercules, CA).

2.3.5 Purification of His-tagged pAcF-tRNA Synthetase

BL21(DE3) harboring pY71-pAcFRS[24] was grown in LB (1 L) to an OD_{600} of 1.0 at 220 rpm and 37°C. pAcF-tRNA synthetase (pAcFRS) was produced by adding isopropyl-β-D-thiogalactopyranoside (IPTG; 0.2 mM, Sigma-Aldrich, St. Louis, MO) for 3 h. Cells were harvested at 5,000 × *g* for 30 min at 4°C, washed with S30 buffer, and stored at -80°C. The frozen cell pellet was thawed in loading buffer (300 mM NaCl, 10 mM imidazole, 50 mM NaH_2PO_4, 5 mM Tris–HCl, pH 8.0)[321], lysed by using a homogenizer at ~138-172 MPa, and centrifuged at 16,000 × *g* and 4°C for 30 min. pAcFRS was purified on a 5 mL Ni-NTA column in a BioLogic DuoFlow FPLC system (Bio-Rad, Hercules, CA). The purified pAcFRS in the elution buffer (300 mM NaCl, 250 mM imidazole, 50 mM NaH_2PO_4, 5 mM Tris–HCl, pH 8.0)[321] was washed three times with S30 buffer by using an Amicon Ultracel YM-30 centrifugal filter and stored at -80°C by adding equal volume of 80% glycerol. The concentration of purified pAcFRS was quantified by Bradford assay.

2.3.6 CFPS Reaction

CFPS reactions were performed to evaluate incorporation of pAcF by using a modified PANOx-SP system[287]. Briefly, a 15 µL CFPS reaction in a 1.5 mL microcentrifuge tube was prepared by mixing the following components: ATP (1.2 mM); GTP, UTP, and CTP (0.85 mM each); folinic acid (34.0 µg/mL); *E. coli* tRNA mixture (170.0 µg/mL); plasmid (13.3 µg/mL); T7 RNA polymerase (100 µg/mL); 20 standard amino acids (2 mM each); nicotinamide adenine dinucleotide (NAD; 0.33 mM); coenzyme-A (0.27 mM); spermidine (1.5 mM); putrescine(1mM); sodium oxalate (4 mM); potassium glutamate (130 mM); ammonium glutamate (10 mM); magnesium glutamate (12 mM); phosphoenolpyruvate (PEP; 33 mM), and cell extract (27% *v/v*). For ncAA incorporation, pAcF (2 mM), pAcFRS (0.5 mg/mL), and linear DNA of o-tRNA^{opt} (10 µg/mL) were additionally added. Linear DNA of o-tRNA^{opt} was amplified from pY71-T7-tz-o-tRNA^{opt} plasmid and transcribed during the cell-free reaction[24]. Furthermore, the o-tRNA was expressed in the source strain during the extract preparation[24]. Each CFPS reaction was incubated for 20 h at 30°C unless noted otherwise. When adding RNase inhibitor, 1 µL (4 U) of inhibitor (Qiagen) was added into the 15 µL cell-free reaction as per the manufacturer's suggestion.

2.3.7 Quantification of Active sfGFP

Active full-length sfGFP protein yields were quantified by measuring fluorescence using a Synergy2 plate reader with λ_{ex} = 485 nm, λ_{em} =528 nm, and cut-off at 510 nm in 96-well half-area black plates (Costar 3694; Corning Incorporated), and the fluorescence units were converted into concentrations by using a standard curve as previously described[24].

2.3.8 Radioactive [^{14}C]Leu Assay

Total and soluble protein yields were quantified by determining radioactive [^{14}C]Leu incorporation by using trichloroacetic acid (TCA)[354]. Radioactivity of TCA-precipitated samples was measured by liquid scintillation counting (MicroBeta2; PerkinElmer).

2.3.9 mRNA Stability Assay

The sfGFP gene was PCR-amplified from the pY71 vector with T7-pro-F and T7-ter-R primers against the T7 promoter and the T7 terminator sequences **(Table 2.2)**. The PCR-amplified linear template was then purified by using a PCR clean-up kit (Promega) and subsequently used as a template for in vitro transcription reactions according to the manufacturer's manual (RiboMAX Large Scale RNA Production System, Promega). The final concentration of mRNA was 1.8 mg/mL. In order to track mRNA stability in our extracts, we replaced the plasmid sfGFP with the mRNA of sfGFP (1800 ng) in the CFPS reaction. For direct measurement of mRNA degradation, 5 μL samples were taken from CFPS reactions during incubation at 30 °C and mixed with equal volumes of RNAprotect Bacteria Reagent (Qiagen, Valencia, CA) and brought to 100 μL with RNase free water. All samples were then purified by using an RNeasy Mini total RNA purification kit (Qiagen) according to the manufacturer's manual. Purified mRNA was visualized on a 2 % formaldehyde agarose gel stained with GelRed (Biotium, Hayward, CA, USA).

2.3.10 DNA Stability Assay

We used pY71-mRFP1-Spinach plasmid (Table S2) to track DNA stability. A preincubation mixture containing 4 μL of cell extract, 12.96 ng/μL of pY71-mRFP1-Spinach plasmid[355], and 6 nM of 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI; Lucerna, New York, NY, USA) was prepared on ice to minimize degradation of plasmid, and then incubated

for 0, 60, and 180 min at 30 °C. CFPS reaction components were added immediately after the preincubation step, and fluorescence of the Spinach aptamer binding to DFHBI was monitored for 180 min by using a CFX96 real-time (RT) PCR module installed on a C1000 Touch Thermal Cycler (Bio-Rad). The excitation and emission wavelengths of the fluorophore were 450–490 nm and 515–530 nm, respectively. The highest fluorescence was detected after 1 h incubation in the RT-PCR machine. For direct assessment of DNA degradation, 15 μL of plasmid DNA and extract mixture was prepared that contained 1 μg of pY71-sfGFP plasmid, 4 μL of extracts from the MCJ.495 or *rEc.E13.ΔprfA* strain, and 3 μL of S30 buffer. After incubation at 30 °C for 0, 15, 30, and 60 min, samples were flash frozen in liquid nitrogen and stored at −20 °C. To remove RNA, 100 μL of RNaseA-containing solution I from an E.Z.N.A. Miniprep kit (Omega Bio-Tek, Norcross, GA, USA) was added to the sample and incubated for 20 min at room temperature, then 200 μL of water was added. Proteins were precipitated by the same volume of phenol/chloroform/isoamyl alcohol (25:24:1) solution, and the plasmid was purified by using a DNA Clean & Concentrator kit (Zymo Research, Irvine, CA, USA). The purified plasmid was digested with BamHI at 37 °C for 90 min to be linearized and was visualized on a 0.7 % agarose gel.

2.3.11 Full-length sfGFP Purification and Mass Spectrometry

To confirm pAcF incorporation at corresponding amber sites, semi-quantitative mass spectrometry (MS) analysis was performed on purified sfGFP with pAcF putatively incorporated. First, full-length sfGFP was purified from CFPS reactions by using C-terminal strep-tags and 0.2 mL gravity-flow Strep-Tactin Sepharose mini-columns (IBA GmbH, Gottingen, Germany) and concentrated by using Microcon YM-10 centrifugal filter columns (Millipore). The purified sfGFP

protein was then analyzed by nanocapillary LC-MS using a 100 mm×75 μm ID PLRP-S column in line with an Orbitrap Elite (ThermoFisher). All MS methods included the following events: 1) FT scan, *m*/z400–2000, 120 000 resolving power and 2) data-dependent MS/MS on the top two peaks in each spectrum from scan event 1 by 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 by using QualBrowser, part of the Xcalibur software packaged with the ThermoFisher Orbitrap Elite (ThermoFisher). In **Figure 4B**, smaller peaks to the right of the colored peaks (Δ*m*=+16 Da) are due to oxidation of the protein—a common electrochemical reaction occurring during electrospray ionization. To remove non-covalent salt and water adducts from intact proteins (in this case, sfGFP), a small level of in-source collision energy (15 eV) was applied. As a result, water loss events from the intact sfGFP (Δ*m*=−18 Da) were detected at minor levels to the left of the major peak.

2.3.12 CAT Plasmid Construction

Gibson assembly was used for seamless construction of plasmids[356]. The wild-type chloramphenicol acetyl transferase (CAT) gene was amplified from pK7CAT[286] by using CAT-FW and -RV primers **(Table 2.2)**, and the pY71 plasmid backbone was amplified from pY71-sfGFP by using pY71-FW and -RV primers **(Table 2.2)**. Both PCR products were cleaned and mixed with Gibson assembly reactants as previously described[356] and incubated at 50 °C for 60 min to construct the pY71-CAT plasmid **(Table 2.1)**. Likewise, CAT-D112Amb, with a single amber site corresponding to Asp112[12], was amplified from pREP-CMD112 to construct the pY71-CAT-D112amb plasmid **(Table 2.1)**.

2.3.13 CAT Activity Assay

Active CAT production was quantified by determining the enzymatic activity of CAT. Cell-free reaction sample (100× diluted, 2 μL) was added to reagent mix (178 μL) containing acetyl-CoA (20 μL, 1 mM) and 5,5′-dithio-bis(2-nitrobenzoic acid) (DTNB; 20 μL, 4 mg/mL). After incubation at 37 °C for 15 min, chloramphenicol (20 μL, 1 mM) was added, and the solution was immediately mixed. The increase in $A_{412\ nm}$ over approximately 5 min was recorded by using the Synergy2 plate reader, and $\Delta A_{412\ nm}$/min was calculated. CAT activity of the cell-free synthesized sample was quantified by comparison to CAT standard activity (C8413, Sigma–Aldrich).

2.3.14 Scaled-up CFPS

Cell-free reaction volumes were increased from 15 to 240 μL in Axygen 1.5 mL polypropylene microcentrifuge tubes (MCT-150-C; Corning, Union City, CA, USA) and a flat-bottom 24-well polystyrene plate (353226; BD Biosciences, San Jose, CA, USA). There was appreciable volume loss due to evaporation in CFPS reactions with volumes less than 100 μL in the flat-bottom 24-well plate; thus, 120 and 240 μL reactions were tested. By filling the outer chambers surrounding the wells with water, which humidified the air, negligible sample evaporation was achieved. Reactions were performed at 30 °C for 20 h.

2.3.15 Semicontinuous Cell-free Reaction

Cell-free reactions (120 μL) were carried out in a microdialysis device (3.5 K MWCO) in a Pierce 96-well Microdialysis Plate (Thermo Fisher Scientific)[224]. The microdialysis device interfaces with 1500 μL of dialysis buffer that contains CFPS reagents as described in section 2.3.6

without T7 RNA polymerase, plasmid, cell extract, o-tRNA^{opt}, or pAcFRS. Time course reactions were monitored at 30 °C for 144 h.

Table 2.1. Strains and plasmids used in this study. Km^R, Sp^R Ap^R, and Cm^R are kanamycin, spectinomycin, ampicillin, and chloramphenicol resistance, respectively. 'Δ' indicates deleted gene, and superscript '-' indicates disabled gene via MAGE.

Strains and plasmids	Genotype/relevant characteristics	Source
Strains		
EcNR2	*MG1655 with λ-prophage::bioA/bioB* and *cmR::mutS*	[25]
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm]* Δ*hsdS λ DE3 = λ sBamHIo* Δ*EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21* Δ*nin5*	New England Biolabs
BL21 Star™ (DE3)	F^- *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm rne131* (DE3)	Life Technologies
rEc.E13.ΔprfA	Δ*prfA* Ω Sp^R, Ap^R, Cm^R, EcNR2 derivative with 13 UAG termination reassigned to UAA at *coaD, hda, hemA, mreC, murF, lolA, lpxK, yafF, pfpA, sucB, fabH, fliN,* and *atpE*	[21]
MCJ.340	*rEc.E13.ΔprfA rna*$^-$	This study
MCJ.435	*rEc.E13.ΔprfA rnb*$^-$	This study
MCJ.436	*rEc.E13.ΔprfA csdA*$^-$	This study
MCJ.437	*rEc.E13.ΔprfA mazF*$^-$	This study
MCJ.495	*rEc.E13.ΔprfA endA*$^-$	This study
MCJ.438	*rEc.E13.ΔprfA rnb*$^-$ *mazF*$^-$	This study
MCJ.527	*rEc.E13.ΔprfA csdA*$^-$ *rnb*$^-$	This study
MCJ.526	*rEc.E13.ΔprfA csdA*$^-$ *mazF*$^-$	This study
MCJ.560	*rEc.E13.ΔprfA rnb*$^-$ *endA*$^-$	This study
MCJ.559	*rEc.E13.ΔprfA csdA*$^-$ *endA*$^-$	This study
MCJ.561	*rEc.E13.ΔprfA mazF*$^-$ *endA*$^-$	This study
MCJ.485	*rEc.E13.ΔprfA csdA*$^-$ *rnb*$^-$ *mazF*$^-$	This study
MCJ.537	*rEc.E13.ΔprfA csdA*$^-$ *rnb*$^-$ *mazF*$^-$ *endA*$^-$	This study
Plasmids		
pY71-sfGFP	Km^R, P_{T7}::super folder *gfp (sfGFP)*, C-terminal strep-tag	[19]
pY71-sfGFP-T216amb	pY71-sfGFP with amber codon at T216	[19]
pY71-sfGFP-2amb	pY71-sfGFP with amber codon at N212 and T216	[24]
pY71-sfGFP-5amb	pY71-sfGFP with amber codon at D36, K101, E132, D190, and E213	[24]

pY71-CAT	Km^R, P_{T7}::CAT (chloramphenicol acetyl transferase)	This study
pY71-CAT-D112amb	pY71-CAT with amber codon at D122	This study
pY71-pAcFRS	P_{T7}::pAcFRS, C-terminal 6x histidine tag	[24]
pY71-mRFP1-Spinach	P_{T7}::mRFP1-Spinach aptamer	[355]
pDULE-o-tRNA	P_{lpp}::o-tRNA, Tet^R	[19]
pY71-T7-tz-o-$tRNA^{opt}$	P_{T7}:: hammer-head ribozyme, o-$tRNA^{opt}$	[24]

Table 2.2. Primers used for MAGE, MASC PCR, and DNA sequencing. Underlined bold text indicates mismatch and frameshift insertion. Four bases at 5'-MAGE oligos were phosphorothioated (*).

Primer Name	Primer Sequence (listed 5' to 3')
MAGE	
rna-MAGE	G*C*T*G*ATTTTCTGACCGTACATGGTCTGTGGCCAGGA**ATGTAAATGCTGATGTA**TTGCCTAAATCGGTTGCTGCCCGTGGTGTTGATGAAC
csdA-MAGE	G*C*T*C*TGGCTGATGTCAGGACGGGTAGTCACGCT**TCAGTTA**AATGCGCACTTCCTGCGGCTCTTTCATAAAGCGGCGGGTAATGCGACGA
rnb-MAGE	A*T*T*T*TGTCACCATCGACAGTGCCAGCACAGAAGATATGGAT**TAACTGA**CTTTTCGCTAAGGCGTTGCCGGATGACAAACTTCAGCTGAT
mazF-MAGE	T*T*G*A*TTGCGTTGTACAAGGAACACACAGACACATACCTGTTTTGTTGTT**TCAGTTA**GAAAGGACTCAGGACAACAGCTGGACGATGTCC
endA-MAGE	C*G*G*T*AAAAGTCCACGCTGACGCGCCCGGTACGTTTTATTGC**TAACTGA**AAAATTAACTGGCAGGGCAAAAAAGGCGTTGTTGATCTGCA
MASC PCR	
rna-wt-F	GTACATGGTCTGTGGCCAGGATTGC
rna-mut-F	GGCCAGGA**ATGTAAATGCTGATGTA**
rna-R	TGGCATGACTTCACTTAGTTTAGC
csdA-wt-F	GCCGCAGGAAGTGCGCATTCAGTCC
csdA-mut-F	GCCGCAGGAAGTGCGCATT**TAACTGA**
csdA-R	CAGTGCGCGGTATTGATCCAGATCGC
rnb-wt-F	CCAGCACAGAAGATATGGATGACGCC
rnb-mut-F	CCAGCACAGAAGATATGGAT**TAACTGA**
rnb-R	TCACTTTCAGGCTGCCAGTCACCGG
mazF-wt-F	CTGTTGTCCTGAGTCCTTTCATGTAC
mazF-mut-F	CTGTTGTCCTGAGTCCTTTC**TAACTGA**
mazF-R	GGCTTTAATGAGTTGTAATTCCTCTG
endA-wt-F	CCCGGTACGTTTTATTGCGGATGT

endA-mut-F	CCCGGTACGTTTTATTGC**TAACTGA**
endA-R	GCTGGCGCTGGTAATTTCGGCGTCA
***prfA* deletion confirmation**	
prfA-F	TGATCTGCAAAGCATCATTTCG
prfA-R	TTGCCTCACGTAACCAGTGTTGATA
prfA-R2	CATAACGGCTGTACATACGGAACAG
CAT plasmid construction	
pY71-FW	GGGCGTAAGTCGACCGGCTGCTA
pY71-RV	TTCTCCATATGTATATCTCCTTCTTAAAGTT
CAT-FW	ATATACATATGGAGAAAAAAATCACTGG
CAT-RV	CGGTCGACTTACGCCCCGCCCTG
DNA sequencing	
rna-seq-F	GTTTCTCTGCTTCCCTTCTCTTCT
csdA-seq-F	CTGCTGGACCACCTGAAACGTGGCA
rnb-seq-F	CTGAAAGGCGATCGTTCTTTCTATG
mazF-seq-F	GTAAAGAGCCCGTATTTACGCTTGC
endA-seq-F	ATGTACCGTTATTTGTCTATTGCTGC
mRNA stability	
T7-pro-F	TCGATCCCGCGAAATTAATACGACTCACTATAGG
T7-ter-R	CAAAAAACCCCTCAAGACCCGTTTA
Orthogonal tRNA amplification	
T7tRNA500-F	CCGAAGGTAACTGGCTTCAGCAGAG
T7tRNAopt-R	TGGTCCGGCGGAGGGGATTTGAACCCCTG

2.4 Results

2.4.1 Strain Construction by MAGE

We targeted the functional inactivation of five nucleases encoded by *rna*, *rnb*, *csdA*, *mazF*, and *endA*. *In vivo*, these nucleases play important roles in regulating DNA and RNA through degradation. However, their presence in crude cell extracts is expected to be deleterious[357], leading to template instability and reaction termination. RNase A (encoded by *rna*) degrades RNA by catalyzing the cleavage of phosphodiester bonds[358], and identification of strains (*e.g.*, MRE600, A19) lacking *rna* was important for early studies in *in vitro* translation. RNase II (encoded by *rnb*) is responsible for mRNA decay by 3’ to 5’ exonuclease activity[359], and cell

extracts lacking RNase II exhibit a 70% increase in CFPS efficiency[357]. MazF (encoded by *mazF*) is a toxin that degrades mRNA by sequence-specific (ACA) endoribonuclease activity, which could affect transcript stability[360]. CsdA (encoded by *csdA*) is part of a cold shock degradosome along with RNase E and induces mRNA decay in cold shock, which the cells experience during harvest prior to extract generation[361]. Finally, DNA-specific endonuclease I (encoded by *endA*) breaks double stranded DNA [362] and its deletion has previously been shown to be important for extending the duration of CFPS reactions[289]. We hypothesized that nuclease deletion could stabilize both the DNA template and mRNA transcripts, leading to improved yields[357-362].

We first engineered the *rEc.E13.ΔprfA* strain by disrupting the RNAase genes, both individually and in combinations, using MAGE **(Figure 2.1.A)**. Specifically, we used MAGE oligos to introduce an internal stop signal (UAA codon) and frame shift mutation at ~1/4 into the open reading frame of the target gene (**Table 2.2**). We generated single disruptions of *rna*, *rnb, csdA*, and *mazF*, as well as multiple disruptions of *rnb*, *csdA*, and *mazF*, in different combinations to create a series of RNase mutants (**Table 2.1**). Gene disruptions were screened by multiplex allele specific PCR that amplified PCR bands specific to each mutation **(Figure 2.1.B)** and confirmed by DNA sequencing. We then measured growth rates for each of the MAGE-modified strains in 2xYTPG media (the media used in preparation of cell lysates) to determine how the gene disruptions might affect cellular fitness. Growth rates of the strains with single or multiple disruptions of *rna, rnb, csdA*, *mazF* were approximately ± 25% relative to the strain *rEc.E13.ΔprfA*, with the exception of *rnb* and *mazF* double disruption (MCJ.438) which displayed a significant growth defect **(Table 2.3)**.

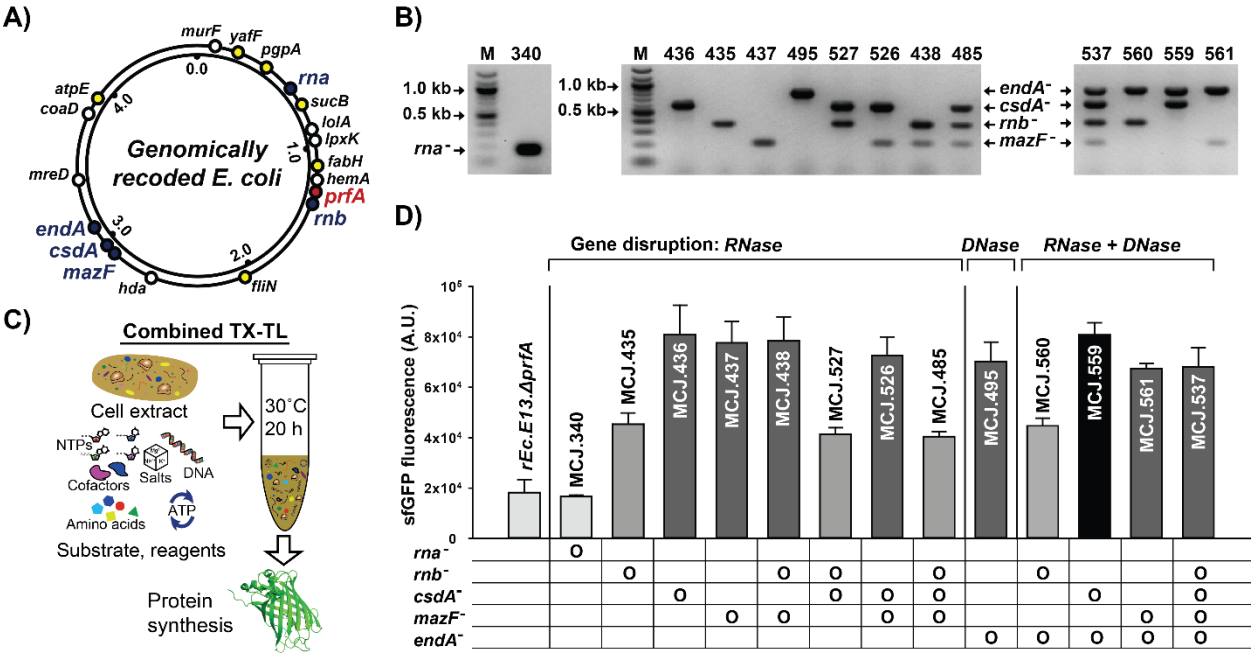


Figure 2.1. Strain construction, verification, and cell-free protein synthesis performance. A) Genomic locations of five nuclease-encoding genes (black circles), *prfA* (gray circle), and 13 genes with stop codons re-coded from UAG to UAA (white circles). Numbers in inner circle indicate millions of bases. B) Verification of nuclease gene mutations by using multiplex allele-specific PCR. Mutant alleles were amplified by using the mutant forward and reverse primer sets (-mut-F and -R; **Table 2.2**). Mutant strain numbers are indicated at the top of the gels. A table in (D) details mutations per each strain. M: DNA ladder. C) Scheme of combined transcription–translation (TX–TL) reaction for sfGFP CFPS. D) Comparison of CFPS efficiency of different cell extracts. Active wild-type sfGFP was synthesized by using cell extracts derived from genomically recoded *E. coli* with single and multiple inactivation of nucleases. At least three independent reactions for each sample were performed for 20 h at 30 °C, and one standard deviation is shown.

Table 2.3. Doubling time of MAGE strains. Cells were grown in 2xYTPG medium at 34°C in a 96-well plate.

Strains	Doubling time (min)	Percentage (%)
rEc.E13.ΔprfA	34.5 ± 0.4	100
MCJ.340	38.1 ± 0.4	110
MCJ.435	43.8 ± 0.8	127
MCJ.436	41.7 ± 0.4	121
MCJ.437	32.4 ± 0.0	94
MCJ.438	52.5 ± 0.4	152
MCJ.527	42 ± 2	121
MCJ.526	40 ± 1	116
MCJ.485	38 ± 1	111

MCJ.495	34 ± 1	97
MCJ.560	39 ± 1	114
MCJ.559	36.6 ± 0.6	106
MCJ.561	33.0 ± 0.8	96
MCJ.537	39.5 ± 0.9	114

2.4.2 Inactivation of RNase II, CsdA, and MazF Improves CFPS by Reducing mRNA Degradation

Lysates from each engineered strain were tested in CFPS to assess their overall protein synthesis capability. For rapid screening, we prepared lysates from shake flask cultures and a syringe-based homogenization method that contrasts our previous work, which used a fermenter for cell growth[24]. CFPS of sfGFP was carried out in 15 µL combined transcription-translation (TX-TL) reactions for 20 h at 30°C **(Figure 2.1.C)**. The *rna* mutation was selected first because of its presence in the commonly used CFPS A19 and MRE600 source strains[363]. However, functional inactivation of *rna* (MCJ.340) in *rEc.E13.ΔprfA* did not impact wild-type sfGFP synthesis, as measured by fluorescence, when compared to the parent strain **(Figure 2.1.D)**. In contrast, single disruption of *rnb*, *csdA*, or *mazF* increased CFPS yields by two- to fourfold **(Figure 2.1.D)**. Next, we investigated the effect of disabling multiple RNase genes together (*rnb*, *csdA*, and *mazF*) in CFPS. CFPS yields were not improved among those combinations of gene disruption and, in fact, decreased in some cases (MCJ.528 and MCJ.485; **Figure 2.1.D**). Taken together, our results show that inactivation of RNase II, CsdA, and MazF are beneficial for CFPS. However, disabling *rnb* in combination with other RNase genes investigated was not beneficial.

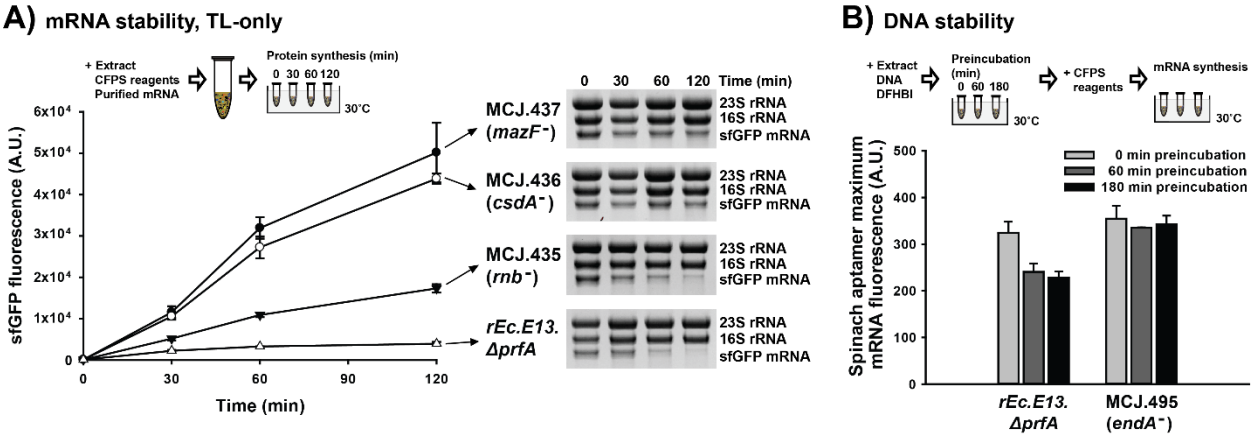


Figure 2.2. The impact of functionally inactivating nucleases on cell-free transcription and translation. A) Cell-free translation (TL)-only reactions of wild-type sfGFP from purified mRNA in different single RNase-deficient cell extracts. At least three independent reactions for each sample were performed for 120 min at 30 °C. sfGFP synthesis was monitored by sfGFP fluorescence (left), and mRNA levels were assessed by an RNA gel (right). **B)** Cell-free Spinach aptamer synthesis by using endonuclease I-deficient (MCJ.495) and -present (*rEc.E13.ΔprfA*) extracts. After preincubation (0, 60, and 180 min) of Spinach aptamer plasmid DNA with cell extract, CFPS reagents were added and incubated at 30 °C. Maximum mRNA synthesis levels from the mRNA synthesis time course **(Figure 2.4)** were compared. At least three independent reactions for each sample were performed, and one standard deviation is shown.

To test our hypothesis that inactivation of RNases would stabilize mRNA in our reactions, we carried out cell-free TL-only reactions for 120 min at 30 °C, priming the reaction with purified mRNA template from the sfGFP gene (600 ng per 15 μL reaction), as opposed to DNA template. Without the ability to replenish mRNA from T7 RNA polymerase, as was possible in combined TX–TL experiments used above **(Figure 2.1.D)**, we could now observe the impact of the genomic changes on RNA stability. For this analysis, we specifically focused on extracts from the single gene disruption strains: MCJ.435 (*rnb*⁻), MCJ.436 (*csdA*⁻), and MCJ.437 (*mazF*⁻). As compared to the extract from the parent strain (*rEc.E13.ΔprfA*), disruption of *mazF*, *csdA*, and *rnb* increased cell-free translation 13-, 11-, and fourfold, respectively **(Figure 2.2.A, left)**. In addition to quantifying sfGFP synthesis by cell-free TL-only reactions in lysates from different genomically

modified strains, we also examined the mRNA degradation profiles by incubating 1800 ng of purified sfGFP mRNA in the cell-free reaction. As expected, mRNA levels were maintained at higher levels in extracts from RNase-deficient strains. Specifically, more than ~60 % of sfGFP mRNA remained after 120 min incubation with the extracts from single disruption of *mazF* or *csdA*, whereas 16 % remained with *rnb* disruption, and mRNA levels in the parent extract derived from *rEc.E13.ΔprfA* were entirely degraded **(Figure 2.2.A, right)**. These results were consistent with the TX–TL reactions **(Figure 2.1.D)** and indicate that inactivating RNases from the lysate source strain reduces mRNA degradation and, in turn, improves CFPS.

2.4.3 Inactivation of Endonuclease I Improves CFPS by Stabilizing the DNA Template

We next investigated the effects of disrupting the DNA-specific endonuclease I (MCJ.495). It has previously been observed that an *endA* deletion strain exhibits increased plasmid DNA production *in vivo*[364], but its role was not clear *in vitro*, as the *endA* deletion was previously assessed only in combination with *recCBD* deletion[289]. In CFPS reactions performed with extracts from source strains lacking endonuclease I (MCJ.495), we observed a greater than fourfold increase in sfGFP synthesis compared to that of rEc.E13.ΔprfA **(Figure 2.1.D)**.

We hypothesized that the improved CFPS yields were a result of plasmid DNA stability. To test this hypothesis, we directly incubated plasmid DNA in the extract alone and monitored plasmid DNA stability by gel electrophoresis. We did not detect differences in plasmid DNA concentrations when comparing extracts with or without endonuclease I **(Figure 2.3)**; however, our results could be confounded by the fact that DNase activity can be inhibited in cell

extracts[365]. Thus, we tried an alternative approach that better mimicked our CFPS conditions. The key idea was to preincubate plasmid DNA with extracts from strains with or without endonuclease I, followed by CFPS **(Figure 2.2.B)**. If the DNA template was degraded during preincubation, or the transcription reaction was inhibited by endonuclease I in some way, less mRNA would be synthesized, which in turn could be responsible for higher CFPS yields when *endA* was disabled. Plasmid DNA containing the mRFP1-Spinach aptamer gene **(Table 2.2)** was pre-incubated with cell extract and a fluorophore molecule, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), for 0, 60, and 180 min. Then, mRNA was synthesized upon addition of CFPS reagents and quantified by measuring the fluorescence of DFHBI-bound Spinach aptamer mRNA[355, 366]. Similar levels of Spinach aptamer mRNA were synthesized in MCJ.495 (*endA*$^{-}$) extracts before and after the preincubation. In contrast, the extract with endonuclease I (*rEc.E13. ΔprfA*) decreased the maximum mRNA synthesis level by 25 % after preincubation **(Figure 2.2.B and 2.4)**. These results support the hypothesis that *endA* disruption improves CFPS by helping stabilize the DNA template.

As inactivating endonuclease I was beneficial for CFPS, we subsequently applied MAGE to create a variety of *endA*- and *RNase*-disrupted strains **(Table 2.1)**. Although we had hoped to observe a synergistic effect from combining these beneficial mutations, we did not observe further improvement **(Figure 2.1.D)**. Most likely, inactivation of a single RNase or endonuclease I facilitates sufficient mRNA in the reaction, resulting in different substrates or protein effectors (e.g., amino acid supply[288], protein degradation[352]) now limiting CFPS.

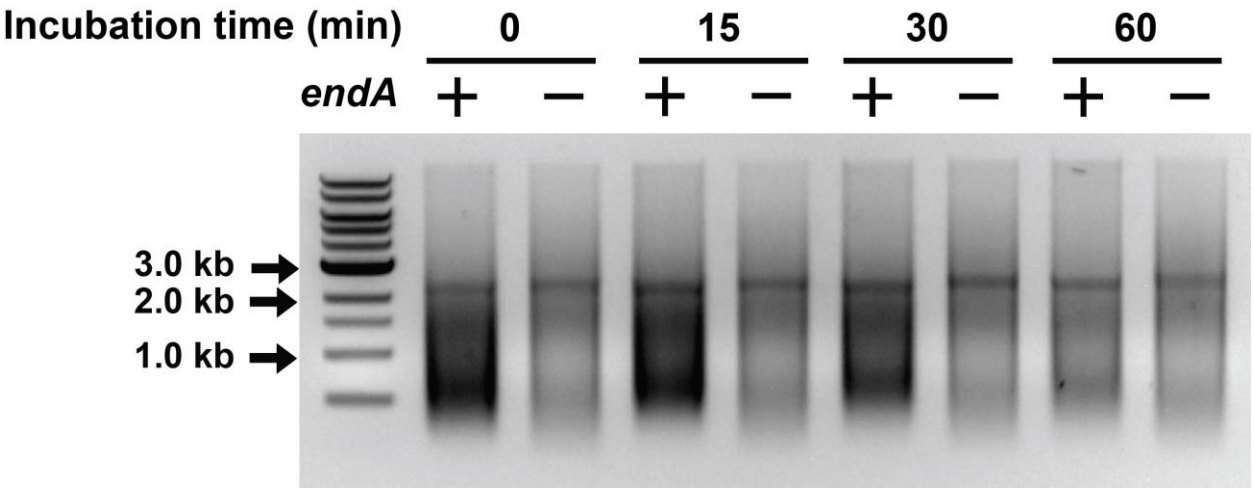


Figure 2.3. DNA gel to assess degradation of plasmid DNA in crude extracts with or without endonuclease I. pY71-sfGFP plasmid DNA was incubated with cell extracts from *rEc.E13.ΔprfA* (*endA*$^{+}$) and MCJ.495 (*endA*$^{-}$) strains for 0, 15, 30 and 60 min at 30°C. Samples were RNase A treated and purified using phenol:chloroform:isoamyl alcohol solution followed by DNA purification kit. The purified plasmid DNA was linearized by BamHI digestion prior to loading on agarose gel (0.7%). Expected band size was 2.5 kb.

2.4.4 Genomically Modified Strains Provide Unique Benefits for CFPS

We then carried out a series of optimization experiments to see if the genomically modified source strains provided unique benefits for CFPS or if yield improvements could have simply been achieved by optimization of reaction conditions using the original host (*rEc.E13.ΔprfA*). For this analysis, we selected the MCJ.559 (*csdA*$^{-}$ *endA*$^{-}$) strain, which had the highest CFPS yields **(Figure 2.1.D)**. We specifically explored the impact of altering the DNA plasmid concentration, changing the T7 RNA polymerase concentration, and adding RNase inhibitor in CFPS reactions with lysates derived from strain MCJ.559 and the parent. Increasing the concentrations of plasmid DNA **(Figure 2.5.A)** or T7 RNA polymerase **(Figure 2.5.B)** with the parent *rEc.E13.ΔprfA* extract did not improve sfGFP synthesis. In fact, CFPS yields decreased. Furthermore, the current concentrations of plasmid DNA (13.3 ng/μL) and T7 RNA polymerase (0.1 μg/μL) used in this

study were already optimal, as similar trends were obtained in the extract from MCJ.559 **(Figure 2.5.A and 2.5.B)**. Addition of RNase inhibitor to CFPS reactions also did not improve protein synthesis **(Figure 2.5.C)**. Together, our data indicates that the improvement of CFPS comes from the nuclease disruptions and is not achievable by simply adjusting the initial cell-free components.

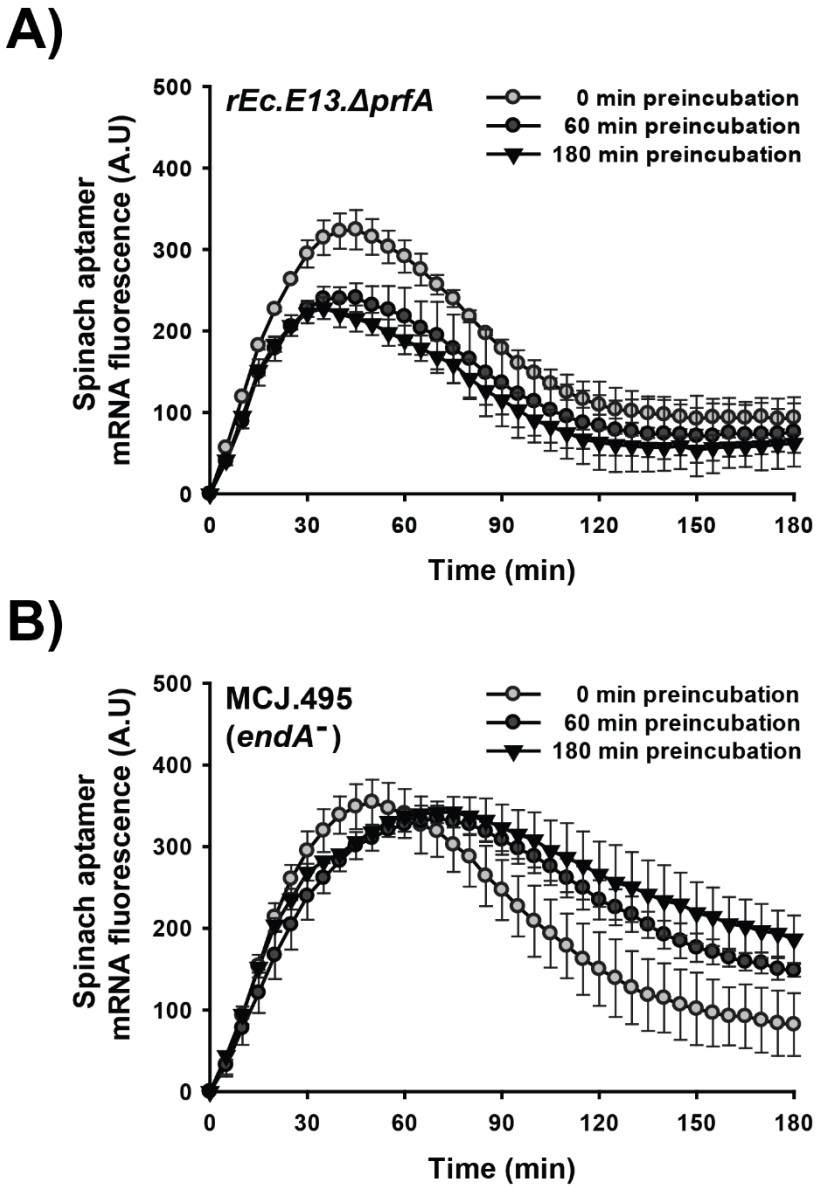


Figure 2.4. Time course Spinach aptamer mRNA synthesis using A) endonuclease I present (*rEc.E13.ΔprfA*) and B) deficient (MCJ.495) extract. Plasmid DNA containing Spinach aptamer gene was preincubated for 0, 60, and 180 min with cell extract and DFHBI. CFPS reactions were performed upon adding CFPS reagents and monitored for 180 min. Maximum mRNA fluorescence was taken for the comparison of mRNA stability in Figure 2B of the main text. At least three independent reactions for each sample were performed at 30°C. One standard deviation is shown.

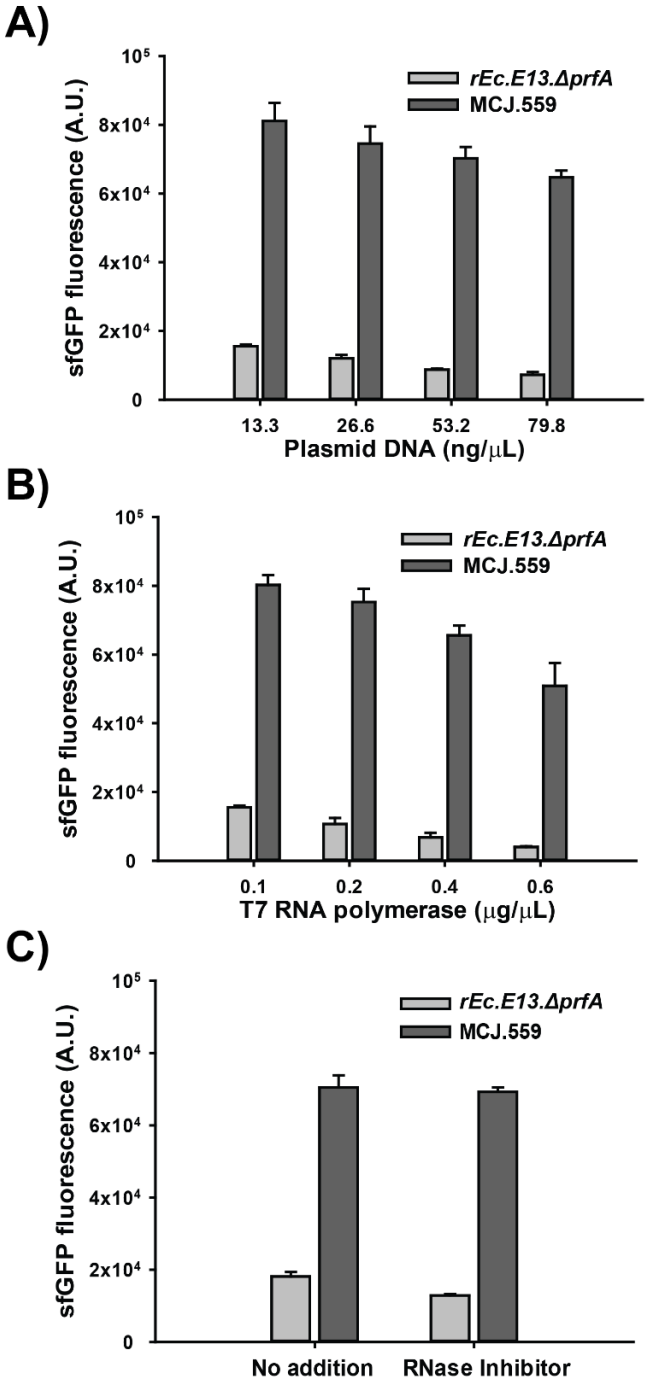


Figure 2.5. Optimization of plasmid DNA and T7 RNA polymerase and the effect of RNase inhibitor in CFPS reactions. CFPS yields observed by increasing the concentration of **A)** plasmid DNA and **B)** T7 RNA polymerase using the extracts from *rEc.E13.ΔprfA* and MCJ.559. **C)** CFPS yields observed in the presence or absence of RNase inhibitor. At least three independent reactions for each sample were performed for 20 h at 30°C. One standard deviation is shown.

2.4.5 MAGE Gene Disruption is Robust and Stable

We also confirmed that the mutations generated by MAGE were not reverted. After 250 generations of the MCJ.559 strain in a rich medium, the cell growth rate was similar (within 4 %) compared to the strain before 250 generations **(Figure 2.6.A)**, the extract performance was the same **(Figure 2.6.B)**, and mutations of *prfA*, *csdA*, and *endA* were preserved, as confirmed by PCR for all three and DNA sequencing for *csdA* and *endA* disruption **(Figure 2.6.C)**. These results highlight that the MAGE gene disruption approach is robust and stable on time horizons that would be associated with a seed train in the laboratory.

2.4.6 MAGE-improved Extract Enhances Single and Multiple Identical ncAA Incorporation

With CFPS improvements from engineered strains in hand, we assessed ncAA incorporation in extracts from our best strain. To do so, we first prepared crude extract from the MCJ.559 strain grown in a 10 L fermenter, which provides exquisite control over growth conditions[367]. In this case, o-tRNA was constitutively expressed during the cell growth as previously reported[19]. As a positive control, we then tested wild-type sfGFP synthesis in a cell-free reaction by using the fermenter-prepared extract. We obtained 660±40 μg/mL of wild-type sfGFP, which was similar to the yields obtained with extract prepared from a shake flask **(Figure 2.7)**. For ncAA incorporation, we quantitatively tested the incorporation of *p*-acetyl-L-phenylalanine (pAcF) into sfGFP with an in-frame amber codon at single and multiple positions **(Table 2.1)**. The necessary components of the orthogonal translation system (OTS) were also added. Specifically, we added 10 μg/mL linear DNA of optimized o-tRNA (o-tRNA^{opt}) in the cell-

free reaction for in situ synthesis of additional o-tRNA (i.e., beyond that overexpressed in the source strain)[24, 322]. The orthogonal pAcF-tRNA synthetase (pAcFRS) was overproduced, purified as previously described[24], and added at a level of 0.5 mg/mL in the cell-free reaction. The ncAA, pAcF, was supplied at a level of 2 mM in each CFPS reaction.

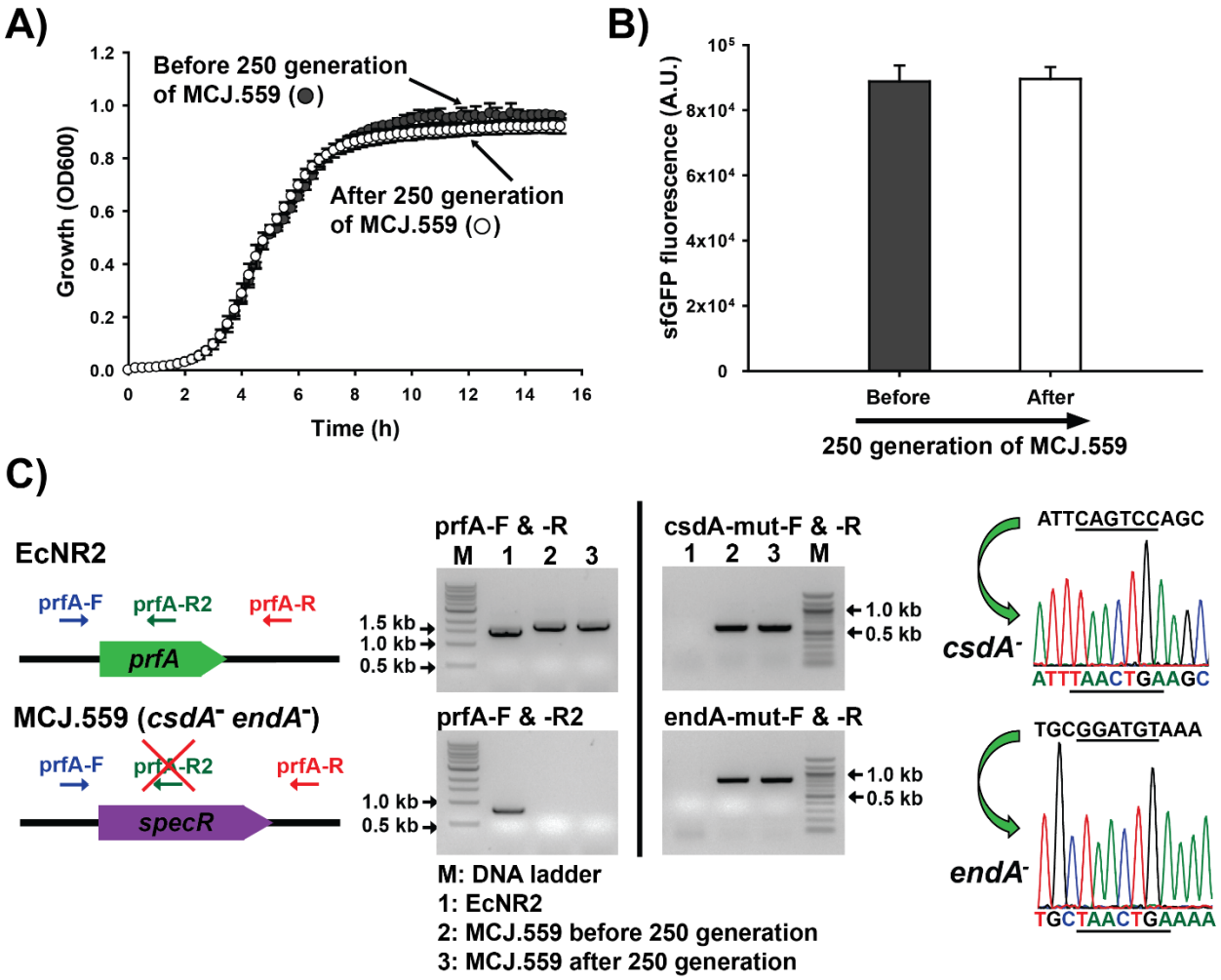


Figure 2.6 (previous page). Assessing stability of the MCJ.559 strain. A) Growth of MCJ.559 before and after 250 generations was assayed in LB medium at 32 °C in 96-well plates. Each data point is the average of ten replicate wells from two independent cultures. **B)** CFPS of sfGFP by using an MCJ.559 crude extract before and after 250 generations shows that extract performance was the same. At least three independent CFPS reactions for each sample were performed for 20 h at 30 °C. C) PCR verification of *prfA*, c*sdA*, and *endA* mutation. "s*pecR*" indicates spectinomycin resistance gene in replacement of *prfA*. On the right is a trace of DNA sequencing results for the *csdA*and *endA* mutation showing the introduction of the UAA stop codon at the desired location.

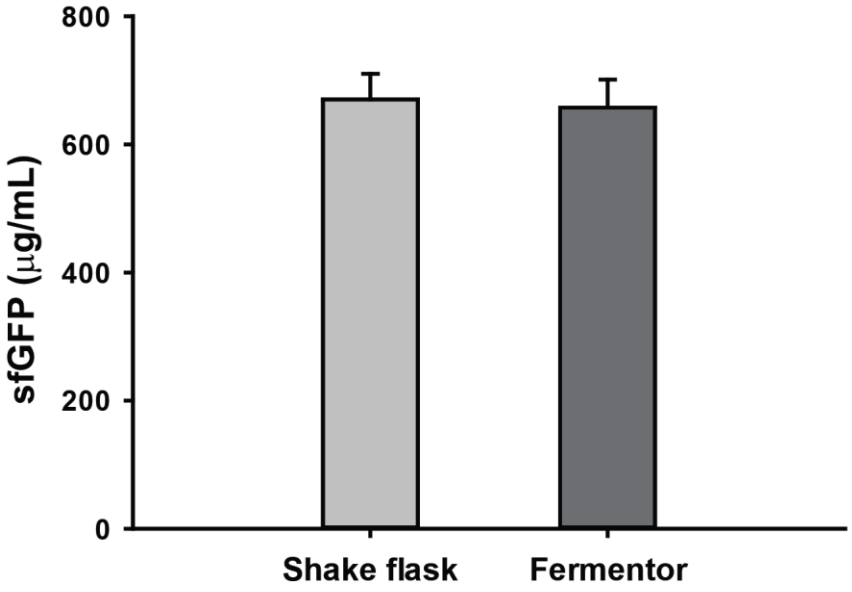


Figure 2.7. Cell extract performance comparison between shake flask- and fermentor-grown MCJ.559. sfGFP synthesis was quantified by measuring the fluorescence in CFPS reactions. Three independent reactions for each sample were performed for 20 h at 30°C. One standard deviation is shown.

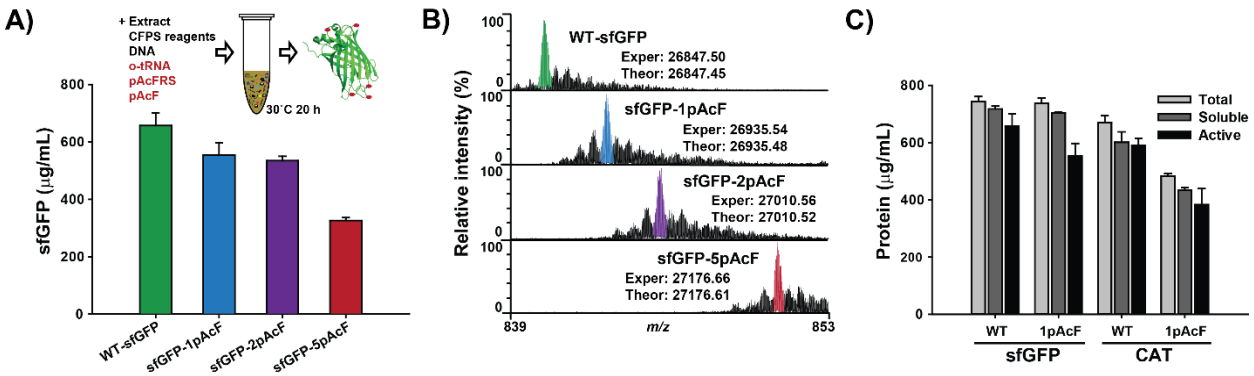


Figure 2.8. pAcF incorporation at single and multiple amber sites by using the improved cell extract from the MCJ.559 strain. A) Yields of active wild-type sfGFP (WT-sfGFP) and modified sfGFP proteins containing one, two, and five pAcFs. **B)** Spectrum of the 32+ charge state of sfGFP, obtained by top-down mass spectrometry and illustrating site-specific incorporation of pAcF at single and multiple sites. Major peaks (gray) in each spectrum coincide with the theoretical peaks for each species (Figure S6). "Exper" indicates experimentally obtained protein mass, and "Theor" indicates theoretically calculated protein mass (Table S4). Smaller peaks to the right of the major peaks are due to oxidation of the protein—a common electrochemical reaction occurring during electrospray ionization. Water loss events from the intact sfGFP were detected at minor levels to the left of the major peaks. **C)** Comparison of total, soluble, and active protein yields of sfGFP and CAT with and without single pAcF. At least three independent CFPS reactions for each sample were performed for 20 h at 30 °C for (A) and (C), and one standard deviation is shown.

We synthesized 550±40 μg/mL of modified sfGFP containing a single pAcF, which represents an ~84 % yield as compared to wild-type sfGFP production **(Figure 2.8.A)**. In addition, we obtained a 540±20 μg/mL yield for modified protein synthesis for two pAcF incorporations (81 % yield) and 330±10 μg/mL for five pAcF incorporations (50 % yield; **Figure 2.8.A**). These results represent a threefold improvement in modified sfGFP synthesis as compared to our previous work on *rEc. E13.ΔprfA* extract[24], as well as more than a tenfold higher protein expression titer (in g/L) compared to recent in vivo ncAA incorporation into GFP[18, 229]. Furthermore, the modified protein synthesis titer was ~2.8 times higher in MCJ.559 extracts than those from a BL21 Star (DE3) extract which contains RF1, whereas the wild-type sfGFP synthesis was similar with both extracts **(Figure 2.9)**. Thus, the MAGE-improved MCJ.559 extract significantly increased the synthesis of proteins containing ncAAs.

We then 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 pAcF into sfGFP. **Figure 2.8.B** shows the 32+ charge state of sfGFP and clearly illustrates mass shifts corresponding to the incorporation of one, two, and five pAcF residues. Site-specific incorporation of pAcF, as detected by MS, was greater than 95 % in all samples **(Figure 2.8.B)**, with less than 3 ppm difference between experimental and theoretical protein masses **(Table 2.4)**. In other words, we achieved efficient, high yielding, and pure site-specific pAcF incorporation into sfGFP. To demonstrate that our observations were not limited to sfGFP, we also examined pAcF incorporation into chloramphenicol acetyl transferase (CAT). Active CAT containing pAcF was synthesized at titers of 380±60 μg/mL, with a ~65 % yield of wild-type protein production **(Figure 2.8.C)**. Hence, the MAGE-enhanced extract provides the synthesis of soluble and active proteins containing pAcF, like wild-type proteins, and may be applied for different types of protein production.

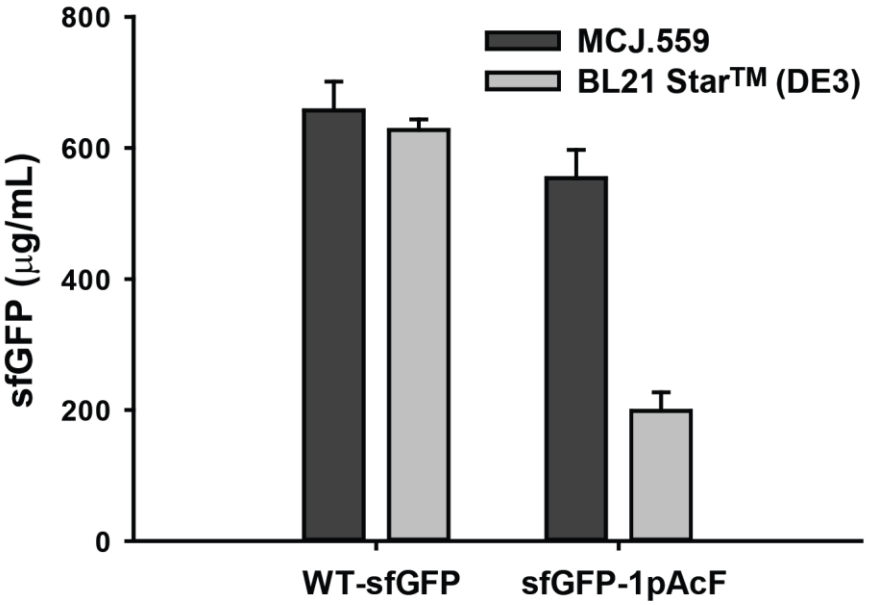


Figure 2.9. Cell extract performance comparison between MCJ.559 and BL21 Star™ (DE3) extract. Active sfGFP synthesis was quantified by measuring fluorescence in CFPS reactions. Single pAcF was incorporated to the amber position corresponding to T216 of sfGFP. Three independent reactions for each sample were performed for 20 h at 30°C. One standard deviation is shown.

Table 2.4. Monoisotopic masses calculated from mass spectrometric data. 'Experiment' indicates experimentally obtained protein mass, and 'Theoretical' indicates theoretically calculated protein mass.

sfGFP species	Mass (Da, Experiment)	Mass (Da, Theoretical)	Error (ppm)	Shift from WT-sfGFP (Da, Experiment)	Shift from WT-sfGFP (Da, Theoretical)
WT-sfGFP	26847.50	26847.45	2.6	-	-
sfGFP-1pAcF (T216)	26935.54	26935.48	2.2	88.04	88.03
sfGFP-2pAcF (N212, T216)	27010.56	27010.52	1.5	163.06	163.07
sfGFP-5pAcF (D36, K101, E132, D190, E213)	27176.66	27176.61	1.8	329.16	329.16

2.4.7 CFPS with MAGE-improved Extract is Scalable

We then set out to demonstrate the potential for scale-up in CFPS reactions that incorporate ncAAs into proteins. Specifically, we tested the effect of increasing the CFPS reaction volume in both a microcentrifuge tube and a flat-bottom 24-well plate. By increasing the reaction volume from 15 to 240 μL, the efficiency of wild-type sfGFP production significantly decreased in the microcentrifuge tube, whereas the same reactions in the flat-bottom 24-well plate did not decrease the protein yields **(Figure 2.10.A)**. We observed similar results for the synthesis of modified protein **(Figure 2.10.A)**. Our results are consistent with those of Voloshin and Swartz, who reported these phenomena previously; specifically, the impact of surface-area-to-volume ratios on CFPS yields[368]. In order to confirm that the decrease in production yield was not specific to active GFP formation (i.e., chromophore maturation), we measured total and soluble sfGFP

production by radioactive [^{14}C]Leu incorporation, which counts all synthesized sfGFP protein. **Figure 2.11** clearly shows that total and soluble sfGFP synthesis was decreased as reaction volume was increased. This is consistent with our observations for active sfGFP synthesis **(Figure 2.10.A)**. We next assessed incubation time to see if the reduced CFPS yield as a function of increasing reaction volume was recovered by increasing incubation time. This was not the case. After 24 h, sfGFP production in all batch reactions was saturated and did not increase with further incubation time **(Figure 2.10.B)**. Modified sfGFP containing single pAcF showed similar results **(Figure 2.10.C)**. In summary, our results indicate that CFPS with RF1-deficient MCJ.559 extract is scalable when accounting for surface-area-to-volume effects.

2.4.8 Semicontinuous CFPS Increases Protein Production Yield

We then applied semicontinuous CFPS[224, 369] using a microdialysis device (3.5 K MWCO) to increase sfGFP production yield. In semicontinuous reactions, substrates and byproducts passively diffuse between the CFPS reaction and a substrate reservoir to sustain small molecule concentrations necessary to keep the reaction active. With a semicontinuous setup, wild-type sfGFP and modified sfGFP production were continually increased until 96 h, yielding titers of about 1900±50 and 1300±100 μg/mL, respectively **(Figure 2.10.B and 2.10.C)**. We expect that these titers could be improved further if the external substrate reservoir was exchanged or the extract condensed, as in the pioneering work of Yokoyama[370]. That said, the modified protein titer from a semicontinuous CFPS setup was >20 times higher than those recently observed in cells on a gram per liter basis[18, 229], joining a growing body of work showing improvement over in vivo titers when using CFPS[19, 24, 322, 371]. Taken together, our results indicate that the yield was significantly improved by using a semicontinuous reaction setup. This is consistent with

previous work from Ozawa et al., which showed the ability to produce high yields of protein with ncAAs by using a semicontinuous CFPS reaction setup and a standard *E. coli* strain[371].

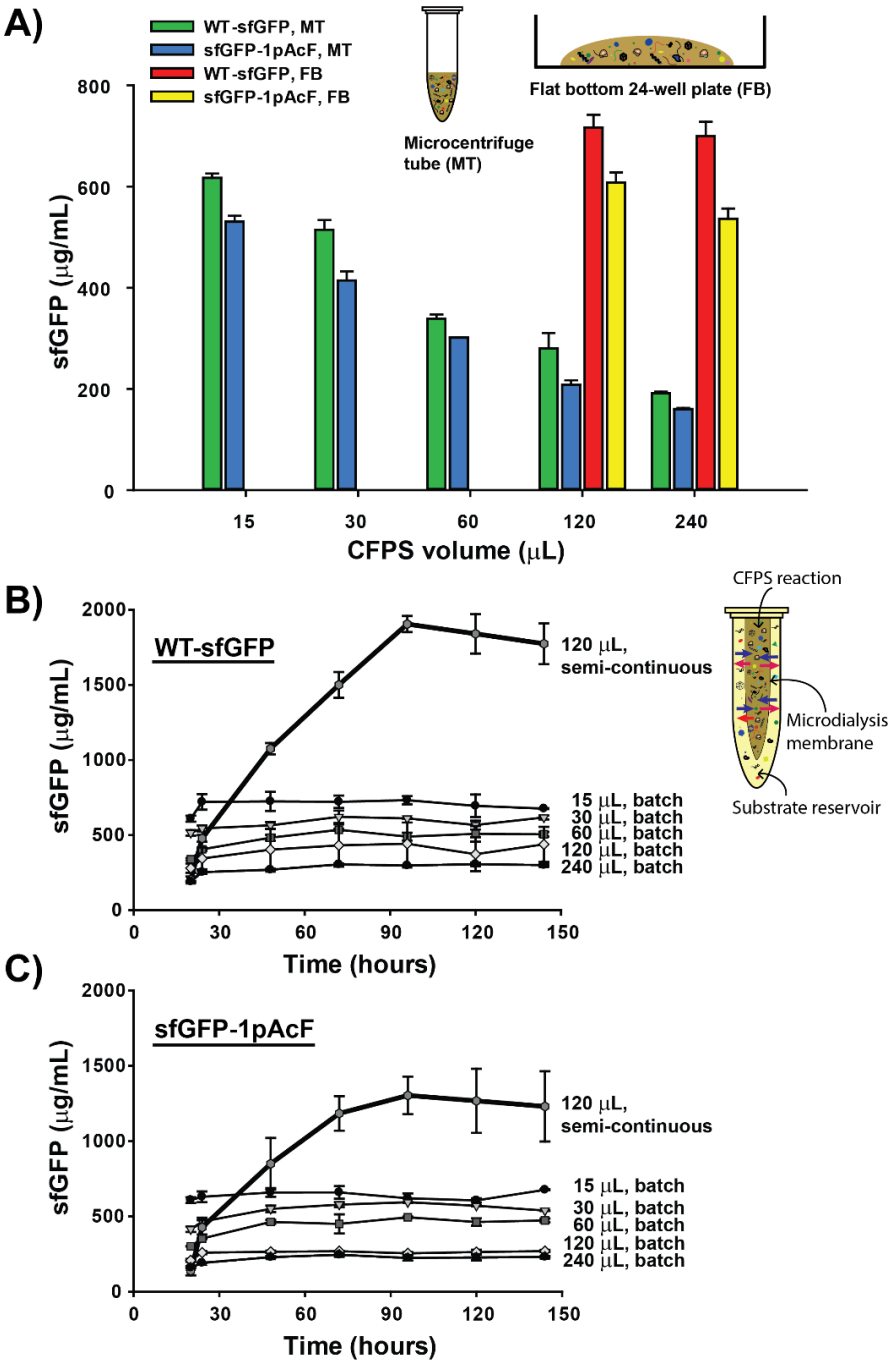


Figure 2.10 (previous page). Scaled-up and semicontinuous CFPS by using an MCJ.559 extract. A) Production of wild-type sfGFP (WT-sfGFP) and modified sfGFP with pAcF (sfGFP-1pAcF) in different reaction volumes by using a microcentrifuge tube (MT) and a flat-bottom 24-well plate (FB). Time course semicontinuous and batch CFPS for **B)** WT-sfGFP and **C)** sfGFP-1pAcF. In the semicontinuous reaction (•), CFPS reagents from the substrate reservoir passively diffuse into the CFPS reaction (inward arrows) through the microdialysis membrane, while by-products are removed from the CFPS reaction (outward arrows). Batches were of 15 (•), 30 (▾), 60 (▪), 120 (◊), and 240 μL (▴). At least three independent reactions for each sample were performed at 30 °C, and one standard deviation is shown.

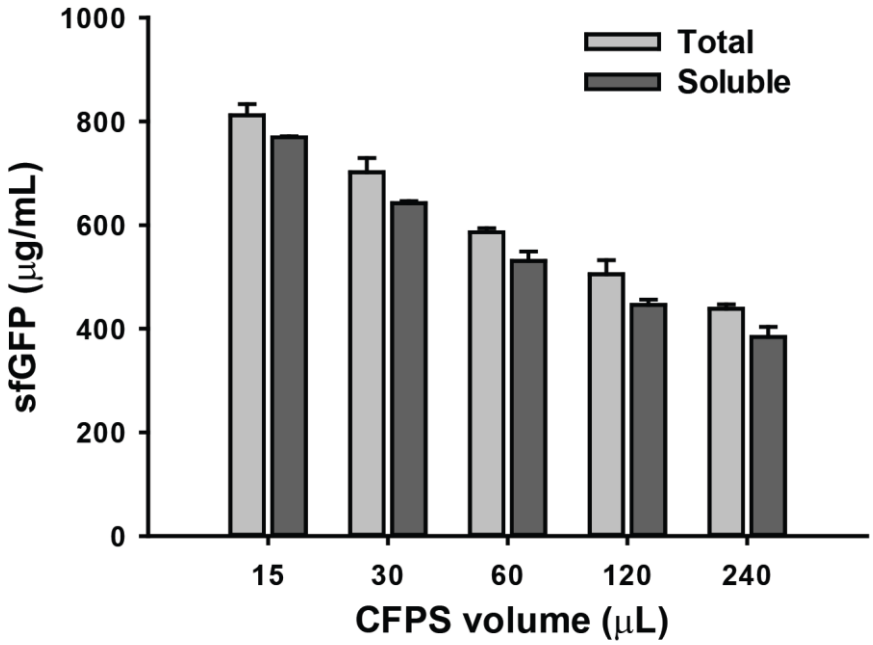


Figure 2.11. Radioactive ^{14}C-Leu incorporation of scale up CFPS reactions using MCJ.559 extract indicates decreasing protein expression yields with increasing reaction volume. Wild-type sfGFP was synthesized in different reaction volumes using a microcentrifuge tube. The total and soluble protein production was assessed by counting radioactive ^{14}C-Leu incorporation. At least three independent reactions for each sample were performed for 20 h at 30°C. One standard deviation is shown.

2.5 Discussion

In this study, we improved RF1-deficient cell extract activity approximately fourfold by functional inactivation of multiple nucleases in extract source strains using MAGE. Using our most productive extract, which stabilizes mRNA and the DNA template by disabling *csdA* and *endA*, we could significantly enhance ncAA incorporation into a target protein. Our results are important for two reasons. First, the strain, which achieved a 1300±100 μg/mL yield of modified protein containing ncAAs in semicontinuous operation, can be a useful resource for the community. Second, this is the first demonstration of using advanced genome engineering methods to develop CFPS chassis strains, effectively combining the inactivation of nucleases and the deletion of RF1. This indicates rich opportunities to understand how the overall catalytic ensemble used to carry out CFPS is altered based on genomic modifications. Looking forward, we expect this work to open the way to novel CFPS technologies for accurate and efficient production of pure proteins and biopolymers containing multiple ncAAs.

2.6 Acknowledgments

I helped conceive the idea, aided in the generation of MAGE mutants, assisted in the implementation of the stability analysis experiments, and aided in manuscript preparation.

2.7 Publication Information

This chapter was published with the following citation information:

Hong S.H., Kwon Y.C., Martin R.W., **Des Soye B.J.**, de Paz A.M., Swonger K.N., Ntai I., Kelleher N.L., and Jewett M.C. Improving cell-free protein synthesis through genome engineering of *Escherichia coli* lacking release factor 1. *ChemBioChem*. 2015 Mar 23;16(5): 844-853. doi: 10.1002/cbic.201402708.

CHAPTER 3

3 Cell-free Protein Synthesis from Recoded Bacteria Enables Multi-site Non-canonical Amino Acid Incorporation at High Yield and Purity

3.1 Abstract

Cell-free protein synthesis has emerged as a powerful approach for expanding the range of genetically encoded chemistry into proteins. Unfortunately, efforts to site-specifically incorporate multiple non-canonical amino acids into proteins using crude extract-based cell-free systems have been limited by release factor 1 competition. Here we address this limitation by establishing a bacterial cell-free protein synthesis platform based on genomically recoded *Escherichia coli* lacking release factor 1. This platform was developed by exploiting multiplex genome engineering to enhance extract performance by functionally inactivating negative effectors. Our most productive cell extracts enabled synthesis of $1{,}780 \pm 30$ mg/L superfolder green fluorescent protein. Using an optimized platform, we demonstrated the ability to introduce 40 identical *p*-acetyl-L-phenylalanine residues site specifically into an elastin-like polypeptide with high accuracy of incorporation ($\geq 98\%$) and yield (96 ± 3 mg/L). We expect this cell-free platform to facilitate fundamental understanding and enable manufacturing paradigms for proteins with new and diverse chemistries.

3.2 Introduction

Cell-free synthetic biology is emerging as a transformative approach to understand, harness, and expand the capabilities of natural biological systems[282]. The foundational principle is that complex biomolecular transformations are conducted without using intact cells. Instead, crude cell lysates (or extracts) are used, which provides a unique freedom of design to control biological systems for a wide array of applications. For example, cell-free protein synthesis (CFPS) systems have been used to decipher the genetic code[255], prototype genetic circuits and metabolic pathways[258, 259, 372-374], enable portable diagnostics[375], facilitate on-demand biomolecular manufacturing[30, 376], and produce antibody therapeutics at the commercial scale[377]. The recent surge of applications has revitalized interest in cell-free systems, especially in areas where limits imposed by the organism may impede progress. One such area is expanding the genetic code to incorporate non-canonical amino acids (ncAAs) into proteins, where the extent of engineering can be limited by the fitness of the organism[6, 338, 378, 379].

Pioneering efforts by Schultz and others have demonstrated it is possible to genetically encode more than 150 ncAAs into proteins and that this encoding can be a powerful tool[202]. For example, site-specific incorporation of ncAAs at a single position in proteins have provided new ways to study protein structure, dynamics, and posttranslational modifications[15], as well as new protein-drug conjugates[205, 380]. However, the task of multi-site ncAA incorporation into proteins with high purity and yields has been limited by inefficiencies associated with the engineered orthogonal translation (TL) machinery (*e.g.*, translation elements that specifically use a ncAA and do not interact with the cell's natural translation apparatus)[27, 220]. A key constraint is that codon re-assignment strategies typically rely on amber suppression[381], where the amber UAG stop codon is re-assigned to encode a ncAA and the orthogonal transfer RNA anticodon is

mutated to CUA. The orthogonal ncAA-tRNA$_{CUA}$ must then outcompete essential TL machinery (*e.g.*, release factor 1, RF1) for the UAG codon. Historically, this competition has led to poor protein expression yields, as premature termination by RF1 exponentially increases with the number of amber codons in the coding sequence[335]. Poor protein expression yields limit applications in both basic and applied science.

Recently, a genomically recoded *E. coli* strain was developed (*C321.*Δ*A*) in which all 321 occurrences of the UAG stop codon were reassigned to the synonymous UAA codon[21] using multiplex automated genome engineering (MAGE)[351] and conjugative assembly genome engineering (CAGE)[26]. This allowed for the genomic deletion of RF1 (*i.e.,* Δ*prfA* or Δ*A*) without affecting cellular physiology, thus freeing the UAG codon for dedicated ncAA incorporation[21]. Precursor RF1-deficient strains in which only a small set of essential genes were recoded have already shown the potential to produce proteins with improved ncAA incorporation efficiencies as compared to strains with RF1[21, 24, 232]; however, the upregulation of natural suppression mechanisms (*e.g., ssrA*) is problematic because they promote the formation of truncation products, especially for tens of incorporation events[21, 24, 382]. The fully recoded *C321.*Δ*A* strain avoids these problems and we recently showed the possibility of using *C321.*Δ*A* coupled with extensively engineered synthetases for multi-site incorporation of up to 30 ncAAs into a single biopolymer *in vivo*[27]. Based on these results, we hypothesized that the fully recoded *C321.*Δ*A* strain would serve as an ideal chassis strain for the development of crude extract cell-free systems capable of highly efficient, multi-site ncAA incorporation into biopolymers. Such a system could serve as a complement to *in vivo* manufacturing strategies, with some advantageous features[23, 280]. For example, the open reaction environment means the supply of orthogonal translation system (OTS)

components and their substrates necessary for high-level ncAA incorporation can be provided and controlled at precise ratios as a way to overcome enzyme inefficiencies. In addition, cell-free systems are not limited by viability requirements, thus avoiding constraints arising from toxic OTS components[24]. Our proposed approach based on the *C321.*Δ*A* strain might also provide cost and ease of use advantages over other cell-free systems that have tried to reduce the effects of RF1 competition by using reconstituted systems[383], antibody inhibitors[349], RF1 depletion by subtractive chromatography[328], or partially recoded *E. coli* strains with elevated natural suppression mechanisms[232].

Here we describe the development of a CFPS platform from the genomically recoded *C321.*Δ*A* to manufacture proteins with tens of identical site specifically introduced ncAAs. Specifically, we use MAGE to improve protein production capacity by inactivating negative effectors in the host strain such that they are not present in the lysate. By testing tens of strain variants, we isolate a CFPS platform capable of synthesizing of 1780 ± 30 mg/L of superfolder green fluorescent protein (sfGFP), as well as modified sfGFP containing up to five *p*-acetyl-L-phenylalanine (pAcF) residues at high purity (≥98%). Using an optimized CFPS platform, we test the ability to synthesize elastin-like polypeptides (ELPs) that contain up to 40 UAG codons. We demonstrate incorporation of 40 ncAAs per pure ELP protein with high yields (~100 mg/L) and high fidelity (≥98%) of site-specific ncAA incorporation.

3.3 Materials and Methods

3.3.1 Strains and Plasmids

The bacterial strains and plasmids used in this study are listed in **Table 3.1**. Carbenicillin (50 µg/mL) was used for culturing *C321.*Δ*A* derivative strains, kanamycin (50 µg/mL) was used for maintaining pY71-based plasmids, and chloramphenicol (34 µg/mL) was used to maintain the pEVOL-pAcF plasmid.

3.3.2 Strain Construction and Verification

The strains in this study **(Table 3.1)** were generated from *C321.*Δ*A*[21] by disrupting genes of interest **(Table 3.2)** with mutagenic oligonucleotides via MAGE **(Table 3.3)**. Cultures were grown in LB-Lennox media (10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl) at 32 ˚C at 250 rpm throughout MAGE cycling steps[351]. MAGE oligonucleotides were designed to introduce an internal stop codon and frameshift ~1/4 of the way into the target gene sequence thereby causing early translational termination as previously reported[232]. Combinatorial disruptions of *endA*, *mazF*, *rna*, *rnb*, *rne*, *gor*, *lon*, *ompT*, *gdhA*, *gshA*, *sdaA*, *sdaB*, *speA*, *tnaA*, and *glpK* were generated to investigate the effects of their inactivation on CFPS. Multiplex allele-specific colony (MASC) PCR was performed to verify gene disruptions using wild-type forward (-wt-F) or mutant forward (-mut-F) primers and reverse primers (-R) **(Table 3.3)**[351]. Wild-type and mutant forward primers were identical except at the 3'-ends of the oligonucleotides, and the reverse primers were used for detection of both wild-type and mutant alleles. The mutant allele could be amplified using the mutant forward and reverse primer set (-mut-F and -R) but not amplified by the wild-type forward and reverse primer set (-wt-F and -R). MASC PCR was performed in 10 µL reactions using a multiplex PCR kit (Qiagen, Valencia, CA). Mutant alleles were screened by running PCR products on a 1.5% agarose gel and confirmed by DNA sequencing using sequencing primers **(Table 3.3)**.

3.3.3 Growth Rate Assessment

Overnight cultures of engineered strains grown in Luria-Bertani (LB) at 250 rpm at 34 ˚C were diluted 1000-fold in 2xYTPG media (31 g/L 2xYT, 7 g/L K_2HPO_4, 3 g/L KH_2PO_4, and 18 g/L glucose; adjusted pH to 7.2 with KOH)[384]. 100 µL of the diluted cultures were added to 96-well polystyrene plates (costar 3370; Corning, Corning, NY). OD was measured at 15 min intervals for 15 h at 34 ˚C in fast shaking mode on a Synergy 2 plate reader (Biotek, Winooski, VT). Growth data of each strain was obtained from three independent cultures, each split into three replicate wells (9 total samples per strain). Doubling time was calculated during the early exponential growth phase (OD_{600} of 0.02 to 0.2).

3.3.4 Cell Extract Preparation

For prototyping engineered strains, cells were grown in 1 L of 2xYTPG media (pH 7.2) in a 2.5 L Tunair® shake flask and incubated at 34 ˚C and 220 r.p.m. to OD_{600} of 3.0. Cells were pelleted by centrifuging for 15 min at 5000 x *g* at 4 ˚C, washed three times with cold S30 buffer (10 mM tris-acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, 2 mM dithiothreitol)[354], and stored at -80°C. To make cell extract, the thawed cells were suspended in 0.8 mL of S30 buffer per gram of wet cell mass and processed as reported by Kwon and Jewett[291]. 1.4 mL of cell slurry was transferred into 1.5 mL microtubes. The cells were lysed using a Q125 Sonicator (Qsonica, Newtown, CT) with 3.175 mm diameter probe at a 20 kHz frequency and 50 % amplitude. To minimize heat damage during sonication, samples were placed in an ice-water bath. For each 1.4 mL sample, the input energy was ~944 Joules and was monitored during sonication. Extract was then centrifuged at 12,000 g at 4 ˚C for 10 min. For strain

derivatives of MG1655, a run-off reaction (37 ˚C at 250 rpm for 1 h) and second centrifugation (10,000 g at 4 ˚C for 10 min) were performed[291]. The supernatant was flash-frozen using liquid nitrogen and stored at -80 ˚C until use. The total protein concentration of the extracts were 40 to 50 mg/mL as measured by Quick-Start™ Bradford protein assay kits (Bio-Rad, Hercules, CA).

3.3.5 Purification of His-tagged Orthogonal tRNA Synthetase

BL21 (DE3) harboring pY71 plasmid encoding either pAcFRS, pAzFRS, or pPaFRS were grown in 1 L of 2xYT to an OD_{600} of 1.0 at 220 rpm and 37°C. Orthogonal synthetase production was induced by adding 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG, Sigma-Aldrich, St. Louis, MO) and cells were allowed to grow for an additional 3 h. Cells were harvested at 5,000 g for 15 min at 4°C, washed with S30 buffer, and stored at -80°C. Frozen cell pellets were thawed in loading buffer (1 mL of 300 mM NaCl, 10 mM imidazole, 50 mM NaH_2PO_4, pH 8.0 solution per gram of wet cells), lysed using sonication as described above and centrifuged at 16,000 g at 4 ˚C for 10 min. The supernatant was diluted 1:1 with loading buffer and incubated at 4 ˚C for 1 h with Ni-NTA beads prewashed with dilution buffer (300 mM NaCl, 50 mM NaH_2PO_4, pH 8.0). The orthogonal synthetase was purified using elution buffer (300 mM NaCl, 250 mM imidazole, 50 mM NaH_2PO_4, pH 8.0) and subsequently dialyzed against S30 buffer and 25% glycerol in a Slide-A-Lyzer™ G2 Dialysis Cassette (Life Technology, Grand Island, NY). Dialyzed synthetase was concentrated using Amicon Ultracel YM-30 centrifugal filter and stored at -80°C. Purified synthetase was quantified by the Quick-Start™ Bradford protein assay kit (Bio-Rad, Hercules, CA).

3.3.6 CFPS Reactions

The PANOx-SP system was utilized for CFPS reactions [287]. Briefly, a 15 µL CFPS reaction in a 1.5 mL microtube was prepared by mixing the following components: 1.2 mM ATP; 0.85 mM each of GTP, UTP, and CTP; 34 µg/mL folinic acid; 170 µg/mL of *E. coli* tRNA mixture; 13.3 µg/mL plasmid; 16 µg/mL T7 RNA polymerase; 2 mM for each of the 20 standard amino acids; 0.33 mM nicotinamide adenine dinucleotide (NAD); 0.27 mM coenzyme-A (CoA); 1.5 mM spermidine; 1 mM putrescine; 4 mM sodium oxalate; 130 mM potassium glutamate; 10 mM ammonium glutamate; 12 mM magnesium glutamate; 57 mM HEPES, pH 7.2; 33 mM phosphoenolpyruvate (PEP), and 27% v/v of cell extract. For ncAA incorporation, 2 mM pAcF, 0.5 mg/mL pAcFRS, and 10 µg/mL of o-tz-tRNA linear DNA were supplemented to cell-free reactions. For multi-site ncAA incorporation, OTS^{opt} levels were increased to 5 mM pAcF, 1 mg/mL pAcFRS, and 30 µg/mL o-tz-tRNA. o-tRNA linear DNA was amplified from pY71-T7-tz-o-tRNA plasmid as described previously and transcribed during the cell-free reaction[24]. Furthermore, the o-tRNA was expressed in the source strain via a plasmid prior to extract preparation. *E. coli* total tRNA mixture (from strain MRE600) and phosphoenolpyruvate was purchased from Roche Applied Science (Indianapolis, IN). ATP, GTP, CTP, UTP, 20 amino acids and other materials were purchased from Sigma (St. Louis, MO) without further purification. T7 RNA Polymerase was purified in house using ion exchange chromatography as described previously[287]. When testing the effect of RNAse inhibitor, 1 µL (4U) of inhibitor (Qiagen, Valencia, CA) was added into each 15 µL reaction as per the manufacturer's suggestion. Each CFPS reaction was incubated for 20 h at 30 ˚C unless noted otherwise.

3.3.7 Fed-batch CFPS Reactions

For fed-batch reactions, 15 µL CFPS batch reactions were prepared as described above. At the specified time, the reactions were removed from the incubator, supplied with 0.5 µL of feeding solution containing the appropriate concentration of the desired amino acid(s), mixed with a pipette, and returned to the incubator. All reactions were incubated at 30 ˚C for 20 h total and assayed.

3.3.8 Scaled-up CFPS

Cell-free reaction volumes were scaled to 255 µL in flat-bottom 24-well polystyrene plate (model 353226; BD Biosciences, San Jose, CA). Remaining wells around the perimeter of the plate were filled with water for internal humidification, which resulted in reduced sample evaporation. By filling the outer chambers surrounding the wells with water, which humidified the air, negligible sample evaporation was achieved. Reactions were performed at 30 ˚C for 20 h while shaking at 300 RPM in a ThermoMixer (Eppendorf, Mississauga, Ontario). Sufficient sfGFP and ELP were purified for mass spectrometry analysis.

3.3.9 Quantification of Active sfGFP

Active full-length sfGFP protein yields were quantified by measuring fluorescence using a Synergy 2 plate reader (BioTek, Winooski, VT) with excitation at 485 nm, emission at 528 nm, and cut-off at 510 nm in 96-well half area black plates (Costar 3694; Corning, Corning, NY). sfGFP fluorescence units were converted to concentration using a standard curve established with ^{14}C-Leu quantified sfGFP as described previously[24].

3.3.10 Quantification of Total and Soluble Protein

Radioactive ^{14}C-Leucine was added into 15 µL CFPS reactions at a final concentration of 10 µM. Reactions were taken at the indicated time and 5 µL of sample was removed for total protein quantitation. The remaining sample was centrifuged at 16,000 g at 4 ˚C for 10 min and the top 5 µL was used to measure the soluble protein. Total and soluble protein yields were quantified by determining radioactive ^{14}C-Leu incorporation into trichloroacetic acid (TCA) -precipitated protein[354]. Radioactivity of TCA-precipitated samples was measured using liquid scintillation counting (MicroBeta2, PerkinElmer, Waltham, MA).

3.3.11 Autoradiogram Analysis

For autoradiogram analysis, 2 µL of each reaction was loaded on 10% NuPAGE SDS-PAGE gel after denaturing the sample. The gel was soaked in Gel Drying solution (Bio-Rad, Hercules, CA) for 30 min, fixed with cellophane films, dried overnight in GelAir Dryer (Bio-Rad, Hercules, CA), and exposed for 3 days on Storage Phosphor Screen (GE Healthcare Biosciences, Pittsburgh, PA). Autoradiograms were scanned using Storm Imager (GE Healthcare Biosciences, Pittsburgh, PA).

3.3.12 Whole Genome Analysis

Because the gene encoding MutS is inactivated in *C321.*Δ*A*, we chose to fully sequence the genomes of six key strains produced during our screening efforts (*C321.*Δ*A*, *C321.*Δ*A.542*, *C321.*Δ*A.705*, *C321.*Δ*A.709*, *C321.*Δ*A.740*, *C321.*Δ*A.759*). One milliliter of cell culture in LB broth was processed with a Qiagen DNeasy Blood and Tissue kit (cat: 69504) to extract genomic DNA. gDNA quality was assessed on a spectrophotometer (assay for A260/280 ratio between 1.8 and 2.0) and by gel electrophoresis (assay for a tight smear at 50 kB). 2.5 µg of gDNA, eluted in

50-µl TE pH 8.0, was sent to the Yale Center Genome Analysis for library prep and analysis as described previously[385]. Genome modification of targeted effectors were confirmed and off target point mutations were limited to regions that were not expected to effect CFPS activity. Genomes have been deposited to NCBI SRA collection, accession number PRJNA361365 (Individual accession numbers: SRX2511757-SRX2511762).

3.3.13 mRNA Stability Assay

mRNA stability was assessed as described previously[232]. Briefly, the sfGFP gene was PCR amplified from the pY71 vector and purified by PCR clean-up kit (Promega, Madison, WI). This template was used for T7-driven *in vitro* transcription reactions. In order to track mRNA stability in our extracts, sfGFP was synthesized using purified mRNA (1,800 ng) in the CFPS reaction. For direct measurement of mRNA degradation, purified mRNA samples from the cell-free reaction were visualized on a 2% formaldehyde agarose gel stained with GelRed (Biotium, Hayward, CA). mRNA band intensities were quantified using ImageJ software (National Institutes of Health, Bethesda, MD) and normalized to rRNA bands.

3.3.14 DNA Stability Assay

DNA stability was assessed as described previously[232]. Briefly, a preincubation mixture containing 4 µL of cell extract, 12.96 ng/µL of pY71-mRFP1-Spinach plasmid **(Table 3.1)**, and 67 µM of 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI; Lucerna, New York, NY) was incubated for 0, 60, and 180 min at 30 ˚C. Then, CFPS reaction components were added immediately after the preincubation step, and fluorescence of the Spinach aptamer binding to DFHBI was monitored for 180 min using CFX96 Real-Time RT-PCR module installed on C1000

Touch Thermal Cycler (Bio-Rad, Hercules, CA). The excitation and emission wavelengths of the fluorophore were 450-490 nm and 515-530 nm, respectively.

3.3.15 Nucleotide and Amino Acid Quantitation Using HPLC

Amino acid and nucleotide concentrations were measured via HPLC. Cell-free reactions were clarified by precipitation with an equal volume 5% w/v TCA. Samples were centrifuged at 12,000 g for 15 minutes at 4 ˚C and the supernatant stored at -80 ˚C until analyzed using an Agilent 1290 series HPLC system (Agilent, Santa Clara, CA).

For amino acid analysis, a Poroshell HPH-C18 column (4.6 x 100 mm, 2.7 µm particle size; Agilent) was used with an automatic precolumn derivitization method using *o*-pthalaldehyde (OPA) and fluorenylmethyl chloroformate (FMOC)[386]. Run times were 16 minutes at a flow rate of 1.5 mL min^{-1} and a column temperature of 40 ˚C. Mobile phase A was comprised of 10 mM sodium phosphate, 10 mM sodium tetraborate, and 5 mM sodium azide at pH 8.2. Mobile phase B was 45% methanol, 45% acetonitrile, and 10% water by volume. The buffer gradient for B was: 0 min, 2%; 0.5 min, 2%; 13.4 min, 57%; 13.6 min, 100%; 15.7 min, 100%; 15.9 min, 2%; 16 min, end. Amino acids were detected at 262 nm except for proline, which was detected at 338 nm. Concentrations were determined by comparison to a standard calibration using NIST standard reference material 2389a. Amino acids not contained in the NIST standard were obtained from Sigma-Aldrich.

Nucleotides were analyzed using a BioBasic AX column (4.6 x 150 mm, 5 µm particle size; Thermo Scientific, West Palm Beach, FL). Separation was carried out at a flow rate of 1 mL min^{-1} and column temperature of 22 ˚C. Mobile phase A and B respectively were 5 and 750 mM potassium phosphate monobasic adjusted to pH 3.30 with phosphoric acid. The buffer gradient for

B was: 0 min, 0%; 6 min, 20%; 11 min, 40%; 20 min, 100%; 25 min, 100%; 25.5 min, 0%; 30 min, end. Nucleotides were detected at 254 nm. Concentrations were determined by comparison to a standard calibration.

3.3.16 ELP Plasmid Construction

ELP genes were codon optimized for *E. coli* expression. ELPs contained three pentapeptides in a monomer unit for 20, 30, and 40 mers and an amber site in a monomer unit for ncAA incorporation **(Figure 3.17.A)**[27]. First, we constructed pY71-KA-sfGFP vector by adding KpnI and ApaI restriction sites at the 5'- and 3'-end of sfGFP gene, respectively **(Table 3.3)**, and then ELP genes were cloned into this vector using the same restriction sites resulting in ELPs with C-terminal sfGFP fusion. Then, we added a Twin-Streptag using ApaI and BlpI restriction sites in place of the sfGFP gene **(Table 3.3)**. All ELP plasmids are listed in **Table 3.1**.

3.3.17 Tandem UAG Plasmid Construction

Eight and nine tandem UAG sequences were added into the sfGFP at position 132 by inverse PCR followed by ligation. The resulting plasmids are pY71-sfGFP-8tdUAG and -9tdUAG **(Table 3.1)**.

3.3.18 Full-length sfGFP and ELP Purification and Mass Spectrometry

To confirm pAcF incorporation at corresponding amber sites, semi-quantitative mass spectrometry (MS) analysis was performed on purified sfGFP and ELP reporter protein with pAcF putatively incorporated. First, full-length sfGFP protein was purified from CFPS reactions using

C-terminal strep-tags and 0.2 mL gravity-flow Strep-Tactin Sepharose mini-columns (IBA GmbH, Gottingen, Germany) and concentrated using Microcon centrifugal filter columns YM-10 (Millipore, Billerica, MA). ELPs were purified using a modified inverse transition cycling (ITC) method in which cell-free reactions were centrifuged at 14,000 g for 3 minutes at room temperature to capture aggregated ELPs in the pellet. The isolated pellet was then resuspended in cold 1xPBS solution to resolubilize the ELP. The resulting mixture is then centrifuged at 14,000 g for 5 minutes at 4°C. To precipitate the ELPs, sodium citrate was added to the mixture at a final concentration of 0.5 M and the resulting mixture was centrifuged at 14,000 g for 3 minutes at room temperature to capture aggregated ELPs in the pellet. The steps were repeated as necessary to purify ELP from contaminants.

Purified reporter protein was then injected onto a trap column (150 µm ID × 3 cm) coupled with a nanobore analytical column (75 µm ID × 15 cm). The trap and analytical column were packed with polymeric reverse phase (PLRP-S, Phenomenex) media (5 µm, 1,000 Å pore size). Samples were separated using a linear gradient of solvent A (95% water, 5% acetonitrile, 0.2% formic acid) and solvent B (5% water, 95% acetonitrile, 0.2% formic acid). Samples were loaded for 10 minutes onto the trap and were subsequently separated using a linear gradient from 5% to 55% of solvent B over 27 minutes followed by washing steps. MS data were obtained on a 12T Velos FT Ultra (Thermo) mass spectrometer fitted with a custom nanospray ionization source. Intact mass spectrometry data were obtained at a resolving power of 170,000 (m/z 400) and in the Velos dual ion trap. FTMS data were deconvoluted using Xtract (ThermoFisher). In the MS figure **(Figure 3.10.C and Figure 3.17.D-F)**, smaller peaks to the right of the colored peaks ($\Delta m = +16$ Da) are due to oxidation of the protein – a common electrochemical reaction occurring during

electrospray ionization. The presence of the initial methionine amino acid residue on a protein will also increase the mass (Δm = +131 Da) and are detected to the right of the major (colored) peak. To remove non-covalent salt and water adducts from intact proteins (in this case sfGFP), a small level of in-source collision energy (15eV) was applied. As a result, water loss events from the intact sfGFP (Δm = -18 Da) are detected at minor levels to the left of the major (colored) peak.

Table 3.1. Strains and plasmids used in this study. Km^R, Ap^R, Zeo^R and Cm^R are kanamycin, ampicillin, zeocin, and chloramphenicol resistance, respectively. 'Δ' indicates deleted gene, and superscript '-' indicates disabled/functionally inactivated gene via MAGE.

Strains and plasmids	Genotype/relevant characteristics	Source
Strains		
EcNR2	*MG1655 with λ-prophage::bioA/bioB* and *cmR::mutS*	[25]
BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs
BL21 Star™ (DE3)	F^- *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm rne131* (DE3)	Life Technologies
MCJ.559	*C13.ΔA. endA*$^-$ *csdA*$^-$ (C13: 13 UAG recoded to UAA)	[232]
C321.ΔA	Δ*prfA* Ω Cb^R, Zeo^R, EcNR2 derivative with all 321 UAG stop codons reassigned to UAA	[21]
C321.ΔA.540	*C321.ΔA. rnb*$^-$	This study
C321.ΔA.541	*C321.ΔA. mazF*$^-$	This study
C321.ΔA.542	*C321.ΔA. endA*$^-$	This study
C321.ΔA.598	*C321.ΔA. rna*$^-$	This study
C321.ΔA.618	*C321.ΔA. ompT*$^-$	This study
C321.ΔA.620	*C321.ΔA. glpK*$^-$	This study
C321.ΔA.626	*C321.ΔA. gshA*$^-$	This study
C321.ΔA.628	*C321.ΔA. tnaA*$^-$	This study
C321.ΔA.644	*C321.ΔA. gdhA*$^-$	This study
C321.ΔA.666	*C321.ΔA. gor*$^-$	This study
C321.ΔA.667	*C321.ΔA. lon*$^-$	This study
C321.ΔA.668	*C321.ΔA. rne*$^-$	This study
C321.ΔA.669	*C321.ΔA. sdaA*$^-$	This study
C321.ΔA.672	*C321.ΔA. sdaB*$^-$	This study
C321.ΔA.674	*C321.ΔA. speA*$^-$	This study

C321.ΔA.544	*C321.ΔA. endA⁻ mazF⁻*	This study
C321.ΔA.678	*C321.ΔA. endA⁻ glpK⁻*	This study
C321.ΔA.679	*C321.ΔA. endA⁻ tnaA⁻*	This study
C321.ΔA.709	*C321.ΔA. endA⁻ gor⁻*	This study
C321.ΔA.711	*C321.ΔA. endA⁻ rne⁻*	This study
C321.ΔA.708	*C321.ΔA. endA⁻ lon⁻*	This study
C321.ΔA.703	*C321.ΔA. endA⁻ gor⁻ glpK⁻*	This study
C321.ΔA.705	*C321.ΔA. endA⁻ gor⁻ rne⁻*	This study
C321.ΔA.706	*C321.ΔA. endA⁻ gor⁻ tnaA⁻*	This study
C321.ΔA.740	*C321.ΔA. endA⁻ gor⁻ mazF⁻*	This study
C321.ΔA.738	*C321.ΔA. endA⁻ gor⁻ lon⁻*	This study
C321.ΔA.759	*C321.ΔA. endA⁻ gor⁻ rne⁻ mazF⁻*	This study
C321.ΔA.619	*C321.ΔA. glpK⁻ ompT⁻*	This study
C321.ΔA.617	*C321.ΔA. glpK⁻ rne⁻*	This study
C321.ΔA.621	*C321.ΔA. ompT⁻ rne⁻*	This study
C321.ΔA.680	*C321.ΔA. endA⁻ glpK⁻ tnaA⁻*	This study
C321.ΔA.664	*C321.ΔA. endA⁻ lon⁻ rna⁻*	This study
C321.ΔA.756	*C321.ΔA. endA⁻ glpK⁻ gor⁻ mazF⁻*	This study
C321.ΔA.755	*C321.ΔA. endA⁻ gor⁻ lon⁻ mazF⁻*	This study
C321.ΔA.758	*C321.ΔA. endA⁻ gor⁻ mazF⁻ tnaA⁻*	This study
C321.ΔA.879	*C321.ΔA. endA⁻ gor⁻ rna⁻ rne⁻*	This study
C321.ΔA.878	*C321.ΔA. endA⁻ gor⁻ rnb⁻ rne⁻*	This study
C321.ΔA.564	*C321.ΔA. endA⁻ gor⁻ mazF⁻ rna⁻ rnb⁻*	This study
C321.ΔA.563	*C321.ΔA. endA⁻ mazF⁻ rna⁻ rnb⁻ rne⁻*	This study
Plasmids		
pY71-sfGFP	Km^R, P_{T7}::super folder green fluorescent protein (sfGFP), C-terminal strep-tag	[19]
pY71-sfGFP-E132	pY71-sfGFP with amber codon at E132	[19]
pY71-sfGFP-T216	pY71-sfGFP with amber codon at T216	[19]
pY71-sfGFP-2UAG	pY71-sfGFP with amber codon at N212 and T216	[24]
pY71-sfGFP-5UAG	pY71-sfGFP with amber codon at D36, K101, E132, D190, and E213	[24]
pY71-sfGFP-8tdUAG	pY71-sfGFP with 8 consecutive amber codons at E132	This study
pY71-sfGFP-9tdUAG	pY71-sfGFP with 9 consecutive amber codons at E132	This study
pY71-CAT	Km^R, P_{T7}::chloramphenicol acetyl transferase (CAT)	[232]
pY71-CAT-D112UAG	pY71-CAT with amber codon at D122	[232]
pK7-mGM-CSF	Km^R, P_{T7}::modified murine granulocyte-macrophage colony-stimulating factor (mGM-CSF)	[387]

pK7-DHFR	Km^R, P_{T7}::dihydrofolate reductase (DHFR)	New England Biolabs
pY71-pAcFRS	P_{T7}::pAcFRS, C-terminal 6x histidine tag	[24]
pY71-pPaFRS	P_{T7}::pPaFRS, C-terminal 6x histidine tag	[19]
pDAK-pAzFRS	P_{T7}::pAzFRS, C-terminal 6x histidine tag	This study
pY71-mRFP-Spinach	P_{T7}::mRFP-Spinach aptamer	[223]
pDULE-o-tRNA	Tet^R, P_{lpp}::o-tRNA	[19]
pEVOL-pAcF	Cm^R, P_{glnS}::*pAcFRS*, P_{araBAD}::*pAcFRS*, P_{proK}::$o\text{-}tRNA^{opt}$	[14]
pY71-T7-tz-o-$tRNA^{opt}$	P_{T7}:: hammer-head ribozyme (tz), $o\text{-}tRNA^{opt}$ (o-tz-tRNA)	[24]
pY71-KA-sfGFP	N-ter KpnI and C-ter ApaI restriction site addition on sfGFP	This study
pY71-ELP-20	ELP-20mer	This study
pY71-ELP-30	ELP-30mer	This study
pY71-ELP-40	ELP-40mer	This study
pY71-ELP-20UAG	ELP-20mer with 20 amber sites	This study
pY71-ELP-30UAG	ELP-30mer with 30 amber sites	This study
pY71-ELP-40UAG	ELP-40mer with 40 amber sites	This study
pY71-ELP-20-TS	ELP-20mer with Twin-Streptag	This study
pY71-ELP-30-TS	ELP-30mer with Twin-Streptag	This study
pY71-ELP-40-TS	ELP-40mer with Twin-Streptag	This study
pY71-ELP-20UAG-TS	ELP-20mer with 20 amber sites, Twin-Streptag	This study
pY71-ELP-30UAG-TS	ELP-30mer with 30 amber sites, Twin-Streptag	This study
pY71-ELP-40UAG-TS	ELP-30mer with 40 amber sites, Twin-Streptag	This study

Table 3.2. Summary of putative negative CFPS effectors functionally inactivated in genomically engineered variants of *C321.*Δ*A*. Individual mutation targets were chosen for their potential to stabilize essential substrates, including DNA, RNA, protein, amino acids, and energy supply.

Function	Gene	Reason for disruption	Reference
DNA Stability	*endA*	Stabilize DNA	[232, 289, 364]
RNA Stability	*mazF*	Stabilize mRNA	[232, 360]
	rna	Stabilize mRNA	[358]
	rnb	Stabilize mRNA	[232, 357, 359]
	rne	Stabilize mRNA	[361]
Protein Stability	*gor*	Decrease reducing power of cell extracts to enhance disulfide bond formation	[310, 357, 388]
	lon	ATP-dependent protease	[352]
	ompT	Outer membrane protease	[352, 389]
Amino Acid Stability	*gdhA*	Stabilize glutamate	[390]
	gshA	Stabilize cysteine	[357, 363]
	sdaA	Stabilize serine	[288]
	sdaB	Stabilize serine	[288]
	speA	Stabilize arginine	[288]
	tnaA	Stabilize tryptophan	[288]
Energy Supply	*glpK*	Conserve ATP	[391, 392]

Table 3.3. Primers used for MAGE, MASC PCR, DNA sequencing, and ELP construction and gblock DNA fragments used for ELP construction. Underlined bold text indicates location of mismatch and insertion of premature stop codon. The first four bases of the 5'-MAGE oligonucleotides were phosphorothioated (*). "/Phos/" indicates 5'-phosphorylated.

Primer Name	DNA Sequence (listed 5' to 3')
MAGE	
rna	G*C*T*G*ATTTTCTGACCGTACATGGTCTGTGGCCAGGA**ATGTAAATG CTGATGTA**TTGCCTAAATCGGTTGCTGCCCGTGGTGTTGATGAAC
rne	C*T*G*T*TGAGCCGCTTCTTCGGCGCACTGAAAGCGCTGTTCAGC**TAAC TGA**GAAGAAACCAAACCGACCGAGCAACCAGCACCGAAAGCAGA
rnb	A*T*T*T*TGTCACCATCGACAGTGCCAGCACAGAAGATATGGAT**TAAC TGA**CTTTTCGCTAAGGCGTTGCCGGATGACAAACTTCAGCTGAT
mazF	T*T*G*A*TTGCGTTGTACAAGGAACACACAGACACATACCTGTTTTGTT GTT**TCAGTTA**GAAAGGACTCAGGACAACAGCTGGACGATGTCC
endA	C*G*G*T*AAAAGTCCACGCTGACGCGCCCGGTACGTTTTATTGC**TAACT GA**AAAATTAACTGGCAGGGCAAAAAAGGCGTTGTTGATCTGCA
ompT	T*G*G*A*CAACTCTCGGCAGCCGAGGTGGCAATATGGTCG**CG**CAGGAC TGGATGGATTCCAGTAACCCCGGAACCTGGACGGATGAAAGTAGA
lon	A*A*C*T*CTGCTTCCGCTTTCTCTTTTGCCTCTTTCGGCATCTT**TCAGTT A**GTCGATTTTGCGCTTCAGGGCTTCGTTTTCGTCCGGCGCGT
gor	G*T*T*T*TGATACCACTATCAATAAATTCAACTGGGAAACGTTG**TAACT GA**AGCCGTACCGCCTATATCGACCGTATTCATACTTCCTATGA
gdhA	T*G*C*G*CTTCCATCCGTCAGTTAACCTTTCCATTCTCAAATTC**TAACTG A**TTTGAACAAACCTTCAAAAATGCCCTGACTACTCTGCCGA
sdaA	T*C*T*A*CAGCAAAACTTATTATTCCATCGGCGGCGGTTTTATC**TAACT GA**GAAGAACACTTTGGTCAGGATGCTGCCAACGAAGTAAGCGT
sdaB	A*C*A*G*CCAGACTTACTACTCTATTGGCGGTGGCTTTATCGTT**TAACT GA**GAGCATTTTGGCCAGCAGGATAGCGCACCGGTTGAAGTTCC
speA	A*T*C*T*TCTCAATGACCAGATAGACCTTGTGCCCCATCTTCTC**TCAGT TA**TAATGCCAGGCGGATATATTCGCGGTCTTTATAACCGTTGC
gshA	G*A*T*G*CACCAAACAGATAAGGAATGACCCAACCGAAACGATA**TCA GTTA**GCGGATAACGCGGAAATAGCCCGCAGAAATTTTCTCTTTGG
tnaA	A*C*C*T*TGAGGGATTAGAACGCGGTATTGAAGAAGTTGGTCCG**TAAC TGA**GTGCCGTATATCGTTGCAACCATCACCAGTAACTCTGCAGG
glpK	G*C*A*C*CAAAGTGAAGTGGATCCTCGACCATGTGGAAGGCTCT**TAAC TGA**CGTGCACGTCGTGGTGAATTGCTGTTTGGTACGGTTGATAC
MASC PCR	
rna-wt-F	GTACATGGTCTGTGGCCAGGATTGC
rna-mut-F	GGCCAGGA**ATGTAAATGCTGATGTA**
rna-R	TGGCATGACTTCACTTAGTTTAGC
rne-wt-F	CACTGAAAGCGCTGTTCAGCGGTGGT
rne-mut-F	CACTGAAAGCGCTGTTCAGC**TAACTGA**
rne-R	GTGCGACTACCGCTTCTTCGGCTAC
rnb-wt-F	CCAGCACAGAAGATATGGATGACGCC
rnb-mut-F	CCAGCACAGAAGATATGGAT**TAACTGA**

rnb-R	TCACTTTCAGGCTGCCAGTCACCGG
mazF-wt-F	CTGTTGTCCTGAGTCCTTTCATGTAC
mazF-mut-F	CTGTTGTCCTGAGTCCTTTC**TAACTGA**
mazF-R	GGCTTTAATGAGTTGTAATTCCTCTG
endA-wt-F	CCCGGTACGTTTTATTGCGGATGT
endA-mut-F	CCCGGTACGTTTTATTGC**TAACTGA**
endA-R	GCTGGCGCTGGTAATTTCGGCGTCA
ompT-wt-F	CAGCCGAGGTGGCAATATGGTCGAT
ompT-mut-F	CAGCCGAGGTGGCAATATGGTCG**CG**
ompT-R	GAGTTCAAAATCTTCATAACGATAAC
lon-wt-F	CTGAAGCGCAAAATCGACGCGGCG
lon-mut-F	CTGAAGCGCAAAATCGAC**TAACTGA**
lon-R	AGCGGGTTTTTCACGCCCACTTTCGC
gor-wt-F	TTCAACTGGGAAACGTTGATCGCC
gor-mut-F	TTCAACTGGGAAACGTTG**TAACTGA**
gor-R	TGCAACATTTCGTCCATACCAAAGC
gdhA-wt-F	CCTTTCCATTCTCAAATTCCTCGGC
gdhA-mut-F	CCTTTCCATTCTCAAATTC**TAACTGA**
gdhA-R	CTGCCGCCAAATGAAAGGCCCTTAC
sdaA-wt-F	ATCGGCGGCGGTTTTATCGTCGAT
sdaA-mut-F	ATCGGCGGCGGTTTTATC**TAACTGA**
sdaA-R	CAAGACCCGCAGCAGCCATTGAACAG
sdaB-wt-F	GGCGGTGGCTTTATCGTTGATGAA
sdaB-mut-F	GGCGGTGGCTTTATCGTT**TAACTGA**
sdaB-R	TACCTGTCCGGCGACCGGGTCACAC
speA-wt-F	GAATATATCCGCCTGGCATTAATTGGC
speA-mut-F	GAATATATCCGCCTGGCATTA**TAACTGA**
speA-R	CGGTGATTACCGTCGGATGCGGCAG
gshA-wt-F	TATTTCCGCGTTATCCGCAATTAC
gshA-mut-F	TATTTCCGCGTTATCCGC**TAACTGA**
gshA-R	AAATCCTCTTCGCGCAGAATTTCCAGC
tnaA-wt-F	GGTATTGAAGAAGTTGGTCCGAATAAC
tnaA-mut-F	GGTATTGAAGAAGTTGGTCCG**TAACTGA**
tnaA-R	CTACCGCCAGACGCTCCATCGCGCC
glpK-wt-F	GACCATGTGGAAGGCTCTCGCGAG
glpK-mut-F	GACCATGTGGAAGGCTCT**TAACTGA**
glpK-R	CAAACAGCGCGGCCTGCTGGTCACC
DNA Sequencing	
rna-seq-F	GTTTCTCTGCTTCCCTTCTCTTCT

rne-seq-F	CAGATGGAAACCCCGCACTACCACG
rnb-seq-F	CTGAAAGGCGATCGTTCTTTCTATG
mazF-seq-F	GTAAAGAGCCCGTATTTACGCTTGC
endA-seq-F	ATGTACCGTTATTTGTCTATTGCTGC
ompT-seq-F	CTGACAACCCCTATTGCGATCAGCTC
lon-seq-F	GTGCTGGTGCGTACTGCAATCAGCC
gor-seq-F	TAAACACTATGATTACATCGCCATC
gdhA-seq-F	GCAAGCCGTTCGTGAAGTAATGACC
sdaA-seq-F	TACTCGCGTTGCCGTGGACGTTTATG
sdaB-seq-F	TGACCCGCGTGGTGGTTGACGTGTAC
speA-seq-F	GTGAAAACTCGTGAAGCACAGGGCC
gshA-seq-F	GAACATATGCTGACCTTTATGCGCG
tnaA-seq-F	CGTAGCTACTATGCGTTAGCCGAG
glpK-seq-F	AGGTTGGGTAGAACACGACCCAATG
tdUAG cloning	
8tdUAG-E132-f	/Phos/**TAGTAGTAGTAGTAGTAGTAGTAG**GATGGCAATATCCTGGGCCATAAACTG
9tdUAG-E132-f	/Phos/**TAGTAGTAGTAGTAGTAGTAGTAGTAG**GATGGCAATATCCTGGGCCATAAACTG
tdUAG-E132-r	/Phos/TTTAAAATCCGTGCCTTTCAGTTCAATGCG
ELP cloning	
TS-BlpI-R	TAGTTATTGCTCAGCGGTGG
TS-ApaI-F	GTCCCGGGTTATGGGCCC
KpnIApaI-f	GGTCGCGGGGTTGGGCCCAGCAAAGGTGAAGAACTGTTTACCG
KpnIApaI-r	TTTGCTCATGGTACCATCTCCTTCTTAAAGTTAAACAAAATTATTTC
gBlock DNA sequence	
gB-Twin-Strep-ApaIBlpI	GTCCCGGGTTATGGGCCCTCGGCGTGGAGCCACCCGCAGTTCGAGAAAGGTGGAGGTTCCGGAGGTGGATCGGGAGGTTCGGCGTGGAGCCACCCGCAGTTCGAAAAATAATAAGTCGACCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTA
	ApaI, BlpI

3.4 Results

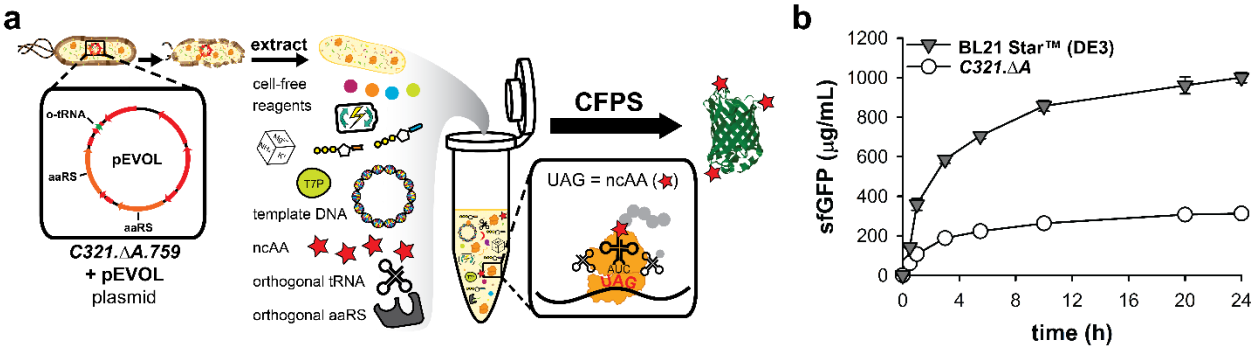


Figure 3.1. Cell-free protein synthesis of modified proteins using genomically recoded organisms. (**a**) Schematic of the production and utilization of crude extract from genomically recoded organisms with plasmid overexpression of orthogonal translation components for cell-free protein synthesis (CFPS). CFPS reactions are supplemented with the necessary substrates (*e.g.*, amino acids, NTPs, etc.) required for *in vitro* transcription and translation as well as purified OTS components to help increase the ncAA incorporation efficiency. (**b**) Time course of superfolder green fluorescent protein (sfGFP) synthesis catalyzed by extracts derived from a genomically recoded organism, *C321.*Δ*A,* and a commercial strain, BL21 Star™ (DE3). Three independent batch CFPS reactions were performed at 30 ˚C for each time point over 24 h, and one standard deviation is shown.

3.4.1 CFPS from Extracts of a Genomically Recoded Organism

To benchmark CFPS protein synthesis activity, we first compared sfGFP yields in extracts from *C321.*Δ*A* and BL21 Star™ (DE3), the standard commercial protein expression strain **(Figure 3.1.A)**. Combined transcription (TX)-TL reactions were carried out in 15 µL volumes for 24 h at 30°C. Protein yields from BL21 Star™ (DE3) extracts were >3-fold higher than those from *C321.*Δ*A* **(Figure 3.1.B)**, highlighting the need to improve protein synthesis yields to take advantage of the benefits of RF1 removal for making modified proteins with ncAAs for preparative purposes. Previously, genomic modifications to the extract source strain to stabilize DNA template[289], amino acid supply[288], and protein degradation[352] have improved CFPS yields from other source strains. For example, we engineered a partially recoded strain of *E. coli*

(*rEc.E13.*Δ*A*) by disrupting genes encoding nucleases (MCJ.559 (*endA*$^{-}$ *csdA*$^{-}$)) to improve protein synthesis yields >4-fold relative to the parent strain[232]. Building on this knowledge, we hypothesized that the genomic disruption of negative protein effectors in *C321.*Δ*A* extracts would help stabilize essential substrates in cell-free reactions, extend reaction durations, and increase CFPS yields.

3.4.2 Strain Engineering for Improved CFPS Performance

We targeted the functional inactivation of five nucleases (*rna*, *rnb*, *mazF*, *endA*, *rne*), two proteases (*ompT*, *lon*), and eight targets shown previously to negatively impact amino acid, energy, and redox stability (*gdhA*, *gshA*, *sdaA*, *sdaB*, *speA*, *tnaA*, *glpK*, *gor*) in *C321.*Δ*A* individually and in combination (see **Table 3.2**). Our effort followed a five-step approach. First, we generated a library of single mutant strains in which we used MAGE to insert an early TL termination sequence into the open reading frames of gene-targets that would functionally inactivate them, as we have done before (**Figure 3.2.A and Tables 3.1 and 3.3**). Second, we confirmed gene disruptions using multiplex allele specific PCR and DNA sequencing. Third, we measured growth rate for each of the MAGE-modified strains, noting that average doubling time increased 9 ± 9% above the parent strain **(Table 3.4)**. Fourth, cell extracts from each strain were generated using a high-throughput and robust extract generation procedure[291]. Fifth, we tested the strains in CFPS to assess their overall protein synthesis capability. We observed that seven single knock-out mutations increased CFPS yields more than 50% relative to extracts from the wild type *C321.*Δ*A* strain; namely, *rne-*, *mazF-*, *tnaA-*, *glpK-*, *lon-*, *gor-* and *endA-* **(Figure 3.2.B)**. These results indicated that some of the protein effectors targeted for inactivation were deleterious to CFPS activity. They also demonstrated the difficulty associated with predicting CFPS productivities from engineered

strains. For example, some mutations identified in previous screens (*e.g.*, *rnb*- in *rEc.E13.ΔA*)[232] were not beneficial in the *C321.ΔA* context, others which reduced cellular fitness enhanced CFPS activity (*e.g.*, *lon*-), and yet others with no impact on cell growth (*e.g.*, *ompT*-) led to poor extract performance **(Figure 3.2.B)**.

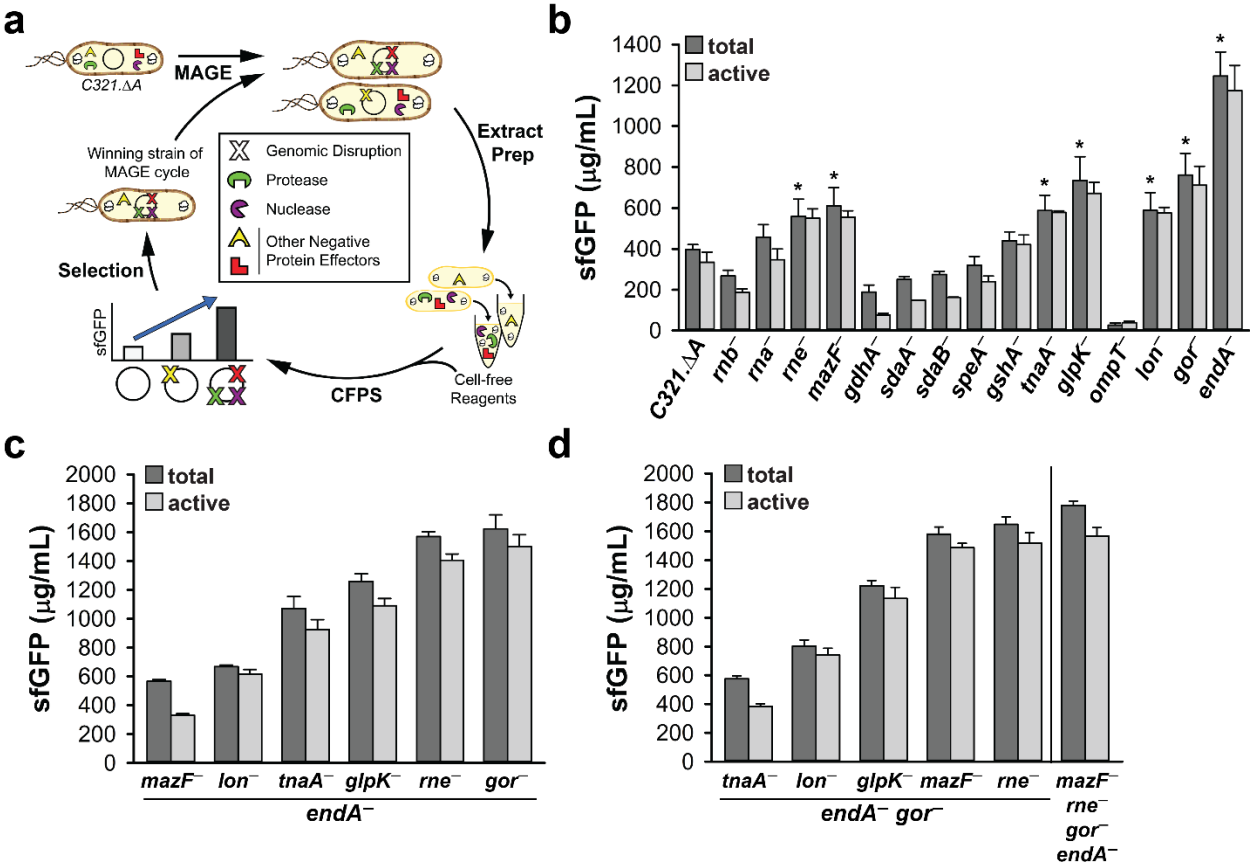


Figure 3.2 (previous page). Engineering and screening *C321.*Δ*A* variants for enhanced protein expression. (**a**) Schematic of design-build-test cycles employing multiplex automated genome engineering (MAGE) to disrupt putative negative protein effectors **(Table 3.2)** in engineered *C321.*Δ*A* strains for producing extracts with enhanced CFPS yields. A more comprehensive schematic of the CFPS reaction is given in **Figure 3.1**. (**b**) Cell extracts derived from *C321.*Δ*A* and genomically engineered strains containing a single putative negative effector inactivated were screened for sfGFP yields. The symbol (*) indicates beneficial mutations that increase active yields ≥50% relative to *C321.*Δ*A* ($p < 0.01$). (**c**) *C321.*Δ*A.542* (*endA*$^-$) was chosen as the new base strain and the following beneficial disruptions were pursued in combination: *rne*, *mazF*, *tnaA*, *glpK*, *lon*, and *gor*. (**d**) *C321.*Δ*A.709* (*endA*$^-$ *gor*$^-$) was selected as the new base strain for triple and quadruple mutant construction. *C321.*Δ*A.759* (*endA*$^-$ *gor*$^-$ *rne*$^-$ *mazF*$^-$) yielded the highest level of CFPS production. Total sfGFP concentration was measured by counting radioactive ^{14}C-Leucine incorporation and active protein was measured using fluorescence. Three independent batch CFPS reactions were performed for each sample at 30 ˚C for 20 h, and one standard deviation is shown.

Table 3.4. Doubling time of MAGE engineered strains. Cells were grown in 2xYTPG medium at 34 ˚C in a 96-well plate. Percentage relative to *C321.*Δ*A* is shown. Each data point is the average of nine replicates from three independent cultures.

Strains	Doubling time (min)	Percentage (%)
BL21 Star™ (DE3)	33.0 ± 0.0	96
*C321.*Δ*A*	34.2 ± 0.7	100
*C321.*Δ*A.540*	40.0 ± 0.7	117
*C321.*Δ*A.541*	34.0 ± 0.3	99
*C321.*Δ*A.542*	35.6 ± 0.3	104
*C321.*Δ*A.598*	33.2 ± 0.9	97
*C321.*Δ*A.618*	34.0 ± 1.2	99
*C321.*Δ*A.620*	37.0 ± 0.7	108
*C321.*Δ*A.626*	37.2 ± 0.6	109
*C321.*Δ*A.628*	35.0 ± 0.3	102
*C321.*Δ*A.644*	35.6 ± 0.7	104
*C321.*Δ*A.666*	35.2 ± 0.3	103
*C321.*Δ*A.667*	40.8 ± 0.6	119
*C321.*Δ*A.668*	37.4 ± 0.9	109
*C321.*Δ*A.669*	40.0 ± 1.2	117
*C321.*Δ*A.672*	34.0 ± 0.3	99
*C321.*Δ*A.674*	33.6 ± 0.6	98
*C321.*Δ*A.544*	34.0 ± 0.3	99
*C321.*Δ*A.678*	37.2 ± 0.6	109
*C321.*Δ*A.679*	35.8 ± 0.9	105
*C321.*Δ*A.709*	37.6 ± 1.2	110

C321.ΔA.711	35.6	±	0.3	104
C321.ΔA.708	43.5	±	1.3	127
C321.ΔA.703	38.0	±	1.2	111
C321.ΔA.705	41.2	±	2.5	120
C321.ΔA.706	38.8	±	3.0	113
C321.ΔA.740	38.0	±	0.3	111
C321.ΔA.738	41.4	±	1.2	121
C321.ΔA.759	44.4	±	1.6	130

With improvements in hand from single mutant strains, we next set out to identify synergistic benefits to CFPS productivity by combining highly productive mutations. We introduced the *rne*-, *mazF*-, *tnaA*-, *glpK*-, *lon*-, and *gor*- mutations to the best performing strain from our initial screen, strain *C321.ΔA.542* (*endA*$^-$) **(Figure 3.2.C)**. The combination of *endA*$^-$ and *gor*$^-$ mutations resulted in an extract capable of synthesizing 1620 ± 10 mg/L of sfGFP (strain *C321.ΔA.709*). We then used *C321.ΔA.709* to generate six additional strains with combined mutations. Although we did not observe synergistic enhancements, our top performing extract chassis strain (*C321.ΔA.759* (*endA*$^-$ *gor*$^-$ *rne*$^-$ *mazF*$^-$)) resulted in yields of 1780 ± 30 mg/L **(Figure 3.2.D)**, representing a 4.5-fold increase in sfGFP yield relative to the progenitor strain (*C321.ΔA*). In addition, we tested 12 additional combinatorial mutants generated throughout our MAGE screening, and although a few demonstrated CFPS yields > 1 g/L of active GFP, none surpassed the CFPS yields observed from *C321.ΔA.759* **(Table 3.5)**. Final strains were fully sequenced to verify functional targeted modifications in the genome. Lastly, we determined that CFPS improvements seen in *C321.ΔA.759* brought on by genomic modifications could not be obtained by simply supplementing *C321.ΔA*-based reactions with RNAse inhibitors **(Figure 3.3)**. Final strains were fully sequenced to verify functional targeted modifications in

the genome. Whole-genome sequences for strains *C321.ΔA*, *C321.ΔA.542*, *C321.ΔA.705*, *C321.ΔA.709*, *C321.ΔA.740*, and *C321.ΔA.759* have been deposited in the NCBI SRA collection under accession code PRJNA361365. Each of the targeted mutations were achieved. MAGE has been shown to induce mutations throughout the genome before, and we observed a number of accumulated polymorphisms in the extract chassis strains. These polymorphisms, along with a specific list of protein-coding genes bearing mutations, are shown in **Supplementary Tables 6 and 7**. In the future, we seek to better understand the systems impact of the non-targeted mutations.

Table 3.5. CFPS yields for additional MAGE engineered strains demonstrate the difficulty in predicting mutation impact on extract activity. Active sfGFP CFPS yields from 12 distinct extracts, each derived from a MAGE-generated mutant in this study. Fold changes are relative to CFPS yields obtained from extracts derived from *C321.ΔA*. Three independent CFPS reactions for each sample were performed.

Strain	Active yield (μg/mL)	Fold change
C321.ΔA	350 ± 6	1.0
C321.ΔA.619	36 ± 4	0.1
C321.ΔA.617	410 ± 14	1.2
C321.ΔA.621	37 ± 1	0.1
C321.ΔA.680	850 ± 16	2.4
C321.ΔA.664	640 ± 6	1.8
C321.ΔA.756	1,340 ± 70	3.8
C321.ΔA.755	700 ± 35	2.0
C321.ΔA.758	1,250 ± 71	3.5
C321.ΔA.879	940 ± 62	2.7
C321.ΔA.878	850 ± 84	2.4
C321.ΔA.564	600 ± 27	1.7
C321.ΔA.563	540 ± 12	1.5

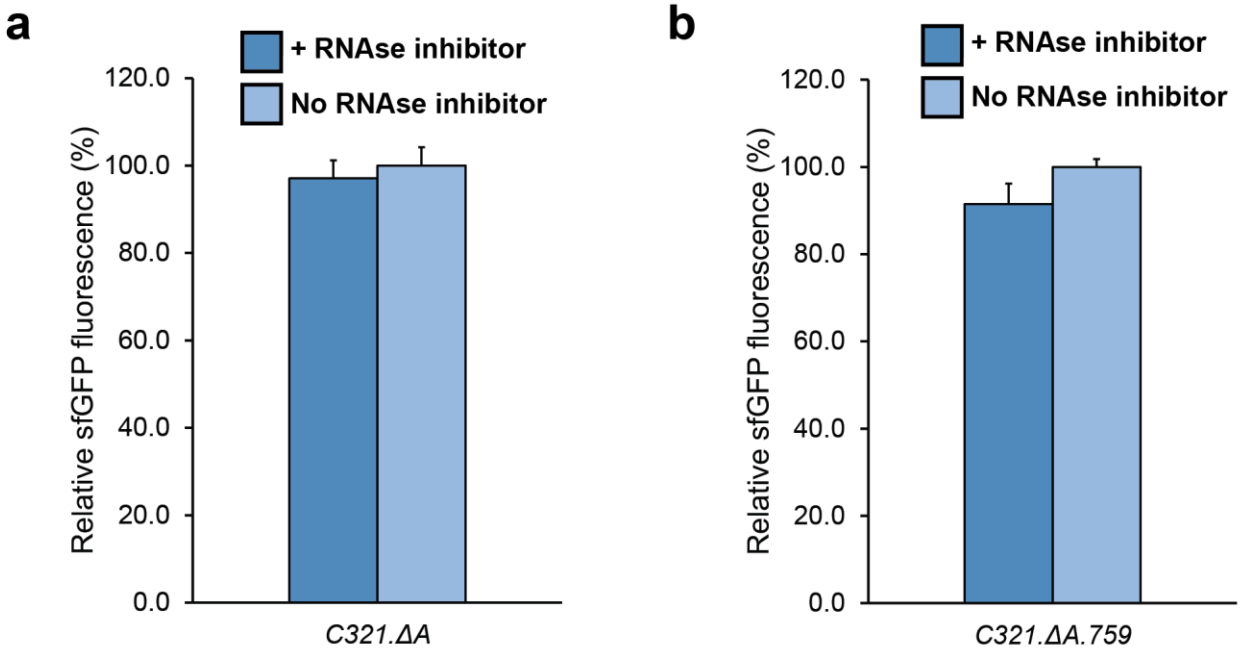


Figure 3.3. Impact of supplementing RNAse inhibitor into CFPS reactions. CFPS reactions were performed using both *C321.ΔA* (**a**) and *C321.ΔA.759* (**b**) lysates, both with and without supplementation with RNAse inhibitor. For each lysate, relative sfGFP fluorescence is shown with the "No RNAse inhibitor" condition set to 100%. Three independent batch mode reactions were performed for each condition, and one standard deviation is shown.

Based on our previous studies using *rEc.E13.*Δ*A*[232], we hypothesized that the beneficial mutations in *C321.*Δ*A.759* reduced mRNA degradation and stabilized the DNA template. To test mRNA stability, we performed TL-only reactions using extracts derived from *C321.*Δ*A.759* and *C321.*Δ*A.* Purified mRNA template coding for sfGFP was used to direct protein synthesis. We observed a twofold increase in mRNA and ~90% increase of active sfGFP using *C321.*Δ*A.759* extracts relative to *C321.*Δ*A* extracts after a 120 minute cell-free reaction **(Figure 3.4)**. To test DNA stability, TX-only reactions were used. Specifically, plasmid DNA containing the modified red fluorescent protein-Spinach aptamer gene **(Table 3.1)** was preincubated with cell extract and a fluorophore molecule, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DHFBI), for 0, 60,

and 180 min. Then, CFPS reagents were added and mRNA was synthesized, then quantified by measuring the fluorescence of DFHBI-bound Spinach aptamer mRNA. After 180 min of preincubation, nearly 50% of Spinach aptamer mRNA was synthesized in *C321.*Δ*A.759* (*endA*$^{-}$) extracts relative to the 0 min control. In contrast, the extract with endonuclease I (*C321.*Δ*A*) decreased the maximum mRNA synthesis level by ~75% **(Figure 3.5)**. Together, our data support the hypothesis that inactivating nucleases in the extract chassis strain stabilized DNA and mRNA to improve CFPS yields.

In addition to confirming added DNA and mRNA stability, we also assessed potential changes in energy and amino acid substrate stability that may have occurred in *C321.*Δ*A.759*– relative to *C321.*Δ*A*–based CFPS. Similar trends in ATP levels, adenylate charge, and amino acid concentrations were observed in CFPS reactions derived from both strains **(Figures 3.6-3.8)**. Supplemental feeding of the amino acids present in the 5 lowest concentrations did not improve yields **(Figure 3.8.E)**. The similar amino acid and energy stability profiles in *C321.*Δ*A.759* compared with *C321.*Δ*A* suggest that our strain engineering efforts did not modulate the availability of these substrates.

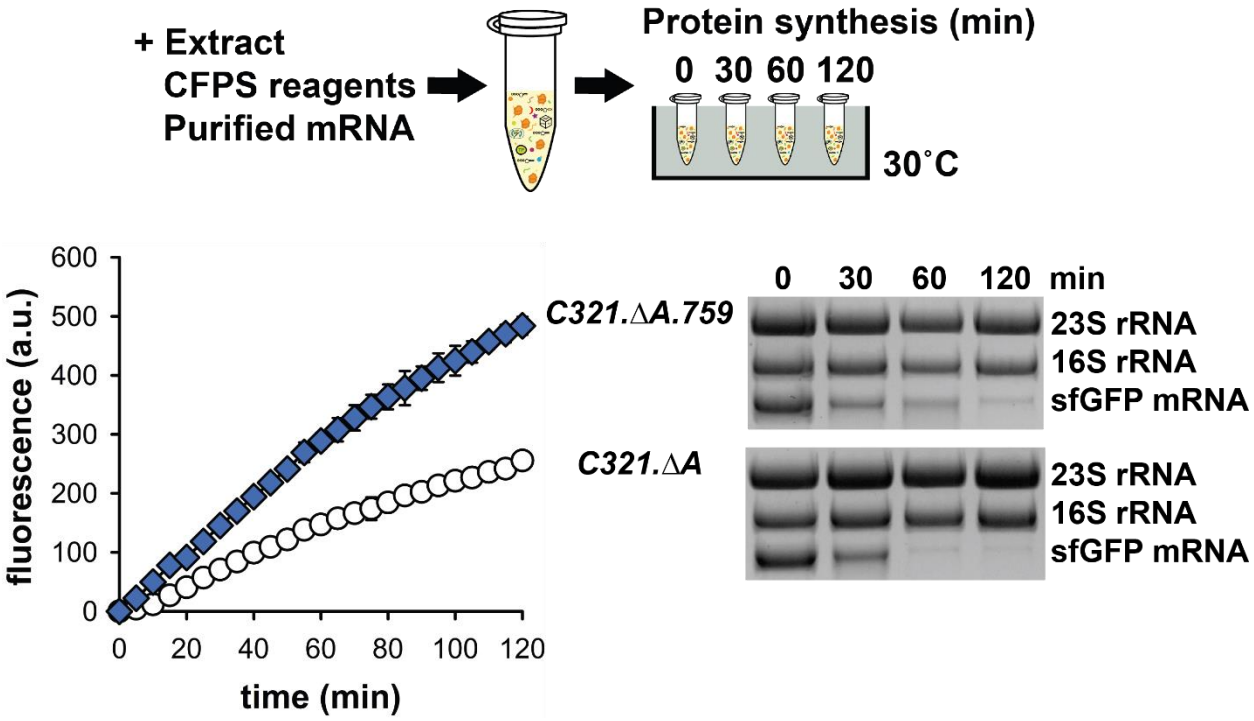


Figure 3.4. Genomic engineering of *C321.ΔA.759* improves CFPS by stabilizing mRNA. Purified mRNA driven cell-free translation (TL)-only reactions of sfGFP in *C321.ΔA.759* and *C321.ΔA* extracts. Three independent reactions for each sample were performed for 120 min at 30 ˚C. sfGFP synthesis was monitored by fluorescence (left), and mRNA levels were assessed by an RNA gel (right) and analyzed using densitometry. For the mRNA gels, 23S rRNA (2904 nucleotides) and 16S rRNA (1541 nucleotides) are shown as loading and size reference controls for each lane. The mRNA with the promoter, gene sequence, and terminator is 917 nucleotides. Error bars indicate standard deviation.

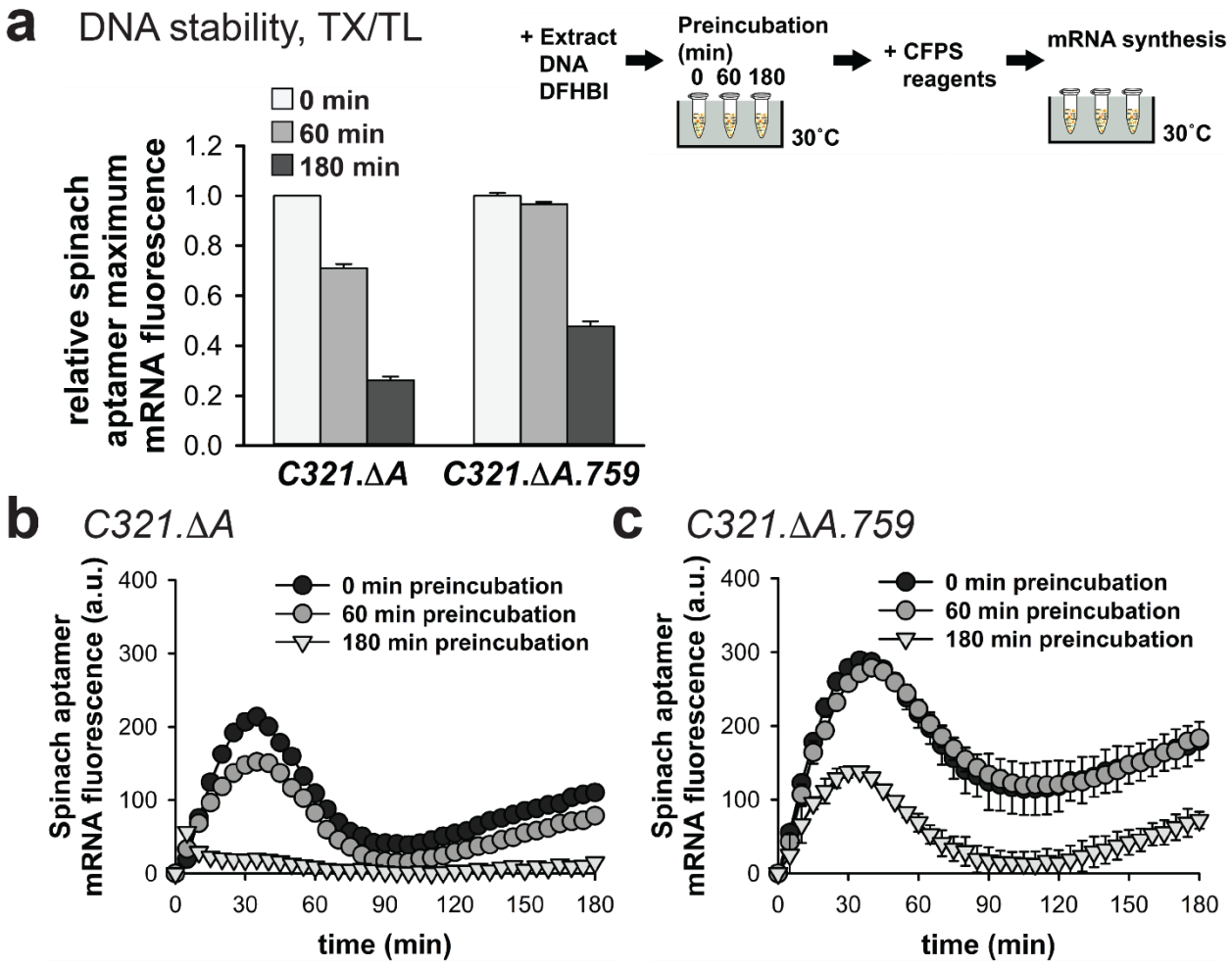


Figure 3.5. Genomic engineering of *C321.ΔA.759* improves CFPS by stabilizing DNA. (**a**) Cell-free Spinach aptamer synthesis using *C321.ΔA.759* and *C321.ΔA* extracts to test plasmid stability. After preincubation (0, 60, and 180 min) of Spinach aptamer plasmid DNA with cell extract, CFPS reagents were added and reactions incubated at 30˚C. Relative maximum mRNA synthesis level time course experiments from transcription (TX) only reactions using (**b**) endonuclease I present (*C321.ΔA*) and (**c**) deficient (*C321.ΔA.759*) extract were compared. While panel **a** captures the relative maximum mRNA level, panels **b** and **c** capture the data spread at 5 minute intervals in a three hour time course incubation. Three independent reactions for each sample were performed. Error bars indicate standard deviation.

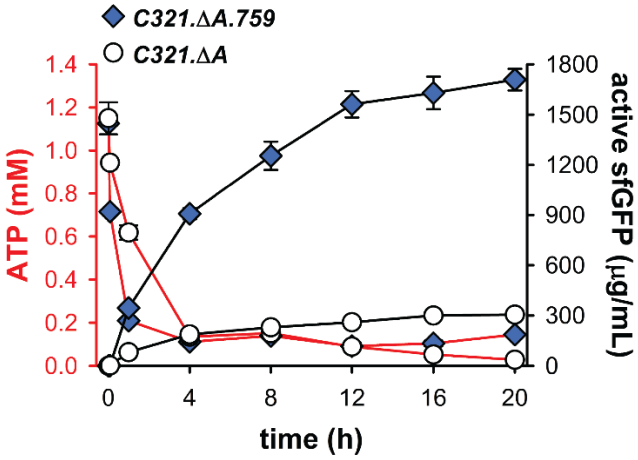


Figure 3.6. Analysis of energy stability and CFPS yields using extract derived from *C321.ΔA.759* and *C321.ΔA*. Time course of active sfGFP production and ATP concentration in cell-free reactions using either *C321.ΔA.759* or *C321.ΔA* extract over 20 h. Three independent batch CFPS reactions were performed. Error bars indicate standard deviation.

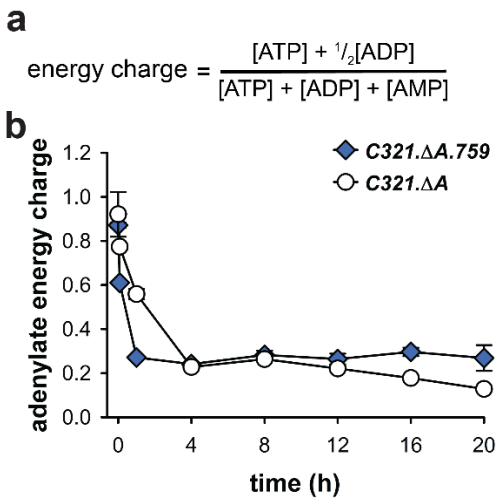


Figure 3.7. Adenylate energy charge of *C321.ΔA.759* and *C321.ΔA*. (**a**) The calculation of the adenylate energy charge was performed as described by Atkinson[393]. (**b**) Energy charge plotted as a function of reaction time for CFPS reactions using extract derived from either *C321.ΔA.759* or *C321.ΔA*. Three independent batch CFPS reactions were performed. Error bars indicate standard deviation.

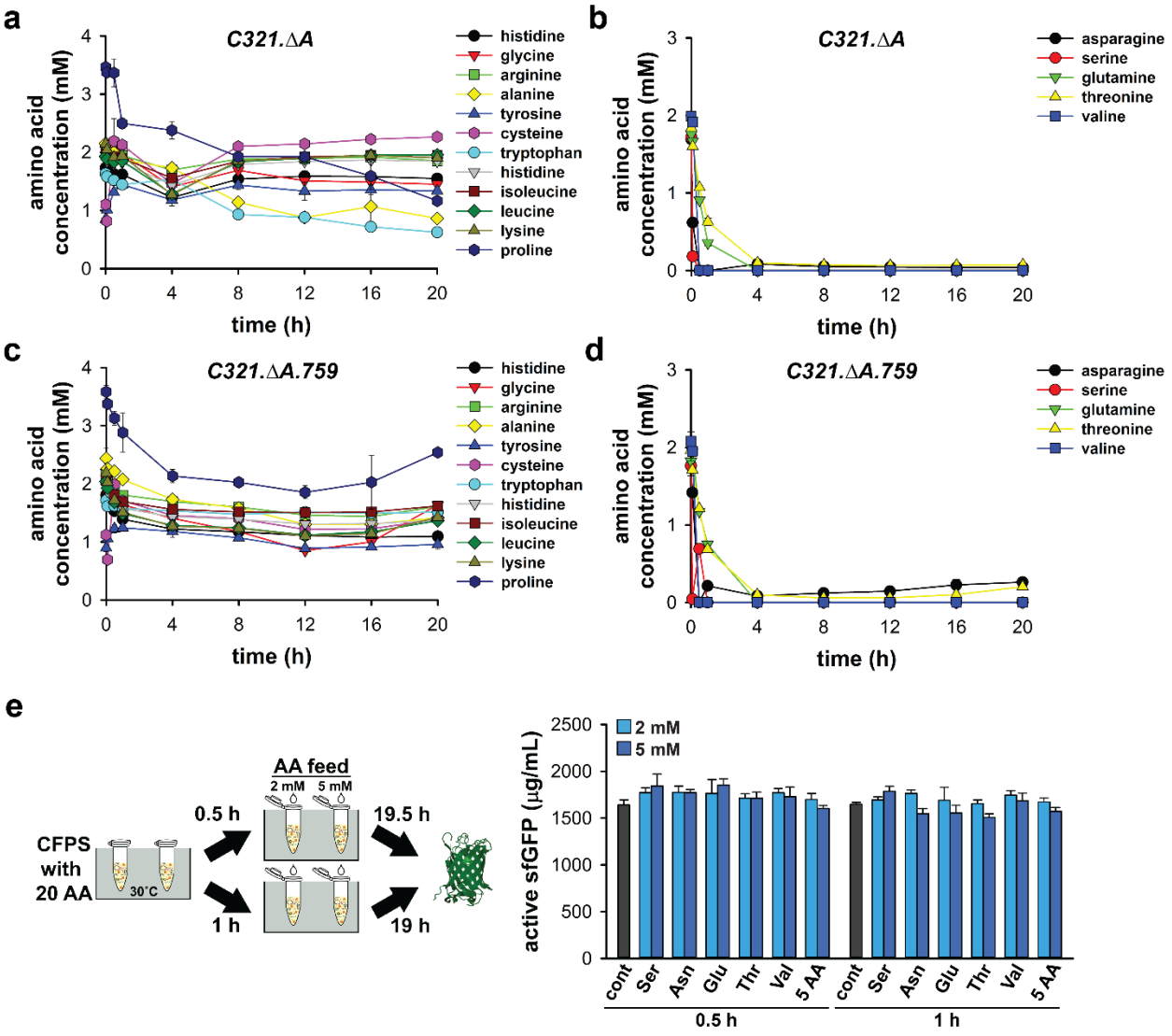


Figure 3.8. Impact of feeding potentially limiting amino acid substrates into CFPS reactions. Concentration of amino acid species in CFPS reactions performed using (**a** and **b**) *C321.ΔA* and (**c** and **d**) C*321.ΔA.759* extract. (**b** and **d**) Amino acids found to be ≥90% depleted after 4 hours of incubation were deemed potentially limiting amino acids. (**e**) Fed-batch reactions with individual potentially limiting amino acids and an equimolar mixture (5 AA) at concentrations of 2 and 5 mM at 0.5 and 1 h time points using 0.5 μL of feeding solution; water was used as a control (left). Three independent batch CFPS reactions were performed and activity of synthesized sfGFP was analyzed for each data point, and one standard deviation is shown.

To generalize CFPS improvements in *C321.ΔA.759*, we next expressed four model proteins that have been previously synthesized in CFPS systems and compared productivities to BL21 Star™ (DE3). We observed a 31-63% increase in soluble and total protein synthesis of sfGFP, chloramphenicol acetyltransferase (CAT), dihydrofolate reductase (DHFR), and modified murine granulocyte-macrophage colony-stimulating factor (mGM-CSF) in our engineered *C321.ΔA.759* extracts as compared to BL21 Star™ (DE3) extracts **(Figure 3.9.A)**. Autoradiograms of proteins produced using *C321.ΔA.759* extract show production of full-length sfGFP, CAT, DHFR, and mGM-CSF **(Figure 3.9.B and 3.9.C)**. In addition, we observed disulfide bond formation in the model mGM-CSF under an oxidizing CFPS environment (-DTT), as has been previously shown **(Figure 3.9.C)**[267, 387]. In sum, the development of enhanced extract source strains by MAGE enabled a general and high-yielding CFPS platform.

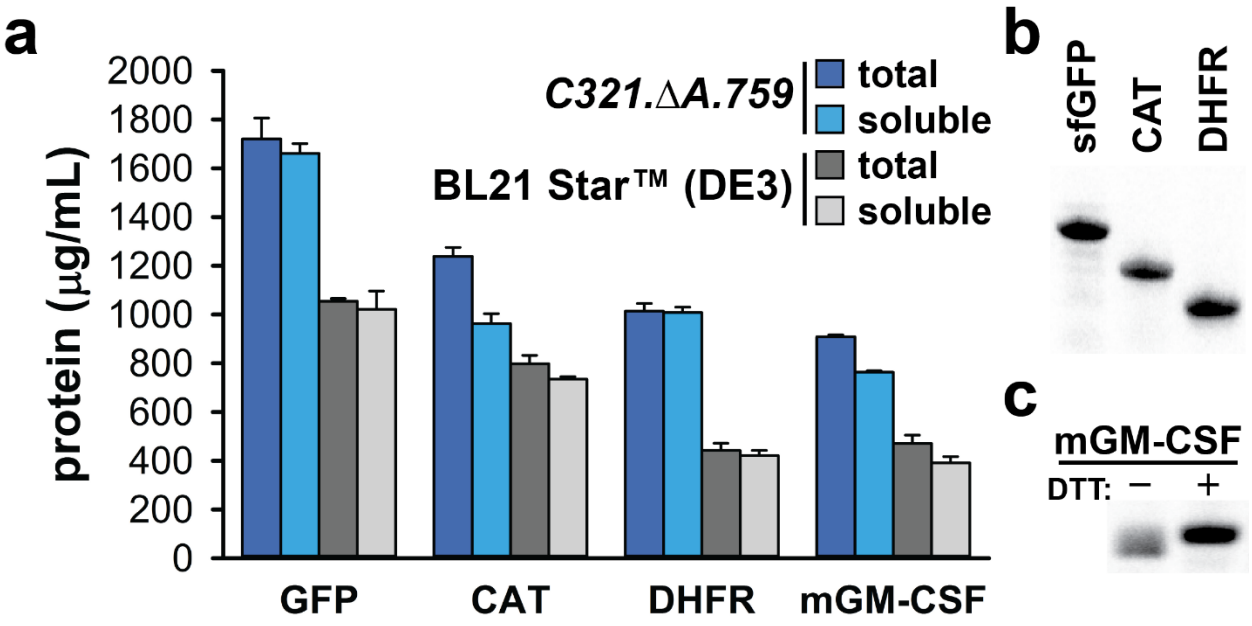


Figure 3.9 (previous page). Genomically engineered *C321.ΔA.759* demonstrates high-yielding, general protein synthesis utility. (**a**) Total and soluble protein yields of sfGFP (26.8 kDa), chloramphenicol acetyltransferase (CAT; 27.7 kDa), dihydrofolate reductase (DHFR; 17.9 kDa), and murine granulocyte-macrophage colony-stimulating factor (mGM-CSF; 16.5 kDa) synthesized by extract from *C321.ΔA.759* and BL21 Star™ (DE3) at 30 ˚C for 20 h. Three independent CFPS reactions for each sample were performed at 30 ˚C for 20 h, and one standard deviation is shown. (**b**) Autoradiogram showing fully reduced and denatured sfGFP, CAT, and DHFR as produced by *C321.ΔA.759* extract in CFPS. (**c**) Autoradiogram of mGM-CSF as produced by *C321.ΔA.759* in CFPS under oxidizing conditions. mGM-CSF was analyzed under oxidizing (-DTT) and reducing (+DTT) conditions. The difference in band position is influenced by the presence of disulfide bonds.

3.4.3 Multi-site ncAA Incorporations into Proteins in CFPS

We next aimed to test site-specific ncAA incorporation into proteins using our new CFPS platform from *C321.ΔA.759*-derived extracts and compare these results to reactions using extracts from BL21 Star™ (DE3) (containing RF1) and a partially recoded RF1-deficient engineered strain MCJ.559 based on *rEc.E13.ΔA*. To do so, we transformed each organism with pEVOL-pAcF plasmid that expresses both orthogonal pAcF synthetase (pAcFRS) and tRNA (o-tRNA^{opt})[14]. Then we quantitatively assessed the incorporation of pAcF into sfGFP variants with up to five in-frame amber codons. CFPS reactions were supplemented with additional OTS components based on our previous work[24]. Specifically, we added 10 μg/mL of linear DNA encoding optimized orthogonal tRNA in the form of a transzyme (o-tRNA^{opt}) for *in situ* synthesis of the tRNA. The orthogonal pAcFRS was overproduced, purified as previously described, and added at a level of 0.5 mg/mL. The ncAA, in this case pAcF, was supplied at a level of 2 mM in each CFPS reaction. Total protein yields were quantified by ^{14}C-leucine radioactive incorporation. Production of wild-type and modified sfGFP containing one UAG codon (sfGFP-UAG) was increased 77% and 92% in *C321.ΔA.759* extracts as compared to BL21 Star™ (DE3), and 120% and 145% as compared to *MCJ.559*, respectively **(Figure 3.10.A)**. Moreover, we observed that sfGFP-UAG was expressed

at 90% the level of wild-type sfGFP. Owing to the absence of RF1 competition, the major protein produced was full-length sfGFP using extracts derived from *C321.ΔA.759* and MCJ.559, whereas truncated sfGFP was visible in reactions catalyzed by BL21 Star™ (DE3) extract, presumably due to RF1 competition **(Figure 3.10.A)**[23, 394]. Similar results were obtained with a second model protein, CAT with an in-frame amber codon at position 112 (CAT-UAG) **(Figure 3.10.A)**. When expressing CAT-UAG using MCJ.559 extract, similar levels of truncated CAT relative to BL21 Star™ (DE3) were observed; however this is most likely due to an upregulation of rescue mechanisms for ribosome stalling in the partially recoded strain[21]. Single pAcF incorporation into CAT-UAG using *C321.ΔA.759* lysate demonstrated only full-length product. Therefore, our completely recoded, genomically engineered *C321.ΔA.759* strain provides benefits for efficient ncAA incorporation without detectable levels of truncation product.

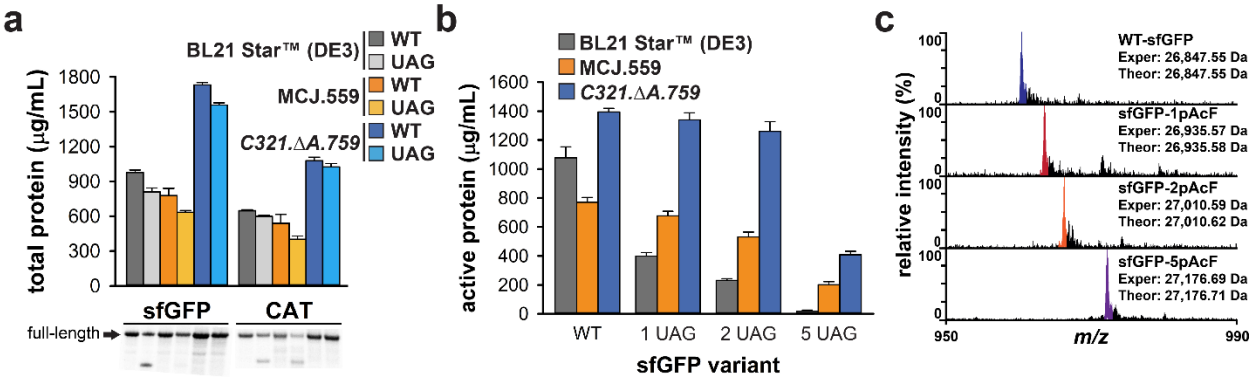


Figure 3.10. Efficient incorporation of pAcF into proteins at multiple amber sites using genomically engineered *C321.ΔA.759* extract. Cell-free pAcF incorporation was compared using extracts derived from BL21 Star™ (DE3), MCJ.559, and *C321.ΔA.759* strains harboring the pEVOL-pAcF vector. (**a**) Total protein yields for wild-type (WT) and 1 UAG versions of sfGFP and CAT are shown (top) along with an autoradiogram of the resulting protein product (bottom). (**b**) Multi-site incorporation of pAcF into sfGFP as quantified by active protein produced. Three independent batch CFPS reactions were performed for each sample at 30 ˚C for 20 h, and one standard deviation is shown. (**c**) Spectrum of the 28+ charge state of sfGFP, obtained by top-down mass spectrometry and illustrating site-specific incorporation of pAcF at single and multiple sites. Major peaks (color) in each spectrum coincide with the theoretical peaks for each species. Smaller peaks immediately to the right of the major peaks are due to oxidation of the protein, a common electrochemical reaction occurring during electrospray ionization. Experimentally determined masses are ≤1 ppm in comparison of theoretical mass calculations. Due to the size of pAcF, misincorporation would result in peaks present at lower *m/z* values relative to the colored theoretical peak.

We then evaluated the ability of our high-yielding CFPS platform to facilitate incorporation of up to five identical ncAAs into sfGFP. For ease of analysis, a fluorescence assay was used, which indicated increased production of sfGFP in extracts from *C321.ΔA.759* **(Figure 3.10.B)**. Results for BL21 Star™ (DE3) extract displayed an exponential decrease in active sfGFP synthesized with an increasing presence of UAG, leading to the production of no detectable active protein for sfGFP-5UAG. Active protein produced by *C321.ΔA.759* extract were ~2-fold greater that that produced by MCJ.559 extract, suggesting that benefits observed in increased yield can be extended to multi-site ncAA incorporation for our enhanced, fully recoded strain. Furthermore, we examined the ability to incorporate consecutive pAcFs into single protein. Protein gel and autoradiogram analysis of sfGFP with eight and nine consecutive amber codons indicated that this is possible, with the percent of full-length product being ~75% and 60%, respectively **(Figure 3.11)**.

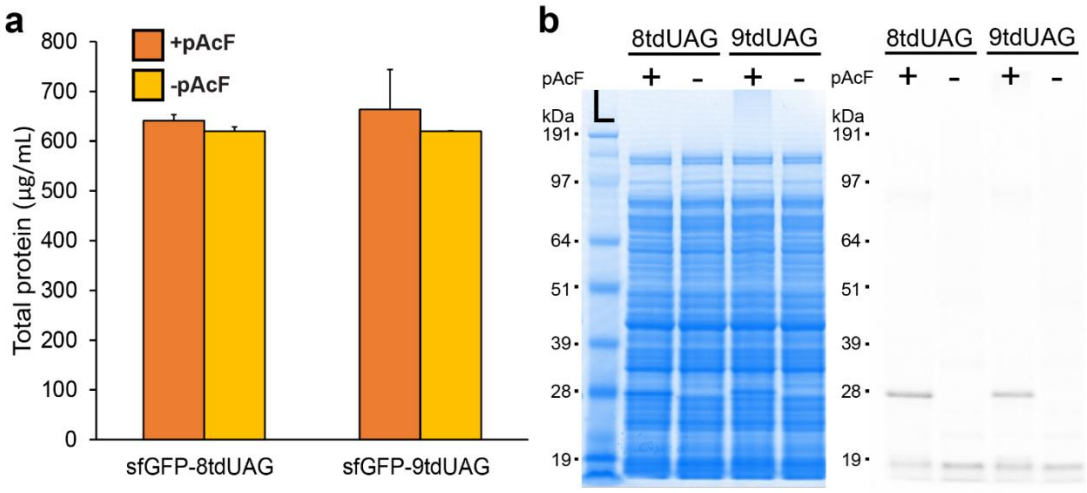


Figure 3.11. Incorporation of multiple consecutive ncAAs. (**a**) Total yields of sfGFP with 8 consecutive UAGs (8tdUAG) and 9 consecutive UAGs (9tdUAG) are shown. Reactions were performed both in the presence and absence of pAcF, and were incubated at 30 °C for 20 h. Quantification was achieved via incorporation of ^{14}C-labeled leucine followed by scintillation counting. Three independent batch reactions were performed for each condition, and one standard deviation is shown. (**b**) SDS-PAGE (left) and autoradiogram (right) analysis of cell-free produced sfGFP-8tdUAG and sfGFP-9tdUAG in the presence and absence of pAcF.

In addition, batch reaction catalyzed by *C321.ΔA.759* extracts could also be scaled 17-fold without loss of productivity provided that proper ratio of surface area to volume ratio is maintained **(Figure 3.12)**[368]. Of note, we believe our reactions could be further scaled to a wide range of volumes to produce larger amounts of protein if accounting for sureface area to volume effects. For example, Sutro Biopharma has applied *E. coli*-based CFPS platforms to clinical manufacturing of therapeutics at the 100 L scale[264], with an expansion factor of 106. In terms of cost, although we use a phosphoenolpyruvate (PEP)-based CFPS system here, cellular metabolism could be used to fuel cost effective, high-level protein synthesis suitable for manufacturing applications[284, 287].

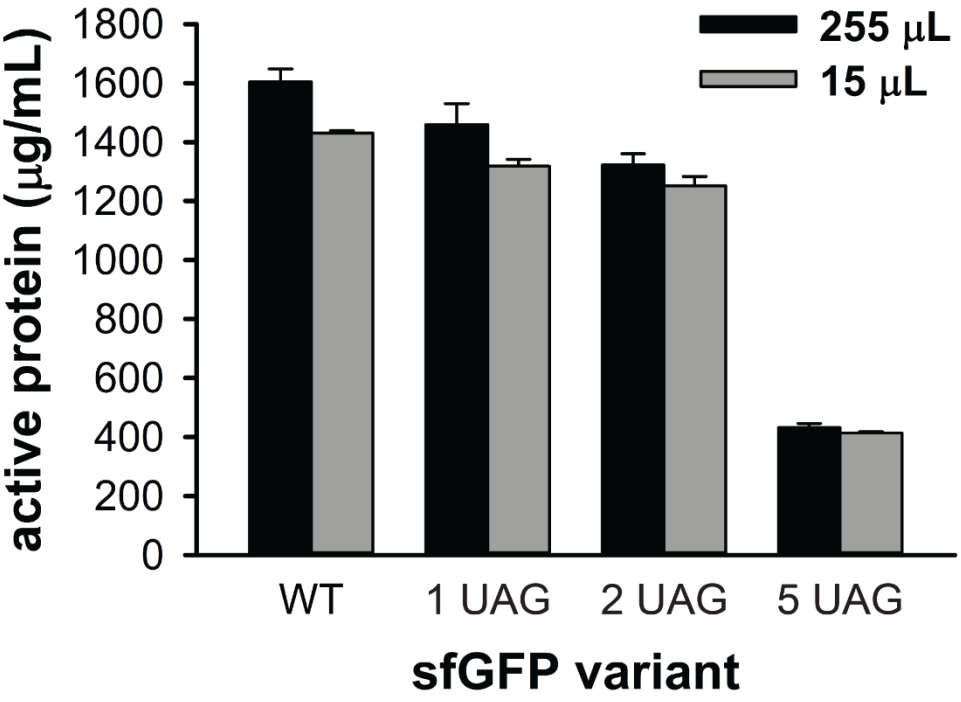


Figure 3.12. Scaled-up synthesis of sfGFP containing multiple identical ncAAs. Active yields of wild-type, 1 UAG, 2 UAG, and 5 UAG sfGFP are shown for 15 µL and 255 µL batch reactions. Reactions were performed at 30 ˚C for 20 h. Three independent batch CFPS reactions were performed for each sample, and one standard deviation is shown.

After demonstrating benefits for protein expression, we carried out top-down mass spectrometry (*i.e.*, MS analysis of whole intact proteins) to detect and provide semi-quantitative data for the incorporation efficiency of pAcF into sfGFP using extract derived from *C321.*Δ*A.759*. **Figure 3.10.C** shows the 28+ charge state of sfGFP and clearly illustrates mass shifts corresponding to the incorporation of one, two, and five pAcF residues. Site-specific incorporation of pAcF, as detected by MS, was ≥ 98% in all samples, with less ≤1 p.p.m. difference between experimental and theoretical protein masses. In other words, efficient, and high yielding site-specific pAcF incorporation into sfGFP was observed when using *C321.*Δ*A.759* extract. We went on to further show that extracts generated from *C321.*Δ*A.759* are compatible with multiple OTSs, showing the incorporation of *p*-propargyloxy-L-phenylalanine (pPaF) and *p*-azido-L-phenylalanine (pAzF) **(Figure 3.13)**.

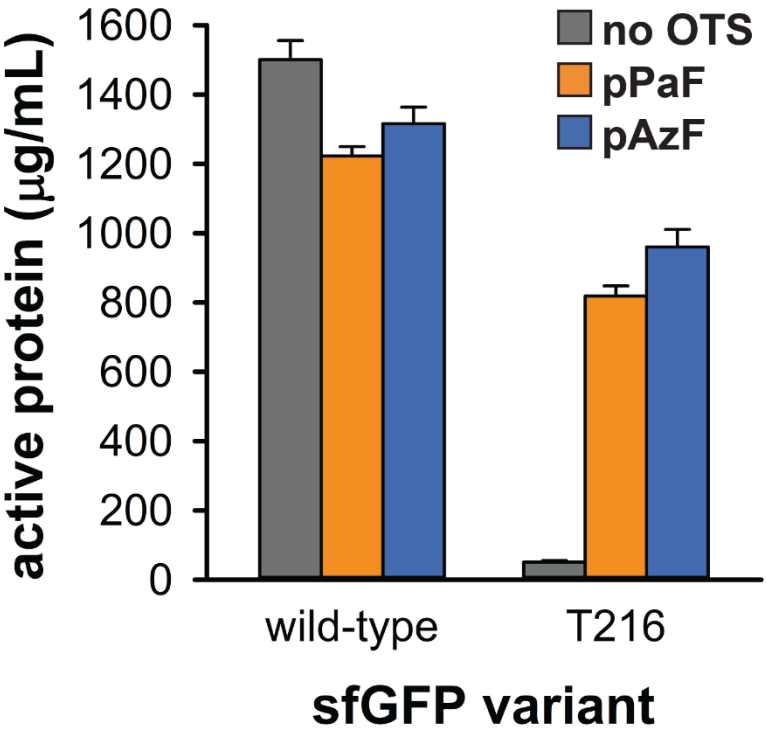


Figure 3.13 (previous page). Multiple orthogonal translation systems are active in crude extracts derived from *C321.*Δ*A.759*. *p*-propargyloxy-L-phenylalanine (pPaF) and *p*-azido-L-phenylalanine (pAzF) were incorporated into position T216 in sfGFP using extracts derived from *C321.*Δ*A.759* exclusively overexpressing o-tRNA from the pDULE plasmid. Reactions were supplemented with the ncAA's corresponding purified orthogonal synthetase. Three independent batch CFPS reactions were performed for each sample, and one standard deviation is shown.

3.4.4 Multi-site ncAA Incorporation into ELPs

We next explored the synthesis of sequence-defined biopolymers containing tens of site-specifically introduced ncAAs using our efficient and tunable CFPS system. As a model biopolymer, we selected ELPs. ELPS are biocompatible and stimuli-responsive biopolymers that can be applied for drug delivery and tissue engineering[395]. Typically, ELPs consist of repeats of the pentapeptide sequence VPGVG, which is known to be a key component in elastin and exhibits interesting self-assembly behavior (random coil to helix) above its transition temperature. The structure and function of elastin is maintained as long as the glycine and proline residues are present; however, the second valine residue is permissive for any amino acid except proline and is therefore also permissive to ncAAs[27]. Previously, ncAAs have been introduced into ELPs by substituting natural amino acids with structurally similar ncAAs in CFPS systems[396]. Conticello and colleagues have also previously produced imperfect ELPs containing up to 22 ncAAs *in vivo* using an *E. coli* strain with an attenuated activity of RF1[397]. We previously incorporated up to 30 ncAAs into ELPs by evolving orthogonal synthetases *in vivo* with enhanced specificities[27]. In this study, we constructed and tested three ELP constructs containing 20, 30, and 40 UAG codons, as well as control proteins with tyrosine codons substituted for UAGs.

Before characterizing ELP yields, we first carried out a series of optimization experiments to enhance CFPS yields of sfGFP with 5 UAG codons, since expression yields for this construct

were reduced in our initial studies **(Figure 3.10.B)**. By testing total and soluble protein yields, we determined that the reduction in yield was a result of loss in sfGFP solubility and activity **(Figure 3.14)**. However, a 31% increase in sfGFP-5UAG production was observed upon increasing pAcFRS levels 2-fold, pAcF levels 2.5-fold, and o-tz-tRNA 3-fold **(Figure 3.15)**. Upon application of these optimized conditions, called OTS^{opt}, to the synthesis of ELP-UAGs containing 20, 30, and 40-mers, total yields increased by 40%, 33%, and 26%, respectively, as compared with supplementing with OTS levels optimized for 1 ncAA incorporation **(Figure 3.16)**. ELP-UAG products were visualized using an autoradiogram, which demonstrated the high percentage of full-length protein and whose band intensities corroborate total yields measured **(Figure 3.16.A)**.

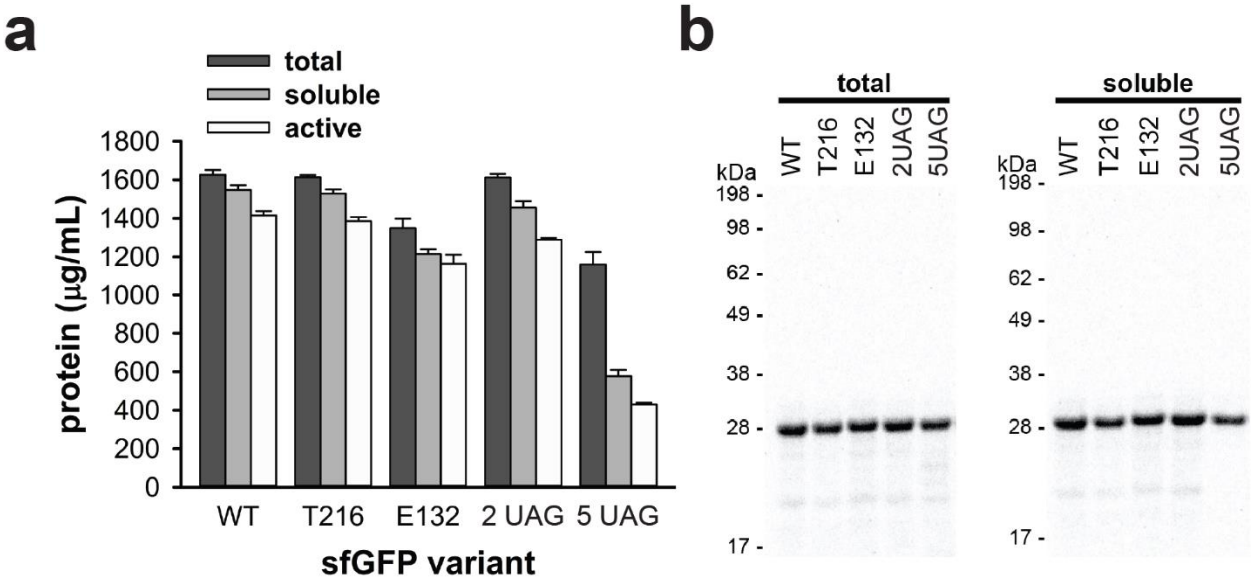


Figure 3.14. Comparison of total, soluble, and active yields for sfGFP containing multiple identical ncAAs. Extract derived from *C321.ΔA.759* harboring pEVOL-pAcF was used to catalyze CFPS reactions synthesizing sfGFP variants. (**a**) Total and soluble yields were measured using radioactive incorporation and active yields were measured using fluorescence. Three independent batch CFPS reactions were performed for each sample, and one standard deviation is shown. (**b**) Total and soluble proteins were visualized via autoradiograms (molecular weight of sfGFP is 27 kDa).

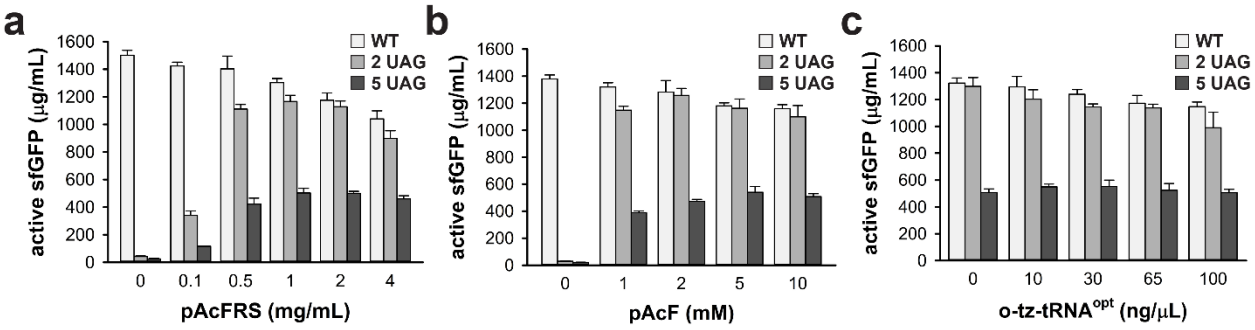


Figure 3.15. Optimization of pAcF incorporation into sfGFP-5UAG. Using extracts derived from *C321.ΔA.759* harboring pEVOL-pAcF, we optimized levels of (**a**) purified orthogonal synthetase (pAcFRS), (**b**) non-canonical amino acid (pAcF), (**c**) and orthogonal transzyme tRNA^{opt} (o-tz-tRNA^{opt}) added into the CFPS reactions. Three independent batch CFPS reactions were performed for each sample, and one standard deviation is shown.

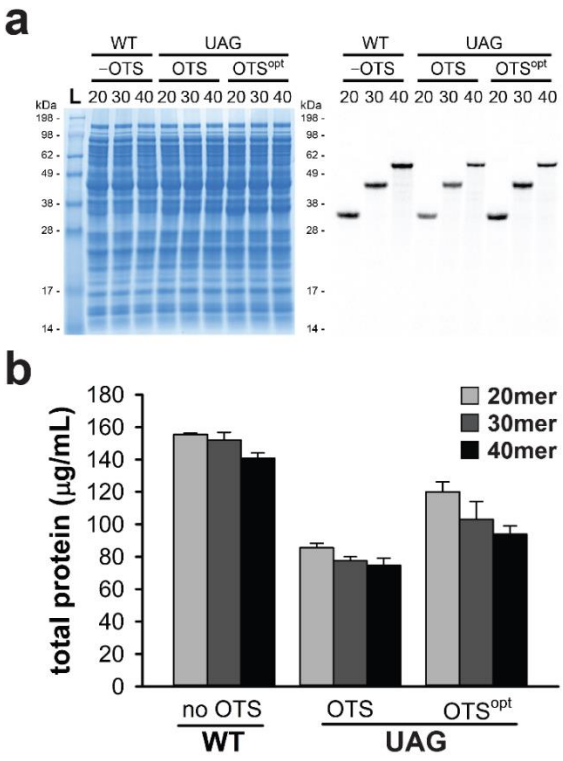


Figure 3.16. Improved full-length ELP-UAG yield utilizing optimized OTS concentrations (OTS^{opt}). ELP-UAG constructs containing 20, 30, and 40 -mers were synthesized in cell-free either using OTS^{opt} (1 mg/mL pAcFRS, 5 mM pAcF, 30 ng/μL o-tz-tRNA^{opt}) levels as identified by results from **Supplementary Figure 14**, or standard OTS (0.5 mg/mL pAcFRS, 2 mM pAcF, 10 ng/μL o-tz-tRNA^{opt}) levels. ELP-WT variants were synthesized in the absence of the OTS as controls. Products were (**a**) visualized by SDS-PAGE (left) and autoradiogram (right) and (**b**) total protein was quantified. Three independent batch CFPS reactions were performed for each sample, and one standard deviation is shown.

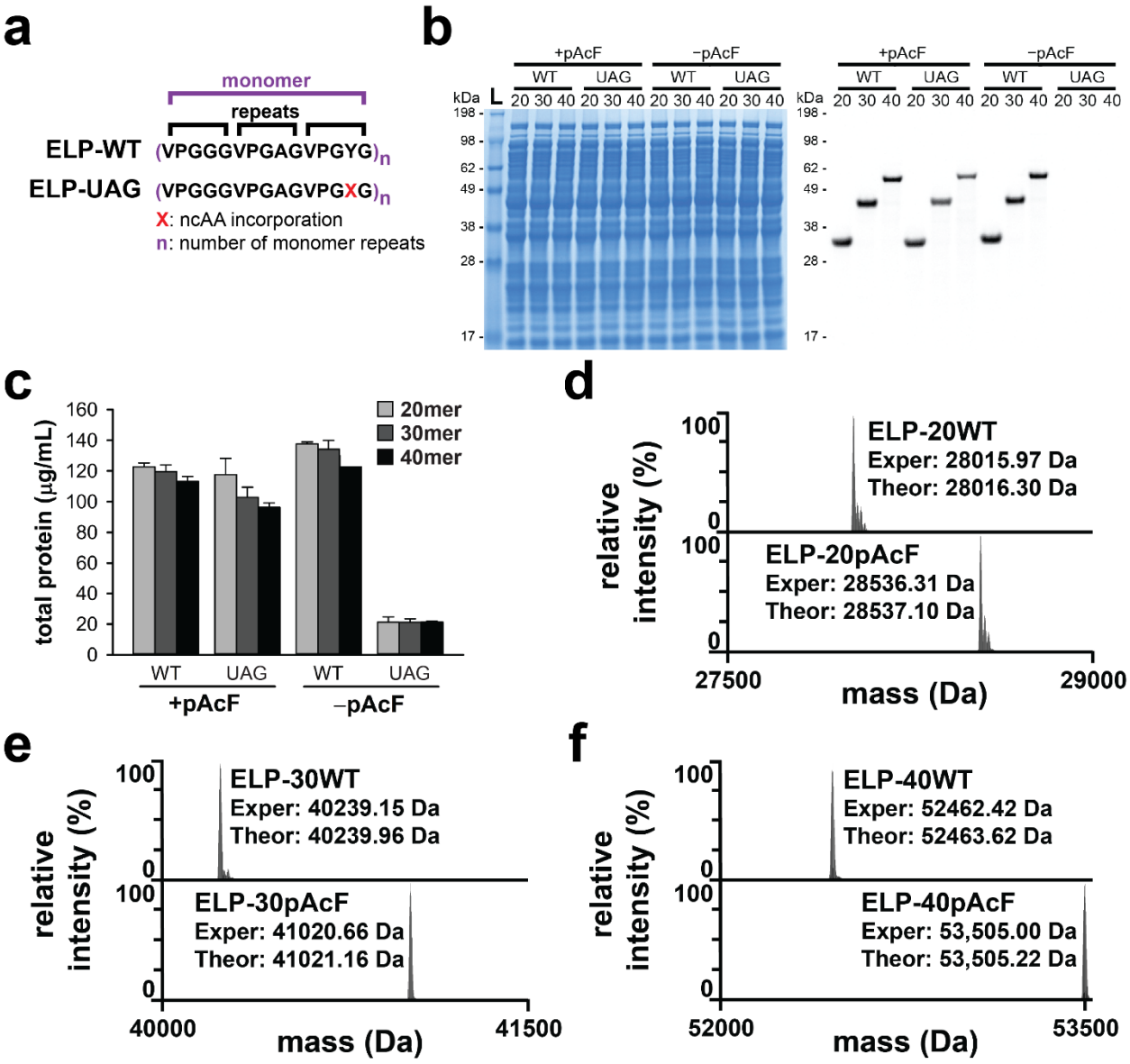


Figure 3.17. Incorporation of multiple identical non-canonical amino acids (ncAAs) into elastin-like polypeptides at high yield and purity using CFPS. (**a**) Schematic of the protein sequences for wild-type ELPs containing three pentapeptide repeats per monomer unit (ELP-WT) and ELPs containing 1 ncAA per monomer unit (ELP-UAG). (**b**) SDS-PAGE (left) and autoradiogram (right) analysis of cell-free produced ELP-WT and ELP-UAG 20, 30, and 40-mers in the presence and absence of pAcF. (**c**) Total protein yields of cell-free synthesized ELPs after incubation at 30 ˚C for 20 h are shown. Three independent batch CFPS reactions were performed for each sample, and one standard deviation is shown. Deconvoluted mass spectra of ELPs, obtained by top-down mass spectrometry illustrate complete, site-specific incorporation of pAcF at (**d**) 20, (**e**) 30, (**f**) and 40 sites. Deconvoluted average masses for the major peaks in each spectrum (Exper) match the theoretical average mass (Theor) for each species within 1.2 Da. Smaller peaks to the right of the major peaks arise from minor oxidation of the protein during electrospray ionization.

We next applied OTS^{opt} to the synthesis of ELP-UAGs with 20, 30 and 40-mers in the presence and absence of pAcF to demonstrate specificity of incorporation. ELP-UAGs were only synthesized in the presence of pAcF without any clear indication of truncation products, whereas no protein was observed in the absence of pAcF **(Figure 3.17)**. We anticipated that yields would decrease as the number of UAG codons increased due to the higher demand of pAcF-charged o-tRNA. In contrast, near wild type yields of ~100 mg/L were obtained for all UAG constructs. We then carefully examined and assayed the efficiency of multi-site ncAA incorporation using top-down liquid chromatography (LC)-MS of intact ELPs. LC-MS analysis showed more than ≥98% site-specific pAcF incorporation in ELP-UAG constructs 20, 30, and 40-mers **(Figure 3.17.D-F)**.

3.5 Discussion

We present a new crude extract-based CFPS platform base on the fully recoded *C321.*Δ*A* that is capable of high-level protein expression. This platform was generated using MAGE to create libraries of improved extract chassis strains by targeting the functional inactivation of multiple negative effectors. A combinatorial disruption of the genes *endA*-, *gor*-, *rne*-, and *mazF*-(*C321.*Δ*A.759*) increased total CFPS yields from 397 ± 24 mg/L in the parent strain to 1780 ± 30 mg/L of sfGFP, which is the highest reported protein yield from RF-1 deficient extracts. These improvements translated to the enhanced yields of proteins harboring site-specifically introduced ncAAs.

By optimizing the cell-free environment for multiple-identical ncAA incorporation, we were able to achieve multi-site ncAA incorporation into multiple model proteins with high yields (~99% wild type sfGFP expression yields for up to 2 ncAAs, ~95% wild type ELP expression

yields for up to 20 ncAAs, and ~85% wild type ELP expression yields for up to 30 and 40 ncAAs) and purity (≥98% accuracy of ncAA incorporation) due to the absence of RF1 in our system. To our knowledge, these are the purest polymers with this many site-specifically introduced ncAAs (*i.e.*, 40) synthesized to date. This exceeds our previous effort in cells that could synthesize ELP constructs with 30 UAG codons with 71% of the proteins having the desired 30 pAcF residues[27]. As such, our approach opens new opportunities to site-specifically modify the dominant physical and biophysical properties of biopolymers. This will allow researchers to go beyond tag-and-modify approaches that have historically been the focus of ncAA incorporation efforts because the field was previously limited to only one or a few instances of site-specific incorporation. Notably, our protein expression yields of up to ~1700 mg/L and 99% suppression efficiency for sfGFP with 2 ncAAs outperform the best expression of proteins with single or multiple ncAAs *in vivo* **(Figure 3.2.D and Figure 3.14)**. For example, previous *in vivo* experiments using an RF1 knockout strain demonstrated the synthesis of enhanced GFP containing one, two, and three pAcFs at 3.5, 3.5, and 5.4 mg/L, respectively, corresponding to amber suppression efficiencies of 23%, 23%, and 36%[229]. Also in a separate report, an *E. coli* strain with attenuated RF1 activity produced 21, 17, and 27 mg/L of sfGFP with 3, ELP with 12, and ELP with 22 ncAAs incorporated, respectively[397]. However, increasing the number of amber codons (*i.e.*, 12 and 22 ncAA incorporation) resulted in numerous truncation products. Here, we demonstrate a cell-free system from engineered recoded bacteria that enables high yields of proteins containing up to 40 ncAAs with no observable truncation products. Thus, our cell-free system will serve as a complement to *in vivo* methods and be useful technology for developing novel orthogonal translation systems for robust synthesis of modified proteins.

Looking forward, incorporating our discovered mutations into a recently published optimized strain, *C321.ΔA.opt* strain, might further increase the protein expression yields[398]. In addition, as new genomically recoded organisms with free codons are constructed[21, 228, 248, 399-401], the development of extract chassis strains enabled by our MAGE-guided approach could aid the generation of highly efficient CFPS systems with two or more distinct ncAAs within a single protein or sequence-defined polymer. We envision that the generalized CFPS platform described here will be applied to on-demand biomanufacturing and biomolecular prototyping to transform biochemical engineering and expand the range of genetically encoded chemistry of biological systems.

3.6 Acknowledgments

Special recognition and thanks are due to my excellent colleague Rey W. Martin (R.W.M.) who spearheaded this project and carried it through to completion. Thanks also to all of the remaining collaborators whose collective efforts made this project a reality: Yong-Chan Kwon (Y.C.K.), Jennifer Kay (J.E.K.), Roderick G. Davis (R.G.D.), Paul M. Thomas (P.M.T.), Natalia I. Majewska (N.I.M.), Cindy X. Chen (C.X.C.), Ryan D. Marcum (R.D.M.), Mary Grace Weiss (M.G.W.), Ashleigh E. Stoddart (A.E.S.), Miriam Amiram (M.A.), Arnaz K. Charna (A.K.C.), Jaymin R. Patel (J.R.P.), Farren J. Isaacs (F.J.I.), Neil L. Kelleher (N.L.K.), Seok Hoon Hong (S.H.H.), and Michael C. Jewett (M.C.J). R.W.M., S.H.H., M.C.J. and I conceived of this study and designed the experiments. R.W.M., R.D.M., M.G.W., A.S., S.H.H., and I performed MAGE genome engineering. R.W.M. and I prepared cell extracts and reagents. R.W.M., N.I.M., M.G.W., A.K.C., S.H.H., and I performed CFPS reactions and ncAA incorporation

experiments. R.G.D., P.M.T. and N.L.K. carried out the quantitative mass spectrometry analysis. Y.C.K. performed mRNA and DNA stability analysis. J.E.K. performed HPLC analysis. S.H.H., M.A. and R.W.M. constructed ELP plasmids, and R.W.M. and C.X.C. performed expression and purifications. J.R.P. carried out genome sequencing and analysis. F.J.I. supervised the construction of the recoded strain, a subset of the ELP constructs, and genome sequence analysis. R.W.M., S.H.H., and M.C.J. all contributed to editing this chapter.

3.7 Publication Information

This chapter was published with the following citation information:

Martin R.W., **Des Soye B.J.**, Kwon Y.C., Kay J., Davis R.G., Thomas P.M., Majewska N.I., Chen C.X., Marcum R.D., Weiss M.G., Stoddart A.E., Amiram M., Charna A.K., Patel J.R., Isaacs F.J., Kelleher N.L., Hong S.H., and Jewett M.C. Cell-free protein synthesis from genomically recoded bacteria enable multisite incorporation of noncanonical amino acids. *Nature Communications*. 2018 Mar 23;(9)1:1203. doi: 10.1038/s41467-018-03469-5.

CHAPTER 4

4 A Highly-productive, One-pot Cell-free Protein Synthesis Platform Based on Genomically Recoded *Escherichia coli*

4.1 Abstract

The site-specific incorporation of non-canonical amino acids (ncAAs) into proteins via amber suppression provides access to novel protein properties, structures, and functions. Historically, poor protein expression yields using amber suppression resulting from release factor 1 (RF1) competition has limited the technology. Here, we address this limitation by developing a high yielding, one-pot cell-free platform for synthesizing proteins bearing ncAAs based on genomically-recoded *Escherichia coli* lacking RF1. A key feature of this platform is the independence on the addition of purified T7 DNA-directed RNA polymerase (T7RNAP) to catalyze transcription. This platform was developed by leveraging λ-Red mediated homologous recombination to genomically integrate T7RNAP into the genome, optimizing the integration location, and exploiting multiplex genome engineering to enhance extract performance by installing mutations into the T7RNAP gene that render the polymerase resistant to proteolysis during cell lysis. Extracts derived from our final strain demonstrate superior productivity *in vitro*, synthesizing 2.15 ± 0.17 g/L superfolder green fluorescent protein (sfGFP) in batch mode. Using an optimized platform we demonstrated high multi-site incorporation efficiency of the ncAA p-acetyl-L-phenylalanine (pAcF) into an elastin-like polypeptide, yielding 50 ± 7 mg/L of a

polypeptide featuring 40 identical ncAA incorporations. Our work has implications for CFPS strain development via genome editing, genetic code reprogramming, and cell-free synthetic biology.

4.2 Introduction

A burst of recent development has transformed cell-free protein synthesis (CFPS) from a niche tool for molecular biology into a new technology platform for manufacturing proteins at scale and for accelerating biological design[280, 281, 402-404]. This rapid advancement has been spurred by the desire to take advantage of the beneficial features unique to CFPS systems, which include easy system access and manipulation, the elimination of competition with cellular growth and adaptation objectives, and a dilute reaction environment that can facilitate folding of complex eukaryotic protein products[284, 287]. Batch CFPS reactions now persist for up to a day with yields exceeding 1.5 g/L[284, 405], and improvements in scalability culminated recently with the successful completion of a 100 L reaction[264]. These impressive advances can be largely attributed to extensive efforts to engineer CFPS systems via chassis organism development, usually by the targeted genetic deletion of genes whose products are known to destabilize key biological substrates (*e.g.* DNA, mRNA, amino acids, and energy) in cell-free reactions [232, 288, 309, 405]. As a result of these transformative efforts, CFPS platforms can now be used to complement protein overexpression *in vivo*, with particular utility in rapid prototyping[256-259], synthesis of toxic products[260-263], the production of proteins that are difficult to solubly express *in vivo*[264-267], manufacture of glycoproteins[30, 147, 268, 269], detection of disease[30, 271, 272], on demand biomanufacturing[30, 265, 273-275], and education[277, 278].

One particularly appealing application of cell-free protein synthesis is the production of proteins containing non-canonical amino acids (ncAAs)[23, 24, 27, 232, 394, 405-408]. To date, more than 150 different ncAAs have been incorporated into (poly)peptides[202], enabling the synthesis of proteins featuring novel structures and functions that would otherwise be difficult or even impossible to obtain using only the twenty canonical amino acids. Typically, site-specific ncAA incorporation is enabled by amber suppression, whereby the amber stop codon (UAG) is recoded as a sense codon designating a ncAA of interest[338]. This process is mediated by orthogonal translation systems (OTSs), which generally consist of the ncAA, an orthogonal suppressor tRNA (o-tRNA) that has been modified to associate with UAG in the ribosomal A-site, and a ncAA-specific aminoacyl-tRNA synthetase (ncAA-RS) that has been evolved to covalently load the ncAA onto the o-tRNA[211, 394], without recognizing natural amino acids. The cytotoxicity of many OTSs [22, 405], membrane impermeability of some ncAAs[19], and the ability to overcome the relatively poor incorporation efficiencies of OTSs via direct supplementation with OTS components[23, 232, 394, 405] makes CFPS an attractive method for the synthesis of peptides featuring ncAAs.

Unfortunately, efforts to apply amber suppression to the incorporation of ncAAs have long been limited by competition with release factor 1 (RF1), which is responsible for terminating translation in response to the ribosome encountering a UAG codon[15]. In attempting amber suppression, functional RF1 can outcompete ncAA-bearing o-tRNAs at UAG codons leading to the production of errant truncated products[23, 24]. Historically, this competition has led to poor protein expression yields, which limits applications in both basic and applied science. Recently, we addressed this limitation. Specifically, we developed a CFPS derived from a genomically-

recoded strain of *Escherichia coli* in which all native instances of the amber codon were changed to the synonymous ochre codon (UAA) followed by elimination of RF1 from the genome (C321.Δ*A*)[21, 405]. Extracts derived from the resulting strain (C321.Δ*A*.759) resulted in CFPS yields ~1,700 mg/L and 99% suppression efficiency for superfolder green fluorescent protein (sfGFP) with 2 ncAAs, which outperform the best expression of proteins with single or multiple ncAAs *in vivo*[405]. While this strain demonstrates high productivity and a new standard for ncAA incorporation *in vitro*, it is limited by its dependence on the addition of purified viral T7 DNA-directed RNA polymerase (T7RNAP) to catalyze transcription. This adds another step to reaction assembly and increasing the cost of the system by requiring the addition of purified polymerase to catalyze robust transcription. In principle, one could create a one-pot CFPS system if T7RNAP could be integrated into the genome, and overexpressed in the source strain prior to lysis. Indeed, extracts derived from T7RNAP-expressing strains (most notably BL21 (DE3)[409] and its derivatives) are innately enriched in polymerase activity and generally do not require (or even benefit from) supplementation[291]. These one-pot CFPS systems, containing all of the biological components necessary to support transcription and translation, are highly attractive due to their convenient, plug-and-play nature.

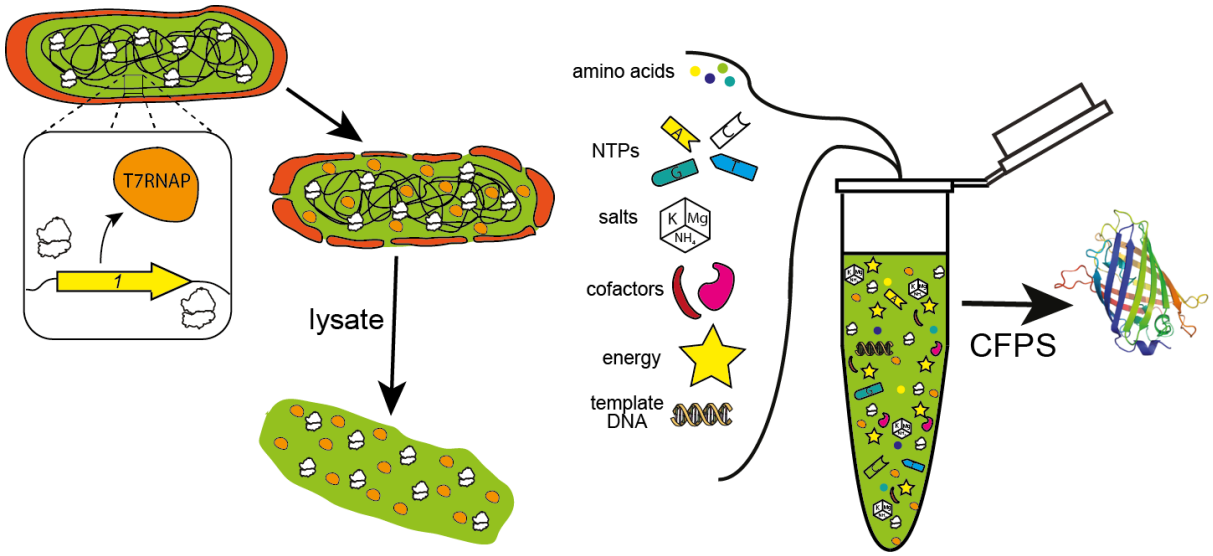


Figure 4.1 Simplified schematic of the production and utilization of crude lysates from *E. coli* cells to catalyze cell-free protein synthesis (CFPS). Reactions are supplemented with enzymatic cofactors, energy, and other substrates required for protein synthesis as well as plasmid DNA template directing the system towards the production of a product of interest. The strain illustrated is shown endogenously expressing T7RNAP to enable orthogonal transcription *in vitro* to generate a one-pot system independent of supplementation with purified protein components.

In this study, we developed a high-yielding one-pot CFPS platform for ncAA incorporation into proteins derived from a genomically-recoded RF1-deficient strain of *E. coli* that has been optimized for productivity in CFPS (C321.Δ*A*.759)[405] **(Figure 4.1)**. Since C321.Δ*A*.759 does not express T7RNAP, we applied λ-Red mediated homologous recombination[410, 411] (λHR) to genomically integrate a series of synthetic constructs featuring the T7RNAP-encoding *1* gene[409] and then assessed the ability of extracts derived from the resulting transformants to catalyze CFPS in the absence of exogenous polymerase supplementation. While native bacterial RNA polymerases and associated sigma factors can be used to catalyze transcription in CFPS reactions[315, 316], we chose to pursue T7RNAP because of its extremely high productivity, orthogonality, and strong sequence preference[315, 409]. Two different genomic loci were targeted for integration, with *1* placed under the regulation of three promoters of different

strengths. A high-performing strain, C321.Δ*A*.759.T7 was capable of synthesizing ~1.4 g/L of superfolder green fluorescent protein (sfGFP) without purified T7RNAP supplementation. We next exploited multiplex automated genome engineering (MAGE)[25] to install mutations in the *1* gene of C321.Δ*A*.759.T7 to make it resistant to proteolytic cleavage during lysate preparation. The resulting strain C321.Δ*A*.759.T7.D yielded ~1.6 g/L sfGFP without T7RNAP supplementation and ~2.2 g/L with. Using an optimized system, we were able to synthesize proteins (elastin-like polypeptide, ELP) bearing up to 40 ncAAs with yields up to ~30 mg/L in the absence of supplemental T7RNAP. When compared to BL21 (DE3) and its derivative strains, one-pot CFPS systems derived from C321.Δ*A*.759.T7.D are highly productive and superior for applications involving ncAAs.

4.3 Materials and Methods

4.3.1 Strains and Plasmids

The bacterial strains and plasmids used in this study are listed in **Table 4.2**. Carbenicillin (50 µg/mL) was used for culturing C321.Δ*A*.759, kanamycin (50 µg/mL) was used for culturing C321.Δ*A*.759 T7RNAP linear insert transformants and maintaining pY71/pJL1-based plasmids, and chloramphenicol (34 µg/mL) was used to maintain the pEVOL-pAcF plasmid.

4.3.2 DNA Gel Electrophoresis

Unless otherwise stated, all DNA electrophoresis was done in 1% agarose gels stained with SYBR® Safe (Thermo Fisher Scientific, Inc., Waltham, MA). Samples were run at 100V for 30-60 minutes on a Mini Gel II Complete Electrophoresis System (VWR, Radnor, PA). 100 bp and 1

kb Quick-load® DNA Ladders (New England Biolabs, Ipswich, MA) were used for fragment size reference.

4.3.3 T7RNAP Linear Insert Construction

The six T7RNAP-encoding inserts used in this study were assembled from PCR products obtained using primers listed in **Table 4.1**. The insert loci were at coordinates 3,986,255 (*asl*) and 805,473 (*int*) in the genome of C321.Δ*A*.759. In brief, each insert was assembled from four segments of linear DNA – a promoter segment featuring 50 base pairs of sequence homology to the genome of C321.Δ*A*.759 upstream of the targeted insert site, a T7RNAP segment containing the *1* gene, a terminator segment encoding the synthetic terminator sequence L3S2P21[412], and a kanR segment featuring the kanamycin resistance cassette from pKD4[410] as well as 50 bp of sequence homology to the genome of C321.Δ*A*.759 downstream of the targeted insert site. All constructs were designed such that coding strand would be integrated into the leading strand during genome replication. Adjacent segments featured at least 20 bp of sequence homology to one another to facilitate their assembly into a single unit of DNA, and novel sequence elements (*e.g.* 6His-tag and synthetic RBS sequences) were built into the 5' tails of primers. Inserts featuring the promoter Lpp5 were assembled via overlap assembly PCR (SOEing)[413] and amplified with end primers. End primer PCR reactions for these species generated a large number of off-target sequences, so full-length insert DNA was separated from other products via electrophoresis and extracted from 1% agarose gel using a DNA gel extraction kit (Product No. D2500; Omega Bio-tek, Norcross, GA). Inserts featuring promoters PtacI and lacUV5 were assembled together with plasmid origins of replication (p15a and pUC, respectively) via Gibson assembly to yield plasmid DNA, and these plasmids were used as template with end primers to yield the linear insert DNA.

We found that the PCR products generated using this approach had a significant reduction in the prevalence of off products observed for the other inserts, accelerating our workflow.

4.3.4 Strain Transformation and Insert Verification

The T7RNAP cassettes were inserted into the genome of C321.Δ*A*.759 via λHR following the protocol of Datsenko and Wanner[410]. In brief, 5 mL cultures of C321.Δ*A*.759 were grown in LB media (10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl)[351] to an OD_{600} of 0.6. 1.5 mL of this culture was washed twice in ice cold, sterile nuclease-free water and resuspended in 30 µL of insert DNA at a concentration of 70 ng/µL. The cell suspension was transferred to a 2 mL electroporation cuvette and DNA was introduced into cells using a Micropulse electroporator (Bio-Rad, Hercules, CA). Immediately following electroporation, cells were resuspended in 1 mL sterile LB media and recovered for 3 hrs at 34°C at 250 rpm. The recovered cell culture was plated on kanamycin selective plates and permitted to grow overnight at 34°C. The following day, colonies to be screened were picked and inoculated into 100 µL of kanamycin media on a 96-well plate (Costar 3370; Corning, Corning, NY) and cultured for 3 hrs at 34°C at 250 rpm. 1 µL of each miniature culture was used as template to detect successful genomic integration of each insert by colony PCR using primers listed in **Table 4.1**. Two primer pairs were used for detection such that if no insert was present, the outermost pair of primers would anneal to the flanking genomic sequence and generate a single ~500 bp product; however, if the insert was present intact at the locus both pairs of detection primers would be able to anneal and generate two products of ~1250 and ~1750 bp (Fig 2C).

4.3.5 Cell Extract Preparation

For rapid prototyping of engineered strains, cells were grown in 1 L of 2xYTPG media (pH 7.2) in a 2.5 L Tunair® shake flask and incubated at 34˚C at 220 rpm. Unless otherwise stated, cultures were inoculated with 1 mM IPTG at an OD_{600} of 0.6 and permitted to continue to grow to an OD_{600} of 3.0. Cells were pelleted by centrifuging for 15 min at 5000 × g at 4°C, washed three times with cold S30 buffer (10 mM tris-acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, 2 mM dithiothreitol)[354], and stored at -80°C. To make cell extract, cell pellets were thawed and suspended in 0.8 mL of S30 buffer per gram of wet cell mass and 1.4 mL of cell slurry was transferred into 1.5 mL microtubes. The cells were lysed using a Q125 Sonicator (Qsonica, Newtown, CT) with 3.175 mm diameter probe at a 20 kHz frequency and 50 % amplitude for three cycles of 45s ON/59s OFF. To minimize heat damage during sonication, samples were placed in an ice-water bath. For each 1.4 mL sample, the input energy was ~844 Joules and was monitored during sonication. Extract was then centrifuged at 12,000 × g at 4˚C for 10 min. For strain derivatives of C321.Δ*A*.759, a run-off reaction (37˚C at 250 rpm for 1 h) and second centrifugation (10,000 × g at 4˚C for 10 min) were performed[291]. The supernatant was flash-frozen using liquid nitrogen and stored at -80˚C until use.

4.3.6 CFPS Reaction

A modified PANOx-SP system was utilized for CFPS reactions testing incorporation of pAcF[287],[414]. Briefly, a 15 µL CFPS reaction in a 2.0 mL microtube was prepared by mixing the following components: 1.2 mM ATP; 0.85 mM each of GTP, UTP, and CTP; 34 µg/mL folinic acid; 170 µg/mL of *E. coli* tRNA mixture; 13.3 µg/mL plasmid; 16 µg/mL T7 RNA polymerase; 2 mM for each of the 20 standard amino acids; 0.33 mM nicotinamide adenine dinucleotide (NAD); 0.27 mM coenzyme-A (CoA); 1.5 mM spermidine; 1 mM putrescine; 4 mM sodium

oxalate; 130 mM potassium glutamate; 10 mM ammonium glutamate; 12 mM magnesium glutamate; 57 mM HEPES, pH 7.2; 33 mM phosphoenolpyruvate (PEP), and 27% v/v of cell extract. For ncAA incorporation, 2 mM pAcF, 1.0 mg/mL pAcFRS, and 10 µg/mL of o-tz-tRNA linear DNA were supplemented to cell-free reactions. o-tRNA linear DNA was amplified from pY71-T7-tz-o-tRNA plasmid as described previously[24] and transcribed during the cell-free reaction. Furthermore, the o-tRNA was expressed in the source strain prior to extract preparation. Each CFPS reaction was incubated for 20 h at 30˚C unless noted otherwise. *E. coli* total tRNA mixture (from strain MRE600) and phosphoenolpyruvate was purchased from Roche Applied Science (Indianapolis, IN). ATP, GTP, CTP, UTP, 20 amino acids and other materials were purchased from Sigma (St. Louis, MO) without further purification. T7RNAP and pAcFRS was purified in house as described previously[405].

4.3.7 Quantification of Active sfGFP

CFPS reactions were diluted 1:25 in nanopure water and active full-length sfGFP protein yields were quantified by measuring fluorescence using a Synergy 2 plate reader (BioTek, Winooski, VT) with excitation at 485 nm, emission at 528 nm, and cut-off at 510 nm in 96-well half area black plates (Costar 3694; Corning, Corning, NY). sfGFP fluorescence units were converted to concentration using a standard curve established with ^{14}C-Leucine quantified sfGFP as described previously[24].

4.3.8 Quantification of Active mRFP

CFPS reactions were diluted 1:25 in nanopure water and active mRFP yields were quantified by measuring fluorescence using a Synergy 2 plate reader (Biotek, Winooski, VT) with excitation

at 503 nm and emission at 607 nm in 96-well half area black plates (Costar 3694; Corning, Corning, NY).

4.3.9 Detection of His-tagged T7RNAP by Western Blot

To visualize polymerase overexpression *in vivo*, cell samples were collected during harvest. Pre-induction cell samples were derived from 1 mL of culture at OD_{600} of 0.6, harvest samples were derived from 200 µL of culture at OD_{600} of 3.0. To prepare samples for gel electrophoresis, cells were pelleted and resuspended in 200 µL of nuclease-free water. 100 µL of this suspension was mixed with 34 µL of 4x NuPAGE® LDS Sample Buffer (Thermo Fisher Scientific, Inc., Waltham, MA) and boiled for 10 minutes. Following the boil, samples were spun at >13,500 x *g*. Samples derived from lysates were prepared by diluting 1 µL of extract in 8 µL of nuclease-free water and boiling for 10 minutes with 3 µL of 4x NuPAGE® LDS Sample Buffer. 12 µL of each samples was loaded into 12% Bis-Tris NuPAGE® gel (Thermo Fisher Scientific, Inc., Waltham, MA) and run at 130 V for 90 min using 1X MOPS running buffer (diluted from 20X MOPS SDS Running Buffer, Thermo Fisher Scientific, Inc., Waltham, MA). For reference, SeeBlue® Plus2 Pre-Stained Protein Standard (Thermo Fisher Scientific, Inc., Waltham, MA) was loaded into wells flanking the samples. Following electrophoresis, gels were washed in nanopure water. Proteins were transferred to Immun-Blot® PVDF membrane (Bio-rad, Hercules, CA) using a semi-dry protocol in 20% methanol/80% 1x MOPS. Transfer proceeded at 80 mA per gel for 55 min using a Trans-Blot® SD Semi-Dry Transfer Cell (Bio-rad, Hercules, CA). Blots were blocked overnight in 5% (m/v) fat-free dry milk at 4°C. Primary antibody (Sigma, Cat. #H1029, St. Louis, MO) was diluted 10,000x in PBS and applied to blots for 2 hrs. Secondary antibody (Bio-rad, Cat. #1708237, Hercules, CA) was diluted 3,000x in PBS-T and applied to blots for 1 hr. Finally, His-

tagged proteins were visualized using the Immun-Blot® Opti-4CN™ Colorimetric kit (Bio-rad, Hercules, CA).

4.3.10 Knockout of *ompT* Locus

In order to use kanamycin resistance to select for successful knockout of the *ompT* locus in C321.Δ*A*.759.T7, the kanR cassette first employed to select for integration of the T7RNAP insert needed to be removed from the genome. This DNA was physically looped out of the genome using the oligo listed in **Table 4.1** for MAGE. Cultures were grown in LB media at 32°C and 250 rpm throughout 8 MAGE cycling steps as previously described[405]. Replica plating[415] was used to identify colonies that regained sensitivity to kanamycin, and colony PCR using the protocol described above confirmed that the kanamycin resistance cassette DNA was no longer present in the genome. The kanamycin resistant cassette from pKD4 was then amplified with primers containing up- and downstream homology to the genomic region targeted for deletion in their 5' tails. The knockout construct was given flanking homology such that coordinates 580,650-592,260 in the genome of C321.Δ*A*.759.T7 would be replaced by the resistance cassette. ΛHR followed by colony PCR detection of the knockout were performed as described above to yield C321.Δ*A*.759.T7.Δ*ompT*.

4.3.11 Generation and Verification of OmpT-resistant T7RNAP-expressing Strains

Nucleotide changes designed to introduce mutations of K183 to glycine/leucine and K190 to alanine were installed into C321.Δ*A*.759.T7's genomic copy of the N-terminally 6His-tagged *1* gene via MAGE using the oligos listed in **Table 4.1.** Cultures were grown in selective LB media

at 32°C and 250 rpm throughout 6 MAGE cycling steps as previously described[405]. Putative mutant colonies were picked and cultured as described above prior to screening for the desired mutation. Multiplex allele-specific colony (MASC) PCR was performed to verify mutations[351] using wild-type forward or mutant forward primers and reverse primers **(Table 4.1)**. Wild-type and mutant forward primers were identical except at the 3'-ends of the oligonucleotides which featured allele-specific sequence such that stable annealing of the end of the primer should only be possible when paired with the corresponding genomic allele. In this way the mutant allele could be amplified using the mutant forward and reverse primer set but not amplified by the wild-type forward and reverse primer set, and vice versa. The reverse primers were used for detection of both wild-type and mutant alleles.

4.3.12 sfGFP Timecourse

CFPS reactions were assembled in the wells of a half-area black 96-well plate (Costar 3694; Corning, Corning, NY) as described above. To minimize the effects of evaporation, the volumes of all components were doubled to scale total reaction volume up to 30 µL. To maintain humidity and preserve reaction volume throughout the CFPS reaction, all unused wells were filled with 100 µL nanopure water and the plate was covered with an evaporation lid and finally sealed with Parafilm M® (Bemis, Neenah, WI). Reactions were run in a Synergy H1 plate reader (BioTek, Winooski, VT) at 30°C, with fluorescence measurements of each experimental well taken every 5 minutes for 20 hrs. Measurements were taken with excitation at 485 nm, emission at 528 nm, and cut-off at 510 nm.

4.3.13 Autoradiogram Analysis

For autoradiogram analysis, samples were prepared as described above from 4 µL of each reaction and loaded on 12% Bis-Tris NuPAGE® gel (Thermo Fisher Scientific, Inc., Waltham, MA). The gel was soaked in Gel Drying solution (Bio-Rad, Hercules, CA) for 30 min, fixed with cellophane films, dried overnight in GelAir Dryer (Bio-Rad, Hercules, CA), and exposed for 3 days on Storage Phosphor Screen (GE Healthcare Biosciences, Pittsburgh, PA). Autoradiograms were scanned using Typhoon FLA 7000 Imager (GE Healthcare Biosciences, Pittsburgh, PA).

4.3.14 ELP Radioactive Quantitation

Radioactive ^{14}C-Glycine was added into 15 µL CFPS reactions. After incubation, yields were quantified by determining radioactive ^{14}C-Gly incorporation into trichloroacetic acid (TCA) -precipitated protein[354]. Radioactivity of TCA-precipitated samples was measured using liquid scintillation counting (MicroBeta2, PerkinElmer, Waltham, MA).

4.3.15 Locus DNA Sequencing

To sequence the genomic T7RNAP inserts, the entire region was PCR amplified using end primers listed in **Table 4.1**. Amplified linear insert DNA was submitted to the NUSeq Core facility along with forward primers spaced ~700 bp apart, and the sequence for each region was determined using traditional Sanger sequencing.

Table 4.1. Primers used for insert component amplification and assembly, *ompT* knockout, MAGE, colony PCR, T7 plasmid cloning, and DNA sequencing. Underlined bold text indicates location of mismatches. The first four bases of the 5'-MAGE oligonucleotides were phosphorothioated (*).

Primer Name	DNA Sequence (listed 5' to 3')
T7 insert assembly	
lacUV5_asl_F	TGTAGGCTGGATAAGATGCGTCAGCATCGCATCCGGCAAAGGCAGA TCTCGCTTCCGGCTCGTATAATGTGT

PtacI_asl_F	TGTAGGCTGGATAAGATGCGTCAGCATCGCATCCGGCAAAGGCAGA TCTCGAGCTGTTGACAATTAATCATCG
Lpp5_asl_F	TGTAGGCTGGATAAGATGCGTCAGCATCGCATCCGGCAAAGGCAGA TCTCATCAAAAAAATATTGACAAC
asl_homology_R	AATATCCACCACGCGCGCAGATTAAATCTGACTAAGCCGGCGCTATC GCTGGTGGAATCGAAATCTCGTGATGG
asl_F	TGTAGGCTGGATAAGATGC
asl_R	AATATCCACCACGCGCGCAG
lacUV5_int_F	TGCTTCTCATAGAGTCTTGCAGACAAACTGCGCAACTCGTGAAAGGT AGGGCTTCCGGCTCGTATAATGTGT
PtacI_int_F	TGCTTCTCATAGAGTCTTGCAGACAAACTGCGCAACTCGTGAAAGGT AGGGAGCTGTTGACAATTAATCATCG
Lpp5_int_F	TGCTTCTCATAGAGTCTTGCAGACAAACTGCGCAACTCGTGAAAGGT AGGATCAAAAAAATATTGACAAC
int_homology_R	ATTTTATGCGCGCACGAAAAGCATCAGGTCTTTCCTTCGAAGGGGAT CCGGGTGGAATCGAAATCTCGTGATGG
int_F	TGCTTCTCATAGAGTCTTGC
int_R	ATTTTATGCGCGCACGAAAAG
lacUV5_promRBS_R	CGATCCTCTCATTTTGTACCTCCTTAGTTGCTTGCAATTGTTATCCGC TCACAATTCC
lacUV5_RBST7_F	CAAAATGAGAGGATCGCATCACCATCACCATCACGGATCCAACACG ATTAACATCGCTAA
Lpp5_promRBS_R	GATCCTCTCATTATGTACCTCCTTACTGTTTTGTTTTAATTGTTATCCG CTCACAATTCC
Lpp5_RBST7_F	CATAATGAGAGGATCGCATCACCATCACCATCACGGATCCAACACG ATTAACATCGCTAA
PtacI_promRBS_R	ATCCTCTCATATATTACCTCCTTAGTAGCGCTGTGTGTAATTGTTATC CGCTCACAATTC
PtacI_RBST7_F	ATATATGAGAGGATCGCATCACCATCACCATCACGGATCCAACACG ATTAACATCGCTAA
T7_synterm_R	CCTGTATCAGGCTGAAAATCTTACGCGAACGCGAAGTCCGACTC
synterm_F	GGACTTCGCGTTCGCGTAAGATTTTCAGCCTGATACAGG
Synterm_kanR_R	CTTTCTACGTGTTCCGCTTATAAAGTGTAAAGCCTGG
kanR_F	AAGCGGAACACGTAGAAAG
Synthetic_term (L3S2P21)	GATTTTCAGCCTGATACAGGATTTTCAGCCTGATACAGCTCGGTACC AAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTTTCGTTTT GGTCCCCTTTTTGCGTTTCTACACCCAGGCTTTACACTTTAT
***ompT* knockout**	
delompT_ F	CGACTACATCCGTGAGGTGAATGTGGTGAAGTCTGCCCGTGTCGGTT ATTGAAGCGGAACACGTAGAAAG
delompT_ R	TAATGGTAAAAAGCTGTCACAATTCATAAAAAACCTTAATATACGCC ACCGGTGGAATCGAAATCTCGTGATG
MAGE	
MAGE_T7_K172L	T*C*G*A*CAACTTGCATAAATGCTTTCTTGTAGACGTGCCCTACGCG **CAG**GTTGAGTTGTTCCTCAACGTTTTTCTTGAAGTGCTTAGCTTCA

MAGE_T7_K172G	T*C*G*A*CAACTTGCATAAATGCTTTCTTGTAGACGTGCCCTACGCG **GCC**GTTGAGTTGTTCCTCAACGTTTTTCTTGAAGTGCTTAGCTTCA
MAGE_T7_double	C*A*G*C*CTCGACAACTTGCATAAATGCTTT**CGC**GTAGACGTGCCCT ACGCG**GCC**GTTGAGTTGTTCCTCAACGTTTTTCTTGAAGTGCTTAG
delkanR_oligo	G*G*T*T*GGGCGTCGCTTGGTCGGTCATTTCGAACCCCAGAGTCCCG CCATGCGAAACGATCCTCATCCTGTCTCTTGATCAGATCTTGATCC
Colony PCR	
K172L_cPCRF_mut	CGTTGAGGAACAACTCAACCTG
K172G_cPCRF_mut	CGTTGAGGAACAACTCAACGG
K172_cPCRF_wt	CGTTGAGGAACAACTCAACAA
172mut_cPCR_R	TTGTTGATTTTCCATGCGGTG
K179A_cPCRF_mut	CGCGTAGGGCACGTCTACGC
K179_cPCRF_wt	CGCGTAGGGCACGTCTACAA
179mut_cPCR_R	TTGGCGACCGCTAGGACTTTC
pUC_T7_cPCRF	GCGATAAGTCGTGTCTTACC
pUC_T7_cPCRR	CATGTAAACGTCTTCGTAGC
delompT_cPCR1_F	GGGACTATTGAGTACGAACG
delompT_cPCR2_R	CGAATCTCATAACGCAAACC
T7 plasmid cloning	
pUC_T7asl_F	CTGCGCGCGTGGTGGATATTGCATGCATCTCCTCAGATTGATTTAAA
pUC_T7asl_R	CGCATCTTATCCAGCCTACAGCATGCATCTCCTCGCTCACTGACTCG CTG
pUC_T7int_F	TTTTCGTGCGCGCATAAAATGCATGCATCTCCTCAGATTGATTTAAA ACTTCAT
pUC_T7int_R	GCAAGACTCTATGAGAAGCAGCATGCATCTCCTCGCTCACTGACTCG CTG
p15a_T7asl_F	CTGCGCGCGTGGTGGATATTGCATGCATCTCCTCTCCCTTAACGTGA GTTTTCG
p15a_T7asl_R	CGCATCTTATCCAGCCTACAGCATGCATCTCCTCTGAGTCAGCAACA CC
DNA Sequencing	
T7_asl_seq1_F	GGCCGCCTGCGGTTGATTGC
T7_asl_seq_R	GTGTGGACCAGACATCCTTC
T7_int_seq1_F	TTCAATTTTGTCCCACTCCCTGC
T7_int_seq_R	ACGAATACCTGAAAATTTATCAAGCAGC
T7_seq2_F	AGCCGGAAGCCGTAGCGTAC
T7_seq3_F	ACGTTTACATGCCTGAGGTG
T7_seq4_F	GCTTCCTTGCGTTCTGCTTTG
T7_seq5_F	TCAAAGATAAGAAGACTGGAG
T7_seq6_F	GGGGCCTTTTTTCGTTTTGG
T7_seq7_F	GCAGCTGTGCTCGACGTTGTC

Table 4.2. Strains and plasmids used in this study. Km^R, Ap^R, and Cm^R are kanamycin, ampicillin, and chloramphenicol resistance, respectively. 'Δ' indicates deleted gene, and '∇' indicates gene(s) inserted into the genome at the specified locus behind the specified promoter. '//' denote amino acid substitution mutations made in the *1* gene open reading frame.

Strains and plasmids	Genotype/relevant characteristics	Source
Strains		
BL21 Star™ (DE3)	F^- *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm rne131* (DE3)	Life Technologies
C321.ΔA.759	*C321.ΔA. endA- gor- rne- mazF-*	[405]
DH5α	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ* ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ^-	Invitrogen
759.T7.int.lacUV5	*C321.ΔA.759.∇1.int.lacUV5,* Km^R	This study
759.T7.int.PtacI	*C321.ΔA.759.∇1.int.PtacI,* Km^R	This study
759.T7.int.Lpp5	*C321.ΔA.759.∇1.int.Lpp5,* Km^R	This study
759.T7.asl.lacUV5	*C321.ΔA.759.∇1.asl.lacUV5,* Km^R	This study
759.T7.asl.PtacI	*C321.ΔA.759.∇1.asl.PtacI,* Km^R	This study
759.T7.asl.Lpp5	*C321.ΔA.759.∇1.asl.Lpp5,* Km^R	This study
759.T7.ΔkanR	*C321.ΔA.759.∇1.asl.Lpp5.ΔkanR*	This study
759.T7.ΔompT	*C321.ΔA.759.∇1.asl.Lpp5.ΔompT,* Km^R	This study
759.T7.K172L	*C321.ΔA.759.∇1.asl.Lpp5/K172L/,* Km^R	This study
759.T7.K172G	*C321.ΔA.759.∇1.asl.Lpp5/K172G/,* Km^R	This study
759.T7.D	*C321.ΔA.759.∇1.asl.Lpp5/K172G, K179A/,* Km^R	This study
Plasmids		
pKD4	Km^R	[410]
pAR1219	Ap^R	[416]
pDPtacIAcRSTT1	Km^R	[417]
pDTT1-Lpp5-EF-Tu	Km^R	[418, 419]
pY71-sfGFP	Km^R, P_{T7}::super folder green fluorescent protein (sfGFP), C-terminal strep-tag	[19]
pY71-sfGFP-T216amb	pY71-sfGFP with amber codon at T216	[19]
pY71-sfGFP-2amb	pY71-sfGFP with amber codon at N212 and T216	[24]
pY71-sfGFP-5amb	pY71-sfGFP with amber codon at D36, K101, E132, D190, and E213	[24]
pY71-pAcFRS	P_{T7}::pAcFRS, C-terminal 6x histidine tag	[24]
pEVOL-pAcF	Cm^R, P_{glnS}::*pAcFRS*, P_{araBAD}::*pAcFRS*, P_{proK}::o-tRNA	[14]
pY71-T7-tz-o-tRNA	P_{T7}:: hammer-head ribozyme (tz), o-$tRNA^{opt}$ (o-tz-	[24]

pY71-mRFP1	P_{T7}::mRFP1	[355]
pY71-Spinach	P_{T7}::Spinach aptamer	This study
pUC-T7-*int.lacUV5*	Km^R, P_{lacUV5}::T7 RNAP w/*int* locus flanking	This study
pUC-T7-*asl.lacUV5*	Km^R, P_{lacUV5}::T7 RNAP w/*asl* locus flanking	This study
pUC-T7-*asl.PtacI*	Km^R, P_{PtacI}::T7 RNAP w/*asl* locus flanking	This study
pUC-T7-*asl.Lpp5*	Km^R, P_{Lpp5}::T7 RNAP w/*asl* locus flanking	This study
P15a-T7-*asl.Lpp5*	Km^R, P_{Lpp5}::T7 RNAP w/*asl* locus flanking	This study
P15a-T7-*asl.PtacI*	Km^R, P_{PtacI}::T7 RNAP w/*asl* locus flanking	This study
P15a-T7-*int.Lpp5*	Km^R, P_{Lpp5}::T7 RNAP w/*int* locus flanking homology	This study
pY71-FI-ELP20	FI-ELP-20mer	[405]
pY71-FI-ELP30	FI-ELP-30mer	[405]
pY71-FI-ELP40	FI-ELP-40mer	[405]
pY71-FI-ELP20X	FI-ELP-20mer with 20 amber sites	[405]
pY71-FI-ELP30X	FI-ELP-30mer with 30 amber sites	[405]
pY71-FI-ELP40X	FI-ELP-40mer with 40 amber sites	[405]

4.4 Results

4.4.1 CFPS Activity of C321.Δ*A*.759 and BL21 Star™ (DE3) With and Without Supplemental T7RNAP

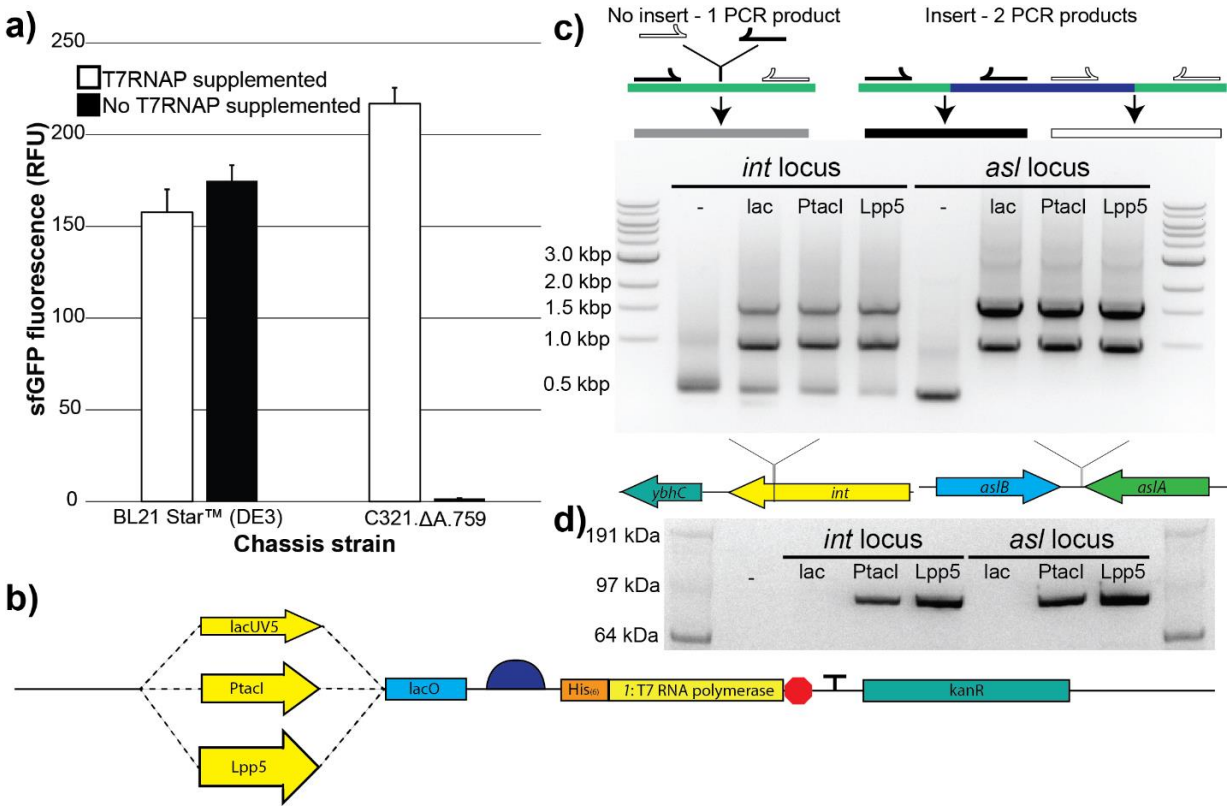


Figure 4.2. Engineering a genomically recoded *Escherchia coli* strain for T7RNAP overexpression. **(a)** sfGFP fluorescence *in vitro* from cell extracts derived from induced BL21 Star™ (DE3) cells as well as C321.Δ*A*.759 cells, both with and without supplementation with purified T7 RNAP. **(b)** Schematic of the synthetic genomic insert used in this study to introduce the gene encoding the T7RNAP into the genome of C321.Δ*A*.759. **(c)** Top: Diagram illustrating the PCR-based detection scheme for successful genomic integration of the synthetic T7RNAP cassette into strain *C321.ΔA.759*. Middle: Shown are the MASC-PCR products generated from C321.Δ*A*.759 and the six T7 RNAP-expressing strains generated in this study, run on an agarose gel. The "-" indicates the unaltered strain without any insert incorporated. Bottom: Depiction of the two insertion loci used. **(d)** α-His Western blot analysis of protein samples derived from IPTG-induced populations of C321.Δ*A*.759 and the six T7RNAP-expressing strains generated in this study. The His-tagged T7RNAP version used has a molecular weight of ~100 kDa.

We first set out to establish the extent to which C321.Δ*A*.759 lysates could perform T7-based transcription. To test this, we prepared batches of crude S12 lysates from C321.Δ*A*.759 as well as BL21 Star™ (DE3) which had T7RNAP expression induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Batch CFPS reactions were performed using these lysates, directed to synthesize sfGFP both with and without direct supplementation of 16 μg/mL of purified T7RNAP[405] **(Figure 4.2.A)**. As expected based on our previous work[405], the yield from C321.Δ*A*.759 lysates with T7RNAP added was ~30% higher than either BL21 Star™ (DE3) condition. There was no observable benefit to supplementing additional T7RNAP into reactions utilizing polymerase-enriched BL21 Star™ (DE3) lysates. Unsurprisingly, essentially no sfGFP was synthesized by the C321.Δ*A*.759 lysates when no T7RNAP was supplemented. Thus, we hypothesized that introducing the *1* gene into C321.Δ*A*.759 would imbue the strain with the ability to synthesize T7RNAP and eliminate its dependence on supplemental polymerase *in vitro*.

4.4.2 T7RNAP Insert Design and Integration

A large body of work has explored various ways of enabling bacteria to produce T7RNAP[409, 416, 420]. Plasmid-based approaches are simple and effective, but expression levels are high enough to impair plasmid maintenance or otherwise place a significant metabolic burden on host cells manifesting itself in the form of drastically increased doubling time[409]. As increases in doubling time are often indicative of reduced ribosome abundance, this phenomenon is extremely undesirable for CFPS chassis strains[421, 422]. Another common scheme for *1* gene introduction is via lysogenization with synthetic DE3 bacteriophage (as in BL21 (DE3) and its derivatives)[409], but as phage insertion occurs site-specifically at a fixed genomic locus and the viral *1* gene is under the control of a fixed set of *cis*-regulatory sequences, this approach suffers

from a lack of tunability and control. An attractive alternative method for genomic integration is λ-Red mediated homologous recombination (λHR), which site-specifically integrates linear DNA constructs into target genomes using flanking sequence homology to direct insertion at the desired site[410, 411]. As C321.Δ*A*.759 natively expresses the requisite λ-Red recombination machinery[21, 405], we elected to proceed via λHR.

Several design criteria were considered for genome integration. First, a challenge in protein expression is tuning the expression level – enough protein must be synthesized to adequately perform the desired function, but aggressive overexpression can place too high of a metabolic strain on the host organism and/or lead to production of inhibitory levels of the protein. Lacking *a priori* knowledge as to how to achieve an ideal level of T7RNAP production in C321.Δ*A*.759, we decided to test a variety of different expression levels. We designed a series of synthetic constructs that placed the *1* gene under the regulation of IPTG-inducible promoters of varying transcriptional strengths, with lacUV5[423], PtacI[417], and Lpp5[418] representing relatively low, medium, and high strength respectively **(Figure 4.2.B)**. Promoter-specific synthetic ribosome binding sites (RBSs) designed for maximal translation using the Ribosome Binding Site Calculator v2.0 were employed for the regulation of translation initiation – in this way, any differences in T7RNAP expression between strains could be predominantly attributed to differences in transcription[424, 425]. Second, in the interest of easy visualization via western blotting, we added a 6-His tag to the N-terminus of the polymerase (a modification that has previously been shown to have little to no effect on polymerase activity[426]). Third, each construct also included the kanamycin kinase (*kanR*) gene from pKD4[410] (which confers resistance to the antibiotic kanamycin) for selection of successful integrants. Finally, to explore influences of genome position on expression, each

construct was designed with 50bp of flanking sequence homology at each end to facilitate integration at one of two genomic loci: the *asl* locus, selected because it was previously identified as a highly-expressing locus in the *E. coli* genome[427], and the *int* locus, selected because it is analogous to the DE3 lysogenization site in BL21 (DE3)[409].

A total of six T7RNAP-expressing constructs were assembled (*int*.lacUV5, *int*.PtacI, *int*.Lpp5, *asl*.lacUV5, *asl*.PtacI, *asl*.Lpp5) and transformed individually into C321.Δ*A*.759 for site-specific genomic integration. Potential integrants were identified by the ability to survive in the presence of kanamycin and verified via screening by multiplex allele-specific colony (MASC) PCR **(Figure 4.2.C)**. Sanger sequencing of all insert loci confirmed that each construct was integrated at the correct locus, fully intact and free of any unwanted mutations. Finally, western blotting with antibodies against the polymerase's N-terminal 6-His tag verified that each insert was indeed promoting expression of T7RNAP **(Figure 4.2.D)**. Polymerase expression as determined by western blot band intensity tracked as expected with promoter strength.

4.4.3 Characterization of T7RNAP-Expressing Strains in CFPS

To assess the ability of these strains to independently catalyze T7RNAP dependent transcription in CFPS, we prepared crude S12 lysates for all six strains for use in cell-free reactions. To promote robust T7RNAP overexpression, all strains were induced with 1 mM IPTG during exponential cell growth. Batch CFPS reactions using each lysate were directed to synthesize sfGFP over 20 h at 30°C both with and without addition of 16 µg/mL of purified T7RNAP to the reactions **(Figure 4.3.A)**. With polymerase supplemented, lysates from all six strains performed within 15% of one another. The strains featuring PtacI- and Lpp5-driven T7RNAP expression demonstrated the ability to perform transcription using only the polymerase expressed by the chassis strain. Not

surprisingly, the amount of sfGFP fluorescence appears to be related to the amount of T7RNAP produced in the cells **(Figure 4.3.A)**. At both insertion loci, the amount of fluorescence increases with increasing promoter strength, and for each promoter more fluorescence was observed from the strains featuring inserts at the highly-expressing *asl* locus. The strain capable of generating the most sfGFP fluorescence without T7RNAP supplementation, C321.Δ*A*.759.*asl*.Lpp5, achieved ~85% as much sfGFP production without supplementation as with. This strain, hereafter referred to as C321.Δ*A*.759.T7, was selected for further characterization and development.

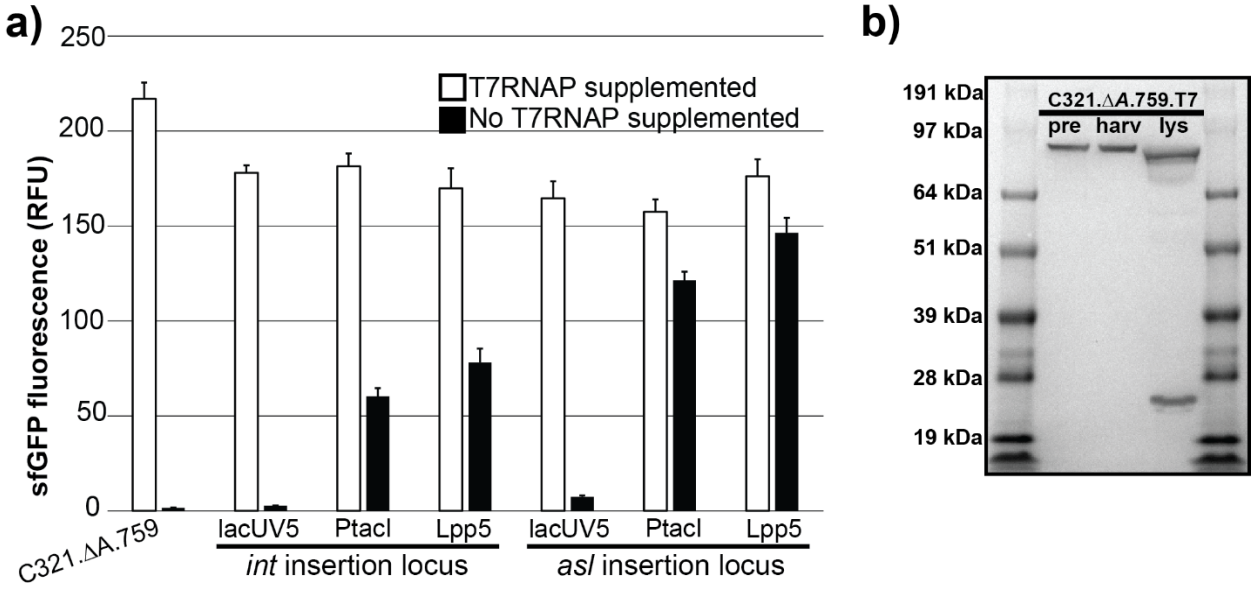


Figure 4.3. Characterization of *C321.ΔA.759* T7RNAP-expressing variants. (a) Characterization of the six *C321.ΔA.759* T7RNAP-expressing variants generated in this study. Extracts derived from *C321.ΔA.759* and its T7RNAP-expressing derivatives were directed to synthesize sfGFP in CFPS both with and without supplementation with purified T7RNAP, and fluorescence was measured after incubation for 20 hrs at 30^0C. Three independent CFPS reactions were performed for each condition, and one standard deviation is shown. **(b)** α-His Western blot characterization of *C321.ΔA.759.T7*.

Curiously, western blot analysis of samples derived from C321.Δ*A*.759.T7 revealed that the T7RNAP produced by the strain is cleaved near the N-terminus to yield a ~21kDa fragment

(Figure 4.3.B). This cleavage is well documented in the literature[416, 428], and previous work has identified the membrane-bound periplasmic protease OmpT as the responsible agent in *E. coli*[429]. The cleaved polymerase itself has been heavily characterized, and prior work has concluded that the nicked enzyme is significantly impaired by a loss in polymerase activity and efficiency[428, 430-432]. Thus, we reasoned that OmpT-mediated proteolysis of the T7RNAP expressed by C321.Δ*A*.759.T7 during cell lysis (when the periplasm and cytoplasm mix) contributed to the reduced capacity of the resulting lysates to support transcription independent of supplemental T7RNAP. To assess this hypothesis, we next sought to inactivate this OmpT activity to protect T7RNAP from proteolysis.

4.4.4 *ompT* Inactivation to Protect T7RNAP During C321.Δ*A*.759.T7 Crude Lysate Preparation

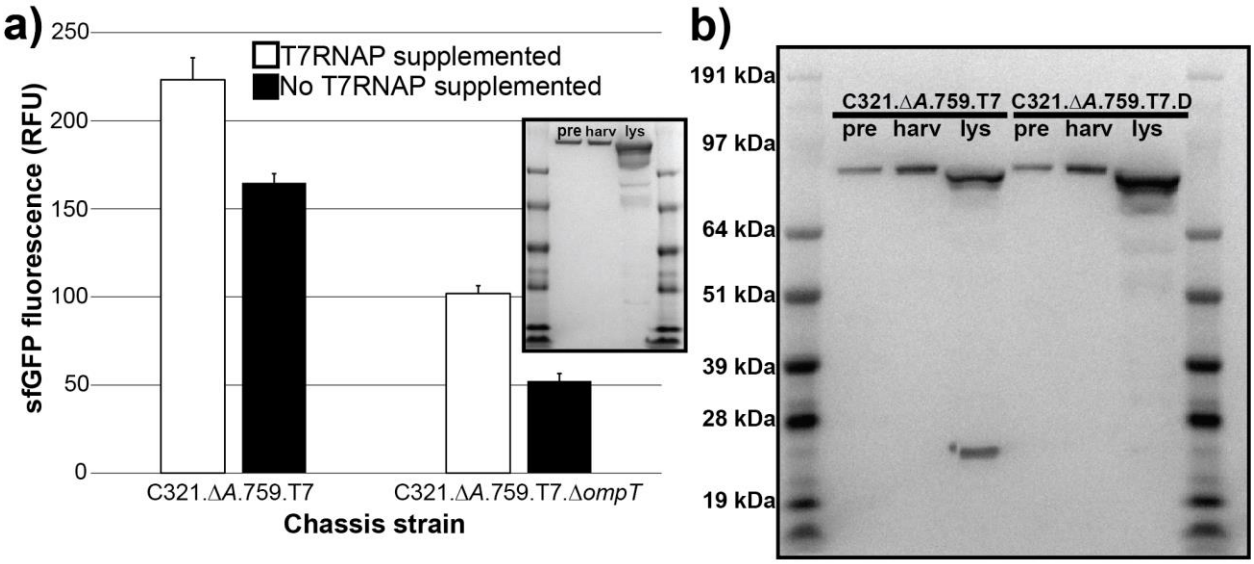


Figure 4.4 (previous page). Engineering an OmpT-resistant T7 polymerase. (a) Characterization of *C321.*Δ*A.759.T7* and *C321.*Δ*A.759.T7.*Δ*ompT*. Extracts from each strain were directed to synthesize sfGFP in CFPS both with and without supplementation with purified T7 RNAP, and fluorescence was measured after incubation for 20 hrs at 30°C. Three independent CFPS reactions were performed for each condition, and one standard deviation is shown. Inset: α-His western blot characterization of *C321.*Δ*A.759.T7.*Δ*ompT*. **(b)** α-His Western blot comparison of *C321.*Δ*A.759.T7* and *C321.*Δ*A.759.T7.D*.pre: samples derived from cells immediately prior to induction. harv: samples derived from mid-exponential phase cells immediately prior to harvest. lys: samples derived from final clarified lysate.

BL21 (DE3) and its derivative strains all feature a deletion at the *ompT* locus which presumably prevents the proteolytic degradation of the T7RNAP produced by those strains[433]. Based on this, we hypothesized that a deletion at the *ompT* locus of C321.Δ*A*.759.T7 would similarly protect strain-synthesized T7RNAP and thus eliminate the strain's partial dependence on supplemental polymerase in CFPS. To test this, we first "looped" *kanR* out of the C321.Δ*A*.759.T7 genome using MAGE[25]. Next, we applied λHR to replace a ~12kbp region of the C321.Δ*A*.759.T7 genome analogous to the spontaneous *ompT* deletion in BL21 (DE3) with a *kanR* cassette to select for successful integrants. MASC-PCR verified the knockout, yielding strain C321.Δ*A*.759.T7.Δ*ompT*.

To assess the CFPS capabilities of the *ompT*-deficient strain, we prepared crude S12 extracts from culture induced with 1 mM IPTG for analysis via both western blot and batch CFPS reactions. As expected, a western blot revealed that in the absence of OmpT the T7RNAP is no longer cleaved (i.e., we did not observe the expected 21 kDa band) **(Figure 4.4.A, inset)**. Unfortunately, batch sfGFP CFPS reactions demonstrated that the strain's ability to perform CFPS suffered significantly overall in response to the *ompT* knockout **(Figure 4.4.A)**. As compared to C321.Δ*A*.759.T7 lysate, C321.Δ*A*.759.T7.Δ*ompT* lysates show a 2-3 fold reduction in CFPS yields both with and without T7RNAP supplementation. This is consistent with earlier work demonstrating that functional OmpT is critical for robust protein synthesis in lysates derived from

C321.Δ*A* and its descendants[405]. Given our interest in designing a one-pot, high yielding CFPS system, we concluded that this was not a viable strategy for improving C321.Δ*A*.759.T7 and discontinued our pursuit of this scheme for preventing T7RNAP cleavage during cell lysis.

4.4.5 Engineering a Protease-resistant T7RNAP

We next considered a chemical biology approach to protecting the T7RNAP produced by C321.Δ*A*.759.T7 from proteolysis during lysate preparation. We reasoned that since the source of the degradation could not be removed without deleterious effects on the strain's productivity *in vitro*, perhaps the T7RNAP could be mutated such that it would no longer be an efficient substrate for OmpT. OmpT binds its substrates at pairs of adjacent basic residues and catalyzes hydrolysis of the amide bond linking them[434]. The requirement of basic residues for OmpT activity at the cleavage site is fairly rigid – in particular, the 1' residue residing immediately upstream of the polypeptide cut site *must* be basic in order for OmpT to facilitate hydrolysis[434]. In T7RNAP, two such sites have been identified proximal to the enzyme's N terminus at K172/R173[431] (K183/R184 in His-tagged mutant polymerase) and K179/K180[430] (K190/K191 in His-tagged mutant polymerase). These sites are relatively close together such that OmpT proteolysis at either would liberate a ~21kDa N-terminal fragment, consistent with what was observed on our C321.Δ*A*.759.T7 western blot.

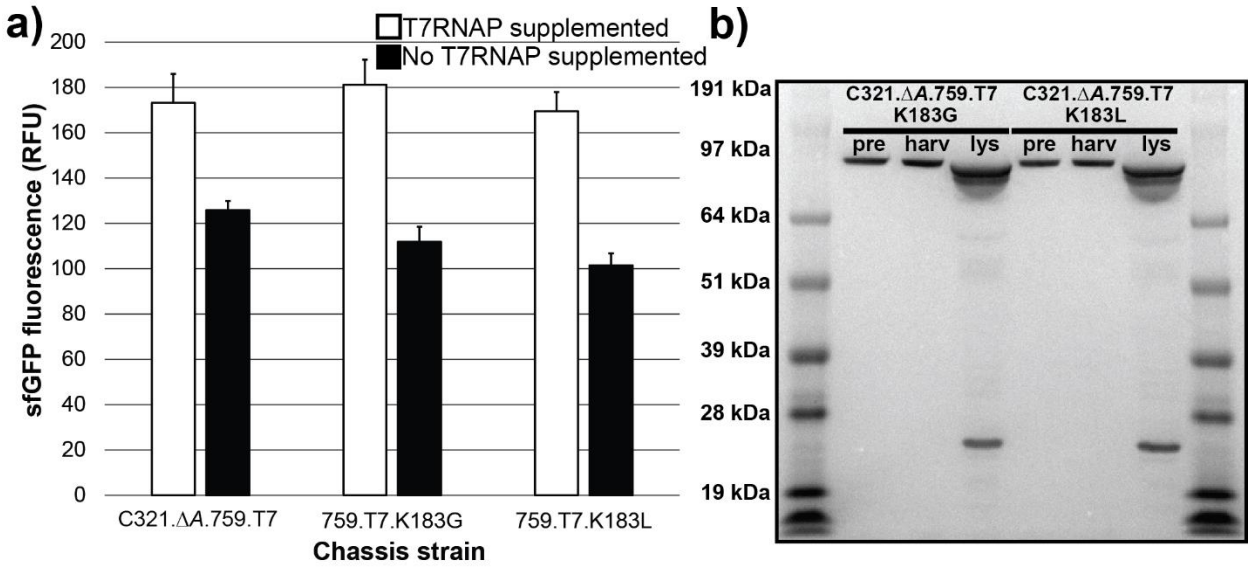


Figure 4.5. Single lysine mutations are insufficient to confer resistance to OmpT proteolysis. Comparison between C321.Δ*A*.759.T7 and mutant strains in which K183 has been mutated to glycine (C321.Δ*A*.759.T7.K183G*)* or leucine (C321.Δ*A*.759.T7.K183L). **(a)** Characterization of the K183 mutants in CFPS. Lysates derived from C321.Δ*A*.759.T7 and the K183 mutant strains were directed to synthesize sfGFP in CFPS both with and without supplementation with purified T7 RNAP, and fluorescence was measured after incubation for 20 hrs at 30°C. Three independent CFPS reactions were performed for each condition, and one standard deviation is shown. **(b)** α-His western blot comparison of the K172 mutant strains.

Because K183/R184 was previously identified as the primary site of OmpT activity in T7RNAP[431], we hypothesized that mutating K183 to a non-basic residue would abolish the target site and thus protect the polymerase from proteolysis despite the presence of fully functional OmpT in the lysate. To test this, we used MAGE to edit the sequence of the *1* gene on the genome of C321.Δ*A*.759.T7 to mutate K183 to either glycine or leucine, as these mutants had previously been shown to retain robust polymerase activity[435]. Mutations were detected using allele-specific primers in MASC-PCR and confirmed by Sanger sequencing. When extracts prepared from each mutant strain were directed to synthesize sfGFP in batch CFPS reactions both with and without supplemental T7RNAP, neither performed better than C321.Δ*A*.759.T7 **(Figure 4.5.A)**.

Analysis of the extracts revealed that despite the installed mutations, the polymerase was still being cleaved during cell lysis **(Figure 4.5.B)**.

Next, we reasoned that while the K183/R184 site may be the preferential site for proteolysis when both sites are present, in the absence of a functional site A OmpT may simply cleave at K190/K191 instead. We hypothesized that the simultaneous elimination of both sites may be necessary to fully prevent the ability of OmpT to bind and cleave the polymerase. To test this, we again exploited MAGE to edit the sequence of *1* on the C321.Δ*A*.759.T7 genome and install the mutations K183G and K190L. Mutations were detected using allele-specific primers in MASC-PCR and confirmed via Sanger sequencing. We prepared crude cell lysates from the resulting strain, C321.Δ*A*.759.T7.D, for western blot and CFPS analysis. The western blot revealed that the double mutant T7RNAP expressed by C321.Δ*A*.759.T7.D is not cleaved despite the presence of active OmpT in the cellular lysate **(Figure 4.4.B)**. In batch mode CFPS reactions, C321.Δ*A*.759.T7.D lysates exhibit a ~15% increase in productivity over C321.Δ*A*.759.T7, producing ~2.2 g/L and ~1.6 g/L of sfGFP with and without T7RNAP supplementation respectively **(Figure 4.6)**. C321.Δ*A*.759.T7.D lysates also significantly outperform both BL21 Star™ (DE3) and C321.Δ*A*.759 lysates regardless of T7RNAP supplementation, establishing it as a robust one-pot CFPS system and one of the most productive CFPS platforms developed to date.

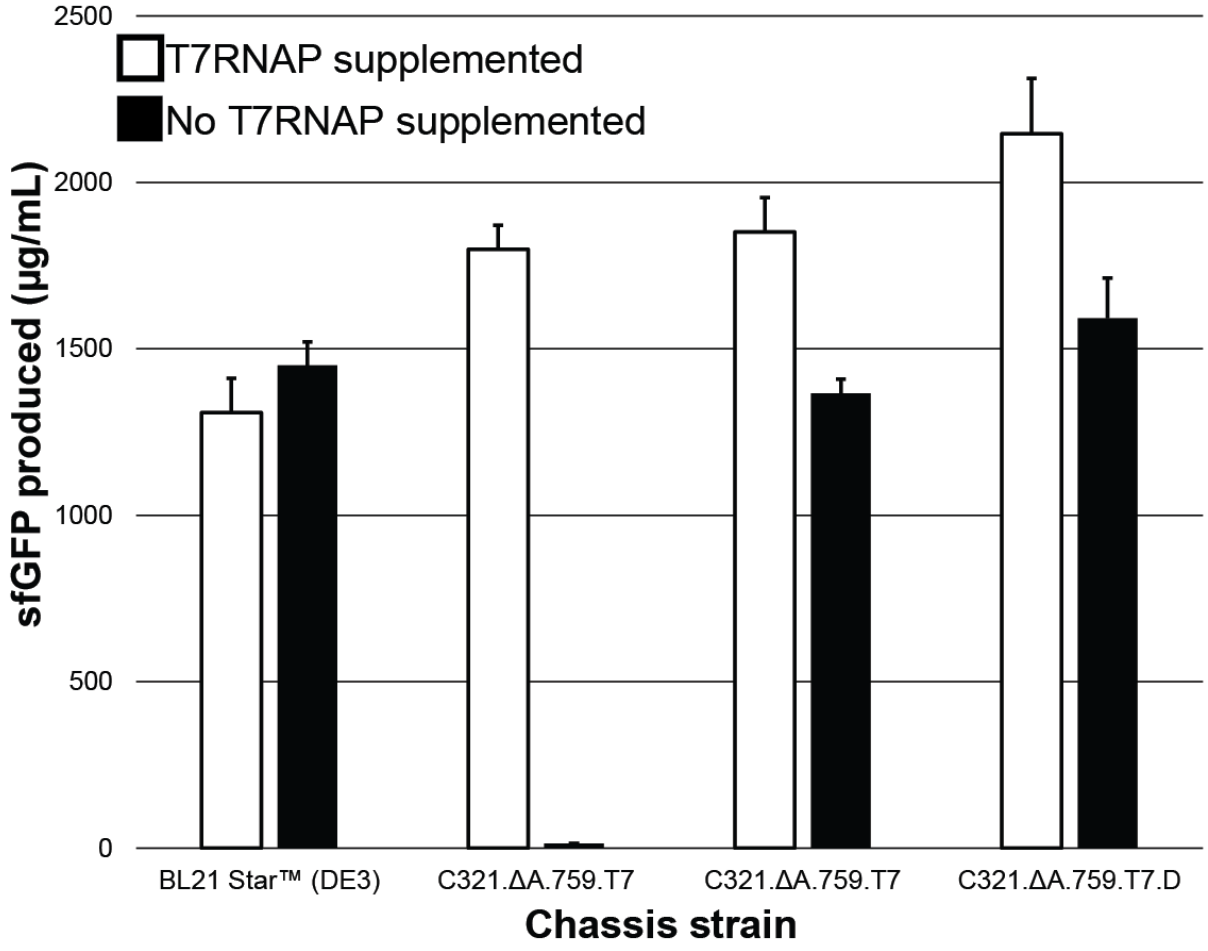


Figure 4.6. C321.Δ*A*.759.T7.D is a highly-productive, one-pot CFPS system. A side-by-side comparison of sfGFP produced in CFPS using crude lysates derived from IPTG-induced BL21 Star™ (DE3), C321.Δ*A*.759, C321.Δ*A*.759.T7, and C321.Δ*A*.759.T7.D cells. Shown are results from CFPS reactions performed both with and without supplementation with purified T7RNAP. At least 3 independent reactions were performed per condition, and one standard deviation is shown.

4.4.6 Characterization of C321.Δ*A*.759.T7.D Lysates *in vitro*

We next set out to characterize the system. To demonstrate generality, batch CFPS reactions were performed using C321.Δ*A*.759.T7.D and BL21 Star™ (DE3) lysates directed to synthesize red fluorescent protein (mRFP) over 20 h at 30°C both with and without T7RNAP supplementation. Under both conditions, C321.Δ*A*.759.T7.D generated significantly higher levels

of RFP fluorescence **(Figure 4.7)**. To explore the induction response of the IPTG-inducible T7RNAP expression cassette, we harvested cell populations of C321.Δ*A*.759.T7.D induced during exponential growth phase with either 0 mM or 1 mM IPTG and prepared lysates from each for testing in CFPS reactions. While these experiments confirmed that maximum IPTG induction yielded the most productive lysates, we observed that even in the complete absence of IPTG induction C321.Δ*A*.759.T7.D lysates still synthesize enough polymerase to facilitate robust protein synthesis *in vitro* **(Figure 4.8.A)**. As Mg^{2+} is a critical factor for both ribosome assembly[436] and T7RNAP function[437], we also optimized the Mg^{2+} content of the cell-free reaction environment **(Figure 4.8.B)**. Exploratory CFPS reactions featuring increased levels of supplemental T7RNAP and plasmid DNA template revealed that even significant increases in the concentrations of these components in the cell-free environment provided negligible benefit to the amount of protein produced **(Figure 4.8.C)**. Finally, we investigated the kinetics of the translational components as compared to the progenitor strains. Batch CFPS reactions using lysates derived from IPTG-induced populations of C321.Δ*A*.759, BL21 Star™ (DE3) (Invitrogen, Carlsbad, CA), C321.Δ*A*.759.T7, and C321.Δ*A*.759.T7.D were performed over 20 h at 30°C, directed to synthesize sfGFP both with and without T7RNAP supplementation. Samples were monitored over the course of the reaction, with protein quantification determined by the emergence of sfGFP fluorescence **(Figure 4.9)**. All systems demonstrated rapid protein production during the first four hrs of the CFPS reaction, which is consistent with previous observations[405]. Consistent with endpoint measurements, C321.Δ*A*.759, C321.Δ*A*.759.T7, and C321.Δ*A*.759.T7.D lysates with supplemental T7RNAP eventually overtake the other conditions to produce high levels of sfGFP fluorescence. C321.Δ*A*.759.T7 and C321.Δ*A*.759.T7.D lysate fluorescence without additional

polymerase trended similarly to both BL21 Star™ (DE3) conditions, while C321.Δ*A*.759 lysate without T7RNAP added produced no fluorescence at all.

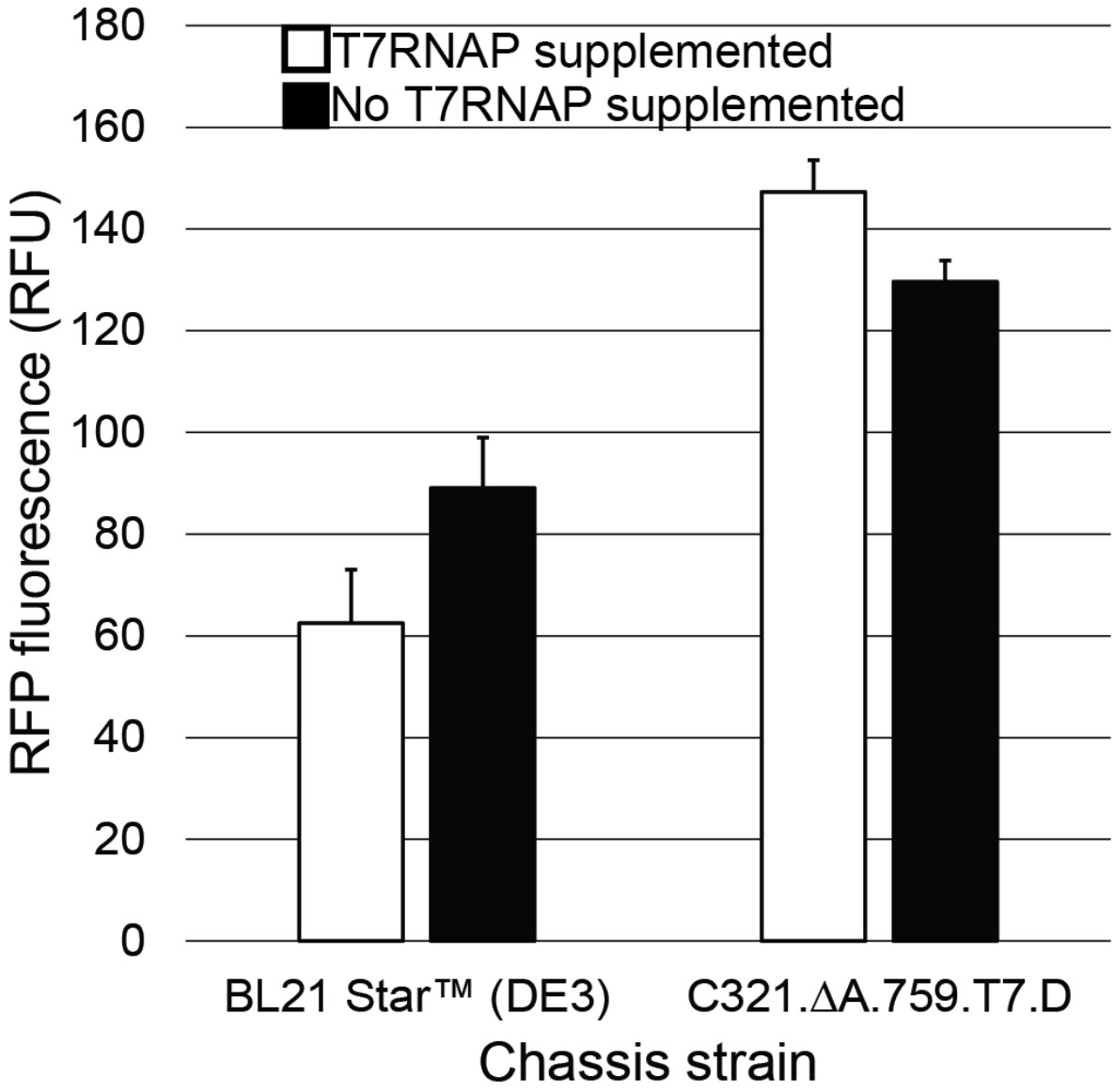


Figure 4.7. C321.Δ*A*.759.T7.D is a highly-productive, one-pot system for general protein synthesis. Comparison of general protein synthesis using BL21 Star™ (DE3) and C321.Δ*A*.759.T7.D. Extracts derived from the listed strains were directed to synthesize RFP with and without supplementation with purified T7RNAP. RFP fluorescence was measured following a 20 hrs incubation at 30°C. Three independent CFPS reactions were performed for each condition, and one standard deviation is shown.

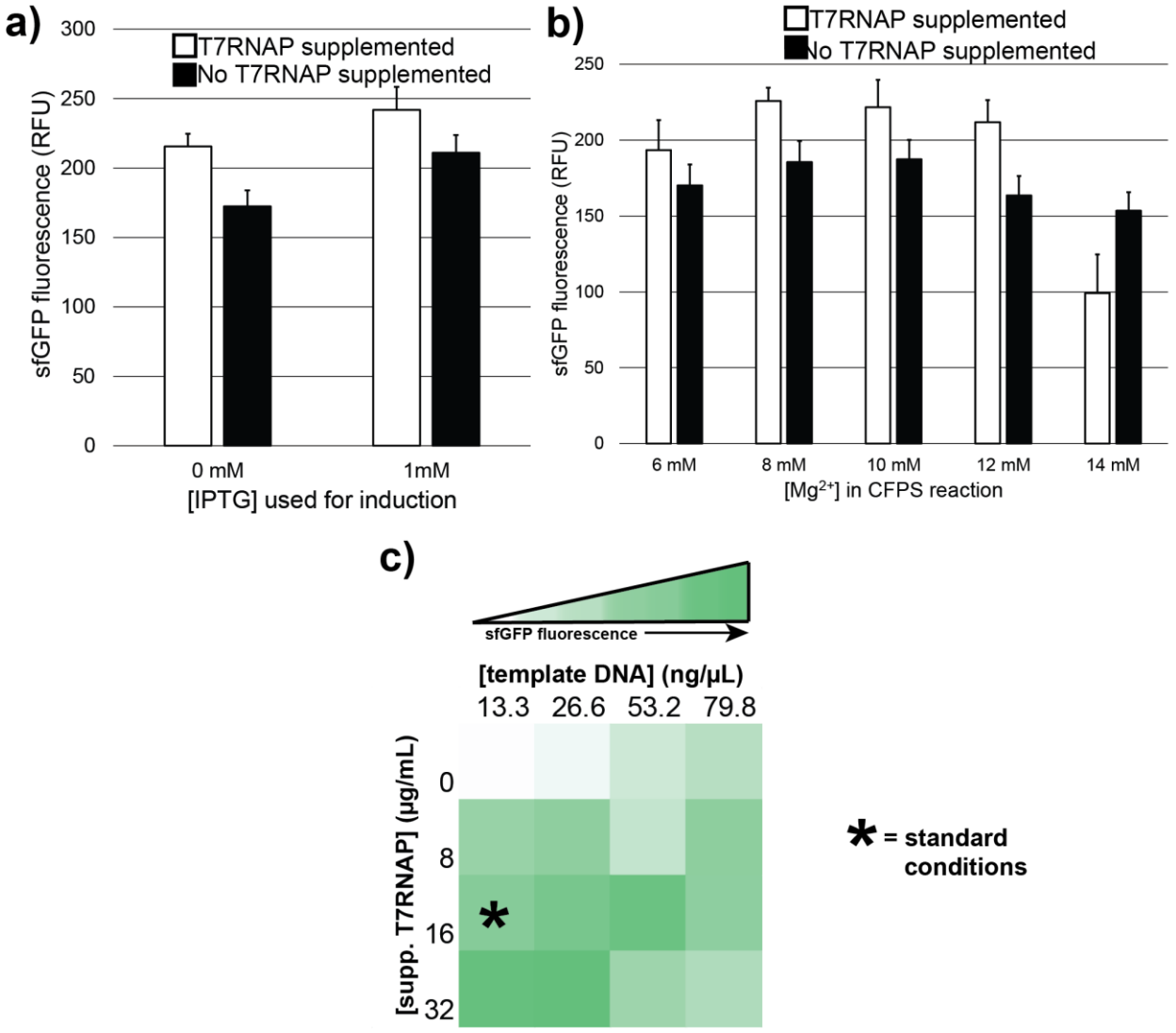


Figure 4.8. Optimization of CFPS reactions conditions for *C321.ΔA.759.T7.D* lysates. (a) Characterization of *C321.ΔA.759.T7.D* lysates derived from cell cultures induced with 0mM or 1mM IPTG. These extracts were directed to synthesize sfGFP in CFPS both with and without supplementation of purified T7 RNAP, and fluorescence was measured after incubation for 20 hrs at 30°C. Three independent CFPS reactions were performed for each condition, and one standard deviation is shown. **(b)** Optimization of [Mg^{2+}] in *C321.ΔA.759.T7.D* CFPS reactions. Reactions containing the indicated concentrations of Mg^{2+} were performed, with the extract directed to synthesize sfGFP in CFPS both with and without supplementation with purified T7RNAP. Fluorescence was measured after incubation for 20 hrs at 30°C. Three independent CFPS reactions were performed for each condition, and one standard deviation is shown. **(c)** Optimization of [supplemental T7RNAP] and [template DNA] in *C321.ΔA.759.T7.D* reactions.

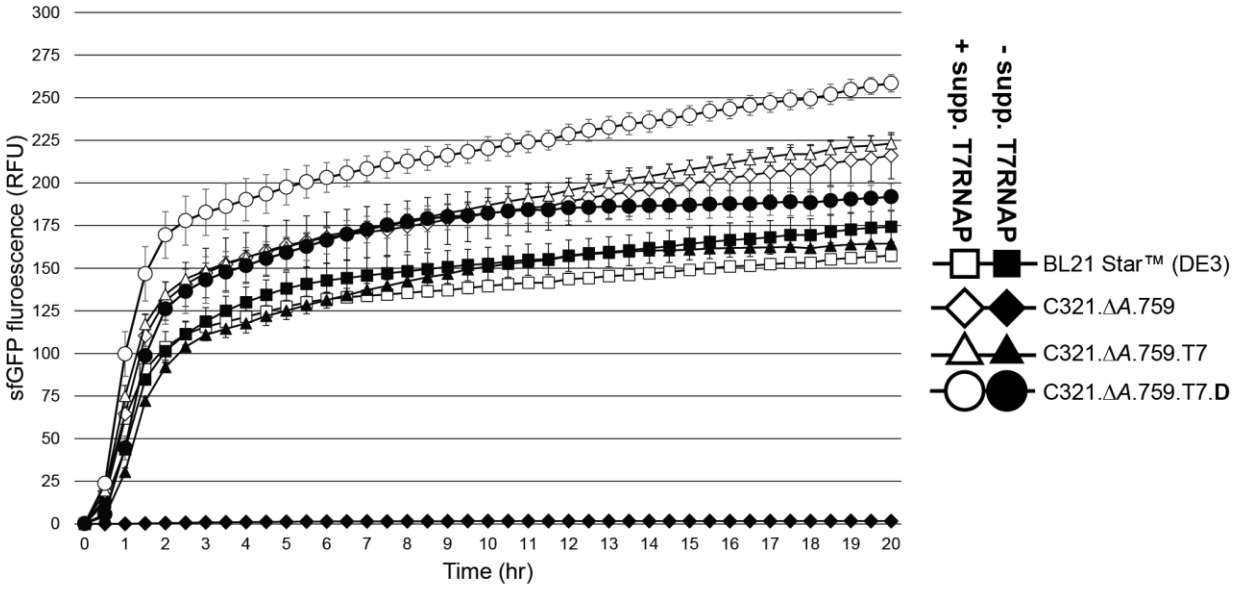


Figure 4.9. Characterization of protein synthesis kinetics using C321.Δ*A*.759.T7.D lysates. Comparison between BL21 Star™ (DE3), C321.Δ*A*.759, C321.Δ*A*.759.T7, and C321.Δ*A*.759.T7.D. Extracts derived from the listed strains were directed to synthesize sfGFP with and without supplementation with purified T7RNAP. sfGFP fluorescence was measured at various time points over the course of a 20 hr incubation at 30°C. Three independent CFPS reactions were performed for each condition, and one standard deviation is shown.

4.4.7 Demonstration of Capacity for Multiple ncAA Incorporations Using T7RNAP-expressing Strains

Finally, we assessed the capacity for our high yielding C321.Δ*A*.759.T7.D lysates to produce proteins featuring ncAAs. Because the parent strain had RF1 removed[21], we expected ncAA incorporation via amber suppression would be highly efficient – indeed, our previous effort using a recoded strain showed up to 40 ncAA incorporations in a single polypeptide[405]. To assess ncAA incorporation into proteins, both C321.Δ*A*.759.T7.D and BL21 Star™ (DE3) were transformed with a pEVOL plasmid encoding the orthogonal translation system (OTS) components for the ncAA *p*-acetyl-L-phenylalanine (pAcF)[14]; namely the pAcF-specific

aminoacyl-tRNA synthetase (pAcFRS) and an orthogonal suppressor tRNA engineered to decode the amber codon (o-tRNA)[438]. Then, we quantitatively assessed the incorporation of pAcF into sfGFP variants with up to five in-frame amber codons. CFPS reactions were supplemented with additional OTS components based on our previous work[405].

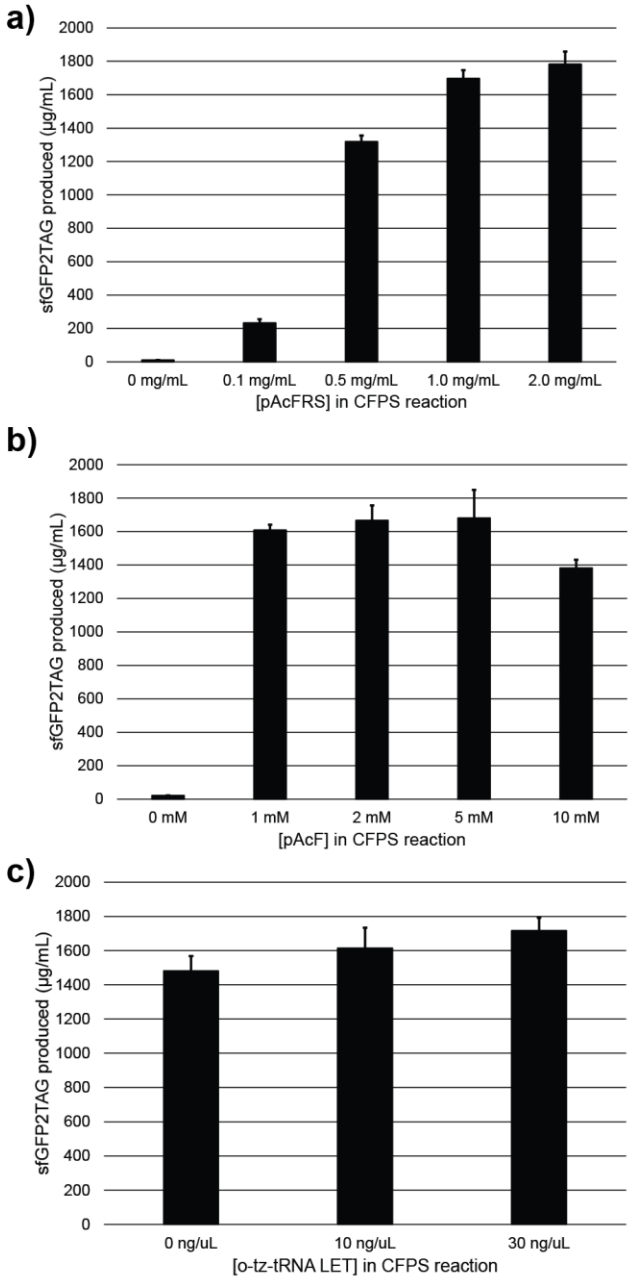


Figure 4.10 (previous page). Optimization of orthogonal translation system (OTS) component supplementation for performing amber suppression using C321.Δ*A*.759.T7.D lysates. C321.Δ*A*.759.T7.D lysates were directed to synthesize sfGFP2UAG in CFPS with T7RNAP supplemented. Concentrations of pAcF OTS components (pAcFRS, **(a)**; pAcF, **(b)**; o-tz-tRNA linear expression template (LET)[232], **(c)**) were titrated to identify the optimal concentration of each component. 3 independent reactions were performed per condition, and one standard deviation is shown.

As an initial demonstration of ncAA incorporation, we directed the lysates derived from these pEVOL-bearing strains to synthesize amber mutant variants of sfGFP. We first established the optimal concentrations of pAcF OTS components to be supplied to these CFPS reactions via a series of CFPS reactions directed to synthesize a sfGFP variant featuring two amber codons (sfGFP-2UAG) **(Figure 4.10)**. Using these conditions, the C321.Δ*A*.759.T7.D and BL21 Star™ (DE3) pEVOL-pAcF lysates were used in CFPS to synthesize wild type sfGFP (sfGFPwt), sfGFP with a single amber codon (sfGFP-T216X), sfGFP with two amber codons (sfGFP-2UAG), or sfGFP with five amber codons (sfGFP-5UAG). Reactions were performed both with and without supplementation with purified T7RNAP **(Figure 4.11.A)**. As expected, regardless of T7RNAP supplementation the RF1-deficient C321.Δ*A*.759.T7.D lysates exhibit a significantly higher capacity for pAcF incorporation than the BL21 Star™ (DE3) lysates, with the difference becoming more pronounced as the number of pAcF incorporations increases. Indeed, C321.Δ*A*.759.T7.D yields ~500 µg/mL of sfGFP-5UAG both with and without T7RNAP supplementation whereas BL21 Star™ (DE3) yields essentially none. C321.Δ*A*.759.T7.D lysates remain highly productive for sfGFP variants bearing up to two pAcFs, yielding >1.6 g/L with and >1.2 g/L without T7RNAP supplementation for sfGFPwt, sfGFP-T216X, and sfGFP-2TAG.

We next explored the synthesis of large polypeptides containing multiple identical ncAAs utilizing our one-pot CFPS platform derived from C321.Δ*A*.759.T7.D. For our model protein, we sought to produce elastin-like polymers (ELPs)[405] containing ncAAs using C321.Δ*A*.759.T7.D

lysates. ELPs are biocompatible and stimuli-responsive biopolymers that can be applied for drug delivery and tissue engineering[395, 439]. Previously, we have introduced multiple, identical ncAAs into ELPs by substituting natural amino acids with ncAAs at a guest position in the repeating pentapeptide unit (VPGVG) that can be modified while maintaining ELP structure and function[27, 405].

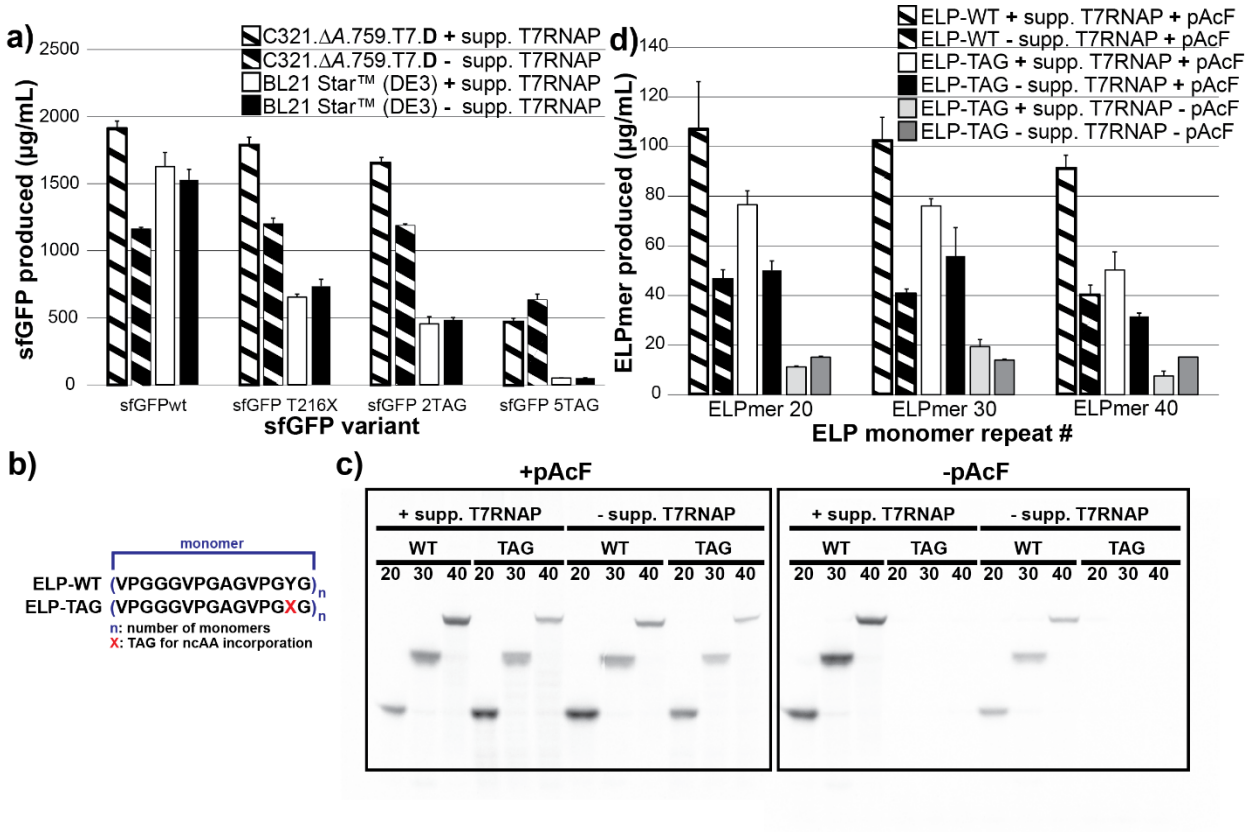


Figure 4.11. C321.Δ*A*.759.T7.D is a highly-efficient platform for one-pot ncAA incorporation. (a) sfGFP produced *in vitro* from cell extracts derived from induced BL21 Star™ (DE3) cells as well as C321.Δ*A*.759.T7.D cells, both with and without supplementation with purified T7 RNAP. The indicated sfGFP amber mutant variants were synthesized in the presence of the complete pAcF OTS. At least 3 independent reactions were performed per condition, and one standard deviation is shown. **(b)** Illustration of both the wild type (WT) and amber mutant (UAG) ELP monomer used in this study. **(c)** Autoradiogram of ELPmers produced by C321.Δ*A*.759.T7.D lysates under the indicated reaction conditions. **(d)** ^{14}C-glycine radioactive count quantification of ELPmers produced by C321.Δ*A*.759.T7.D lysates under the indicated conditions. At least 3 independent reactions were performed per condition. and one standard deviation is shown.

Our ELP construct consisted of three pentapeptide repeats per monomer unit with a single valine codon per monomer changed to UAG in amber mutants[405] **(Figure 4.11.B)**. Lysates derived from bearing C321.Δ*A*.759.T7.D bearing pEVOL-pAcF were directed to synthesize wild type (ELP-WT) and amber mutant (ELP-UAG) ELPs with 20, 30, and 40 monomer units in the presence of pAcF, both with and without supplementation with purified T7RNAP. Products were visualized using an autoradiogram, demonstrating that a high percentage of the protein produced under each condition is full-length **(Figure. 4.11.C, left)**. When the experiments were repeated without the addition of pAcF to reactions, synthesis of full-length ELP-WT species was unaffected but full-length ELP-UAG protein was no longer observed **(Figure. 4.11.C, right)**. These results, when combined with our earlier work validating this OTS system[405], demonstrate the relatively high fidelity of the pAcF OTS, showing that the percent of full-length product is ~100%. Absolute yields for each of the various ELPs were quantified via ^{14}C-glycine radioactive scintillation counting **(Figure 4.11.D)**. The lysates generated 91 ± 5 mg/L and 40 ± 0.2 mg/L of each ELP-WT with and without T7RNAP supplementation, respectively. For the ELP-UAG constructs featuring up to 30 amber codons, the C321.Δ*A*.759.T7.D lysates yielded 76 ± 3 mg/L and 55 ± 11 mg/L with and without T7RNAP supplementation, respectively. Similar yields were observed with the construct with 20 amber codons. Synthesis of the 40 amber codon construct was reduced relative to the 20- and 30 amber codon variants, experiencing a decrease in productivity of ~35% and ~45% with and without T7RNAP supplementation, respectively. Yields from reactions not containing pAcF were low (<15 mg/L), suggesting that the ELP-UAG species quantified when pAcF is present is composed of a high percentage of full-length protein featuring pAcF at the majority of UAG codons. Taken together, these results demonstrate that C321.Δ*A*.759.T7.D lysates are

capable of catalyzing the production of proteins bearing multiple ncAAs independent of supplementation with T7RNAP.

4.4.8 Improving Endogenous T7RNAP Productivity Via His-tag Removal

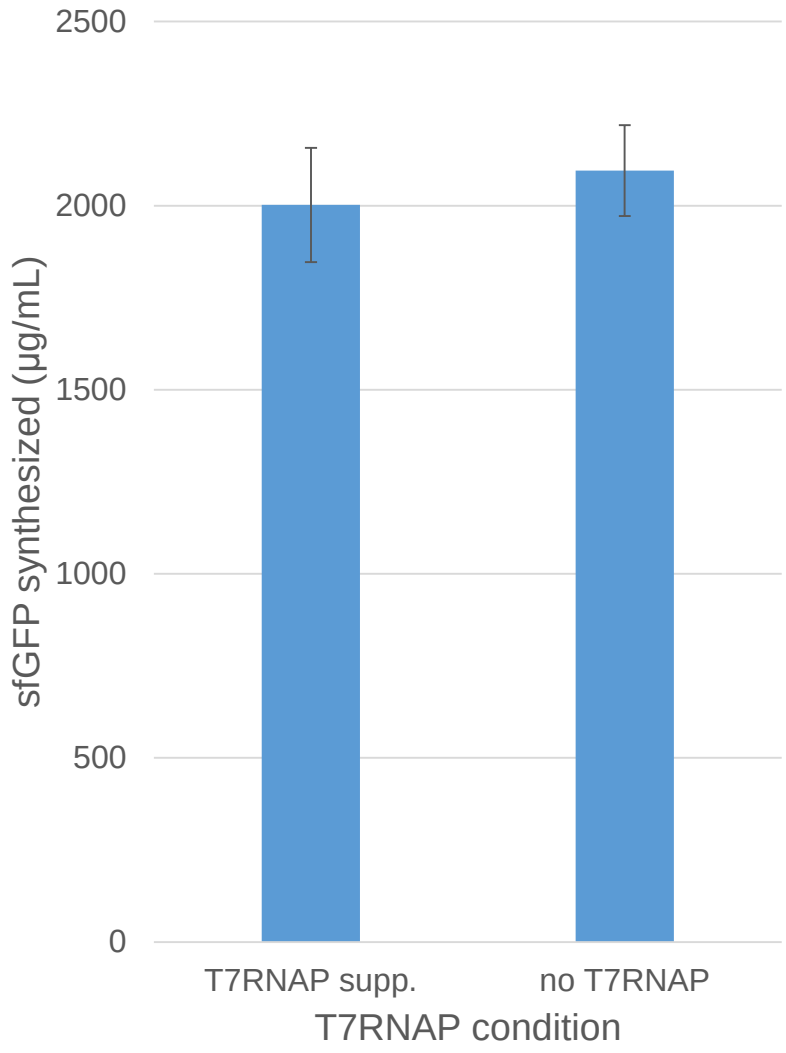


Figure 4.12. Characterization of lysates derived from C321.Δ*A*.759.T7.D.ΔHis. Shown is sfGFP synthesized by lysates derived from C321.Δ*A*.759.T7.D.ΔHis both with and without supplementation with purified T7RNAP (T7RNAP supp. and no T7RNAP, respectively). For each condition, three independent reactions were performed and one standard deviation is shown.

Finally, we sought to identify and address the fundamental feature limiting the productivity of the T7RNAP synthesized in C321.Δ*A*.759.T7.D such that the lysates still depend partially on supplementation with purified polymerase **(Figure 4.6)**. Because the presence of a his-tag has been shown to negatively impact the function of other proteins[440, 441], we reasoned that the N-terminal tag used throughout this effort could be reducing the productivity of the T7RNAP expressed by C321.Δ*A*.759.T7.D and thus causing lysates derived from the strain to be transcription limited. To test this, we applied a combination of CRISPR/Cas9 and MAGE to remove the his-tag from the *1* gene of C321.Δ*A*.759.T7.D to yield C321.Δ*A*.759.T7.D.ΔHis (see Chapter 6 for details). Lysates were prepared from an induced cell culture of C321.Δ*A*.759.T7.D.ΔHis and used in CFPS to synthesize sfGFP both with and without supplementation with purified T7RNAP. As expected, the removal of the N-terminal his-tag from the polymerase in the lysate improved productivity in the absence of supplemental polymerase **(Figure 4.12)**. Indeed, C321.Δ*A*.759.T7.D.ΔHis lysates generated identical yields of sfGFP both with and without purified polymerase supplementation.

4.5 Discussion

One-pot systems, such as those derived from the state-of-the-art protein overexpression strain BL21 Star™ (DE3), are highly desirable for CFPS due to their enrichment with critical enzymes such as T7RNAP. In the specific case of T7RNAP, such systems reduce the cost of CFPS and make the system easier to put together. While robust and versatile, existing one-pot platforms based on BL21 Star™ (DE3) struggle with the production of proteins containing multiple ncAAs due to the competitive action of RF1 in the reaction environment. In this study, we describe the generation and utilization of a highly productive one-pot CFPS platform beginning with

C321.Δ*A*.759, a genomically recoded RF1-deficient strain that was previously optimized for CFPS. We integrated a series of DNA constructs into the genome of C321.Δ*A*.759, each of which featured the T7RNAP-encoding gene *1* under the control of one of three different promoter sequences of varying potency. The construct featuring *1* regulated by strong promoter Lpp5 integrated at a previously identified high-expression genomic locus *asl*[427] yielded C321.Δ*A*.759.T7, which was capable of supporting *in vitro* transcription independent of supplementation with purified T7RNAP. When used in CFPS, C321.Δ*A*.759.T7 lysates yielded ~85% as much sfGFP without T7RNAP supplementation as with. In an effort to address the continuing partial dependence of the system on polymerase supplementation, we explored different strategies to protect the T7RNAP expressed in C321.Δ*A*.759.T7 from OmpT-mediated proteolysis during lysate preparation. By mutating two lysine residues proximal to the N-terminus of *1*, we were able to abolish the putative OmpT target sites and establish an OmpT-resistant mutant version of T7RNAP in our final strain C321.Δ*A*.759.T7.D. C321.Δ*A*.759.T7.D lysates are highly productive, yielding ~2.2 g/L sfGFP with and ~1.6 g/L sfGFP without T7RNAP supplementation. We also demonstrated the merits of RF1-deficient systems for ncAA incorporation, highlighting the significantly increased capacity for amber suppression in C321.Δ*A*.759.T7.D lysates as compared to BL21 Star™ (DE3). Furthermore, we were able to confirm the synthesis of full-length polypeptides containing up to 40 ncAAs without the addition of purified T7RNAP using C321.Δ*A*.759.T7.D lysates. Finally, by removing the his-tag from the N-terminus of the T7RNAP expressed in the strain, we were able to generate a truly-independent system that does not benefit at all from supplemental polymerase addition into reactions.

Looking forward, one intriguing avenue is the continued development of C321.Δ*A*.759.T7.D for improved productivity and enhanced functionality in CFPS. This might be achieved by correcting some of the potentially-harmful off-target mutations incurred during the initial recoding of the strain. Additionally, upregulation of other positive effectors of CFPS (chaperones, elongation factors, energy regeneration enzymes, *etc*.) could be achieved via genomic integration using a strategy similar to that employed in this study. In particular, the development of orthogonal synthetases with improved kinetics and substrate specificity is a critical hurdle that must be overcome before these enzymes can be overexpressed in CFPS chassis strains without deleterious effects on cellular health and lysate performance[22].

Developing efficient CFPS systems specialized for ncAA incorporation is important for synthetic biology for various emerging applications. Powerful one-pot production platforms will support the large-scale synthesis of protein products featuring novel structures and functions, in turn promoting the mass production of potent therapeutics and materials. We anticipate that CFPS platforms such as that derived from C321.Δ*A*.759.T7.D are promising for these and other synthetic biology applications.

4.6 Acknowledgments

Thanks to my colleagues Seok Hoon Hong, Rey W. Martin, and Yong Chan Kwon for thoughtful consultation and discussions regarding the content of this chapter. Professor Michael C. Jewett (M.C.J.) and I conceived of this study. I performed all experiments and analysis. M.C.J. contributed to editing this chapter.

4.7 Publication Information

As of this writing, this chapter was submitted with the following citation information:

Des Soye B.J. and Jewett M.C. A Highly Productive, One-Pot Cell-free Protein Synthesis Platform Based on Genomically Recoded *Escherichia coli*. In revision at *Cell Chemical Biology*.

CHAPTER 5

5 Establishing a High-yielding Cell-free Protein Synthesis Platform Derived from *Vibrio natriegens*

5.1 Abstract

A new wave of interest in cell-free protein synthesis (CFPS) systems has shown their utility for producing proteins at high titers, establishing genetic regulatory element libraries (*e.g.,* promoters, ribosome binding sites) in non-model organisms, optimizing biosynthetic pathways before implementation in cells, and sensing biomarkers for diagnostic applications. Unfortunately, most previous efforts have focused on a select few model systems, such as *Escherichia coli*. Broadening the spectrum of organisms used for CFPS promises to better mimic host cell processes in prototyping applications and open up new areas of research. Here, we describe the development and characterization of a facile CFPS platform based on lysates derived from the fast-growing bacterium *Vibrio natriegens*, which is an emerging host organism for biotechnology. We demonstrate robust preparation of highly active extracts using sonication, without specialized and costly equipment. After optimizing the extract preparation procedure and cell-free reaction conditions, we show synthesis of 1.6 ± 0.05 g/L of superfolder green fluorescent protein in batch mode CFPS, making it competitive with existing *E. coli* CFPS platforms. To showcase the flexibility of the system, we demonstrate that it can be lyophilized and retain biosynthesis capability, that it is capable of producing antimicrobial peptides, and that our extract preparation procedure can be coupled with the recently described Vmax™ Express strain in a one-pot system. Finally, to further increase system productivity, we explore a knockout library in which putative

negative effectors of CFPS are genetically removed from the source strain. Our *V. natriegens*-derived CFPS platform is versatile, and simple to prepare and use. We expect it will facilitate expansion of CFPS systems into new laboratories and fields for compelling applications in synthetic biology.

5.2 Introduction

Cell-free systems have recently enjoyed a technical renaissance that has transformed them into robust platforms for the synthesis of a wide variety of useful and interesting products[280, 307, 405, 442]. Such platforms combine crude cell lysates or purified components with substrates in a test tube, enabling the activation and use of cellular processes *in vitro* **(Figure 5.1)**. Cell-free protein synthesis (CFPS) systems in particular have made significant advances in reaction volume, duration, and productivity, now reaching g/L quantities of protein[264, 284, 287, 306, 405, 414, 443-445]. These systems provide several unique advantages for understanding, harnessing, and expanding the capabilities of natural biological systems. Reactions are open, and are therefore easily accessible for sample extraction and substrate feeding. Dilute reaction environments facilitate the folding of complex eukaryotic protein products which may otherwise express poorly in bacterial systems[442]. Importantly, the removal of genomic material from the chassis organism directs reaction substrates and machinery towards the desired synthesis reaction at high rates. Exploiting these features, CFPS platforms enjoy increasingly widespread use as a complement to *in vivo* expression for applications including biomolecular breadboarding[256-259], expression of toxic products[260-263], production of complex protein products that are poorly soluble *in vivo*[264-267], manufacture of glycoproteins[147, 268-270], detection of disease[30, 271, 272], on demand biomanufacturing[30, 265, 273-276], and education[277, 278].

Despite the emergence of cell-free systems as a prominent research tool for fundamental and applied biology, the vast majority of previous efforts have focused on a select few model systems such as *Escherichia coli*, *Saccharomyces cerevisiae*, and Chinese Hamster Ovary cells, among others[287, 299, 306, 307, 445, 446]. However, we and others hypothesize that developing cell-free systems composed of extracts derived from relevant chassis organisms that better mimic the natural physicochemical environment might enhance predictive power for synthetic biology applications. This idea motivates the development of new cell-free systems. In this context, several new CFPS systems have been developed, including some from *Streptomyces* species and *Bacillus* [311, 408, 447-449]. For example, an elegant study by Freemont and colleagues characterized new DNA parts from the non-model bacterium *Bacillus megaterium* by combining automated CFPS and Bayesian models[311].

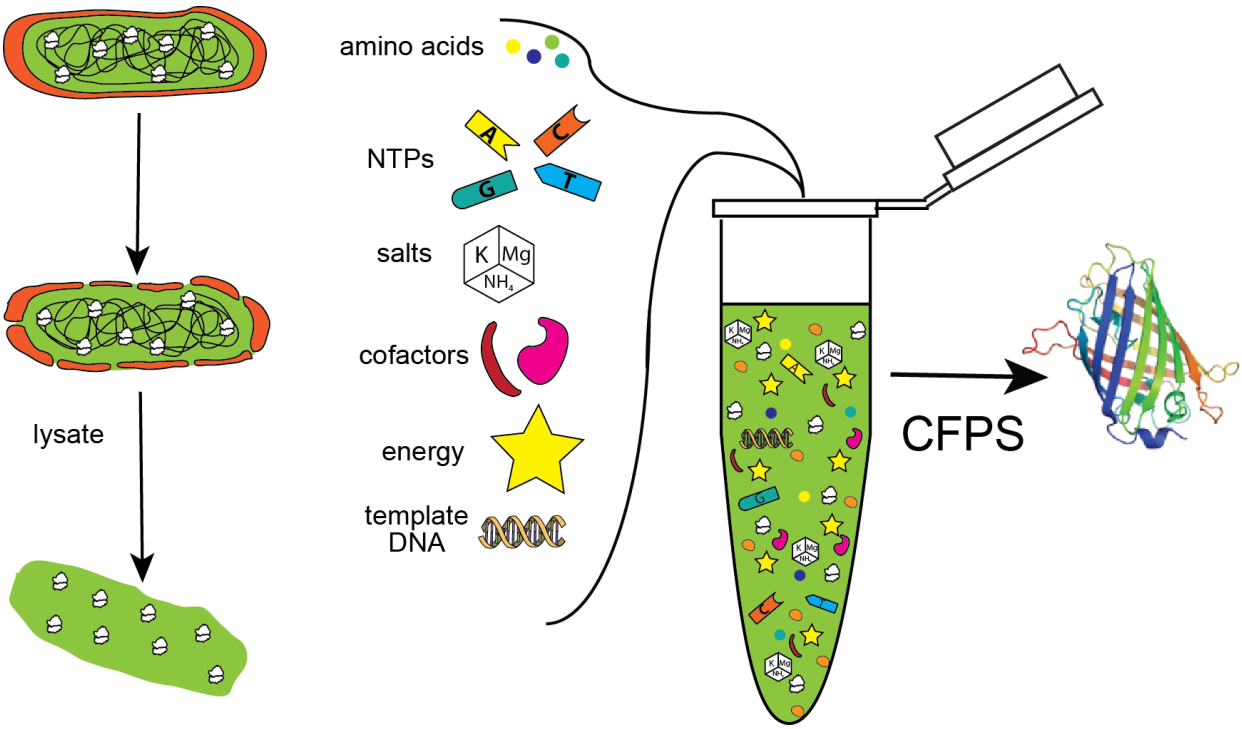


Figure 5.1 (previous page). Simplified schematic of the production and utilization of crude lysates from bacterial chassis cells to catalyze cell-free protein synthesis (CFPS). Reactions are supplemented with enzymatic cofactors, energy, and other substrates required for protein synthesis, as well as plasmid DNA template directing the system towards the production of a product of interest.

A particularly exciting chassis organism for developing a new cell-free system is the fast-growing halophilic marine bacterium *Vibrio natriegens*. First discovered in a Georgia salt marsh in 1958, *V. natriegens* (originally classified as *Pseudomonas natriegens*) was identified as the fastest-growing bacterium known to date when it was discovered that cell populations in liquid culture double approximately once every 10 minutes[450, 451]. Despite this noteworthy trait, *V. natriegens* went largely unstudied for decades. Recently, interest in this organism has been renewed, largely out of a desire to leverage its rapid generational time to accelerate molecular biology efforts and improve recombinant protein production[28, 29].

V. natriegens' rapid doubling time is particularly interesting for potential CFPS system development, as in the context of CFPS it is generally accepted that lysate productivity loosely correlates to chassis organism growth rate[421, 422]. After all, proper cell division relies on the coordinated activities of a large suite of proteins, so it is reasonable to infer that rapidly dividing cells require high protein synthesis rates, and by extension possess highly active protein translation machinery. This is very likely the case for *V. natriegens*—the species features 12 rRNA operons as compared to the 7 found in *E. coli* strain MG1655[29]. Furthermore, it has been suggested that exponentially-growing *V. natriegens* cells contain ~115,000 ribosomes/cell, significantly higher than the ~70,000 ribosomes/cell observed in *E. coli*[452]. Considering these advantages, we

hypothesized that lysates derived from *V. natriegens* would be enriched in active protein translation machinery and thus compose a highly productive CFPS platform.

In this study, we describe the development of a facile CFPS platform derived from *V. natriegens*. Key design criteria were to make the system robust, easy to use, and accessible to all. Therefore, since cell lysis procedures using homogenization or French press can be expensive, time and labor intensive, and hard to standardize, we focused on developing an extract preparation procedure using standard sonication equipment[291]. Previously, we have shown that sonication offers a simple strategy to reduce cost and variability in crude *E. coli* extract preparation, while eliminating the need for specialized and expensive growth and lysis equipment[291]. We first showed the ability to create a protein synthesis competent cell-free system. We then optimized the extract preparation process by modifying growth media, cultivation time, cell disruption conditions, and lysate clarification conditions. This led to a 400% increase from the non-optimized case, resulting in a cell-free system capable of synthesizing ~1 mg/mL of superfolder green fluorescent protein (sfGFP) using a typical *E. coli* CFPS reagent mix. Further optimization of key reagent concentrations increased the productivity of the system to ~1.6 mg/mL. We lyophilized fully-assembled *V. natriegens* CFPS reactions and found that, if prepared in the presence of trehalose, reactions retain 100% productivity after one week of storage at room temperature. Next, we applied homologous recombination-based genome engineering to prepare a small library of knockout strains in which the genes encoding putative negative effectors of CFPS have been removed. Lysates were prepared from each of these knockout strains in an attempt to identify a strain background with improved productivity *in vitro*. The *V. natriegens* CFPS system described here

is productive, robust, and facile to prepare. We expect it will lower the barrier for entry into the use of CFPS systems.

5.3 Materials and Methods

5.3.1 Strains and Plasmids

The bacterial strains and plasmids used in this study are listed in **Table 5.1**. *V. natriegens* was purchased from the American Type Culture Collection (ATCC® 14048™). Vmax™ Express was purchased from SGI-DNA, a subsidiary of Synthetic Genomics, Inc. pJL1 plasmids encoding the antimicrobial peptides were synthesized and assembled by Twist Bioscience. Assembled plasmids were submitted to the NUSeq Core facility along with forward primers, and sequences were confirmed using traditional Sanger sequencing. Kanamycin (50 µg/mL) was used for maintaining pJL1-based plasmids. Chloramphenicol (34 µg/mL) was used to select for all negative effector knockout mutants.

5.3.2 Cell Culture

V. natriegens cells were grown in BHI media supplemented with v2 salts (204 mM NaCl, 4.2 mM KCl, 23.14 mM $MgCl_2$) unless noted otherwise. For confirmation of *V. natriegens* growth rate, 100 µL cultures were assembled in a clear 96-well plate (Costar 3370; Corning, Corning, NY) and shaken at 250 RPM at 37°C for 20 hr in a Synergy H1 plate reader (BioTek, Winooski, VT) which continuously monitored the OD_{600} of each sample. To minimize sample evaporation, plates were covered and sealed with Parafilm®. Doubling times were calculated using timepoints corresponding to OD_{600} values between 0.02 and 0.2. For cultures performed at 1 L scale, cells

were grown in a 2.5 L Tunair® shake flask and incubated at 37˚C at 250 RPM. Except for experiments performed to identify optimal harvest OD_{600}, cultures were grown until the onset of stationary phase (an approx. OD_{600} of 6.5-7.5). 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to cultures of Vmax™ Express cells between OD_{600} 0.6-0.8 to induce expression of T7 RNA polymerase. In all cases, cells were pelleted by centrifuging for 15 min at 5000 × g at 4°C, washed three times with cold S30 buffer (10 mM tris-acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, 2 mM dithiothreitol)[354], and stored at -80°C until lysed.

5.3.3 Extract Preparation

Unless otherwise noted, cell pellets were thawed and suspended in 0.8 mL of S30 buffer per gram of wet cell mass. Prior to the optimization of lysis parameters, cell pellets were instead resuspended in 1.0 mL buffer per gram of cells. Following suspension, 1.4 mL of cell slurry was transferred into 1.5 mL microtubes. The cells were lysed using a Q125 Sonicator (Qsonica, Newtown, CT) with 3.175 mm diameter probe at a 20 kHz frequency and 50% amplitude. Sonication was continued for three cycles of 45s ON/59s OFF unless stated otherwise. To minimize heat damage during sonication, samples were placed in an ice-water bath. For each 1.4 mL sample, the input energy was ~270 Joules/sonication cycle. The lysate was then centrifuged at 12,000 × *g* at 4˚C for 10 min. The supernatant was flash-frozen using liquid nitrogen and stored at -80˚C until use. For preparations including a runoff reaction, following the first clarifying spin supernatant was transferred to a new tube and subjected to a 1 hr incubation at 30˚C or 37˚C, either stationary or with shaking at 250 RPM. Following this incubation, samples were centrifuged at 10,000 x *g* at 4˚C for 10 min after which supernatant was flash-frozen and stored at -80˚C until use.

5.3.4 CFPS Reaction

A modified PANOx-SP system was utilized for CFPS reactions. Briefly, a 15 µL CFPS reaction in a 2.0 mL microtube was prepared by mixing the following components: 1.2 mM ATP; 0.85 mM each of GTP, UTP, and CTP; 34 µg/mL folinic acid; 170 µg/mL of *E. coli* tRNA mixture; 13.3 µg/mL plasmid; 16 µg/mL T7 RNA polymerase; 3 mM for each of the 20 standard amino acids; 0.33 mM nicotinamide adenine dinucleotide (NAD); 0.27 mM coenzyme-A (CoA); 1.5 mM spermidine; 1 mM putrescine; 4 mM sodium oxalate; 290 mM potassium glutamate; 10 mM ammonium glutamate; 6 mM magnesium glutamate; 57 mM HEPES, pH 7.2; 67 mM phosphoenolpyruvate (PEP), and 4 µL (27% v/v) of cell extract. Each CFPS reaction was incubated for 20 hr at 30˚C unless noted otherwise. Experiments performed prior to determining an optimal reagent mix for *V. natriegens* lysates used the above mix with the following changes: 2 mM for each of the 20 standard amino acids, 130 mM potassium glutamate, 10 mM magnesium glutamate, and 33 mM PEP were used instead. As individual reagent concentrations were optimized, their optimal value listed above were used for all reactions from that point onward. For reactions involving pAzF incorporation, 2 mM pAzF was added to reactions as applicable. *E. coli* total tRNA mixture (from strain MRE600) and phosphoenolpyruvate was purchased from Roche Applied Science (Indianapolis, IN). ATP, GTP, CTP, UTP, 20 amino acids and other materials were purchased from Sigma (St. Louis, MO) without further purification. T7RNAP was purified in house as described previously[405]. To direct synthesis of a specific product, 200 ng of pJL1 template plasmid encoding the product was added to each reaction.

5.3.5 Quantification of Active sfGFP

CFPS reactions were diluted 1:25 in nanopure water and active full-length sfGFP protein yields were quantified by measuring fluorescence using a Synergy 2 plate reader (BioTek, Winooski, VT) with excitation at 485 nm, emission at 528 nm, and cut-off at 510 nm in 96-well half area black plates (Costar 3694; Corning, Corning, NY). sfGFP fluorescence units were converted to concentration using a standard curve established with ^{14}C-Leu quantified sfGFP as described previously[232].

5.3.6 CFPS Lyophilization

Samples were assembled in 2 mL microtubes and lyophilized overnight using a VirTis BenchTop Pro Freeze Dryer (SP Scientific, Warminster, PA). Lyophilized samples were stored at room temperature under vacuum in a dessicator with Drierite dessicant. For reconstitution of fully-assembled reactions, template plasmid DNA plus nuclease-free water were added to each sample. Samples consisting of only lyophilized lysate were reconstituted with the complete CFPS reagent mix. Where indicated, lyophilized reactions were supplemented with 2.5% (m/v) trehalose.

5.3.7 Quantification of Antimicrobial Peptide Yield in CFPS

Radioactive ^{14}C-Leucine was added into 15 µL CFPS reactions to a final concentration of 25 µM. After incubation, yields were quantified by determining radioactive ^{14}C-Leu incorporation into peptides precipitated in 15% (m/v) trichloroacetic acid (TCA)[354]. Radioactivity of TCA-precipitated samples was measured using liquid scintillation counting (MicroBeta2, PerkinElmer, Waltham, MA).

5.3.8 Generation of Negative Effector Knockout Library

The *V. natriegens* ATCC 14048 genome sequence (GenBank accessions CP016345 and CP016346, corresponding to chromosomes I and II, respectively) was screened for the presence of homologs to a subset of negative effectors known to limit the productivities of *E. coli*-based CFPS lysates (*endA*, *lon*, *mazF*, *ompT*, *rna*, *rnb*, *glpK*, *gor*, *gshA*, *tnaA*) using a tBLASTn search with *E. coli* gene sequences as the query **(Sequence List 5.1)**. As an additional confirmation of the identity of potential homologs, the gene annotation engine present in the Archetype® Genomics Discovery Suite software package (Synthetic Genomics, Inc.) was used to assign putative functions to the identified *V. natriegens* genes. Using this approach, likely homologs to *endA, lon, rnb, glpK, gor, gshA,* and *tnaA* were identified in *V. natriegens* **(Sequence List 5.2)**. Natural competence-mediated homologous recombination was used (essentially as described in Dalia, *et al*[453]) to generate the knockout strains. Briefly, *V. natriegens* cells rendered naturally competent were transformed with DNA cassettes composed of a chloramphenicol resistance gene bounded by 3 kb homology arms to the genetic loci of interest, resulting in the replacement of the putative negative effector gene with the chloramphenicol resistance marker. The resulting chloramphenicol-resistant transformants were screened by colony PCR to confirm the desired knockout.

Table 5.1. Strains and plasmids used in this study. Km^R and Cm^R are kanamycin and chloramphenicol resistance, respectively.

Strains and plasmids	Genotype/relevant characteristics	Source
Strains		
BL21 Star™ (DE3)	F^- *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm rne131* (DE3)	Life Technologies
V. natriegens	ATCC® 14048™	ATCC
Vmax™ Express	Δ*Dns**, IPTG-inducible T7 RNAP (Cat. No. CL1100)	Synthetic Genomics, Inc.
SGI-*VnatΔendA*	ATCC® 14048™ *endA*::Cm^R	This study
SGI-*VnatΔlon*	ATCC® 14048™ *lon*::Cm^R	This study
SGI-*VnatΔrnb*	ATCC® 14048™ *rnb*::Cm^R	This study
SGI-*VnatΔglpK*	ATCC® 14048™ *glpK*::Cm^R	This study
SGI-*VnatΔgor*	ATCC® 14048™ *gor*::Cm^R	This study
SGI-*VnatΔgshA*	ATCC® 14048™ *gshA*::Cm^R	This study
SGI-*VnatΔtnaA*	ATCC® 14048™ *tnaA*::Cm^R	This study
Plasmids		
pJL1-sfGFP	Km^R, P_{T7}, super folder green fluorescent protein (sfGFP), C-terminal strep-tag	[19]
pY71-sfGFPT216X	Km^R, P_{T7}, super folder green fluorescent protein with an amber codon at position 216, C-terminal strep-tag	[19]
pJL1-cecropinA	Km^R, P_{T7}, cecropin A	This study
pJL1-cecropinP1	Km^R, P_{T7}, cecropin P1	This study
pJL1-opistoporinI	Km^R, P_{T7}, opistoporin I	This study
pJL1-pyrrhocoricin	Km^R, P_{T7}, pyrrhocoricin	This study
pEVOL-pAzF.2.t1	Cm^R, P_{glnS}::*pAcFRS*, P_{araBAD}::*pAcFRS*, P_{proK}::o-$tRNA^{opt}$	[27]

**Dns* is the conventional name for the *endA* homolog in *Vibrio* species

Sequence List 5.1. Protein sequences of known *E. coli* effectors used to identify putative homologs in *V. natriegens*.

>endA
MYRYLSIAAVVLSAAFSGPALAEGINSFSQAKAAAVKVHADAPGTFYCGCKINWQGKKGVVDL
QSCGYQVRKNENRASRVEWEHVVPAWQFGHQRQCWQDGGRKNCAKDPVYRKMESDMHNLQ
PSVGEVNGDRGNFMYSQWNGGEGQYGQCAMKVDFKEKAAEPPARARGAIARTYFYMRDQYN
LTLSRQQTQLFNAWNKMYPVTDWECERDERIAKVQGNHNPYVQRACQARKS

>lon
MNPERSERIEIPVLPLRDVVVYPHMVIPLFVGREKSIRCLEAAMDHDKKIMLVAQKEASTDEPGV
NDLFTVGTVASILQMLKLPDGTVKVLVEGLQRARISALSDNGEHFSAKAEYLESPTIDEREQEVL
VRTAISQFEGYIKLNKKIPPEVLTSLNSIDDPARLADTIAAHMPLKLADKQSVLEMSDVNERLEYL
MAMMESEIDLLQVEKRIRNRVKKQMEKSQREYYLNEQMKAIQKELGEMDDAPDENEALKRKID
AAKMPKEAKEKAEAELQKLKMMSPMSAEATVVRGYIDWMVQVPWNARSKVKKDLRQAQEIL
DTDHYGLERVKDRILEYLAVQSRVNKIKGPILCLVGPPGVGKTSLGQSIAKATGRKYVRMALGG
VRDEAEIRGHRRTYIGSMPGKLIQKMAKVGVKNPLFLLDEIDKMSSDMRGDPASALLEVLDPEQ
NVAFSDHYLEVDYDLSDVMFVATSNSMNIPAPLLDRMEVIRLSGYTEDEKLNIAKRHLLPKQIER
NALKKGELTVDDSAIIGIIRYYTREAGVRGLEREISKLCRKAVKQLLLDKSLKHIEINGDNLHDYL
GVQRFDYGRADNENRVGQVTGLAWTEVGGDLLTIETACVPGKGKLTYTGSLGEVMQESIQAAL
TVVRARAEKLGINPDFYEKRDIHVHVPEGATPKDGPSAGIAMCTALVSCLTGNPVRADVAMTGE
ITLRGQVLPIGGLKEKLLAAHRGGIKTVLIPFENKRDLEEIPDNVIADLDIHPVKRIEEVLTLALQN
EPSGMQVVTAK

>mazF
MVSRYVPDMGDLIWVDFDPTKGSEQAGHRPAVVLSPFMYNNKTGMCLCVPCTTQSKGYPFEV
VLSGQERDGVALADQVKSIAWRARGATKKGTVAPEELQLIKAKINVLIG

>ompT
MRAKLLGIVLTTPIAISSFASTETLSFTPDNINADISLGTLSGKTKERVYLAEEGGRKVSQLDWKF
NNAAIIKGAINWDLMPQISIGAAGWTTLGSRGGNMVDQDWMDSSNPGTWTDESRHPDTQLNYA
NEFDLNIKGWLLNEPNYRLGLMAGYQESRYSFTARGGSYIYSSEEGFRDDIGSFPNGERAIGYKQ
RFKMPYIGLTGSYRYEDFELGGTFKYSGWVESSDNDEHYDPGKRITYRSKVKDQNYYSVAVNA
GYYVTPNAKVYVEGAWNRVTNKKGNTSLYDHNNNTSDYSKNGAGIENYNFITTAGLKYTF

>rna
MKAFWRNAALLAVSLLPFSSANALALQAKQYGDFDRYVLALSWQTGFCQSQHDRNRNERDEC
RLQTETTNKADFLTVHGLWPGLPKSVAARGVDERRWMRFGCATRPIPNLPEARASRMCSSPETG
LSLETAAKLSEVMPGAGGRSCLERYEYAKHGACFGFDPDAYFGTMVRLNQEIKESEAGKFLAD
NYGKTVSRRDFDAAFAKSWGKENVKAVKLTCQGNPAYLTEIQISIKADAINAPLSANSFLPQPHP
GNCGKTFVIDKAGY

>rnb
MFQDNPLLAQLKQQLHSQTPRAEGVVKATEKGFGFLEVDAQKSYFIPPPQMKKVMHGDRIIAVI
HSEKERESAEPEELVEPFLTRFVGKVQGKNDRLAIVPDHPLLKDAIPCRAARGLNHEFKEGDWA
VAEMRRHPLKGDRSFYAELTQYITFGDDHFVPWWVTLARHNLEKEAPDGVATEMLDEGLVRE

DLTALDFVTIDSASTEDMDDALFAKALPDDKLQLIVAIADPTAWIAEGSKLDKAAKIRAFTNYLP
GFNIPMLPRELSDDLCSLRANEVRPVLACRMTLSADGTIEDNIEFFAATIESKAKLVYDQVSDWL
ENTGDWQPESEAIAEQVRLLAQICQRRGEWRHNHALVFKDRPDYRFILGEKGEVLDIVAEPRRIA
NRIVEEAMIAANICAARVLRDKLGFGIYNVHMGFDPANADALAALLKTHGLHVDAEEVLTLDG
FCKLRRELDAQPTGFLDSRIRRFQSFAEISTEPGPHFGLGLEAYATWTSPIRKYGDMINHRLLKAV
IKGETATRPQDEITVQMAERRRLNRMAERDVGDWLYARFLKDKAGTDTRFAAEIVDISRGGMR
VRLVDNGAIAFIPAPFLHAVRDELVCSQENGTVQIKGETVYKVTDVIDVTIAEVRMETRSIIARPV
A

>glpK
MTEKKYIVALDQGTTSSRAVVMDHDANIISVSQREFEQIYPKPGWVEHDPMEIWATQSSTLVEV
LAKADISSDQIAAIGITNQRETTIVWEKETGKPIYNAIVWQCRRTAEICEHLKRDGLEDYIRSNTGL
VIDPYFSGTKVKWILDHVEGSRERARRGELLFGTVDTWLIWKMTQGRVHVTDYTNASRTMLFNI
HTLDWDDKMLEVLDIPREMLPEVRRSSEVYGQTNIGGKGGTRIPISGIAGDQQAALFGQLCVKE
GMAKNTYGTGCFMLMNTGEKAVKSENGLLTTIACGPTGEVNYALEGAVFMAGASIQWLRDEM
KLINDAYDSEYFATKVQNTNGVYVVPAFTGLGAPYWDPYARGAIFGLTRGVNANHIIRATLESIA
YQTRDVLEAMQADSGIRLHALRVDGGAVANNFLMQFQSDILGTRVERPEVREVTALGAAYLAG
LAVGFWQNLDELQEKAVIEREFRPGIETTERNYRYAGWKKAVKRAMAWEEHDE

>gor
MTKHYDYIAIGGGSGGIASINRAAMYGQKCALIEAKELGGTCVNVGCVPKKVMWHAAQIREAI
HMYGPDYGFDTTINKFNWETLIASRTAYIDRIHTSYENVLGKNNVDVIKGFARFVDAKTLEVNG
ETITADHILIATGGRPSHPDIPGVEYGIDSDGFFALPALPERVAVVGAGYIAVELAGVINGLGAKT
HLFVRKHAPLRSFDPMISETLVEVMNAEGPQLHTNAIPKAVVKNTDGSLTLELEDGRSETVDCLI
WAIGREPANDNINLEAAGVKTNEKGYIVVDKYQNTNIEGIYAVGDNTGAVELTPVAVAAGRRLS
ERLFNNKPDEHLDYSNIPTVVFSHPPIGTVGLTEPQAREQYGDDQVKVYKSSFTAMYTAVTTHR
QPCRMKLVCVGSEEKIVGIHGIGFGMDEMLQGFAVALKMGATKKDFDNTVAIHPTAAEEFVTM
R

>gshA
MIPDVSQALAWLEKHPQALKGIQRGLERETLRVNADGTLATTGHPEALGSALTHKWITTDFAEA
LLEFITPVDGDIEHMLTFMRDLHRYTARNMGDERMWPLSMPCYIAEGQDIELAQYGTSNTGRFK
TLYREGLKNRYGALMQTISGVHYNFSLPMAFWQAKCGDISGADAKEKISAGYFRVIRNYYRFG
WVIPYLFGASPAICSSFLQGKPTSLPFEKTECGMYYLPYATSLRLSDLGYTNKSQSNLGITFNDLY
EYVAGLKQAIKTPSEEYAKIGIEKDGKRLQINSNVLQIENELYAPIRPKRVTRSGESPSDALLRGGI
EYIEVRSLDINPFSPIGVDEQQVRFLDLFMVWCALADAPEMSSSELACTRVNWNRVILEGRKPGL
TLGIGCETAQFPLPQVGKDLFRDLKRVAQTLDSINGGEAYQKVCDELVACFDNPDLTFSARILRS
MIDTGIGGTGKAFAEAYRNLLREEPLEILREEDFVAEREASERRQQEMEAADTEPFAVWLEKHA

>tnaA
MENFKHLPEPFRIRVIEPVKRTTRAYREEAIIKSGMNPFLLDSEDVFIDLLTDSGTGAVTQSMQAA
MMRGDEAYSGSRSYYALAESVKNIFGYQYTIPTHQGRGAEQIYIPVLIKKREQEKGLDRSKMVA
FSNYFFDTTQGHSQINGCTVRNVYIKEAFDTGVRYDFKGNFDLEGLERGIEEVGPNNVPYIVATIT
SNSAGGQPVSLANLKAMYSIAKKYDIPVVMDSARFAENAYFIKQREAEYKDWTIEQITRETYKY
ADMLAMSAKKDAMVPMGGLLCMKDDSFFDVYTECRTLCVVQEGFPTYGGLEGGAMERLAVG

LYDGMNLDWLAYRIAQVQYLVDGLEEIGVVCQQAGGHAAFVDAGKLLPHIPADQFPAQALACE
LYKVAGIRAVEIGSFLLGRDPKTGKQLPCPAELLRLTIPRATYTQTHMDFIIEAFKHVKENAANIK
GLTFTYEPKVLRHFTAKLKEV

Sequence List 5.2. **Gene sequences for putative negative effectors in *V. natriegens* genome** (From GenBank accessions CP016345 and CP016346 corresponding to the *V. natriegens* chromosomes I and II, respectively).

>Vnat_endA
ATGAAATACCTGTTCTCTTTATTCATTCTTGCACTATCCAGTGCCGCCGTGGCCGCGCCACCA
AGTTCATTTTCAGCCGCTAAGCGCGAAGCGGTAAAAATCTATCAAGATCATCCCACCAGCTT
TTATTGCGGCTGTGATATTCAATGGCAAGGCAAGAAAGGCTTACCTGATCTTTCCTCTTGTGG
TTACCAGGTTCGCAAACAAGAAAAGCGTGCTTCACGCATCGAGTGGGAACATGTCGTTCCAG
CTTGGCAATTTGGGCACCAGCTGCAATGCTGGCAAAGCGGTGGTCGTAAAAACTGCTCGCGT
AATGACAAAACATTCCGCTCAATGGAAGCCGATCTGCACAACCTGACTCCTGCGATTGGTGA
GGTAAATGGTGATCGCTCTAACTACAATTTCAGTCAGTGGAATGGGATCGATGGCGCAACCT
ATGGTCGTTGTGAAGTCCAGGTAAACTTCAAGCAACGCAAAGTCATGCCACCCGATCGAGC
ACGCGGCTCCATCGCTCGTACCTATCTTTATATGAGCAAGGAGTACGGCTTCAAACTGTCCA
AGCAACAAACTCAGTTAATGAGTGCATGGAACAAAACCTACCCAGCCGATAAATGGGAATG
CGAACGCGATAAGCGCATTGCCAAAGTACAAGGCAACCATAATCCATTCGTTCAAGAGGCC
TGTCGCGCACTGTAA

>Vnat_lon
ATGAACTTGGAACGTTCCGAGAGTATCGAGATCCCGGTACTACCTCTACGTGACGTAGTGGT
CTACCCGCACATGGTCATTCCATTGTTTGTTGGTCGTGAAAAATCGATTAGCTGTCTAGAAAC
GGCGATGGAAACAAACAAACAAGTTCTGCTTGTGGCACAAAAGCAAGCGGATACAGACGAG
CCAACGGTTGACGACCTATTTGAGGTAGGTACGGTAGCCACCATTCTTCAGCTTTTAAAGCT
TCCTGATGGCACAGTAAAAGTACTGGTTGAAGGTCAGCAGCGTGCGAAGATTAATCACTTTA
AAGAGAGCGATTTTTTCTTAGCTGAAGCTGAATTCATCGTGACACCTGAGCTGGATGAGCGT
GAGCAAGAAGTTATTGTTCGTAGTGCGATCAACCAGTTCGAAGGCTTTATCAAGCTGAACAA
AAAAATCCCACCAGAAGTTCTGACTTCGCTAAATGGGATTGATGAAGCCGCGCGTCTAGCCG
ATACCATCGCAGCTCACATGCCTTTGAAATTGGTCGACAAGCAACAAGTACTTGAGATCATA
GACGTTACCGAGCGTTTGGAATTCCTGATGGGCCAAATGGAGTCGGAAATCGATCTGCTGCA
AGTTGAAAAACGCATCCGTGGCCGCGTTAAAAAGCAAATGGAGAAGTCTCAGCGCGAGTAC
TATCTGAATGAGCAAATGAAAGCGATTCAGAAAGAGCTAGGTGAGATGGAAGATGCGCCGG
ATGAATTTGAAACGCTGCAGAATAAAATCGAAGAGTCCAAAATGCCTCAAGAGGCGCGCGA
AAAGACAGAGCAAGAGCTACAAAAGCTTAAGATGATGTCTCCAATGTCTGCGGAAGCAACA
GTGGTGCGTAGCTACATCGATTGGATGGTGAGCGTTCCTTGGGCTAAGCGTTCTAAAGTGAA
AAAGAACCTGGCTAAGGCAGAAGAGATTCTAAACGAAGATCATTACGGTCTGGAGCGCGTC
AAAGAGCGCATTCTGGAATACTTGGCAGTACAAAACCGTATTAACAAGCTGAAAGGCCCAA
TCCTTTGTCTTGTTGGTCCTCCAGGTGTGGGTAAAACCTCTCTTGGCCGTTCGATCGCATCTG
CGACTGGGCGTAAATACGTGCGTATGGCGCTTGGTGGTGTGCGTGACGAAGCTGAGATTCGT
GGCCACCGTCGTACTTACATTGGCTCACTACCGGGTAAGCTTATCCAGAAAATGTCTAAAGT
TGGGGTTAAAAACCCGCTATTCCTTCTTGATGAGATCGATAAGATGTCTTCTGACATGCGCG
GTGATCCAGCATCTGCACTGCTAGAAGTTCTGGATCCTGAACAAAACAACTCGTTCAACGAT
CACTACTTAGAAGTAGATTACGATCTGTCGGATGTAATGTTCGTTGCGACGTCTAACTCGAT

GAATATCCCAGGTCCGCTTCTTGACCGTATGGAAGTTATCCGTCTTTCTGGTTACACAGAAG
ATGAAAAACTGAACATCGCGAAACGCCACTTGGTAGATAAGCAGTTGAAGCGTAACGGACT
GAAGCCAAACGAGATTGTTATCGAGGACTCAGCGATTGTCGGCATTATTCGTTACTACACTC
GTGAAGCGGGTGTACGTAACCTAGAGCGTGAAATTTCTAAGATCTGTCGTAAGGCGGTGAA
GAATATCCTGCTGGATAAAGATATCAAGTCTGTGACCGTATCTATGGACAACCTGAAAGAGT
ACCTGGGTGTTCAGCGTTTTGACTATGGTAAAGCTGATGAAAGCAACCGTATTGGTCAGGTG
ACGGGTTTGGCTTGGACTGAAGTCGGTGGTGATTTGCTAACTATTGAAACTCAGTCTATGCC
GGGTAAAGGTAAACTGACACAGACGGGGTCTCTTGGCGACGTGATGCAAGAGTCTATCCAA
GCGGCCATGACAGTTGTTCGCTCTCGTGCTGAAAAGCTGGGTATCAACACCGATTTCTACGA
GAAGAAAGACATCCACGTGCATGTTCCTGAAGGTGCGACACCAAAAGATGGCCCAAGTGCT
GGTACAGCAATGTGCACTGCTTTGGTTTCTGCGTTAACTGGTAACCCAGTGAAAGCGGAAGT
GGCAATGACGGGTGAAATCACACTACGTGGTGAAGTTTTGCCTATCGGTGGCCTAAAAGAA
AAATTACTTGCGGCACATCGTGGCGGCATTAAAACGGTACTGATTCCAAAAGATAACGAGC
GTGATTTGGAAGAGATTCCAGAGAATGTTATCGCAGATCTGACAGTTATCCCGGTTCAGTGG
ATTGATGAAGTACTGAAAGTTGCACTCGAGCGAGACCCGACGGGCGTTGAGTTTGAAGCTA
AAAAATAG

>Vnat_rnb
ATGTTTCAAGATAACCCGCTATTAGCCCAACTTAAGCAGCAAATCCAAGAAAACCTTCCTAA
AAAAGAAGGGTCAATCAAAGCAACAGATAAAGGCTTTGGTTTCCTTGAAGTGGACAGCAAA
ACCAGTTTCTTCATTCCACCTGCGTACATGAAAAAGTGCATTCATGGCGATAAAGTAGTCGC
CATTATTCGTACAGAGAATGAACGTGAAGTTGCAGAGCCACAAGAGCTGCTAGAACAGTCA
CTTACTCGCTTTATTGGCCGTGTAAAAATGTTCAAAGGTAAGCTGAATGTTGTTCCTGACCAC
CCTCAACTGAAAAAGTTGTCGCTAAAAGCAAAACTAAAGAAAGGCCAAAAGCCAGATAACT
TCAATGAAGGTGATTGGGTTGTCGGCCATCTTATTCGCCACCCTCTTAAAGGTGACAACTCTT
TCTTCGTAGAAATCTCTGAAAAAATCACGGATGCTGACGATAAGATCGCACCATGGTGGGTA
ACCTTGGCACAAAACGATCTTCCAAACTCTGAGCCTGCTGGTATTGAGAACTGGGAGTTAAA
AGACGACGCGGAATTAGACCGCGTTGATTTAACTCACGTTCCATTCGTAACTATCGATGGCG
AGTCCACCAAAGATATGGACGATGCGCTGCATGCGAAGAAAACAGAATCTGGCGATTTTGA
ACTGACCATCGCGATTGCCGATCCTACCGCATACATTACGCCAGAAGATGAAATGGACAAA
GTCGCTCGTGAGCGTGGCTACACGATCTACTTGCCAGGGCGCAACATCCCAATGCTGCCTCG
CGATCTTGCTGATGACCTATGTTCTCTTATTGAAGGTGAAACTCGCCCTGCCCTTTGCTGTAC
CGTAAGCGTAAGTAAAGATGGTGTGATTGGTGATAACATCAACTTCTTTGCTGCGAACATTA
AGTCTCACGCTCGTCTTGCTTATGACCACGTATCAGACTGGATCGAAAATGGCAGCTCTGAT
AAATGGCAGCCATCGGAAGATATCGCAACTATCGTACGTGACCTGTACGACTTCTCCGTTGC
TCGTGCCGACTGGCGTGAAAAGAACGCGGTTGTATTCCCGGACCGTCCGGACTACCGTTTCG
AACTTAGCGAAGACAATGACGTTATTGCCATTCATGCAGACATGCGTCGCAGTGCAAACCGT
CTTGTAGAAGAGTCGATGATCACAGCGAATATCTGTGCTGGCCGCACGCTTCGCGATAAGTT
TGAAACTGGCGTGTTCAACACTCACTCTGGTCTCAAAGCAGAGAAGATTGAAGAAGTTGTAC
AATTAGTTGACCCAGAAGGTACACATGGCTTCACAGCGGACACTATCGCGACACTGGAAGG
CTTTGCAGCATTACGTCGTTGGTTGTCTACTCAGGAAACTTCATACTTGGATAACCGCATCCG
CAAGTTCCAGGCGTACAGTGAAGTCGGCAATCAGCCACTTCCTCACTACGCGATGGGCCTAG
ATATTTACGCGACTTGGACATCTCCAATTCGTAAATACGGCGATATGATCAACCACCGTATG
CTAAAAGCGGTGATTCTTGACAAAGAACCTGTTCAAAAGCCTGATGATCAAGTCGGTGAAG
AACTGGCATTACACCGTAAGCACCACAAAATTGCGGAGCGTAATGTGTCTGACTGGCTTTAC

GCTCGTACTCTTGCTGAGGAGCCAAGCAAACAAACTCGCTACATCGGTGAAATTTTCGATAT
CAATCGTGCTGGTGCGCGTGTTCGCCTGCTTGAAAACGGCGCCGCGGCGTTTATTCCTGGCT
CACTCATTGTTGACAATAAAGAACGCATTGAGTGCAACGGTGACAACGGTACTATCTCCATC
GATAAAGAAGTGGTGTACAAGCTGGGCGATACGCTAGAAGTTGTCTTGGCAGATGTGAACC
AGGAAAACCGCAGTTTAGTTGCTAAACCGACGCAAGTGTTCGCTGAGCCGCCAAAAGCACA
GACGGAACAAACGGTTGAATAA

>Vnat_glpK
ATGACTGAGCAAAAATACATTGTTGCCCTAGACCAAGGTACGACAAGCTCTCGCGCAGTAAT
TCTTGATCACGACGCGAATATCGTCAGCGTCGCTCAGCGCGAATTTACTCAGATTTATCCGC
AAGCAGGTTGGGTTGAGCACGATCCAATGGAAATCTGGGCAACACAAAGTTCAACCTTAGT
TGAAGCTTTAGCTAAGTCAGGCATCCGCAGCGATCAATTAGCGGCTATTGGCATCACAAACC
AACGTGAAACAACGATCGTCTGGAACAAAGAAACAGGTAAACCGGTTTACAACGCTATCGT
CTGGCAATGTCGACGCACCGCAGAAATTTGCGAAGACTTAAAACGCCGCGGCTTAGAAGAC
TACGTACGTGACAACACAGGCTTAGTGCTTGACCCTTACTTCTCTGGCACCAAAGTGAAGTG
GATTCTTGATAACGTCGAAGGTGCCCGCGAAGATGCCGAAGCAGGAAAACTGTTATTTGGTA
CGGTTGATACCTGGCTGGTGTGGAAAATGACACAAGGCCGTGTGCACGTAACGGATTACAC
CAACGCTTCACGTACGATGCTATTTAATATCAATGACCTATGTTGGGACCAAAAGCTGTTGG
ATGAGTTGGGTATTCCTGCGTCAATGATGCCTGAAGTAAAACGTTCATCTGAAATCTATGGC
AAAACCAACATTGGTGGTAAAGGTGGTACGCGTATTCCTATTGCGGGTATTGCTGGTGACCA
ACAAGCTGCGCTATACGGCCAGATGTGTGTTGAAGCAGGTCAGGCAAAGAACACCTACGGT
ACAGGTTGCTTCTTGTTGATGAATACTGGTCAGGAAAAAGTGACGTCTACACACGGCCTGCT
GACAACACTGGCATGTGGCCCGAAAGGTGAACCGGCATACGCACTAGAAGGCGCGGTGTTC
ATGGGTGGTGCGTCTATTCAGTGGCTGCGTGACGAGCTTAAGATTCTTAATGGCGCAGAAGA
CTCTGAATACTTCGCAACAAAAGTAGACACATCCAATGGTGTGTACGTCGTGCCAGCCTTTA
CTGGCCTTGGCGCACCATACTGGGATGCGTACGCACGAGGAACCATTGTTGGCTTAACTCGA
GGCGTTAACTCAAACCATATTATTCGTGCGACTTTGGAAGGTATCGCTTACCAAACCCGTGA
CGTATTGGATGCAATGCAAGCCGACTCTGGAATCCAGTTGGCTAACCTAAGAGTTGACGGCG
GCGCAGTAGCAAACAACTTCTTGATGCAATTCCAGTCTGACGTACTAAATACGCAAGTACTT
CGCCCTGAAGTCACTGAAGTAACCGCTCTGGGTGCAGCTTATCTGGCAGGCTTAGCGGTAGG
ATATTGGGATAGCCTCGATGAACTGCAAGGTAAAGCGGTAATTGATCGTACATTTGAGCCTC
ATGATGATGAAGAGAAGCGTAATCGTCGCTACAAAGGCTGGAAGCGTGCAGTGAAATGTGC
TCAAACTTGGTCAGAACTTCACGACGAAGACGATTAA

>Vnat_gor
ATGGCGACTCATTTTGACTATATCTGTATCGGTGGCGGCAGTGGCGGCATCGCATCTGCAAA
CCGTGCAGCCATGTACGGCGCGAAAGTAGCGCTGATTGAAGCACAAGACCTTGGTGGCACC
TGTGTAAACGTAGGTTGTGTACCAAAGAAAGTGATGTGGCATGGCGCGCAAATCGCAGAAG
CAATGAACCTGTACGCGGAAGATTATGGGTTTGATGTCGACGTGAAAGGTTTCGACTGGAGC
AAGCTGGTAGAGAGTCGTCAGGCGTACATTGGTCGTATTCACCAATCTTACGACCGTGTTCT
AGGCAACAACAAAGTAAATGTTATCAAAGGCTTTGCTAAGTTTGTTGACGAAAAAACCGTTG
AAGTAAACGGTGAACACTACACGGCCGATCATATCCTGATCGCTGTTGGTGGCCGTCCAACT
ATTCCAAACATCCCGGGCGCAGAATACGGCATCGATTCAAACGGCTTCTTCGACCTGGCTGA
GCAACCAAAACGCGTAGCGGTTGTTGGTGCAGGCTACATCGCAGTTGAAATCGCGGGCGTG
CTACACGCACTAGGTACAGAAACACACCTGTTCGTACGTAAAGAATCGCCACTGCGTAGCTT

CGATCCAATGATCATCGAAACGTTGGTTGAAGTGATGGACGCTGAAGGTCCAAAACTGCAC
ACCCATTCTGTACCAAAAGAAATCATTAAAGAAGCAGATGGCACTCTGACTCTGCACCTGGA
AAATGGTGAAAGCCAAAACGTTGACCAGCTAATCTGGGCAATCGGCCGTCACCCAACAACA
GACGCTATCAACCTAGCATCAACTGGCGTTGCAACCAACGACAAAGGCTACATCAAGGTAG
ACGAATACCAAGAAACCAACGTGAAAGGCATCTACTGTGTAGGTGACATCATGGAAGGTGG
TATCGAGCTAACACCTGTAGCAGTGAAAGCAGGTCGTCAGCTTTCTGAGCGTTTGTTCAACA
ACAAGCCAAACGCGAAGATGGATTACGACCTCGTTCCGACTGTCGTATTCAGCCACCCACCA
ATTGGTACCATCGGTCTGACTGAGCCAGAAGCGATTGCGAAGTACGGCGAAGACAACGTGA
AAGTGTACACCTCTGGCTTCACTGCAATGTACACCGCGGTCACTAAGCACCGTCAACCATGT
AAGATGAAGCTGGTATGTGCTGGTGAAGAAGAAACGGTAGTCGGCCTGCACGGCATCGGTT
TCACGGTAGACGAAATGATTCAGGGCTTTGGCGTAGCAATGAAGATGGGTGCAACTAAAGC
AGACTTCGACTCTGTTGTAGCAATTCACCCAACGGGCTCGGAAGAGTTCGTTACTATGAGAT
AG

>Vnat_gshA
TTGACTGATTTTGCTGCGCGACTGGAAAAAGTTGCATCAAACCCGGAAGTATTTAAACAGTT
TGGACGCGGTGTTGAGCGTGAAACGTTACGCTATCGTCAGGATGGACAGCTAGCAACAACA
CCTCATCCAGAGGGATTGGGCTCAGCGTTCACAAACCAGTGGATTACCACGGACTTTTCAGA
GTCGTTACTGGAGTTCATTACTCCGGTTTCTCATGATGTTCCGGAGCTAATGGCACAACTGAA
AGATATTCATCACTTTACTCAAACTAAAATGGGTGAAGAAAAAATGTGGCCGCTTTCTATGC
CATGTTATGTCGCCAGTGAAGATAATATTAATCTGGCACAGTACGGATCGTCTAACGCAGCT
CGAATGAAAACGCTCTACCGAGAGGGTTTGAAACGCCGTTATGGCAGCTTAATGCAGATCAT
TTCGGGTGTTCACTTCAACTTCTCGTTCCCAGAATCGTTTTGGGATGCCCTATATGGTGAGCA
GGATGAAGAGGCTCGTCAGGAAACCAAATCCGATGCCTATTTTGCCCTCATTCGTAACTACT
ATCGTTTTGGTTGGATGATTCCCTACTTCTTTGGTGCTTCGCCTGCGTTGTGTGGTTCGTTTAT
TCAAGGCCGAGAAACAAGCTTACCGTTTGAGAGCTTGGGTGGAACGTTATTCTTACCGAAAT
CAACGTCTTTGCGTCTGAGCGACCTTGGTTACACGAATAATGCACAGAGTTCGCTAAAGATT
GGCTTCAATAGTATTGACCAGTATCTGGAAGGTTTAAGTGATGCTATTCGTCGTCCGTCAGA
AGAGTTCGCAAAAATTGGTGTGAAAGTTGATGGCGAGTACCGTCAGCTCAATTCGAATGTGT
TGCAAATAGAAAACGAATTGTACGCGCCAATTCGCCCTAAACGAGTGGCTAAAAGTGGTGA
GAAACCATCAGAAGCGTTAAAGCGTGCGGGTGTTGAATATATTGAAGTTCGTTCATTGGACG
TGAACCCATTCAGCCCAGTAGGCATAACTGAAGAGCAAGTTCGCTTCCTCGACCTGTTCCTG
ACTTGGGCAGCACTGTCAGACTCAGAACCAATGGATAACTGTGAACTGGAGTGTTGGCGTG
ATAACTGGAACAAAGTCATTGTATCGGGCCGTGAAAAAGGCTTGATGCTTCAGATCGGTTGC
CAAGGTGAGCGCTTACCTCTACAAGAGTGGGCTCACCGCGTGTTTGCAGATCTACGCCAAAT
TGCTGTGATGATGGATGAAGTGAACGGTGATAATGCTTACCAAGAGGTTTGCGATAAGTTAA
CTGGCTGGATTGATGAACCTGAATTGACGACTTCTGGTCAATTGCTGGAACTGACCAAGGAG
TTGGGCGGCTTAGGTAAAGTAGGTTGTTCACTTGGTATGAAGCATCGTGAAGATAACCTGAA
TCACGGCTACCAGCATTACTCACAAGAAGTGATGGAACAGGAAGCGTTAGCATCGGTTGAA
AAGCAAAAGCAAGCAGAGCTAAGTGACACGATGTCTTTTGATGACTTCCTGGAAGACTATTT
TTCTTATTTAAAACAATAA

>Vnat_tnaA
ATGGAAAACTTTAAACACTTACCTGAGCCGTTTCGTATTCGCGTTATTGAACCAGTAAAACG
CACTACACGTGAATATAGAGAAGAAGCGATTCTAAAAGCGGGTATGAACCCTTTTCTTCTGG

ATAGCGAAGATGTGTTTATCGACCTTCTTACCGATAGTGGTACAGGCGCTATTACGCAAGAG
ATGCAAGCAGCCATGTTGCGCGGCGATGAAGCATACAGCGGCAGCCGAAGCTACCATGCAC
TTCAACAGGCCGTTGAGGATATCTTTGGCTACCAATACACCATTCCAACTCACCAAGGACGT
GGTGCAGAGCAAATTTACATTCCGGTTCTGATTAAAAAACGTGAGAAAGAGAAAGAACTAG
ATAGAAGCAAGATGGTTGCGCTGTCGAATTACTTTTTTGACACCACACAAGGCCATACACAG
TTAAACTGCTGTGTTGCGAAAAACGTTTATACCGAAGAAGCATTCGACACTTCAATCGAAGC
AGACTTTAAAGGTAACTTTGATCTGGAGAAACTAGAACAGGCAATCCTAGAAGCTGGCGCA
GCAAATGTACCTTATATTGTCAGCACCATCACATGTAACTCTGCCGGCGGCCAGCCAGTATC
GCTTGCTAACCTAAAAGCGGTTTATGAAATTGGTCAGAAATACGATATTCCAGTGATCATGG
ATTCGGCTCGTTTTGCCGAGAATGCTTACTTCATTCAACAACGTGAAGCTGGCTACGCTGACT
GGTCAATCGAAGCAATCACCAAAGAGAGCTACAAATACGCGGATGGTCTGGCGATGTCGGC
AAAGAAAGATGCCATGGTTCAAATGGGCGGCTTACTTTGCTTTAAAGACGATTCGTTAATGG
ACGTGTACACGGAATGCCGCACGCTGTGTGTTGTTCAGGAAGGCTTTCCGACCTATGGTGGT
CTGGAAGGTGGCGCTATGGAGCGTCTTGCCGTAGGCCTTTATGATGGCATGCGTCAGGACTG
GCTGGCATACCGAATCGGTCAGGTTCAGTACCTGGTTGATAAGCTGGAAGCGATTGGCATCG
TTTGTCAGCAAGCGGGTGGCCATGCTGCCTTCGTTGACGCAAGCAAACTCCTGCCTCACATT
CCTGCAGAGCAATTCCCGGCACATGCACTGGCGTGCGAGCTTTATAAAGTCGCTGGTATTCG
AGCGGTAGAAATAGGCTCCCTTCTATTGGGCCGAGATCCTGCGACAGGTAAACAACACCCGT
GTCCGGCTGAACTGCTTCGCCTTACCATTCCTCGCGCTACCTACACACAAACGCATATGGATT
TCATTGTCGAAGCATTCCAGGCGGTAAAAGAAAATGCAGCCAATGTAAAAGGTCTGGACTTT
ACTTATGAACCGGAAGTTTTACGTCACTTCACCGCGCGTCTAAAAGAGGTTGAGTCAGAGAC
TCAAGCTCCATCCATCACGACAGCAGAAACAGTATAA

5.4 Results

5.4.1 Identifying Extract Preparation Conditions for *V. natriegens* CFPS

We began our study by trying to develop extract preparation procedures. Because CFPS exploits an ensemble of catalytic proteins prepared from the crude lysate of cells, the cell extract (whose composition is sensitive to growth media, lysis method, and processing conditions) is the most critical component of extract-based CFPS reactions. In recent years, systematic optimization of

each step in extract preparation for *E. coli* CFPS has improved extract robustness and productivity[280, 291]. Similar advances have been made in *S. cerevisiae, Streptomyces,* and *Pseudomonas* CFPS systems[299, 447, 454]. Based on these successes, we chose to vary extract preparation conditions in search of parameters that might improve reproducibility between extract preps, increase the level of protein synthesized, and allow for potential downstream scalability. A key focus was to generate a large volumetric yield of lysate even when chassis cells are cultured at volumes that can be accommodated by shake flasks.

Generally, the extract preparation process includes the following major steps: cell cultivation, cell disruption, lysate clarification, and some optional steps like run-off reaction and dialysis. We decided to explore each of these steps. First, we wanted to confirm the previously reported doubling times for *V. natriegens* (*Vnat*). To accomplish this, we prepared liquid cultures of wild type *Vnat* cells in several different liquid growth media. Each growth medium was tested with and without v2 salt supplementation[29], and culture growth was monitored via plate reader over a 20 hour incubation **(Figure 5.2.A)**. Under the conditions tested, *Vnat* was found to double approximately every 12-14 minutes in salt-supplemented growth medium. This range aligns with previously reported data and outpaces standard *E. coli* laboratory strains[28, 29]. Unsurprisingly, in most of the media investigated, the growth of *Vnat* is severely inhibited in the absence of salt. This held true at culturing volumes of 1 L incubated in shake flasks, which are typical volumes and conditions used to culture cells for lysate preparation **(Figure 5.2.B)**.

a)

Growth media	+ v2 salts	- v2 salts
LB	14.1	no growth
BHI	**12.4**	no growth
2xYTPG	14.5	25.6

b)

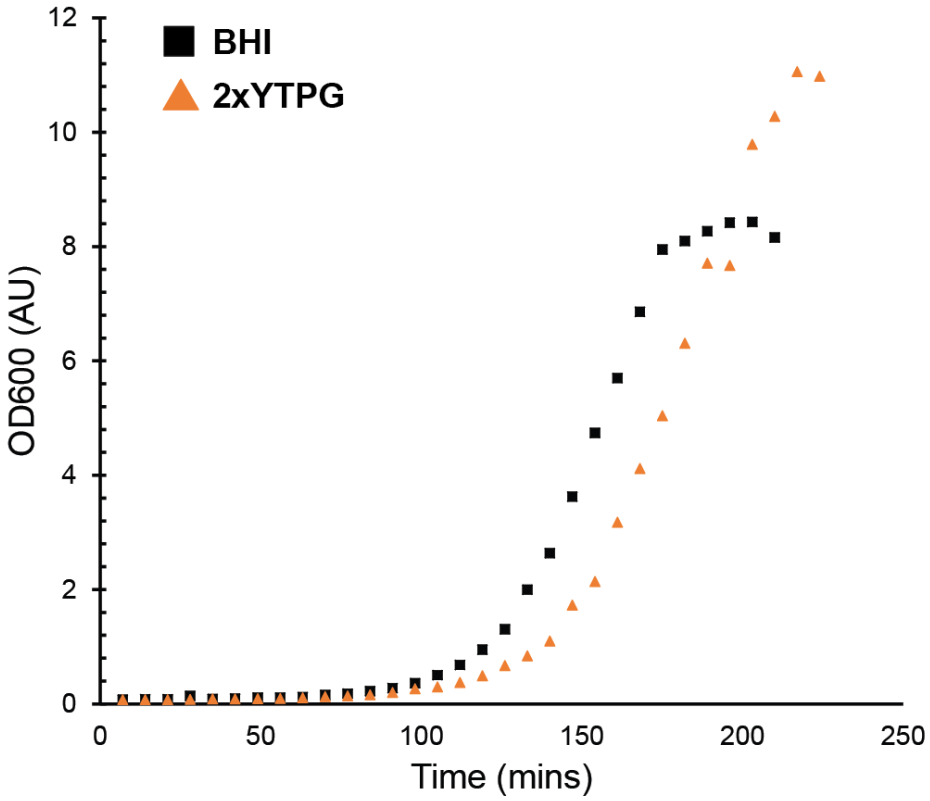


Figure 5.2. Characterization of *V. natriegens* growth rate. **(a)** Doubling time, in minutes, of *V. natriegens* grown in the specified media and salt conditions at 37°C. v2 salts = 204 mM NaCl, 4.2 mM KCl, 23.14 mM $MgCl_2$. **(b)** Timecourse of *V. natriegens* growth in liquid culture at 1L scale using the listed growth media. Cells were grown in a 2.5 L Tunair® shake flask and incubated at 37˚C at 250 RPM.

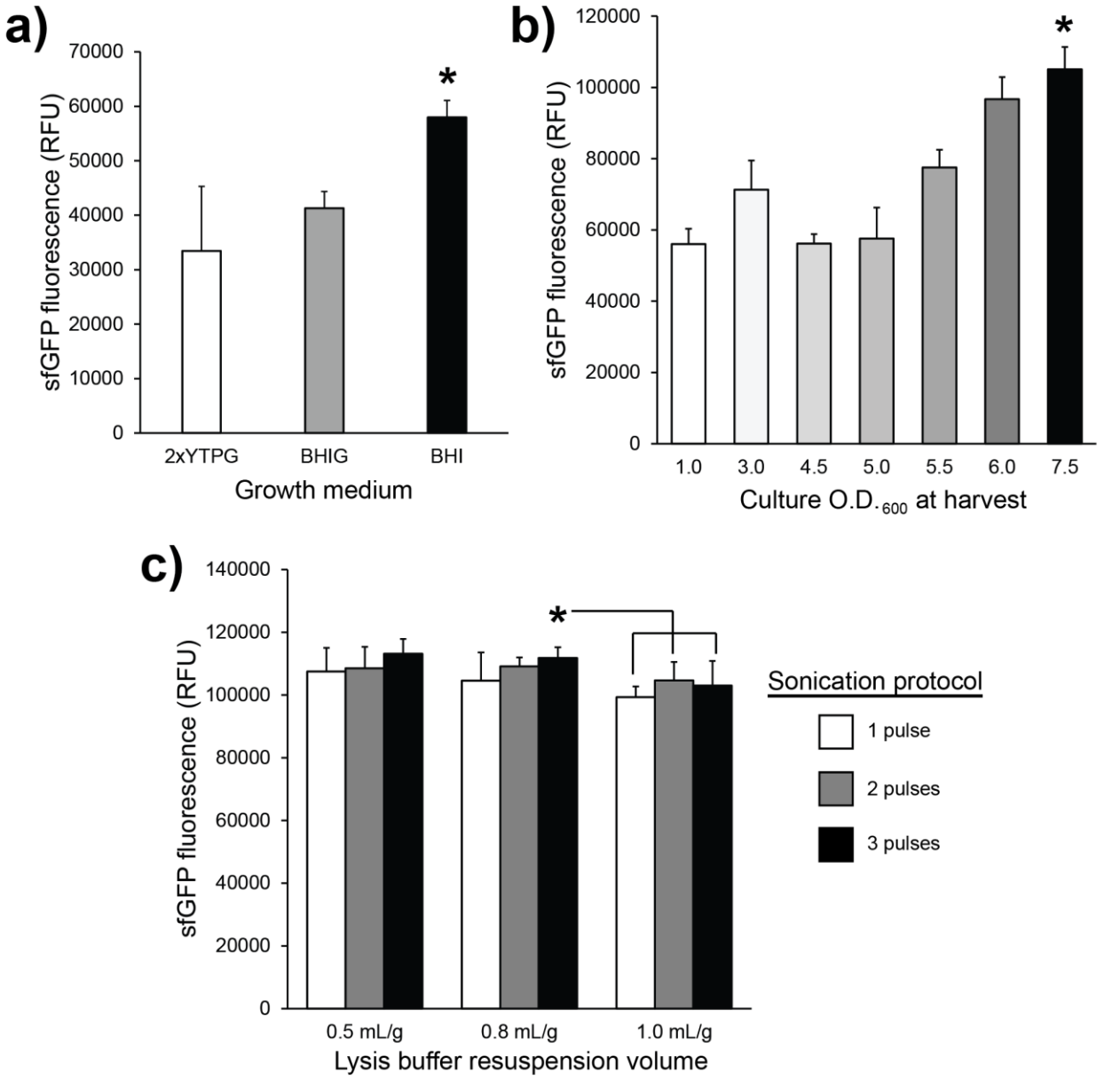


Figure 5.3. Optimization of *V. natriegens* harvest and lysis procedures. **(a)** sfGFP fluorescence *in vitro* from cell extracts derived from *V. natriegens* cultured in the specified liquid medium supplemented with v2 salts (204 mM NaCl, 4.2 mM KCl, 23.14 mM $MgCl_2$). Cells were harvested at OD_{600} = 4.5 **(b)** sfGFP fluorescence from cell extracts derived from *V. natriegens* harvested at the specified OD_{600}. Cell culture was performed using BHI medium plus v2 salts. For both **(a)** and **(b)**, established *E. coli* protocols were used for lysate preparation. **(c)** Lysis optimization. *V. natriegens* cell pellets were suspended in the specified volume of S30 lysis buffer and subjected to the indicated number of sonication pulses to achieve lysis. sfGFP fluorescence *in vitro* from the resulting lysates is shown. All CFPS reactions used an existing *E. coli* reagent mix. For all conditions, three independent reactions were performed and one standard deviation is shown. * = statistically significant for $p < 0.05$.

We next set out to establish the growth medium that is optimal for culturing cells for *Vnat* CFPS lysate preparation. 1 L cultures of *Vnat* were grown in 2xYTPG (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L K_2HPO_4, 3 g/L KH_2PO_4, 18 g/L glucose), BHI (brain-heart infusion), and a BHI variant supplemented with 1.8% (m/v) glucose (BHIG), in all cases supplemented with v2 salts. Cell pellets were collected at OD_{600} = 4.5 to mimic the best practice of harvesting *E. coli* cultures during mid-exponential phase, and protocols previously established for *E. coli* were used for lysate preparation and CFPS synthesis of sfGFP. Specifically, we assembled CFPS reactions with the sfGFP template and carried out 15 µL batch reactions for 20 h at 30°C. Lysates derived from cells cultured in BHI were significantly more productive than those from cells grown in either of the other two media types **(Figure 5.3.A)**. Based on these cumulative results, we selected BHI for use in *Vnat* cultures going forward.

Next, we investigated the ideal time point at which to collect *Vnat* cell pellets for CFPS lysate production. As previously mentioned, in *E. coli*-based systems, the most productive lysates are derived from cells gathered during mid-exponential phase growth[232, 291, 405]. It is generally accepted that this is because the pool of active ribosomes is most enriched during this phase of the growth cycle. We therefore hypothesized that the most productive *Vnat* lysates would similarly be derived from cells harvested in mid-exponential phase. To test this, we harvested *Vnat* cultures at a range of optical densities ranging from lag phase, through exponential phase, and even into stationary phase. Lysates were prepared from each cell pellet and directed to synthesize sfGFP in CFPS **(Figure 5.3.B)**. We were surprised to observe that *Vnat* lysate productivity increases with increasing OD_{600} past the exponential phase of growth—indeed, the most productive lysate identified was prepared from cells in early stationary phase. Lysates retain ~85% productivity

when prepared from cultures at stationary phase for several hours **(Figure 5.4.A)** and ~75% productivity when prepared from overnight cultures **(Figure 5.4.B)**. These surprising results run counter to what is typically observed with other CFPS chassis organisms. In *E. coli*, for instance, it is believed that stationary phase cells experience a reduction in active ribosomes in response to the reduced demand for protein synthesis; this effect propagates to lysates, resulting in a severe reduction in CFPS productivity[455]. Going forward, *Vnat* cultures were harvested for lysate preparation immediately upon entry into stationary phase (OD_{600} ~ 7.5).

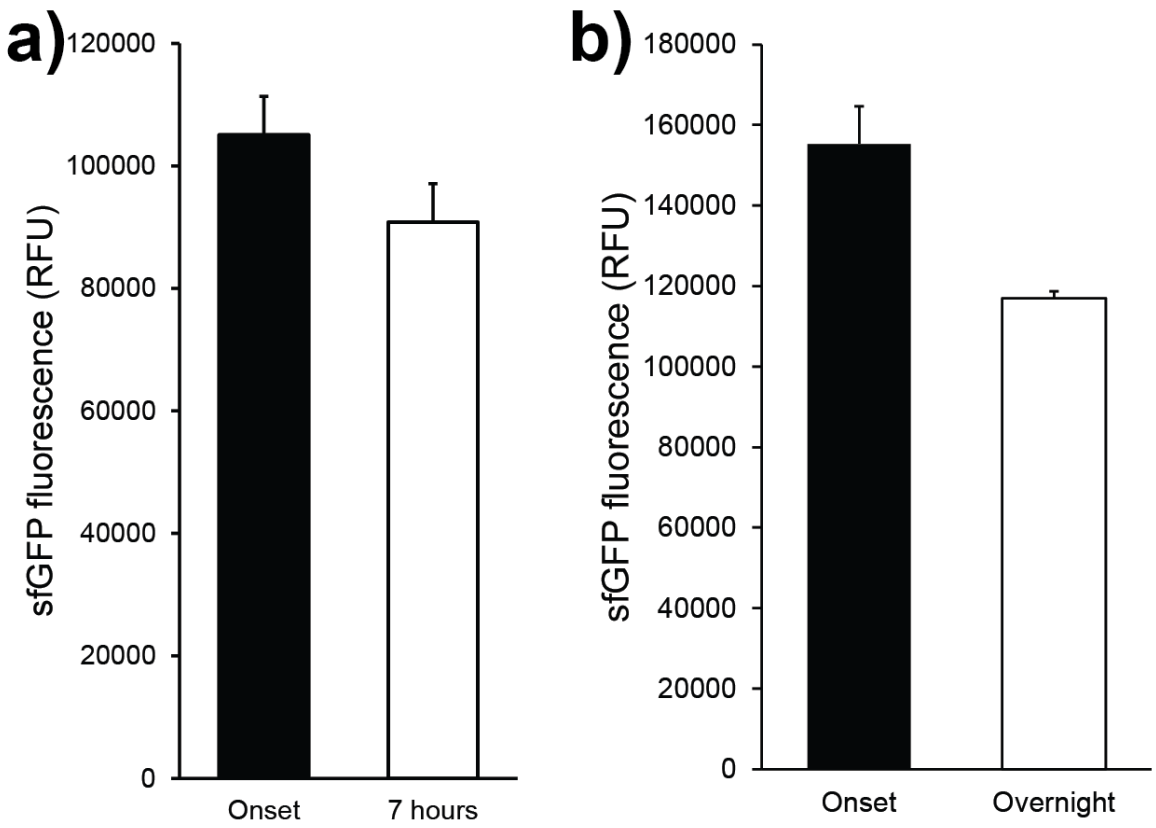


Figure 5.4. CFPS yields from *V. natriegens* cells harvested at late stages of growth. (a) sfGFP fluorescence values from a lysate that was prepared from *V. natriegens* cells that remained in stationary phase for 7 hours, compared to a lysate prepared from cells gathered right at the onset of stationary phase. **(b)** sfGFP fluorescence values from a lysate that was prepared from *V. natriegens* cells that were inoculated near the end of the day and allowed to grow at 37°C overnight (total culture time = 16 hours). The result is compared to a lysate prepared from cells gathered right at the onset of stationary phase. For each condition 3 independent reactions were performed, and one standard deviation is shown.

5.4.2 Identification of Optimal Procedures for Preparation of *V. natriegens* Lysates

Having established *Vnat*-specific cell culture and harvest parameters, we proceeded to identify conditions for preparing *Vnat* lysates via sonication that maximized CFPS yields. We focused our investigation on two key factors pertaining to lysate preparation. The first of these, cell pellet resuspension volume, describes the volume of lysis buffer used to resuspend a cell pellet prior to lysis. Modulating this volume has a direct influence on the concentration of cellular components in the final lysate, which in turn affects lysate productivity. The second factor considered was energy delivery to the cells during sonication. During lysis, enough energy must be delivered to ensure adequate rupturing of cellular compartments but must then be constrained to prevent denaturation of ribosomes, enzymes, and other fragile cellular components required for robust protein synthesis. Both of these parameters were simultaneously varied for lysis optimization, as we have done before when developing a similar protocol for *E. coli*[291]. *Vnat* cell pellets were resuspended in 0.5, 0.8, or 1.0 mL of S30 buffer per gram of wet cell mass. Each suspension was then sonicated, with lysis achieved using one, two, or three 45-second pulses. Each resulting lysate was directed in CFPS to synthesize sfGFP **(Figure 5.3.C)**.

Collectively, the results reveal that *Vnat* cells are relatively agnostic to both resuspension volume and lysis energy. Productivities of all nine lysates tested were within 10% of one another. The more concentrated suspensions demonstrate a modest (but statistically significant) increase in productivity relative to the samples prepared from cells resupended in 1.0 mL buffer/g cells, likely a consequence of slightly more concentrated translation components. There is no appreciable difference between samples suspended in 0.5 mL buffer/g cells vs. 0.8 mL buffer/g cells. Due to

the relative difficulty of resuspending cells in 0.5 mL buffer/g as well as the accompanying reduction in the volume of lysate yielded under that condition, we selected 0.8 mL buffer/g cells as our resuspension density going forward. Due to the insignificant differences in productivity when two or three sonication pulses are used for lysis, we settled on the use of three pulses to remain consistent with several recently-reported *E. coli* lysis protocols[232, 405].

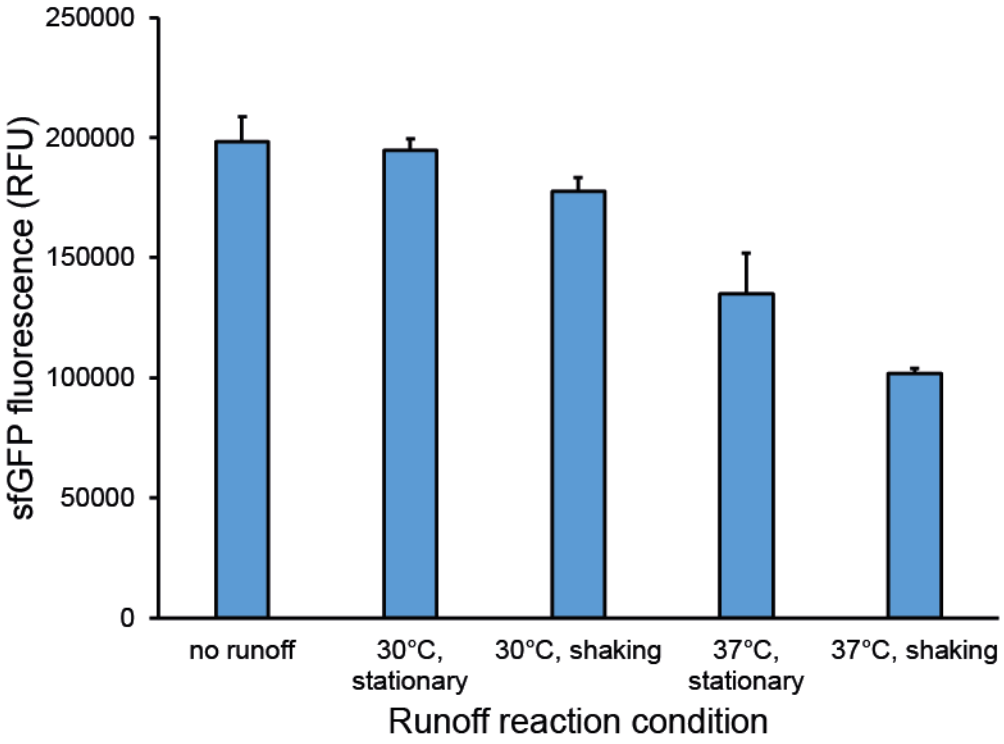


Figure 5.5. Characterization of *V. natriegens* run-off reaction. sfGFP fluorescence from lysates prepared using run-off reactions performed under the specified conditions is shown. Run-off reactions, if performed, ran for 1 hr. Shaking = agitated at 250 RPM. For each condition 3 independent reactions were performed, and one standard deviation is shown.

After we defined a reproducible cell lysis strategy to generate highly active extracts, we decided to investigate a post-lysis extract preparation step. Specifically, we evaluated the effect of the run-off reaction. Lysates derived from some strains of *E. coli* benefit tremendously from a run-off reaction, whereby clarified lysate is shaken in an incubator followed by a second clarifying

spin to yield the final extract[232, 291, 405]. It is believed that this incubation allows ribosomes to complete translation of native mRNAs that they were bound to at the moment of lysis and subsequent degradation of those mRNAs by native RNAses. In this way, ribosomes are made available for synthesis of a target CFPS product[456, 457]. To see if *Vnat* lysate productivity could be improved in this way, we prepared a panel of lysates subjected to run-off reactions at either 30°C or 37°C, both with and without shaking at 250 RPM **(Figure 5.5)**. This analysis revealed no benefit to performing any sort of run-off reaction. Indeed, overall productivity suffers when *Vnat* lysates are subjected to prolonged agitation or elevated temperatures. Since pre-incubation is not necessary, we chose to not include the run-off reaction step in our *Vnat* cell extract preparation procedure.

5.4.3 Optimization of Reagent Concentrations and Reaction Conditions in *Vibrio* CFPS

With extraction preparation procedures in hand, we turned our attention to optimizing several reaction conditions which are known to impact CFPS yields. These optimizations were essential, since all reactions performed up until this point had been done using reagent and substrate conditions previously developed for *E. coli* CFPS. Because *Vnat* in nature are found in a different environment than *E. coli*, we reasoned that *Vnat* lysates may have different demands for small molecules and other reagents. Previous studies have demonstrated that CFPS performance can be improved by supplying reagents at concentrations similar to what is encountered by the chassis organism in nature[222, 287]. Thus, we hypothesized that yields of the system could be increased by modifying the existing *E. coli* reagent mix to have a more *Vnat*-specific composition.

To test this, we varied the concentrations of several key components and observed how each change affected the yields of the system in CFPS.

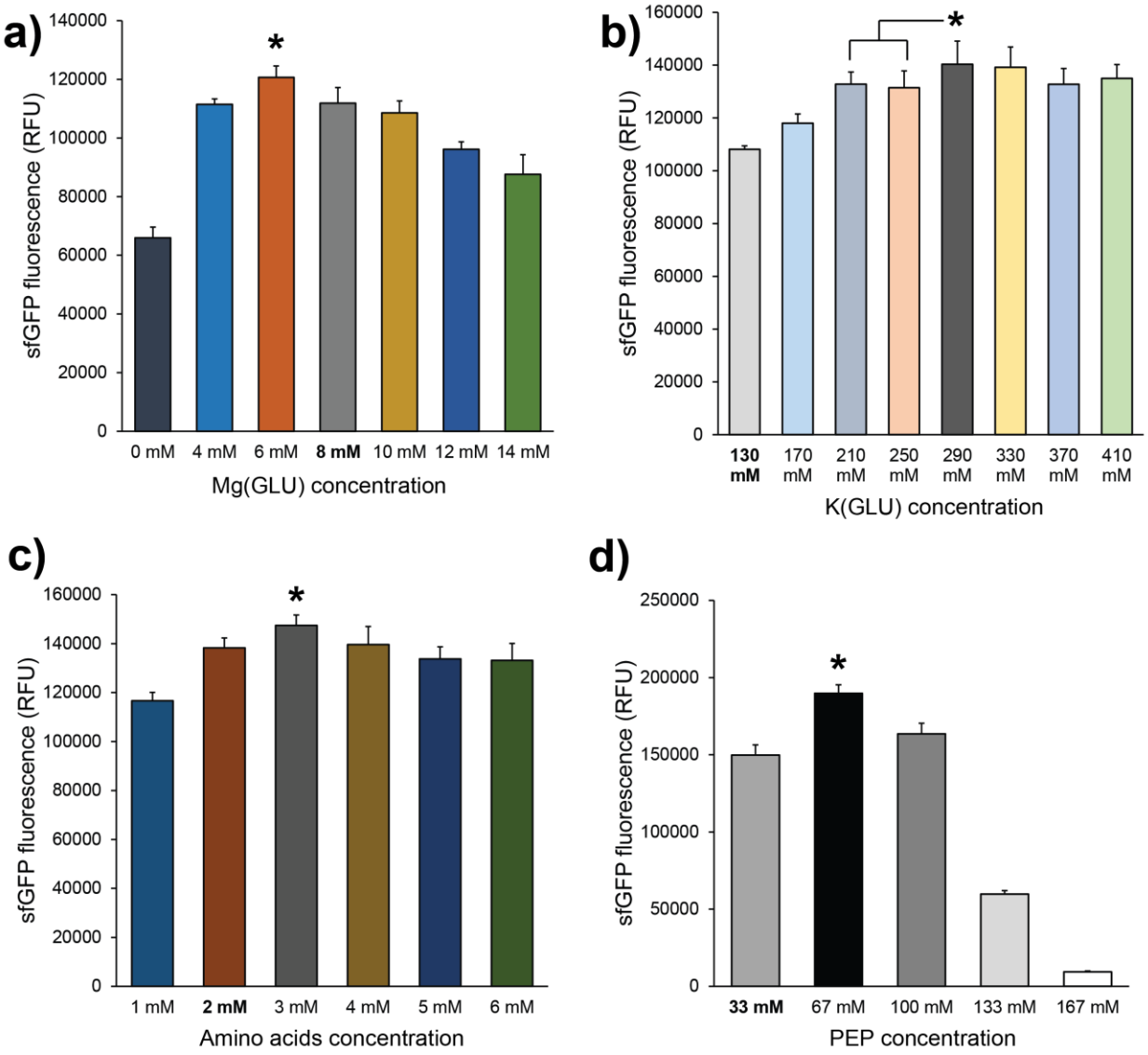


Figure 5.6. Optimization of CFPS reagent mix. A single *V. natriegens* lysate was prepared using optimized parameters identified in **Figure 5.3**. Shown are sfGFP fluorescence values obtained using the lysates *in vitro* when the specified concentration of each of the following reagents is used: **(a)** Mg(GLU), **(b)** K(GLU), **(c)** amino acids, **(d)** PEP. As a reference, for each reagent the concentration typically used in *E. coli* CFPS is indicated in bold. For all conditions, three independent reactions were performed and one standard deviation is shown. * = statistically significant for $p < 0.05$.

One notable difference between *E. coli* and *Vnat* is that the latter is halophilic[28, 29, 450, 451]. Indeed, exponential phase *Vnat* requires significant amounts of cations including Mg^{2+}, K^{+}, and Na^{+}[458]. In CFPS, Mg^{2+} is a particularly important reagent as it is a critical cation required for proper ribosome assembly[436]. Thus, we reasoned that the salt content of the extant reagent mix could be improved for use with *Vnat* lysates. Our approach to address this was twofold. First, we varied the concentration of magnesium glutamate [Mg(GLU)] in the reaction mix **(Figure 5.6.A)**. The system performed best when supplemented with 6 mM Mg(GLU). We also varied the concentration of potassium glutamate [K(GLU)] in the reaction mix **(Figure 5.6.B)**. System productivity increased steadily with increasing K(GLU) concentration, up to a maximum beginning at 290 mM K(GLU). This is more than double the 130 mM K(GLU) used in the *E. coli* mix, and is in agreement with a previous study documenting the higher demand for K^{+} in *Vnat* growth[458]. We also tested various concentrations of sodium glutamate, but observed no significant improvements to the system **(Figure 5.7)**.

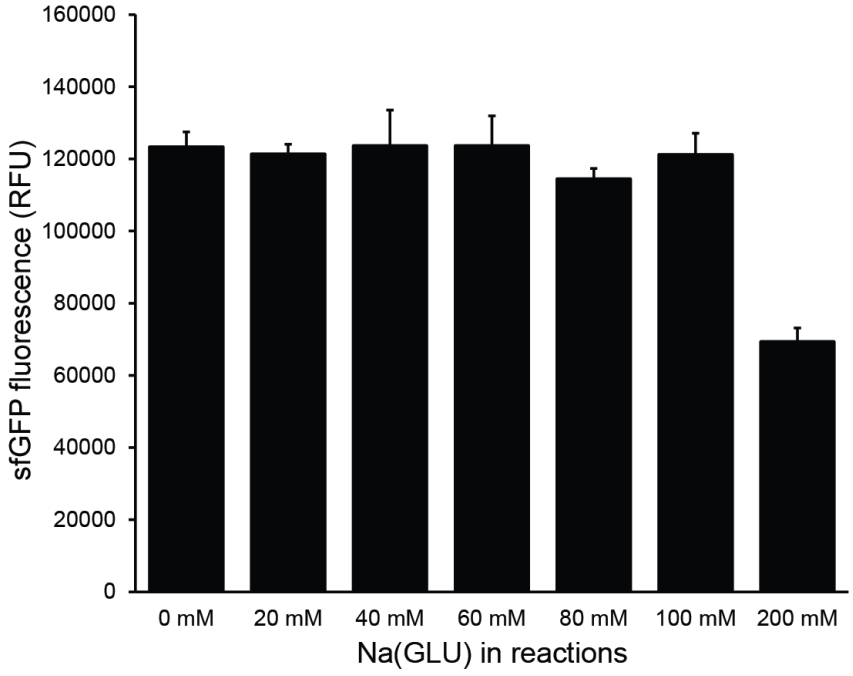


Figure 5.7 (previous page). *V. natriegens* Na(GLU) optimization. sfGFP fluorescence is shown from CFPS reactions supplemented with the indicated concentrations of Na(GLU). For each condition, 3 independent reactions were performed, and one standard deviation is shown.

Next, we looked at optimizing the concentration of amino acids in the reaction mix. As the monomeric building blocks of proteins, amino acids are a key CFPS reagent. Besides their central involvement in protein synthesis, some amino acids are also active participants in central metabolic pathways – consequently, CFPS productivity can be impaired by any non-productive consumption of amino acids by metabolic pathways still active in the lysate[405, 459]. We therefore reasoned that due to potential differences in central metabolism, *Vnat* lysates may have different amino acid demands than *E. coli* systems. To test this, we varied the concentrations of all 20 amino acids added to the CFPS reagent mix **(Figure 5.6.C)**. Increasing the concentration of each amino acid from 2 mM to 3 mM yielded a modest, but significant increase in the productivity of the *Vnat* CFPS system.

The final reagent examined was phosphoenol pyruvate (PEP). A secondary metabolite used in the latter half of glycolysis, PEP is the primary source of ATP in the PANOx-SP energy regeneration system[287]. Similar to amino acids, ATP is utilized by a plethora of cellular processes, and in cell-free systems is consumed to an extent by non-productive side pathways. Thus, as with amino acids, we reasoned that *Vnat* lysates may have different PEP requirements than *E. coli* lysates. To test this, we varied the concentration of PEP added to *Vnat* CFPS reactions **(Figure 5.6.D)**. We found that doubling the PEP concentration from 33 mM (the optimal concentration previously reported for *E. coli*-based CFPS) to 66 mM yielded a 34% increase in the productivity of the system.

Cell-free reaction temperature is another key factor that can be optimized, because it affects enzyme activities and protein folding. We therefore next sought to establish the optimal reaction incubation temperature for *Vnat* CFPS. All prior reactions up to this point had been incubated at 30°C as per *E. coli* protocols[232, 405]. However, as *V. natriegens* evolved in an environment with an ambient temperature of only ~23°C, we reasoned that its cellular components may operate more efficiently at lower temperatures. To test this, we incubated *Vnat* CFPS reactions at several temperatures ranging from 16-37°C **(Figure 5.8)**. We observed that *Vnat* CFPS reactions are most productive at 30°C, which could also represent the optimum for our reporter protein sfGFP to fold into an active confirmation. Still, the system only experiences a 10% reduction in productivity when incubated at room temperature (~23°C), and retains >60% productivity at 16°C.

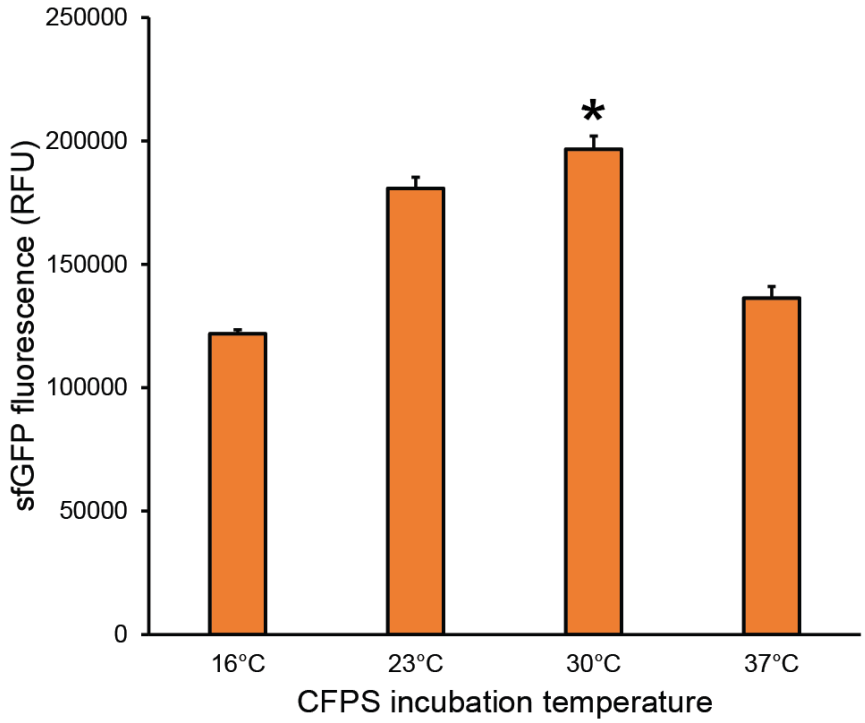


Figure 5.8. Characterization of *V. natriegens* CFPS reaction incubation temperature. Shown are sfGFP fluorescence values for CFPS reactions left to incubate for 20 hrs at the indicated temperature. For each condition, 3 independent reactions were performed, and one standard deviation is shown. * = statistically significant for $p < 0.05$.

The final, optimized *Vnat* CFPS platform described here is capable of synthesizing ~1.6 mg/mL of sfGFP in 20-hour batch mode reactions. **Figure 5.9** captures the step wise yield increases achieved per process optimization. Overall, the yield is comparable to state-of-the-art systems derived from *E. coli*, which have been improved over the last two decades[280, 284, 405], and to our knowledge is the highest-yielding CFPS system derived from this relatively understudied non-model organism.

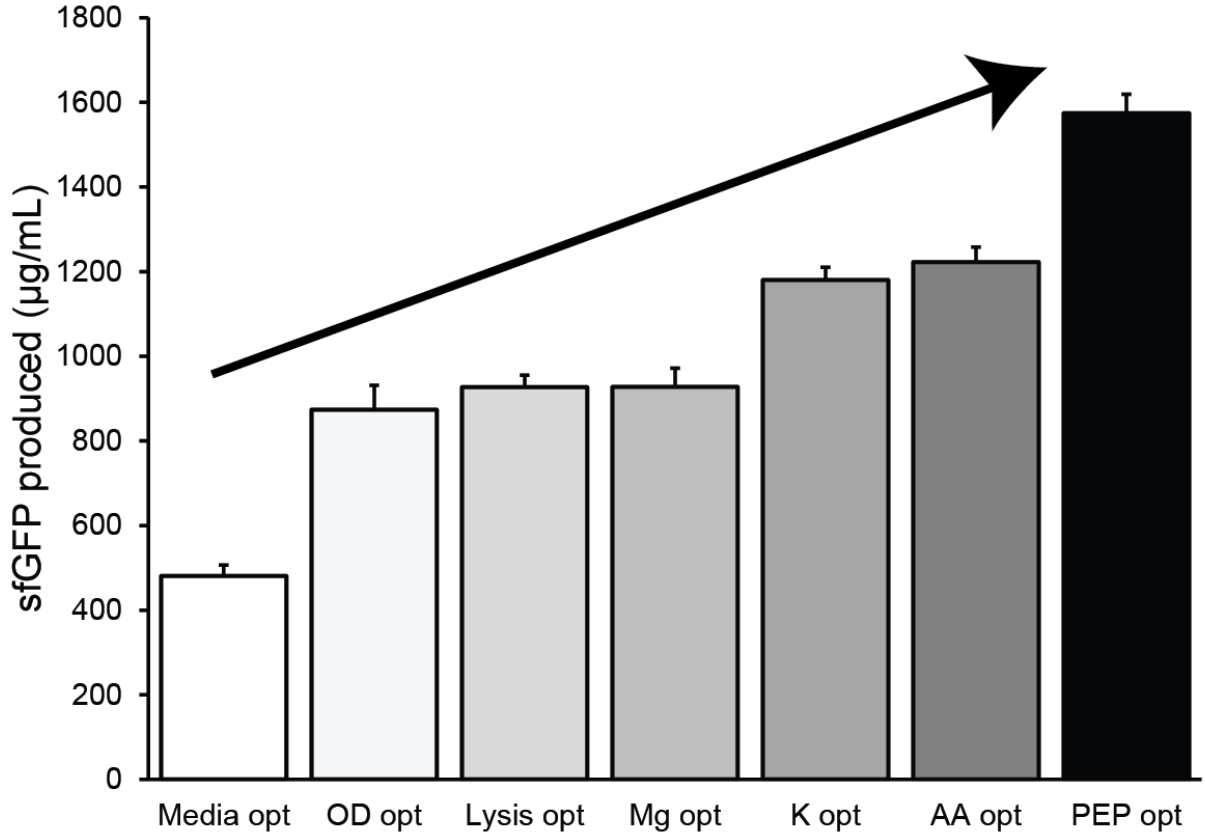


Figure 5.9. Summary of the development and optimization of *V. natriegens* CFPS. Shown are the step-wise and cumulative improvements to the system following each of the indicated optimizations. sfGFP yield from the best condition identified in each optimization is shown. For each condition, three independent reactions were performed and one standard deviation is shown. Media opt = identification of preferred liquid culture medium; OD opt = identification of optimal OD_{600} at harvest; Lysis opt = identification of optimal sonication procedure; Mg opt = optimization of Mg(GLU) in CFPS reaction mix; K opt = optimization of K(GLU) in CFPS reaction mix; AA opt = optimization of amino acids in CFPS reaction mix; PEP opt = optimization of PEP in CFPS reaction mix.

5.4.4 Assessing the Capabilities of *V. natriegens* CFPS

After a systematic optimization of the *V. natriegens*-based CFPS system, we wanted to assess its capabilities. The ability to lyophilize CFPS reactions for storage at room temperature greatly expands the potential user base for a CFPS platform by removing the requirement of storing the materials in freezers at cold temperatures[274-276]. In *E. coli*-based systems, this flexibility often comes at the expense of productivity, as even in the presence of cryoprotectants the productivity of lyophilized samples generally decreases[274, 313]. To assess the ability of our *Vnat* CFPS platform to support robust protein synthesis even after lyophilization, we freeze-dried fully assembled reactions both with and without cryoprotectant supplementation **(Figure 5.10.A)**. Samples reconstituted with water immediately following lyophilization performed quite well, experiencing only a small loss of activity likely as a result of the lyophilization process. Addition of 5% trehalose to reactions fully preserves reaction efficacy after a week of room temperature storage, with samples experiencing no detectable loss in productivity after this time. These data suggest freeze-dried strategies developed in other CFPS systems could be applied to our *V. natriegens*-based CFPS system.

To further demonstrate the ease-of-use of our system and to demonstrate possible applications, we tested CFPS reactions using lysates derived from a strain of *Vnat* recently developed and commercialized by Synthetic Genomics, Inc. that includes a genomic insert encoding the T7 RNA polymerase under the control of an IPTG-inducible promoter (Vmax™ Express)[29]. Such a chassis strain circumvents the need to supply this polymerase to CFPS reactions in purified form, partially addressing limitations imposed on reaction volume scale-up related to high costs of reaction substrates[459]. To test the ability of this strain to compose a one-

pot CFPS platform, we prepared lysates from cells in which polymerase expression was induced and performed synthesis reactions both with and without supplementation with purified polymerase **(Figure 5.10.B)**. This analysis revealed that, while overall productivity of the engineered strain is reduced relative to the wild type strain (~21% reduction), the lysates were enriched with enough T7 polymerase to catalyze more than 1 mg/mL of sfGFP synthesis.

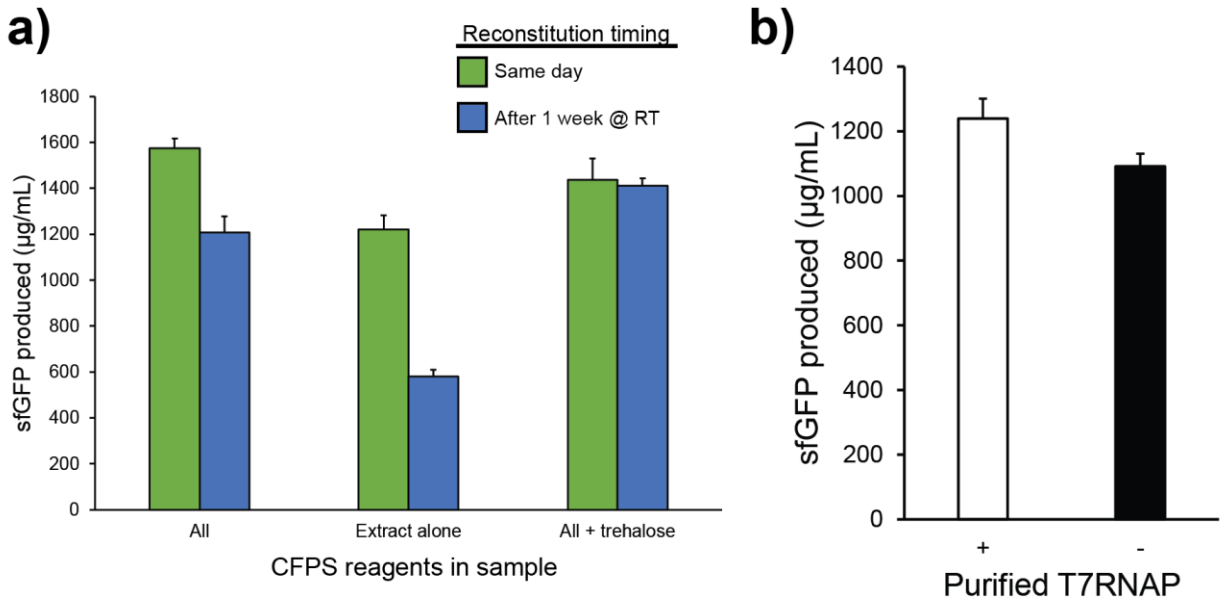


Figure 5.10. Demonstration of the capabilities of *V. natriegens* CFPS. (a) Yields of sfGFP from lyophilized *V. natriegens* CFPS reactions. Reactions were assembled both without (All) and with supplementation with 2.5% trehalose (All + trehalose). These reactions, along with samples of just *V. natriegens* lysate (Extract alone) were lyophilized overnight. Lyophilized samples were reconstituted either immediately (Same day) or after incubation at room temperature (~23°C) for 1 week. For each condition 3 independent reactions were performed, and one standard deviation is shown. **(b)** One-pot CFPS using *V. natriegens* lysates. A lysate was prepared from engineered *V. natriegens* strain Vmax™ Express (Synthetic Genomics, Inc.), which features an IPTG-inducible genomic insert expressing T7 RNA polymerase. sfGFP yields from this lysate in CFPS performed both with and without supplementation with purified polymerase is shown. For each condition, 3 independent reactions were performed, and one standard deviation is shown.

Next, we aimed to expand the targets of our CFPS reactions beyond our reporter protein (sfGFP). Short peptides (<10 kDa) have emerged as important agents in biological engineering and synthetic biology. These small biomolecules are widely used as protein mimics for

interrogating protein-protein interactions and assessing enzyme substrate preferences[268, 460-462], and bacteria-killing antimicrobial peptides (AMPs) are increasingly being considered for use as next-generation antibiotics as we rapidly approach a post-antibiotic era[30, 31, 463]. The use and study of peptides has historically been limited by our ability to synthesize usable amounts of these molecules – solid phase peptide synthesis is generally applicable only to peptides shorter than 30 amino acids[464], and recombinant expression in bacterial hosts is opposed by the degradation of peptide products by host proteases[30, 31]. We reasoned that protease activity in *Vnat* may be reduced since unwanted proteins could simply be diluted out by rapid cell divisions, which in turn may make this organism well suited for peptide synthesis. We thus set out to see if our *Vnat* CFPS platform could catalyze robust expression of peptides. To test this, *Vnat* CFPS was applied towards the synthesis of the AMPs cecropin A[465], cecropin P1[30], and opistoporin I[30] **(Table 5.2)**. Peptide products were quantified using ^{14}C-leucine radioactive incorporation **(Table 5.3)**. Opistoporin I expression in particular surpassed 250 µg/mL, suggesting that our *Vnat* CFPS platform might have utility for the recombinant expression of peptides.

Table 5.2. Antimicrobial peptides used in this study.

Name	DNA Sequence	Source
Cecropin A	atgAAGTGGAAATTGTTTAAAAAGATCGAAAAGGTGGGGCAAAATATCCGCGACGGGATCATTAAGGCAGGTCCGGCTGTGGCGGTCGTTGGTCAGGCAACGCAAATCGCAAAATAA	[465]
Cecropin P1	atgAGCTGGCTGAGCAAAACCGCGAAAAAACTGGAAAACAGCGCGAAAAAACGCATTAGCGAAGGCATTGCGATTGCGATTCAGGGCGGCCCGCGCTAA	[30]
Opistoporin I	atgGGCAAAGTGTGGGATTGGATTAAAAGCACCGCGAAAAAACTGTGGAACAGCGAACCGGTGAAAGAACTGAAAAACACCGCGCTGAACGCGGCGAAAAACCTGGTGGCGGAAAAAATTGGCGCGACCCCGAGCTAA	[30]

Peptide name	Length (AAs)	Mass (Da)	Yield (μg/mL)
cecropin A	38	4136	22.4 ± 1.4
cecropin P1	32	3470	96.8 ± 12.8
opistoporin I	45	4968	278.4 ± 9.1

Table 5.3. Yields of antimicrobial peptides using *V. natriegens* CFPS. Shown are yields of the indicated antimicrobial peptides using *V. natriegens* CFPS, quantified via ^{14}C-leucine incorporation followed by scintillation counting of 15% TCA-precipitated peptides. For each peptide 3 independent reactions were performed, and one standard deviation is indicated.

5.4.5 Increasing System Productivity Via Genome Engineering

Next, in an effort to further enhance protein synthesis yields, we pursued increases in *Vnat* CFPS productivity via genome engineering. Recent efforts in *E. coli* have seen tremendous success in increasing *in vitro* productivity by genetically inactivating or removing genes whose products may destabilize key substrates in CFPS[232, 405, 459]. These so-called "negative effectors" have broadly included targets such as DNAses, RNAses, proteases, and central metabolic enzymes that might non-productively consume ATP and/or amino acids throughout the CFPS reaction. Here, we prepared a short list of genes that were previously identified in *E. coli* as being beneficial knockout targets for increasing CFPS yields **(Table 5.3)**[405]. We performed a bioinformatics search to identify close homologs of these genes in the genome of *Vnat*. Then, we applied an approach to leverage *Vnat*'s natural propensity for homologous recombination to replace each gene individually with a selectable marker to generate a library of *Vnat* knockout strains[29, 453]. Lysates were prepared from each strain using optimal practices identified here, and assessed in CFPS **(Figure 5.11)**.

Table 5.4. Description of negative CFPS effector knockout targets.

E. coli* gene	***E. coli* gene function**	**Negative effector hypothesis**	***V. natriegens* homolog (tBLASTn)†**		***V. natriegens* putative annotation‡**	**likely homolog**
			E-value	**bit score**		
endA	endonuclease	degrades template DNA[232, 289, 364]	1.12E-91	292.35	endonuclease I	yes
lon	protease	consumes ATP and degrades protein product[352]	0.00	1,284.63	ATP-dependent protease La (LON) domain	yes
mazF	RNAse	degrades mRNA[232, 360]	1.05	26.18	-	no
ompT	protease	degrades protein product[352, 389]	4.70	26.56	-	no
rna	RNAse	degrades mRNA[358]	1.12	28.11	-	no
rnb	RNAse	degrades tRNA and mRNA[232, 357, 359]	0.00	645.58	exoribonuclease 2	yes
glpK	glycerol kinase	consumes ATP[391, 392]	0.00	815.07	glycerol kinase	yes
gor	glutathione reductase	reduces glutathione disulfide to sulfhydryl form[310, 357]	0.00	656.37	glutathione-disulfide reductase	yes
gshA	glutamate-cysteine ligase	consumes cysteine, glutamate, and ATP[357, 363]	0.00	568.54	glutamate--cysteine ligase	yes
tnaA	tryptophanase	consumes tryptophan[288]	0	847.42	tryptophanase	yes

* *E. coli* protein sequences are provided in Supplementary Sequence List 1.

† *E. coli* sequences were searched (tBLASTn algorithm) against the published *V. natriegens* chromosomes (GenBank accessions CP016345 (chrI) & CP016346 (chrII)). Gene sequences of the *V. natriegens* homologs are provided in Supplementary Sequence List 2.

‡ Annotations were generated using the Archetype® Genomics Discovery Suite software package (Synthetic Genomics, Inc.)

Unfortunately, we were unable to recapitulate the beneficial effects of any of these knockouts in our *Vnat* system. This could be a result of the pronounced difference in strain backgrounds—*Vnat* is significantly diverged from *E. coli* such that knockout targets which work well for one species may not necessarily inform engineering efforts in the other. In this case, a more comprehensive screen of knockouts in *Vnat* may be required to identify effectual knockouts. It is also possible that system productivity is being limited by something more fundamental, such as ribosome or energy availability, which could mask any observable benefit of these knockouts to overall protein yield. Further studies are needed.

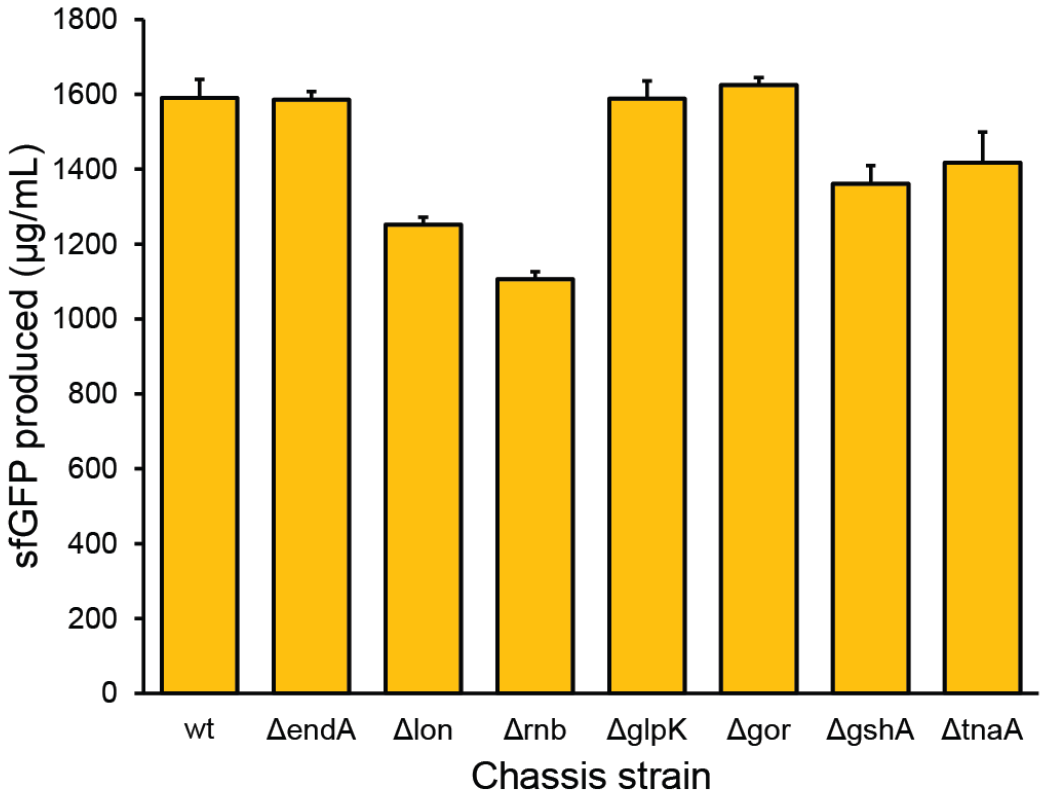


Figure 5.11. Characterization of *V. natriegens* negative effector knockout strains. Lysates were prepared from *V. natriegens* strains in which homologs of the listed *E. coli* genes were genetically removed. sfGFP yields from each lysate in CFPS is shown. For each condition, 3 independent reactions were performed, and one standard deviation is shown.

5.4.6 Assessing ncAA Incorporation Using *V. natriegens* CFPS

Finally, we set out to assess the ability of *Vnat* CFPS to facilitate cotranslational incorporation of non-canonical amino acids (ncAAs). The gold-standard by which this is achieved is a process known as amber suppression, which hijacks the amber stop codon (UAG) and repurposes it as a coding channel to encode an ncAA to allow orthogonal translation system (OTS) components developed specifically for an ncAA of interest to mimic the natural processes underlying elongation and catalyze cotranslation incorporation of the ncAA[394]. Notably, many recent works have sought to combine this approach with CFPS for robust synthesis of proteins featuring ncAAs[23, 24, 136, 232, 394, 405]. To test this process using our *Vnat* CFPS platform, we expressed the OTS components for the ncAA *p*-azidophenylalanine (pAzF)[203] during cell growth and prepared a lysate enriched with the orthogonal machinery. We then directed the OTS-enriched as well as a plain lysate (lacking OTS components) in CFPS to synthesize wild type sfGFP as well as sfGFP with an amber codon at position 216 (sfGFPT216X, **Figure 5.12**), both with and without inclusion of the ncAA pAzF in the reaction volume. These experiments revealed that templated incorporation of pAzF is achievable using our *V. natriegens* platform. Above-background synthesis of sfGFPT216X is only observed from the lysate enriched with the pAzF OTS, and in this lysate the fluorescence level is elevated when pAzF is included in reactions suggesting that the ncAA is being incorporated. The yield of sfGFPT216X is dramatically reduced relative to wild type sfGFP. This is not surprising, as amber suppression is limited by competition with the amber codon-specific release factor 1 (RF1) which is still present and active in these lysates[23, 394]. Going forward, improvements to OTS-mediated ncAA incorporation using this platform will almost certainly require the removal of RF1 from the system.

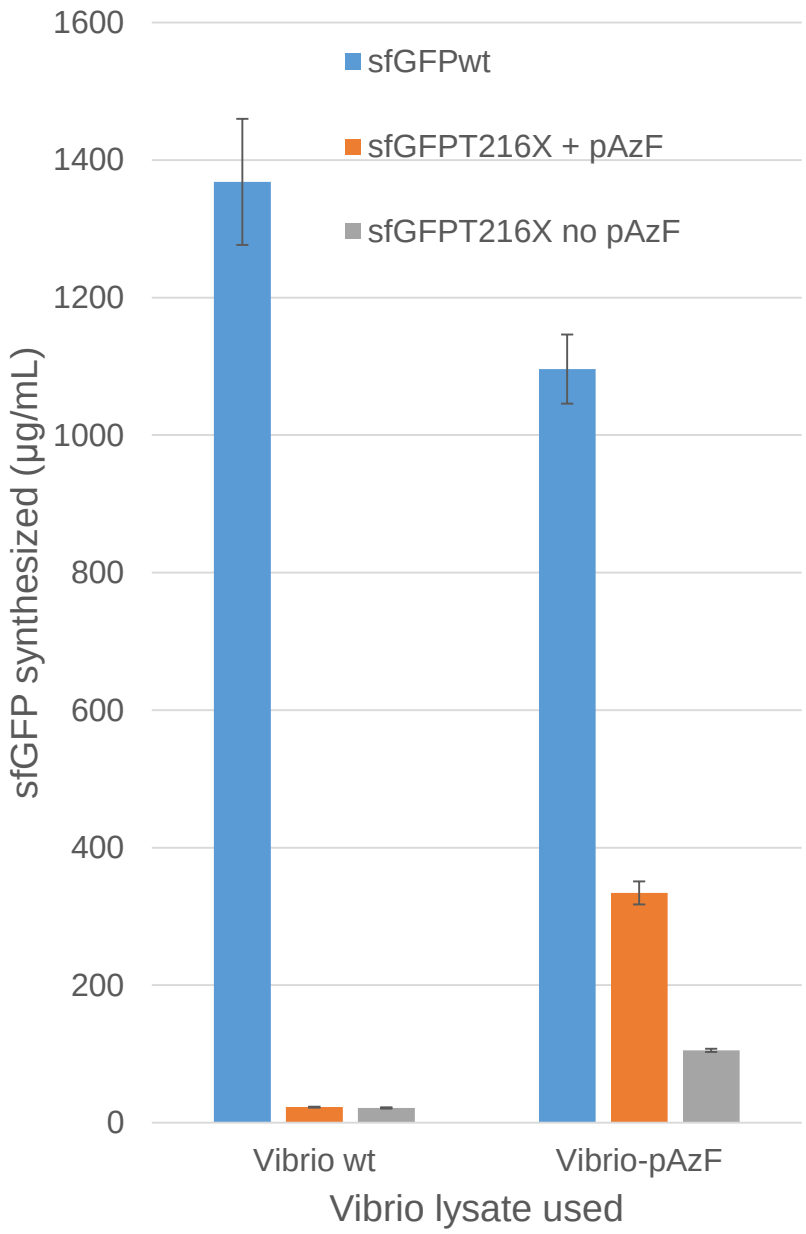


Figure 5.12. ncAA incorporation in *V. natriegens* CFPS. Shown are the yields of wild type sfGFP (sfGFPwt) or sfGFP with an amber codon at position 216 (sfGFPT216X) from *Vnat* lysate either enriched with OTS components for pAzF (Vibrio-pAzF) or lacking these components (Vibrio wt). Reactions synthesizing sfGFPT216X were performed both in the presence and absence of pAzF in the reaction mix (+ pAzF and no pAzF, respectively). For each condition three independent reactions were performed, and one standard deviation is shown.

5.5 Discussion

In this study, we present the development of a novel CFPS platform composed of crude lysates derived from the non-model bacterium *V. natriegens* **(Figure 5.12)**. We identified optimal culture harvest conditions for this organism, demonstrating that culturing in BHI media supplemented with v2 salts and harvesting at stationary phase is optimal for the generation of productive lysates. This feature is convenient, as it eliminates the need to collect cell pellets within a tightly specific OD_{600} window in order to preserve lysate quality. Researchers can "set and forget" *V. natriegens* cultures and harvest whenever it is convenient without worrying about spoiling the resulting extracts. Attempts to optimize lysis parameters revealed that *V. natriegens* lysates are highly-resistant to damage via overlysis, and overall the system is relatively agnostic to both lysis buffer resuspension volume and lysis energy delivery. Surprisingly, the system is much less sensitive to overlysis during lysate preparation as compared to *E. coli*-based systems, a huge benefit for first-time or inexperienced CFPS users.

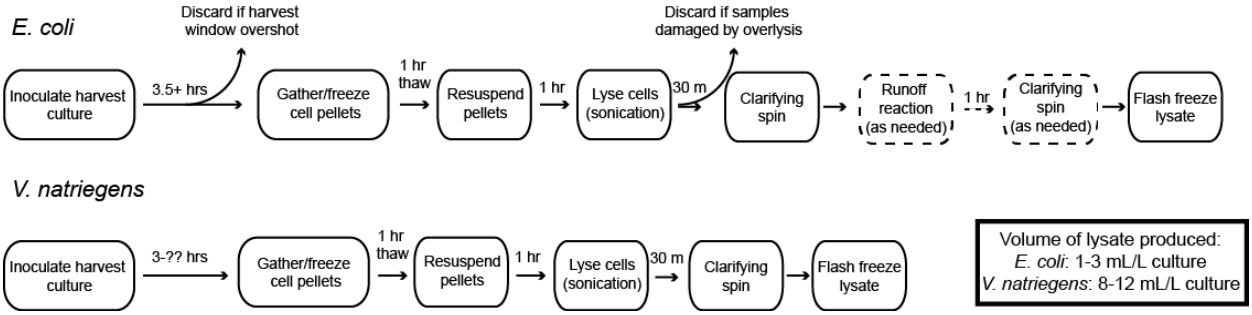


Figure 5.13. Comparison of *E. coli* and *V. natriegens* lysate preparation workflows. As compared to *E. coli*, *V. natriegens* cells are easier to handle both in the context of culture harvest and lysis. It is extremely difficult to spoil a *V. natriegens* extract preparation. *V. natriegens'* overall workflow is faster due to the organism's rapid growth rate and omission of any run-off reaction step. Finally, because *V. natriegens* pellets are gathered at stationary phase, they are significantly larger than those gathered for *E. coli*—consequently, the volume of *V. natriegens* lysate derived from a single 1 L harvest is significantly higher.

Optimization of the CFPS components specifically for use with *V. natriegens* lysates significantly increased the overall productivity of the system to ~1.6 mg/mL sfGFP, comparable to popular platforms based on lysates derived from *E. coli* and the highest-yielding CFPS system derived from a non-model organism, to our knowledge. This is also higher than a recent study published during the preparation of this manuscript which demonstrated expression titers of 0.4 mg/mL of green fluorescent protein in a *V. natriegens* CFPS system[466]. Further, our system is stable at room temperature if lyophilized in the presence of trehalose, is capable of synthesizing small peptide products, and can be carried out in a one-pot system with the use of the Vmax™ Express chassis strain.

Looking forward, we anticipate that the CFPS system described here will find use in the identification and characterization of *V. natriegens* genetic parts. Recent publications have developed a suite of tools for using *V. natriegens* for cloning and recombinant protein expression[28, 29], but the knowledge base for this organism still lags behind the more established *E. coli* with regards to regulatory sequences such as promoters, terminators, and ribosome binding sites (RBSs). Due to its open and easily-accessible nature, the CFPS platform developed here could be used to interrogate many such parts in parallel for rapid characterization. The resulting data could in turn be used to inform construct design for use *in vivo*, supporting the increasing interest in using this organism as an alternative to *E. coli* for molecular biological applications. This idea has already been briefly explored using a *V. natrigens*-based CFPS platform[466].

Perhaps the most surprising finding to arise from this effort was that the most productive *V. natriegens* extracts are derived from stationary phase cells. This not only contradicts what is usually found in other bacterial CFPS systems, but is also is overall difficult to rationalize

alongside the generally-accepted notion that ribosomes are downregulated in stationary phase[455]. It is possible that sequestration of ribosomes by native mRNAs in lysates derived from exponential phase cells accounts for the relatively low productivity of these lysates, despite the presence of a larger ribosome pool. In other systems this issue is alleviated by subjecting the lysate to a run-off reaction[232, 291, 405, 456, 457], but this approach was not successful here.

Going forward, we expect the efficacy of *V. natriegens* CFPS to improve rapidly. Indeed, the system development and characterization described here accomplished in a very short amount of time what took decades of research and development in *E. coli*. Exploration of the use of alternative energy regeneration systems is one obvious future direction. Indeed, development of an entirely novel energy regeneration system (perhaps using sucrose as the starting substrate[28]) might ultimately be required to fully optimize *V. natriegens* CFPS. Another fruitful direction is in the screening of a larger, more comprehensive library of negative effector knockout strains. This could perhaps be informed by a time course analysis of small molecule concentrations in CFPS reactions, with knockouts targeted towards metabolic pathways that might be siphoning away critical substrates.

In conclusion, the *Vnat* cell-free platform is excellent for early forays into the use of CFPS systems, as the cells are fast and easy to grow, easy to lyse, and a high volume of active lysate is generated from as little as 1 L of cell culture. Collectively, these features reduce the need for specialized knowledge and equipment that have limited the use of CFPS. We expect that this reduced entry barrier will facilitate the spread of these systems into new areas for use on exciting, novel applications in synthetic biology.

5.6 Acknowledgments

Thanks to my collaborators and co-authors Samuel R. Davidson (S.R.D.), Matthew T. Weinstock (M.T.W.), and Dan G. Gibson (D.G.G.) for their contributions to this project. Professor Michael C. Jewett (M.C.J.) and I conceived of this study. S.R.D. provided support and feedback in experimental design and was a second set of hands for many of the experiments described in this chapter. M.T.W. and D.G.G. performed bioinformatics analysis to identify putative negative effectors in the genome of *V. natriegens* and generated the library of knockout strains. I performed all other experiments and analysis. M.C.J., S.R.D., M.T.W., and D.G.G. all contributed to editing this chapter.

5.7 Publication Information

This chapter was published with the following citation information:

Des Soye B.J., Davidson S.R., Weinstock M.T., Gibson D.G., and Jewett M.C. Establishing a high-yielding cell-free protein synthesis platform derived from *Vibrio natriegens*. *ACS Synth. Biol.* 14 Aug 2018; doi: 10.1021/acssynbio.8b00252.

CHAPTER 6

6 Towards an Improved Platform for Cotranslational Phosphorylation in Cell-free Protein Synthesis

6.1 Abstract

Protein phosphorylation is one of the most ubiquitous and important polypeptide modifications in nature, with thousands of modified sites identified across the proteomes of several model organisms. Unfortunately, understanding the structural and mechanistic basis underlying how individual phosphorylation events modulate the activity of their associated protein remains a significant challenge, largely due to difficulties in generating homogeneous samples of proteins featuring phosphorylations at specific sites. Here, I describe my effort to address this limitation by working towards a cell-free protein synthesis (CFPS) system optimized for templated, cotranslational incorporation of the non-canonical amino acid (ncAA) *o*-phosphoserine (Sep). I present efforts to generate a chassis strain optimized specifically for use in Sep incorporation and initial characterization of lysates from the completed strain. In parallel, I describe the assembly and initial characterization of a novel plasmid system for in-strain overexpression of improved orthogonal translation system (OTS) components specific for Sep. Finally, I apply CFPS and Sep incorporation to the synthesis of the glycolytic enzyme triosephosphate isomerase (TPI) featuring specific serine phosphorylations to enable robust characterization of the proteoform variants *in vitro*. I show that phosphorylation of serine 20 increases the catalytic speed of TPI by more than

4-fold, potentially due to improved substrate access to the active site caused by the phosphorylation. While these efforts are not yet complete, the work presented in this chapter provides strong support for the notion that applying OTSs for the templated synthesis of proteins featuring specifiable phosphorylations will be a transformative tool in fully understanding individual instances of this critical post-translational modification.

6.2 Introduction

Phosphorylation is the reversible attachment of a phosphate group (R-**O-PO$_3^{2-}$**) to a protein. Usually observed at the side chains of serine, threonine, and tyrosine, this modification is regulated by the opposing activities of kinase enzymes (which catalyze the transfer of a γ-phosphate from adenosine triphosphate to the amino acid sidechain) and phosphatase enzymes (which facilitate removal of phosphate groups via hydrolysis)[8]. Phosphorylation is one of the most important polypeptide chain modifications in nature and plays a pivotal role in protein folding, function, stability, and localization[8]. This ubiquitous modification is present in all domains of life: approximately one-third of all eukaryotic proteins are predicted to be phosphorylated[43] and phosphorylated proteins have been found in archaea[44] and bacteria[45]. Protein phosphorylation is critical for the development, growth, function and survival of all organisms and is involved in the regulation of a myriad of biological processes, including cell cycle control, receptor-mediated signal transduction, differentiation, proliferation, and metabolism[10, 46-50]. Moreover, perturbation of phosphorylation patterns and breakdown of associated control systems is often a hallmark of disease states such as cancer, inflammation, Alzheimer's disease, and congenital disorders[51-54].

Accordingly, a significant amount of effort has gone towards the development of methods for identifying locations where phosphorylation naturally occurs and subsequently characterize the ways in which individual modifications affect the structure, function, and behavior of the protein to which they are attached. Many of the earliest approaches developed, such as immunoprecipitation (IP) and western blotting, leverage the specific binding capabilities of antibodies for the isolation, enrichment, and identification of phosphorylated proteins in complex protein mixtures[134, 136] and are still regularly used in phosphorylation studies. However, these approaches suffer from some significant limitations, including the requirement that an antibody actually exist for a target of interest[5, 148], occurrences of false positive results caused by errant off-target antibody binding[5, 149, 150], occurrences of false negative results caused by the structural occlusion of modified residues[150], and a general inability to specifically identify the residue(s) at which phosphorylation is occurring in the identified targets. Recent mass spectrometry-based proteomics approaches address many of these limitations and are able to identify phosphorylation sites on proteins with high confidence at high throughput[166, 170].

In terms of characterization, much progress has been achieved by a combination of approaches both *in vivo* and *in vitro* to investigate the roles played by specific phosphorylation events. *In vivo,* a common approach is to install mutations into targets of interest utilizing canonical amino acids to mimic phosphorylation[183, 186, 187], but in many cases the structural changes underlying observed changes in function cannot be ascertained using these techniques[180]. Additionally, common phosphorylation mimetics (glutamic acid, aspartic acid[183]) differ both structurally and electrostatically from the modified amino acids they are replacing such that they cannot always accurately replicate the effects of a phosphate group addition [136, 188]. Methods

in vitro typically seek to isolate a phosphoproteoform of interest for controlled interrogation of parameters such as efficiency, substrate specificity and binding affinity, and even structure[190, 191]. These sorts of studies provide the most comprehensive characterization of individual phosphorylations, but are severely limited by the need to obtain homogeneous samples of a specific proteoform for use in exploratory assays. Because most PTMs exist only briefly during the lifetime of a protein at substoichiometric levels, it is difficult-to-impossible to purify most phosphoproteoforms of interest from cells[158, 176, 178]. The simplest way to obtain highly pure samples of a specific proteoform is to express the unmodified protein and phosphorylate it *in vitro*; however, this approach requires knowledge of the kinase(s) responsible for performing those modifications and the ability to obtain them in pure, active configurations, which is not always possible[176]. Furthermore, kinases are often difficult to control and direct *in vitro*[192, 193]. Simply stated, there is an unmet need for a synthesis platform enabling preparative scale synthesis of homogeneous protein populations featuring specific phosphorylation events.

Amber suppression-mediated incorporation of non-canonical amino acids (ncAAs) presents one approach whereby designer phosphoproteoform synthesis is possible. This method seeks to hijack the amber stop codon (UAG), repurposing it as an open coding channel that can be used to genetically encode a ncAA of interest[394]. Through the coordinated activities of an orthogonal mutant aminoacyl-tRNA synthetase specific to a ncAA (ncAARS) and an orthogonal suppressor tRNA engineered to decode the amber codon (collectively referred to as an orthogonal translation system, or OTS) the biological processes underlying the addition of natural amino acids to polypeptide chains can be mimicked to catalyze the templated, cotranslational incorporation of the ncAA[394]. Historically, this process has been limited by competition with ribosomal release

factor 1 (RF1) at amber codons; however, this limitation was recently addressed with the creation of *Escherichia coli* strains in which amber codon recoding has enabled RF1 knockout[21, 394]. Using amber suppression, phosphorylated amino acids can be genetically encoded for site-specific, cotranslational incorporation into proteins - to that end, OTSs for *o*-phosphoserine (Sep)[6, 225], phosphothreonine[20], and phosphotyrosine[227] have been developed and shown to catalyze incorporation of their respective phosphorylated residues in cells, though their utility has been limited by the relatively poor membrane permeability of the phospho-ncAAs coupled with toxicity effects associated with the OTSs[20, 21, 23, 136].

The bioavailability and toxicity limitations described above have in similar situations been circumvented by the use of cell-free protein synthesis (CFPS) systems for protein production. These systems harness cellular machinery responsible for catalyzing protein synthesis and energy regeneration from purified components or crude lysates derived from chassis organism cells, and contain the necessary elements for translation and protein folding (*e.g.* ribosomes, aminoacyl-tRNA synthetases, translation initiation and elongation factors, ribosome release factors, chaperones, foldases, *etc.*) and energy regeneration (*e.g.* enzymes involved in glycolysis, oxidative phosphorylation) such that upon being combined with raw materials (amino acids, nucleotide triphosphates, energy, enzyme cofactors, buffer, salts, DNA template) these biological processes can be coordinately activated in a test tube in the absence of intact cells to catalyze synthesis of a specified protein[279-281]. The lack of physical barriers facilitates the use of substrates with poor membrane permeability (*e.g.* phospho-ncAAs), and in the absence of living cells the use or synthesis of toxic species (*e.g.* OTSs) is no longer deleterious[20, 23, 136].

Though the use of amber suppression in CFPS has been successfully leveraged for the synthesis of specific phosphoproteoforms (specifically, of proteins featuring Sep)[136], the current platform is limited by two fundamental flaws. First, these efforts thus far have utilized the original Sep OTS described by Park *et al*[6], which demonstrates polysubstrate specificity such that it commonly misincorporates native amino acids such as glutamine at amber codons. Second, in *E. coli* systems the Sep-specific phosphatase SerB must be inactivated in the source strain in order to succeed at Sep incorporation[6, 136]. This deletion consistently leads to a significant growth defect in extract source cells, with an accompanying reduction in lysate productivity. Here, I present the development of a CFPS platform optimized specifically for Sep incorporation that seeks to address these limitations via a two-fold approach. First, I pursued a scheme whereby SerB activity in CFPS could be eliminated while retaining activity *in vivo*. To accomplish this, next generation genome engineering approaches were used to install multiple basic residues in an external loop of SerB in an effort to make it a more efficient substrate for the periplasmic protease OmpT. I show that lysates derived from the final strain show a modest reduction in CFPS productivity, but are still superior to complete OmpT knockout strains. Second, I set out to assemble a novel Sep OTS expression plasmid encoding improved components recently reported in the literature[225]. By reverse engineering the component sequences from the reporting manuscript, I was able to obtain linear DNA fragments that were ultimately assembled to yield the final correct construct. Finally, in a separate (but related) effort, Sep incorporation using CFPS was used to study specific phosphoproteoforms of the glycolytic enzyme triosephosphate isomerase (TPI). I show that phosphorylation of serine 20 increases the speed of TPI catalysis more than 4-fold, and present a molecular model that suggests a structural basis for this improved enzyme performance. **Though**

all lines of inquiry presented in this chapter remain incomplete, the work presented here suggests a path forward for improving Sep incorporation for designer phosphoproteoform synthesis.

6.3 Materials and Methods

6.3.1 Strains and Plasmids

The bacterial strains and plasmids used in this study are listed in **Table 6.1.** Carbenicillin (50 µg/mL) was used for culturing C321.Δ*A*.759 and to maintain plasmid pMA7CR_2.0. Kanamycin (50 µg/mL) was used for culturing C321.Δ*A*.759.T7.D and for maintaining B40OTS, pSepOpt, all pMAZ plasmids, and all pJL1 plasmids.

6.3.2 pMAZ Plasmid Assembly

Most pMAZ plasmids were assembled from two parts: linear plasmid backbone derived from PCR amplification (primers listed in **Table 6.2**) of pMAZ-ScaI that had been linearized by ScaI digestion, and annealed oligos encoding the target gRNA sequence. Insertion of the annealed oligos into the plasmid backbone was performed with USER® Enzyme (New England Biolabs, Ipswich, MA), after which products were transformed into cells. pMAZ-Cure was assembled via Gibson assembly[356] to insert the anti-pMA7CR_2.0 gRNA cassette into the correct locus using enzyme mix prepared in house. Sanger sequencing confirmed correct sequences of all pMAZ plasmids.

6.3.3 CRMAGE Engineering of rEcoli.759.T7.D

All primers and oligos described in this section are listed in **Table 6.2**. In preparation for CRMAGE, genomically-encoded resistances to ampicillin and kanamycin were removed from rEcoli.759.T7.D to allow for the CRMAGE plasmids to be selected for. This was achieved via MAGE[25] to loop the resistance genes out of the genome followed by replica plating to identify sensitized colonies[415], yielding strain rEcoli.759.T7.D.*ΔAbR*. To begin CRMAGE, pMA7CR_2.0 was transformed into rEcoli.759.T7.D.*ΔAbR*. For each cycle of CRMAGE, the strain was grown to an $OD_{600}$of ~0.6-0.8 and made electrocompetent via washing with cold nuclease-free water. Next, the pMAZ plasmid encoding the gRNA of interest and its associated MAGE oligo were transformed into the strain simultaneously. After electroporation, cells were allowed to recover for 1 hour in 5 mL SOC media at 34°C at 250 RPM. After 1 hr, 50 µg/mL kanamycin was added to the recovery media to begin selecting for the pMAZ plasmid and the culture was left to continue shaking for 2 hours. Next, a total of 3 hours post-transformation, 1 mL of recovery culture was transferred to 4 mL of fresh SOC media with 50 µg/mL carbenicillin, 50 µg/mL kanamycin, and 400 ng/mL anhydrotetracycline (aTc), and put back at 34°C, 250 RPM for another 3 hours. Finally, the culture was diluted 1:100 in nanopure water and plated on LB plates with 50 µg/mL carbenicillin, 50 µg/mL kanamycin, and 400 ng/mL anhydrotetracycline. Plates recovered at 34°C overnight. Individual colonies were sequenced to verify incorporation of the MAGE oligo. Once the desired mutation was detected, cells were grown to an OD_{600} of ~0.6-0.8 and induced with 400 ng/mL aTc and 2% (m/v) rhamnose to induce destruction of the pMAZ plasmid. Replica plating identified cells that had lost the pMAZ plasmid, enabling the next cycle to begin. A total of three CRMAGE cycles were performed. After the last pMAZ plasmid was cured from the strain, plasmid pMAZ-Cure was introduced into the cells. pMAZ-Cure encodes a

gRNA cassette targeting the selectable marker and origin of replication of pMA7CR_2.0 such that induction of the system with both aTc and rhamnose would trigger elimination of both plasmids, finally yielding strain rEcoli.759.T7.Sep. Sanger sequencing of all modified loci confirmed that the desired mutations had been installed in all cases.

6.3.4 Cell Extract Preparation

For rapid prototyping of engineered strains, cells were grown in 1 L of 2xYTPG media (pH 7.2) in a 2.5 L Tunair® shake flask and incubated at 34˚C at 220 rpm. Cultures were induced with 1 mM IPTG at an OD_{600} of 0.6, and cells carrying a Sep OTS expression vector were supplemented with 2 mM Sep at this point. After induction, cultures were permitted to continue to grow to an OD_{600} of 3.0. Cells were pelleted by centrifuging for 15 min at 5000 × g at 4°C, washed three times with cold S30 buffer (10 mM tris-acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, 2 mM dithiothreitol)[354], and stored at -80°C. To make cell extract, cell pellets were thawed and suspended in 0.8 mL of S30 buffer per gram of wet cell mass and 1.4 mL of cell slurry was transferred into 1.5 mL microtubes. The cells were lysed using a Q125 Sonicator (Qsonica, Newtown, CT) with 3.175 mm diameter probe at a 20 kHz frequency and 50 % amplitude for three cycles of 45s ON/59s OFF. To minimize heat damage during sonication, samples were placed in an ice-water bath. Extract was then centrifuged at 12,000 × g at 4˚C for 10 min. For strain derivatives of C321.Δ*A*.759, a run-off reaction (37˚C at 250 rpm for 1 h) and second centrifugation (10,000 × g at 4˚C for 10 min) were performed[291]. The supernatant was flash-frozen using liquid nitrogen and stored at -80˚C until use.

6.3.5 CFPS Reaction

A modified PANOx-SP system was utilized for CFPS reactions testing incorporation of Sep[287]·[414]. Briefly, a 15 µL CFPS reaction in a 2.0 mL microtube was prepared by mixing the following components: 1.2 mM ATP; 0.85 mM each of GTP, UTP, and CTP; 34 µg/mL folinic acid; 170 µg/mL of *E. coli* tRNA mixture; 13.3 µg/mL plasmid; 16 µg/mL T7 RNA polymerase; 2 mM for each of the 20 standard amino acids; 0.33 mM nicotinamide adenine dinucleotide (NAD); 0.27 mM coenzyme-A (CoA); 1.5 mM spermidine; 1 mM putrescine; 4 mM sodium oxalate; 130 mM potassium glutamate; 10 mM ammonium glutamate; 12 mM magnesium glutamate; 57 mM HEPES, pH 7.2; 33 mM phosphoenolpyruvate (PEP), and 27% v/v of cell extract. For Sep incorporation, 2 mM Sep was supplemented to cell-free reactions. Each CFPS reaction was incubated for 20 h at 30˚C unless noted otherwise. *E. coli* total tRNA mixture (from strain MRE600) and phosphoenolpyruvate was purchased from Roche Applied Science (Indianapolis, IN). ATP, GTP, CTP, UTP, 20 amino acids and other materials were purchased from Sigma (St. Louis, MO) without further purification. For scaled up synthesis of TPI proteoforms, 250 µL reactions were assembled in a 24-well plate (Falcon 351147; Corning, Corning, NY). To reduce loss of reaction volume by evaporation, unused wells were filled with 2 mL water, and the chamber was sealed with Parafilm M® (Bemis, Neenah, WI). Plates were shaken at 300 RPM for 20 h at 30°C.

6.3.6 pSepOpt Assembly

All primers and oligos described in this section are listed in **Table 6.2**. To assemble pSepOpt, synthetic linear DNA constructs encoding the optimized SepRS2 and tRNASepB4 [225] were synthesized by Life Technologies (Grand Island, NY). Pieces of the B40OTS plasmid encoding *lacI* and the plasmid origin were obtained via restriction digest with SphI/MscI and

SacI/SacII, respectively. The *lacI* fragment and synthetic tRNASepB4 fragment were combined into a single piece using overlap extension PCR. Gibson assembly [356] was used to join this fragment with the synthetic SepRS2 fragment and the plasmid origin fragment. Sanger sequencing of the resulting vector revealed that while it was mostly correct, sections of sequence from the SepRS2 locus were missing. To correct this, both the faulty plasmid as well as the SepRS2 synthetic fragment were digest with MscI/SacI, allowing for the correct SepRS2 sequence to be ligated in. After this ligation, Sanger sequencing confirmed the correct assembly of pSepOpt.

6.3.7 Sep OTS Component Overexpression Vector Assembly

All primers and oligos described in this section are listed in **Table 6.2**. To insert the improved Sep OTS components individually into pJL1, PCR was first used to add C-terminal his-tags and flanking sequence homology to SepRS2 and EF-Sep. These pieces were inserted into pJL1 and/or pET-28a backbones using Gibson assembly[356]. For tRNASepB4, PCR was used to both add flanking sequence homology as well as to add the sequence for the hammerhead ribozyme [232] to the 5’ end of the tRNA, after which the part was inserted into pJL1 via Gibson assembly[356]. Sanger sequencing confirmed the correct assembly of all plasmids.

6.3.8 Western Blot

To confirm full-length expression of his-tagged SepRS2 and EF-Sep, sequence-confirmed plasmids were transformed into BL21 Star (DE3) and cells were allowed to grow overnight at 37°C in the presence of 1 mM IPTG to induce expression of these components. 200 µL of saturated overnight culture were collected. To prepare samples for gel electrophoresis, cells were pelleted and resuspended in 200 µL of nuclease-free water. 100 µL of this suspension was mixed with 34

µL of 4x NuPAGE® LDS Sample Buffer (Thermo Fisher Scientific, Inc., Waltham, MA) and boiled for 10 minutes. Following the boil, samples were spun at >13,500 x *g*. 12 µL of each sample was loaded into 12% Bis-Tris NuPAGE® gel (Thermo Fisher Scientific, Inc., Waltham, MA) and run at 130 V for 90 min using 1X MOPS running buffer (diluted from 20X MOPS SDS Running Buffer, Thermo Fisher Scientific, Inc., Waltham, MA). For reference, SeeBlue® Plus2 Pre-Stained Protein Standard (Thermo Fisher Scientific, Inc., Waltham, MA) was loaded into wells flanking the samples. Following electrophoresis, gels were washed in nanopure water. Proteins were transferred to Immun-Blot® PVDF membrane (Bio-rad, Hercules, CA) using a semi-dry protocol in 20% methanol/80% 1x MOPS. Transfer proceeded at 80 mA per gel for 55 min using a Trans-Blot® SD Semi-Dry Transfer Cell (Bio-rad, Hercules, CA). Blots were blocked overnight in 5% (m/v) fat-free dry milk at 4°C. Primary antibody (Sigma, Cat. #H1029, St. Louis, MO) was diluted 10,000x in PBS and applied to blots for 2 hrs. Secondary antibody (Bio-rad, Cat. #1708237, Hercules, CA) was diluted 3,000x in PBS-T and applied to blots for 1 hr. Finally, His-tagged proteins were visualized using the Immun-Blot® Opti-4CN™ Colorimetric kit (Bio-rad, Hercules, CA).

6.3.9 Purification and Protease Treatment of TPI

TPI synthesized in CFPS was purified using Qiagen Ni-NTA agarose (Qiagen, Valencia, CA) according to the manufacturer's manual. Following washes, bound TPI was eluted in 500 mM imidazole. 5 µL of Ulp1 protease produced in-house was added to the sample, and the entire volume was dialyzed against 50 mM Tris-HCl, 150 mM NaCl, and 1 mM DTT in a Slide-A-Lyzer™ Dialysis Cassette with a MWCO of 3.5 kDa. Dialysis proceeded for up to 48 hours at 4°C. Following dialysis, the cleaved sample was removed from the cassette for further use.

6.3.10 Native Mass Proteomics

Samples analyzed by native top down mass spectrometry (nTDMS) were buffer exchanged into 150 mM ammonium acetate with a nominal protein concentration of 10-20 μM. A customized Q Exactive HF mass spectrometer with Extended Mass Range (Thermo Fisher Scientific, Waltham, MA) was utilized for nTDMS analyses (*50*). The nTDMS platform employs direct infusion of sample into a native electrospray ionization (nESI) source held at +2 kV and coupled to a three-tiered tandem MS process. The process [467] first includes the analysis of the intact protein complex (MS^1), which provides the total mass (reported as a neutral average mass value). In stage two, the complex is collisionally activated with nitrogen gas to eject monomers (MS^2), thereby liberating the subunits that comprise each intact complex. In stage three, further vibrational activation of the ejected subunits via collisions with neutral gas yields backbone fragmentation products from each monomer (MS3) that were recorded at isotopic resolution (120,000 resolving power at *m/z* 400). These fragments were used to characterize the primary sequence of the protein and localize posttranslational modifications. Intact mass values for TPI complexes and ejected subunits were determined by deconvolution to convert data from the *m/z* to the mass domain using mMass (www.mmass.org) and MagTran [468]. Fragmentation data were processed using ProSight Lite [469] or ProSightPC 4.0 (Thermo Fisher Scientific, Waltham, MA) to assign recorded fragment ions to the primary sequence of the subunits. Unexplained mass shifts (Δm) observed at the MS^1 and MS^2 levels for the intact complex and subunits, respectively, were manually interrogated using the UNIMOD database as a reference to provide candidate modifications.

6.3.11 TPI Activity Assay

Kinetic rates of DLD1 fractionated cell lysates were compared using a commercial activity assay (ID#: ab197001; Abcam, Cambridge, United Kingdom) based on the enzyme-coupled assay published by Plaut and Knowles [470]. Fractions of interest contained equal concentrations of triose phosphate isomerase (TPI), but differing ratios of phosphorylated and unmodified TPI. The assay couples TPI and its downstream partner glyceraldehyde-3 phosphate dehydrogenase (GAPDH), which reduces NAD+ to NADH. Given sufficient concentration of TPI substrate and GAPDH (roughly x5000 times more concentrated than TPI), the formation of NADH will be solely dependent on the concentration and activity of TPI. NADH formation was monitored via UV-Vis absorbance at 450 nm using a Neo Synergy2 plate reader (Biotek, Winooski, VT) in kinetic mode for 40 minutes. All assays were run at 37°C.

6.3.12 Molecular Dynamics Simulations

Using the physics-based docking simulation programs implemented in Schrodinger platform, the small molecule ligand docking of the substrate dihydroxy acetone phosphate was carried out in the active site of human TPI (4POC.pdb) crystal structure to obtain the binding energy (BE) or Kd of the substrate. Then the TPI structure was phosphorylated at Serine 20 residue and molecular dynamics simulations at 5 - 50 ns were performed to visualize the conformational change to the protein structure. Considering the energy-minimized structure of the phosphorylated protein, again the substrate was docked into the active site of the protein and the BE/Kd was computed.

Table 6.1. Strains and plasmids used in this study. Km^R, Ap^R, and Cm^R are kanamycin, ampicillin, and chloramphenicol resistance, respectively. 'Δ' indicates deleted gene/DNA sequence, and '∇' indicates gene(s)

inserted into the genome at the specified locus behind the specified promoter. '//' denote amino acid substitution mutations made in the *1* gene open reading frame. '+' denotes DNA sequence added to the preceding gene open reading frame. '[]' denote amino acid substitution mutations made in the *serB* gene open reading frame.

Strains and plasmids	Genotype/relevant characteristics	Source
Strains		
BL21 Star™ (DE3)	F⁻ *ompT hsdS*$_B$ ($r_B^-m_B^-$) *gal dcm rne131* (DE3)	Life Technologies
*C321.*Δ*A.759*	*C321.*Δ*A. endA⁻ gor⁻ rne⁻ mazF⁻*, Ap^R	[405]
*C321.*Δ*A.759.*Δ*serB*	*C321.*Δ*A. endA⁻ gor⁻ rne⁻ mazF⁻ serB⁻*, Ap^R	This study
DH5α	F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM 15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ⁻	Invitrogen
B-95.Δ*A*		[471]
759.T7.D	*C321.*Δ*A.759.*∇*1.asl.Lpp5/K172G, K179A/*, Ap^R Km^R	See Ch. 4
*759.T7.D.*Δ*AbR*	*C321.*Δ*A.759.*∇*1.asl.Lpp5/K172G, K179A/*	This study
*759.T7.D.*Δ*His*	*C321.*Δ*A.759.*∇*1.asl.Lpp5/K172G, K179A/.*Δ*His*	This study
*759.T7.D.*Δ*His.serB*Strep	*C321.*Δ*A.759.*∇*1.asl.Lpp5/K172G, K179A/.*Δ*His.serB*+N-terminal Strep-II tag	This study
*759.T7.D.*Δ*His.serB*Strep *.serB*sens	*C321.*Δ*A.759.*∇*1.asl.Lpp5/K172G, K179A/.*Δ*His.serB*+N-terminal Strep-II tag.[D216R, L218R]	This study
Plasmids		
B40OTS	Km^R, SepRS, 5x *o*-tRNASep, EF-Sep	[136]
pSepOpt	Km^R, SepRS2, *Lpp5*::o-tRNASepB4, EF-Sep	This study
pMA7CR_2.0	Ap^R, *Tet*::Cas9, *ARA*::λβ,*dam*	[472]
pMAZ-SK	Km^R, *Tet::*gRNA	[472]
pMAZ-SacI	Km^R, *Tet::*gRNA+SacI site	This study
pMAZ-Δ*His*	Km^R, *Tet::* Δ*His* gRNA	This study
pMAZ-*serB*Strep	Km^R, *Tet:: serB*Strep gRNA	This study
pMAZ-*serB*sens	Km^R, *Tet:: serB*Strep gRNA	This study
pMAZ-CureAll	Km^R, *Tet::* pMA7CR_2.0 gRNA	This study

pY71-sfGFP	Km^R, P_{T7}::super folder green fluorescent protein (sfGFP), C-terminal strep-tag	[19]
pY71-sfGFP-T216amb	pY71-sfGFP with amber codon at T216	[19]
pY71-sfGFP-2amb	pY71-sfGFP with amber codon at N212 and T216	[24]
pJL1-TPI-MetON	Km^R, P_{T7}::His-TEV-TPI with leader methionine	This study
pJL1-TPI-MetOFF	Km^R, P_{T7}::His-TEV-TPI without leader methionine	This study
pJL1-TPI-S20X-MetOFF	Km^R, P_{T7}::His-TEV-TPI without leader methionine, amber codon at S20	This study
pJL1-TPI-S20X-MetON	Km^R, P_{T7}::His-TEV-TPI with leader methionine, amber codon at S20	This study
pJL1-TPI-S4X-MetON	Km^R, P_{T7}::His-TEV-TPI with leader methionine, amber codon at S4	This study
pJL1-TPI-SUMO-MetOFF	Km^R, P_{T7}::His-SUMO-TPI without leader methionine	This study
pJL1-TPI-SUMO-MetOFF-S20X	Km^R, P_{T7}::His-SUMO-TPI without leader methionine, amber codon at S20	This study
pJL1-TPI-SUMO-MetOFF-S20E	Km^R, P_{T7}::His-SUMO-TPI without leader methionine, S20E	This study
pJL1-TPI-SUMO-MetOFF-E104D	Km^R, P_{T7}::His-SUMO-TPI without leader methionine, E104D	This study

Table 6.2. Primers used in TPI plasmid assembly/sequencing, pMAZ plasmid assembly/sequencing, CRMAGE oligos, SepOTS component expression vector assembly/sequencing, and pSepOpt assembly/sequencing. The first four bases of the 5'-MAGE oligonucleotides were phosphorothioated (*).

Primer Name	DNA Sequence (listed 5' to 3')
TPI plasmid assembly and sequencing	
TPI_E104D_F	AGCGTCGCCACGTCTTTGGGGATTCCGACGAGCTTATTGGACA
TPI_E104D_R	CCCAAAGACGTGGCGACGCTCGCTGTGACCTAAAACAACC
TPI_string_F	TTTAAGAAGGAGATATACAT
TPI_string_R	TTTGTTAGCAGCCGGTCGAC
TPI_cPCR_F	GGTGTGATTGCGTGCATCGG
pJL1_cPCR_R	ctcgctcaggcgcaatcacg
pJL1_TPI_seq2_F	GGGAAATTGGAAGATGAACG

pMAZ plasmid gRNA oligos and primers

T7delHis_oligo1	gagcacTAATCGTGTTGGATCCGTGAgttttagagctagaaat
T7delHis_oligo2	ctaaaacTCACGGATCCAACACGATTAgtgctcagtatctct
serBHis_oligo1	gagcacATCTTCAGGCAGGTCGCACCgttttagagctagaaat
serBHis_oligo2	ctaaaacGGTGCGACCTGCCTGAAGATgtgctcagtatctct
serBOmpT_oligo1	gagcacCAGTTCATTGGCTACCACGGgttttagagctagaaat
serBOmpT_oligo2	ctaaaacCCGTGGTAGCCAATGAACTGgtgctcagtatctct
pMAZbb_F	AGCTAGAAAUAGCAAGTTAAAATAAGGC
pMAZbb_R	AGTATCTCUATCACTGATAGGGATGTCA
T7delHisMASC_HisF	GCATCACCATCACCATCACG
T7delHisMASC_wtF	GGAGGTACATAATGaacacg
T7delHisMASC_R	tttccatgcggtgttttgcg
serBHisMASC_HisF	AGCCTTAATGCACCACCACC
serBHisMASC_wtF	AGCCTTAATGCCTAACATTAC
serBMASC_R	CAAACAAAGGCGCGTTTTGG
serBsensMASC_wtF	AATACCTGCGCGACAAGCTG
serBsensMASC_mutF	AATACCTGCGCCGTAAGCGT
serBStrepMASC_mutF	AGCCTTAATGTGGAGCCACC
serB_seqPCR_F	TGACTAAACCATGACGCTCC
serB_seq2_F	GTGGCGGAAGTAACCGAACG
serB_seqPCR_R	ATCAGAATGGCACTTCCTGG
aCas9_cassette_F	tccctatcagtgatagagattgac
aCas9_cassette_R	tgtgaccgtgtgcttctc
pMAZbb_F	gagaagcacacggtcacac
pMAZbb_R	gtcaatctctatcactgataggga
cureALL_cPCR_F	ggcgataagtcgtgtcttac
cureALL_cPCR_R	ttgccatcctatggaactgc
pMAZ_seq	caattcagcaaattgtgaacatcatc

MAGE oligos

T7delHis_MAGE	c*a*g*t*tcgatgtcagagaagtcgttcttagcgatgttaatcgtgttCATTATGTACCTCCTTACTGTTTTGTTTTAATTGTTATCCGCTCA
serBHis_MAGE	A*A*A*G*AGACATCTTCAGGCAGGTCGCACCATGTAATGTTAGGATGGTGATGGTGGTGGTGCATTAAGGCTCCTGTAAAATCGTTCGAAGCA

serBOmpT_MAGE	A*C*C*G*TCCATGATCTCCAGTTCATTGGCTACCACGGCTGTCAG GCGACGCTTACGGCGCAGGTATTCAGCAAAGAAAGTAAAGCCGC CGGA
serBStrep_MAGE	A*A*A*G*AGACATCTTCAGGCAGGTCGCACCATGTAATGTTAGGT TTTTCGAACTGCGGGTGGCTCCACATTAAGGCTCCTGTAAAATCG TTCGAAGCA
kanR_KO_MAGE	G*G*T*T*GGGCGTCGCTTGGTCGGTCATTTCGAACCCCAGAGTCC CGCCATGCGAAACGATCCTCATCCTGTCTCTTGATCAGATCTTGAT CC

Sep OTS individual component plasmid assembly/sequencing

tRNA_Sep_str_F	acatttccccgaaaagtgc
tRNA_Sep_str_R	ttcgatggtgtcaacgtaaatgc
SepRS2_pY71_R	tttgttagcagccggtcgacttattagtgatggtgatggtgatgttcaatttttacttcgacatttaaGC
SepRS2_pY71_F	tttaagaaggagatatacatatgtttaaaagagaagaaatcattg
pY71_tztRNASep_bb_R	taccgtttcgtcctcacggactcatcaggccgtctccctatagtgagtcgtattaagatctgaattcg
pY71_tztRNASep_bb_F	ccggcggtagttcagcctgg
tRNASeptz_ins_F	tccgtgaggacgaaacggtacccggtaccgtcGCCGGGGGTAGTCTAGGGG
tRNASeptz_ins_R	ccaggctgaactaccgccggTGGAGCCGGGGGGTGGGATTTG
SepRSOpt_R	taatatactttaggtaccgagctcttattcaatttttacttcgacatttaaGC
EFSep_pY71_F	tttaagaaggagatatacatatgtctaaagaaaagtttgaacg
EFSep_pY71_R	tttgttagcagccggtcgacttattagtgatggtgatggtgatggctcagaacttttgctacaacg
SepRS_cPCR_F	AATTGCAGGTGAAGGCGTGG
EFSep_cPCR_F	cctggaactggctggcttcc
tztRNASep_cPCR_F	ACGTGGGTTCAAATCCCACC
tztRNASep_bb_R_v2	taccgtttcgtcctcacggactcatcaggccgtctccctatagtgagtcgtattaagatctgaattcg
tRNASepLET_R	TGGAGCCGGGGGTGGGATTTG
tRNASepLET500_F	CCGAAGGTAACTGGCTTCAGCAGAG
tRNASepLET300_F	GACTCAAGACGATAGTTACCGGATAAG
SepRS2_pET_F	tttaagaaggagatataccatatgtttaaaagagaagaaatcattg
SepRS2_pET_R	gctttgttagcagccggatcTTAgtgatggtgatggtgatgttcaatttttacttcgacatttaaGC

pSepOpt assembly, screening, and sequencing

tRNA_lacI_F	gtcaattcagggtggtgaatgtgaaaccagtaacgttatacg
lacI_R	tgccatgggtctgtttcc
lacI_SepRS_F	ggaaacagacccatggcaatgtttaaaagagaagaaatcattgaaatgg
SepRS_R	ctaatatactttaggtaccgagctcttattcaatttttacttcgacatttaaGC
EF_Sep_F	gagctcggtacctaaagtatattag
EF_Sep_R	agcttgcatgcctgcaggtc
ori_kanR_F	tacaaactctttttgtttgc
ori_kanR_R	gtatccgctcatgaattaattcttagaaaaactcatcgagcatc
lacI_F	tttacgttgacaccatcgaatgg
lacI_R	atttcttctctttttaaacattgc
SepRS2_F	atgtttaaaagagaagaaatc
SepRS2_R	ctaatatactttaggtaccgagc
B40_Chin_cPCR_F	tcaaagttaaagttATGTGG
B40_Chin_cPCR_R	gtatcaggctgaaaatcttc
BJD_SepOptSeq_1	aaagtaatcgacgttagagg
BJD_SepOptSeq_2	cctggaactggctggcttcc
BJD_SepOptSeq_3	tctccttcattacagaaacg
BJD_SepOptSeq_4	gcagccatcggaagctgtgg
BJD_SepOptSeq_5	ttgaaatttcaagcgagatgg
BJD_SepOptSeq_6	aaatccccttgtgaattagc
SepOptScrn1_F	ttagcgggacttccaagacc
SepOptScrn1_R	tttaacatagtcggcgtgcc
SepOptScrn2_F	acaggattttcgcctgctgg
SepOptScrn2_R	agcatctatccaaaactgcc
SepOptScrn3_F	cccagtcacgtagcgatagc
SepOptScrn3_R	gcgccacttatttttgatcg

6.4 Results

6.4.1 Development of a Recoded Chassis Strain Optimized for Sep

I began by setting out to develop a chassis strain whose lysates would enable preparative scale synthesis of proteins featuring site-specific Sep incorporations using CFPS. To be capable of this, I reasoned that the strain would need to possess three critical features. First, the strain would need to lack the amber codon-specific release factor 1 (RF1) to eliminate competition at the amber codons used to encode Sep in targets of interest[394]. Second, to enable preparative scale synthesis of targets the strain's lysates would need to be highly productive in CFPS. Third, the Sep-specific phosphatase SerB would need to be inactive in the strain's lysates[6, 136]. Based on these requirements, I decided to use fully-recoded, RF1-deficient, CFPS optimized strains developed by our group (see Chapters 3 and 4) as the starting point for my development efforts as they thoroughly satisfy the first two requirements described above. Thus, going forward my efforts mostly revolved around eliminating SerB activity in the strain's lysates.

I first looked to assess the overall impact of a *serB* deactivation/knockout on the health of the organism and the productivity of its lysates. To test this, I used multiplex advanced genome engineering (MAGE)[25] to install inactivating point mutations in the coding region of *serB* in the genome of rEcoli.759 (see Chapter 3) to yield rEcoli.759.Δ*serB*. As expected, inactivation of SerB activity in the cells caused a significant growth defect, and lysates prepared from the modified strain suffered a ~75% reduction in productivity as compared to the parent strain to yield ~450 μg/mL of wild type superfolder green fluorescent protein (sfGFP, **Figure 6.1**). Based on the overly deleterious effects of completely eliminating SerB activity from the strain, we discontinued this strategy for SerB inactivation in CFPS.

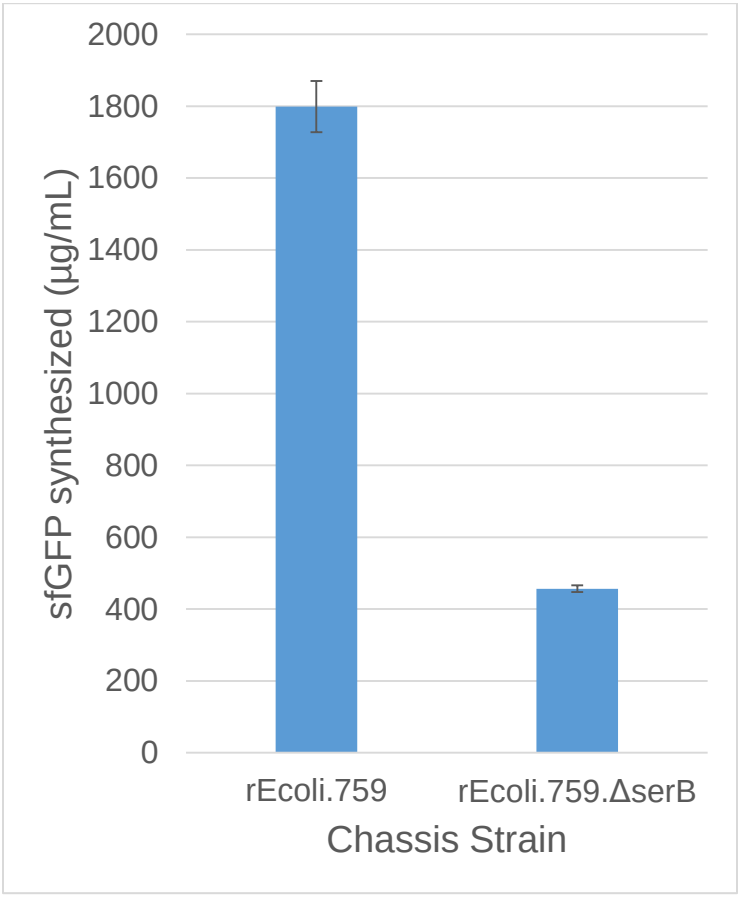


Figure 6.1. Assessment of the impact of *serB* inactivation on CFPS productivity. Shown is sfGFP production in CFPS using lysates derived from the indicated strains. For each condition, three independent reactions were performed and one standard deviation is shown.

Next, I pursued a scheme that would leave SerB activity intact while the chassis cells were growing but then eliminate the troublesome phosphatase during CFPS lysate preparation. I reasoned that this approach would mitigate the growth defect and associated reduction in lysate productivity observed with complete *serB* knockouts while still achieving the goal of SerB inactivation in CFPS. My scheme involved two key modifications to *serB* that could be used individually or in combination to selectively remove the enzyme from CFPS lysates. First, I would

install an affinity tag at the N-terminus of the protein that could be used to both visualize it on a western blot and selectively pull it out of lysates, an approach that has been successfully employed for the removal of other targets[328, 329]. Second, I would mutate *serB* to install a run of 5 basic residues at an external loop in order to make it a more efficient substrate for the periplasmic protease OmpT[434], a strategy that has been used before to conditionally inactivate protein targets in lysates[330]. Because OmpT operates primarily in the periplasm and SerB operates in the cytoplasm, the two proteins would be physically separated until cellular rupturing during lysate prep causes the two compartments to mix. I hypothesized that, in this way, SerB would remain intact and active during cell growth and then during lysis be rapidly processed upon exposure to OmpT. To test this, I leveraged an approach that combines CRISPR/Cas9 and MAGE (CRMAGE)[472] to cyclically make edits to the genome of rEcoli.759.T7.D (see Chapter 4), first removing the N-terminal His-tag from the genomic copy of the *1* gene[409] (yielding strain rEcoli.759.T7.D.ΔHis), then installing a Strep-tag II on the N-terminus of *serB* (yielding strain rEcoli.759.T7.D.ΔHis.*serB*Strep), and finally installing the mutations D216R/L218R to generate a region with 5 tandem basic residues (RRKRR) within SerB (yielding strain rEcoli.759.T7.D.ΔHis.*serB*Strep.*serB*sens, or simply rEcoli.759.T7.Sep).

I next wanted to assess the extent to which the series of genomic edits made to rEcoli.759.T7.D impacted its overall productivity in CFPS. To test this, I prepared a lysate from rEcoli.759.T7.Sep and applied it in CFPS toward the synthesis of sfGFP. Encouragingly, the lysate still composed a one-pot system, and were able to synthesize ~1.4 g/L of sfGFPwt in the absence of supplemental T7 RNA polymerase **(Figure 6.2)**. This represents only a modest 12.5% reduction in productivity relative to the parental strain (see Chapter 4) and shows great promise for the use

of rEcoli.759.T7.Sep as a production platform for the preparative synthesis of proteins featuring Sep.

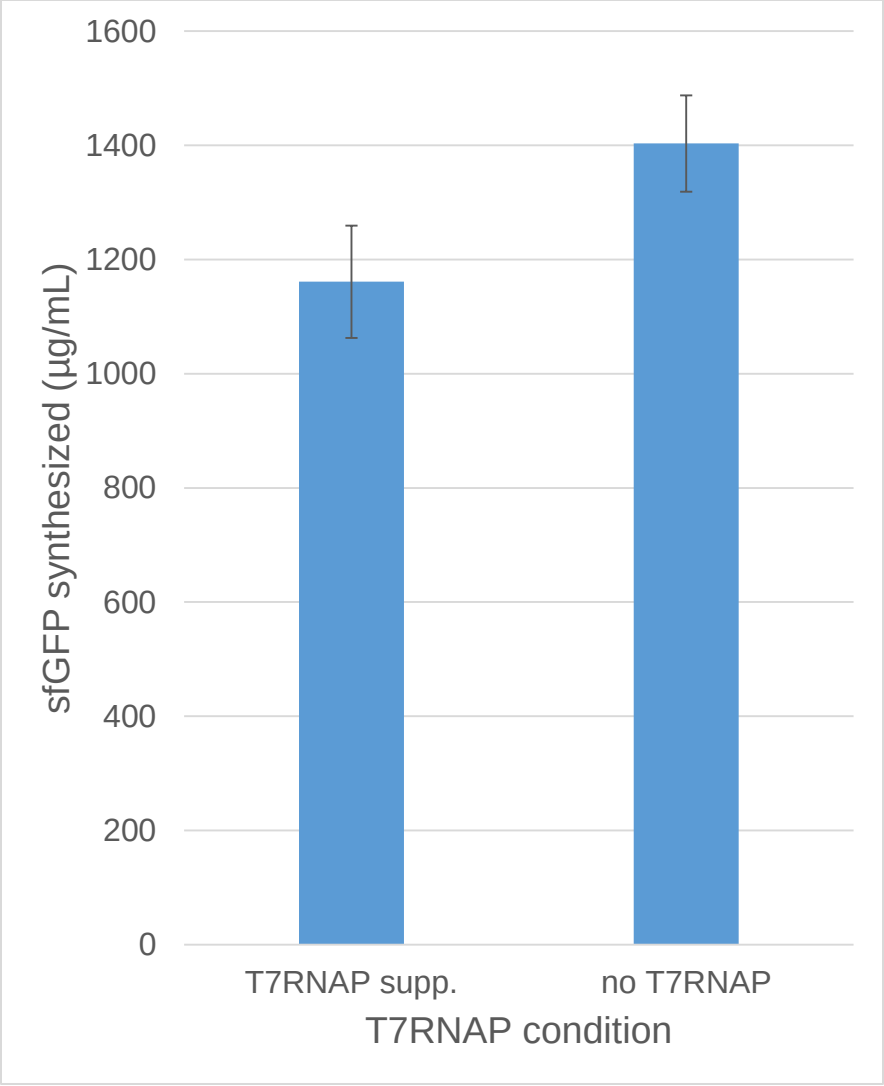


Figure 6.2. Characterization of rEcoli.759.T7.Sep lysates in CFPS. Shown is sfGFP production in CFPS using a lysate derived from rEcoli.759.T7.Sep. For these reactions, the lysate was tested both with (left) and without (right) supplementation with 16 µg/mL purified T7 RNA polymerase (T7RNAP). For each condition, three independent reactions were performed and one standard deviation is shown.

Further characterization of the capabilities of lysates derived from rEcoli.759.T7.Sep remain incomplete. Some investigation into the ability of this system to facilitate Sep incorporation is described in the next section, but more experiments are currently needed.

6.4.2 pSepOpt – an Improved Sep OTS Expression Vector

A past effort to incorporate Sep using CFPS systems induced the plasmid pKD-SepRS-EFSep-5xtRNASep (B40OTS, Jesse Rinehart, unpublished) in the source strain in order to overexpress the OTS components such that they would be enriched in the resulting lysate[136]. This vector features the original Sep-specific aminoacyl-tRNA synthetase (SepRS), tRNA (tRNASep), and elongation factor mutant (EF-Sep) first described by Park *et al*, and while it is capable of enabling Sep incorporation *in vitro*, occurrences of errant misincorporation at amber codons are high using these orthogonal components. A recent effort by Rogerson *et al* sought to address this limitation by co-evolving SepRS and tRNASep, resulting in the identification of a novel orthogonal pair (SepRS2 and tRNASepB4) that demonstrated an 18-fold improvement in Sep incorporation efficiency[225]; however, the authors provide relatively few details about the exact constructs used and did not make their plasmids available for use in other laboratories. Thus, in order to use these improved components it was necessary for me to construct a completely novel Sep OTS expression plasmid encoding the improved components SepRS2 and tRNASepB4.

Conceptually, my approach was to essentially "copy/paste" the sequences for each improved component into B40OTS in place of the original component. In this way, the plasmid architecture and genetic regulatory sequences of B40OTS could be retained, and simply be directed to instead drive expression of the improved components. To achieve this, I first identified the specific mutations present in SepRS2 and tRNASepB4, after which I purchased linear DNA

constructs encoding the improved components. Initially, I ordered the tRNASepB4 construct as a cassette featuring 5 identical tandem copies of the tRNA, as this is how the original tRNA is encoded in B40OTS. However, this construct proved impossible to synthesize. As scarless assembly of repetitive DNA sequences in tandem is notoriously difficult[473], I changed my approach. My new design included only a single copy of tRNASepB4, and sought to overcome the reduction in copy number by putting the single copy under the control of the strong constitutive promoter Lpp5 [418].

My initial plan was to assemble the B40OTS derivative encoding the improved Sep OTS components (which I had come to refer to as pSepOpt) via ligation of four linear pieces. Two of the pieces would be the synthetic linear constructs encoding the improved Sep OTS components already discussed. The other two pieces would be excised from B40OTS via restriction digest, and collectively include the origin of replication, EF-Sep, and the selectable marker. All four pieces would be digested by restriction endonucleases to generate compatible sticky ends and ligated together to yield pSepOpt. I attempted assembly using this approach several times, but every attempt was a failure.

Next, I slightly shifted my approach to use seamless Gibson assembly[356] instead of ligation to construct pSepOpt. Assembly would essentially use the same four DNA pieces as above, though this time PCR would be required to install regions of sequence homology at the ends of adjacent pieces for assembly using Gibson's method. As before, attempts to assemble pSepOpt in this way failed. I reasoned that reducing the number of pieces involved in the assembly reactions would increase efficiency and lead to a successful outcome. To test this, I used overlap extension PCR to assemble multiple combinations of neighboring pieces prior to Gibson assembly

and reduce the number of pieces being assembled to three. In this way, I was able to assemble a very-nearly-correct version of pSepOpt that was only missing sections of sequence in the SepRS2 coding region. Excision of the incorrect portion of the plasmid via restriction digest and ligation of the correct sequence finally yielded sequence-confirmed pSepOpt.

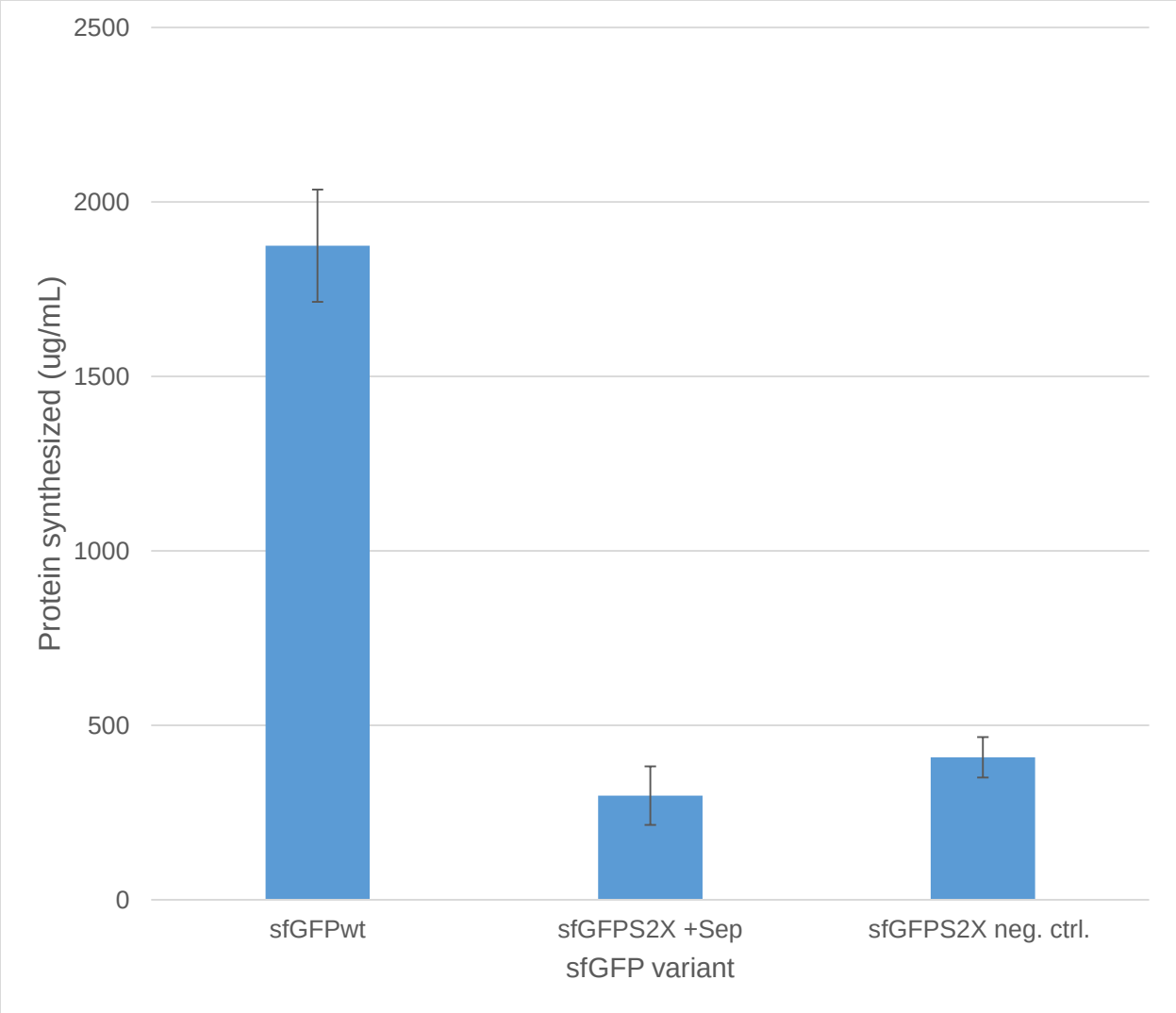


Figure 6.3. Attempted Sep incorporation using pSepOpt in B95 lysates. Shown is sfGFP production in CFPS using a lysate derived from B95 carrying pSepOpt. This lysate was directed to synthesize the indicated sfGFP variants. For sfGFP with an amber codon at position 2 (sfGFPS2X) reactions were either supplemented with 2mM Sep (+Sep) or had Sep withheld (neg. ctrl.). For each condition, three independent reactions were performed and one standard deviation is shown.

I next wanted to assess the extent to which pSepOpt could enable Sep incorporation in CFPS. To test this, I first attempted to transform the plasmid into a *serB*-deficient variant of a strain that had previously been optimized for high productivity in CFPS[405]. No successful transformants were ever obtained, likely a result of the cumulative cytotoxic stress imposed by both the *serB* knockout and the expression of the Sep OTS components. I next transformed the plasmid into the partially-recoded release factor 1 (RF1)-deficient strain B95[471]. Though *serB* is intact in this strain, I reasoned that its interference may be mitigated in the presence of highly-efficient Sep OTS components. B95 was able to grow while carrying pSepOpt, allowing me to prepare a lysate. Overall productivity of the lysate was quite high, with wild-type superfolder green fluorescent protein (sfGFPwt) yields of 1875 ± 161 μg/mL **(Figure 6.3)**. Unfortunately, Sep incorporation was below the negative control, suggesting that incorporation was not actually occurring. Sequencing revealed that upon being introduced into B95 pSepOpt had developed an inactivating point mutation in tRNASepB4, accounting for this failure. Finally, I transformed pSepOpt into rEcoli.759.T7.Sep. The strain was able to carry the plasmid, and sequencing confirmed that no inactivating point mutations were obtained post-transformation. I prepared a lysate from the strain to again attempt Sep incorporation in CFPS **(Figure 6.4)**. Initial results were mixed – unsurprisingly, overall productivity of the lysate in the plasmid-bearing strain was reduced, and it showed essentially no ability to incorporate Sep into sfGFP with an amber codon at position 216 (sfGFPT216X). However, this codon position in sfGFP has been problematic for Sep incorporation in the past (data not shown), so I instead applied the lysate towards synthesis of sfGFP with an amber codon at position 2 (sfGFPS2X). In this case, near wild type yields of the amber mutant were observed both in the presence and absence of the Sep OTS components,

suggesting that errant readthrough of the amber codon accounts for most of the full-length product observed.

Characterization of pSepOpt in CFPS remains incomplete as of this writing.

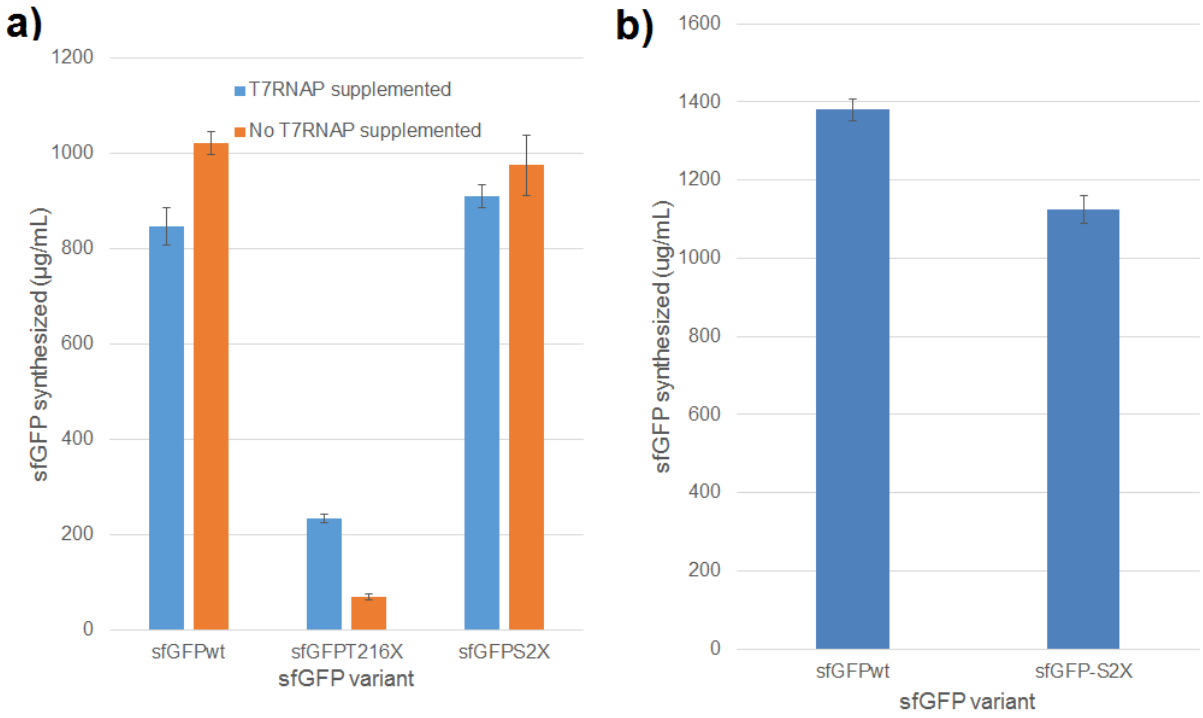


Figure 6.4. Attempted Sep incorporation using pSepOpt in rEcoli.759.T7.Sep lysates. Shown is sfGFP production in CFPS using a lysates derived from rEcoli.759.T7.Sep. **(a)** Synthesis of the indicated variants of sfGFP using a lysate derived from rEcoli.759.T7.Sep carrying pSepOpt such that the lysate is expected to be enriched with Sep OTS components. Reactions were performed both with and without supplementation with T7 RNA polymerase (T7RNAP). **(b)** Synthesis of the indicated variants of sfGFP using a lysate derived from rEcoli.759.T7.Sep *without* pSepOpt (*i.e.* Sep OTS components are not present). For each condition, three independent reactions were performed and one standard deviation is shown.

6.4.3 Supplementation of Improved Sep OTS Components in CFPS

The boundary-free nature of CFPS reactions permits the direct supplementation of limiting substrates into the reaction volume[280]. This feature makes CFPS particularly well-suited for the use of OTSs to catalyze ncAA incorporation, since purified OTS components (*e.g.* orthogonal

synthetase) can be purified and added directly to reactions to help overcome the poor catalytic efficiency generally exhibited by these enzymes[23]. Thus, I set out to encode the optimized Sep OTS components (SepRS2, tRNASepB4, and EF-Sep) individually into plasmid backbone pJL1 for T7 RNA polymerase-driven overexpression and purification *in vivo*. To achieve this, His-tagged variants of SepRS2 and EF-Sep were assembled into pJL1 using Gibson assembly[356]. Additionally, a variant of tRNASepB4 featuring a self-cleaving hammerhead ribozyme at the 5' end of the coding sequence was inserted into this backbone[232]. Sequencing and anti-histag western blot analysis identified a functional expression vector for EF-Sep, and sequencing confirmed correct assembly of the plasmid encoding the tRNA. Functional pJL1-SepRS2 (in terms of His-tag detection) was not achieved, although sequencing suggested that the plasmid was assembling correctly. In an attempt to troubleshoot the nonfunctional SepRS2 plasmid, I cloned the construct into the *in vivo* expression vector pET-28a. Again, though sequencing confirmed that the plasmid was correctly assembled, no samples generated a positive result via anti-histag western blot. A functional overexpression vector for SepRS2 remains elusive. As of this writing, none of these purified components have been assessed in CFPS.

6.4.4 Direct expression of TPI phosphoproteoforms

Next, I entered into a collaboration to investigate the functional roles played by phosphorylation in modulating activity of the enzyme triosephosphate isomerase (TPI). TPI is a key enzyme in glycolysis that isomerizes dihydroxyacetone phosphate into glyceraldehyde-3-phosphate, with a k_{cat}/K_M ranging from 10^7 to 10^{10} $s^{-1}M^{-1}$ [474, 475]. TPI functions endogenously as a dimer composed of two subunits which can be differentially modified to give rise to a multi-proteoform complex (MPC). Due to its high catalytic efficiency, it is the original example of the

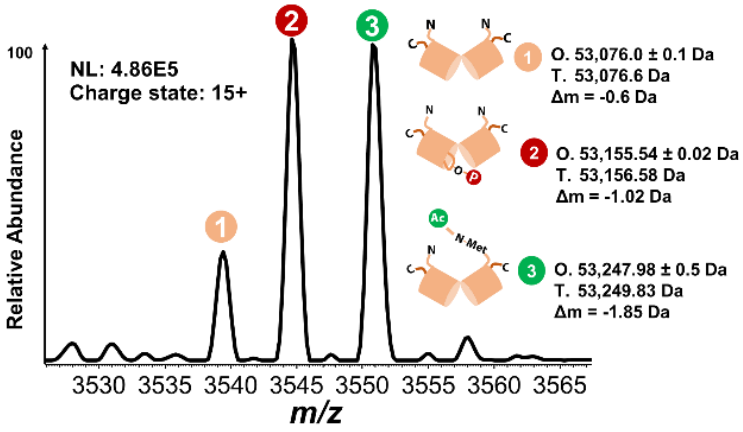
"perfect enzyme" as its rate of reaction is close to the limit of diffusion[476, 477]. Several hypotheses have been advanced to explain why TPI is so catalytically efficient, including the intense evolutionary pressure to instantly generate muscle ATP through glycolysis, the fact that TPI acts near thermodynamic equilibrium, and the relative simplicity of the proton-transfer reaction it catalyses[478]. Given its status as a "perfect enzyme", it was thus imperative to determine the role of the observed PTMs on the function of TPI.

To better understand the ways in which the TPI complex can be regulated by PTMs, we first set out to identify and characterize the presence and distribution of TPI MPCs naturally occurring in cells. To achieve this, we applied native Top-Down Mass Spectrometry (nTDMS) to protein samples derived from cell lines. Unlike many structural biology techniques, nTDMS can interrogate the regulatory role of PTMs because it can characterize the endogenous modifications and non-covalent cofactors of protein complexes[172, 479, 480]. Our approach, which does not rely on digestion and/or denaturation of the protein complexes like in common proteomic practices, begins by measuring the intact masses of the various modified states of a single intact complex. Subsequent stages of MS/MS then liberate the subunits and ultimately yield fragment ions that can be used to localize PTMs on the protein sequence[467]. This technique has enabled us to track changes in the PTM stoichiometry of endogenous MPCs as a function of cell type[172].

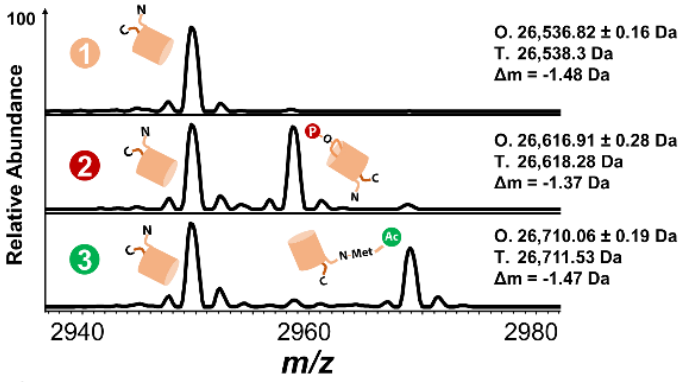
The targeted analysis of TPI in multiple cell-lines, including colorectal cancer, HEK, and HeLa cells, revealed the predominance of three major MPCs **(Figure 6.5)**. One MPC, present in all cell lines, was characterized as a homodimer of TPI without the initiator methionine (Met-off). The second MPC had one subunit phosphorylated near the N-terminus at serine 20, and the third had one subunit that retained the initiator methionine (Met-on) and contained an N-terminal

acetylation. Importantly, the abundances of the different MPCs did not reflect a random distribution of possible subunit combinations in any of the tested cell lines. In fact, homodimers of the modified subunits were not observed in measurable amounts suggesting some form of stoichiometric control. The activity of the endogenous TPI MPCs was measured **(Figure 6.6)**, demonstrating that the fraction containing phospho-TPI had nearly 2-fold higher overall activity, suggesting a functional role for this previously uncharacterized phosphorylation in regulating the catalytic efficiency of the MPC.

MS¹: Native mass spectrum of intact TIM multi-proteoform complexes



MS²: Native Top Down Mass Spectra of TIM subunits ejected from complex



MS³: Graphical fragment maps of Ob3b subunits

Met-Off - Unmodified

Met-Off Phospho-Ser20

Met-On + N-terminal acetylation

Figure 6.5 (previous page). Native Top Down Mass Spectrometry (nTDMS) performed on TPI from HEK cells. The nTDMS platform first consists of the measurement of the mass of the intact multi-proteoform complexes (MPC) of TPI (MS^1, top panel), reported as average mass. Shown is a zoom-in of a single charge state of TPI indicating the three MPCs present. In the second step, a single charge state of each MPC is isolated and collisionally activated with nitrogen gas to eject monomers (MS^2, middle panel). In this way, the subunits that make up each intact MPC are liberated and characterized (calculated average masses are provided). Shown in the figure are single charge states of the liberated subunits for each MPC activated. When activating MPC (1), only one subunit is ejected, indicating that the intact MPC is a homodimer. Upon activation of MPC (2) and (3), two different subunits are ejected indicating that MPCs (2) and (3) are each heterodimers. In the third step, each ejected subunit is isolated and subjected to further vibrational activation via collisions with neutral gas. This process yields backbone fragmentation products from each monomer (MS^3, bottom panel), which are depicted as blue flags in the graphical fragment maps of the lower panel. Each flag indicates the segments of the protein sequence that are accounted for in mass by fragmentation, thereby enabling the identification of proteoforms and the characterization of their modifications (blue, phosphorylation; red, acetylation). Met-Off indicates removal of initiator methionine; Met-On indicates retention of initiator methionine.

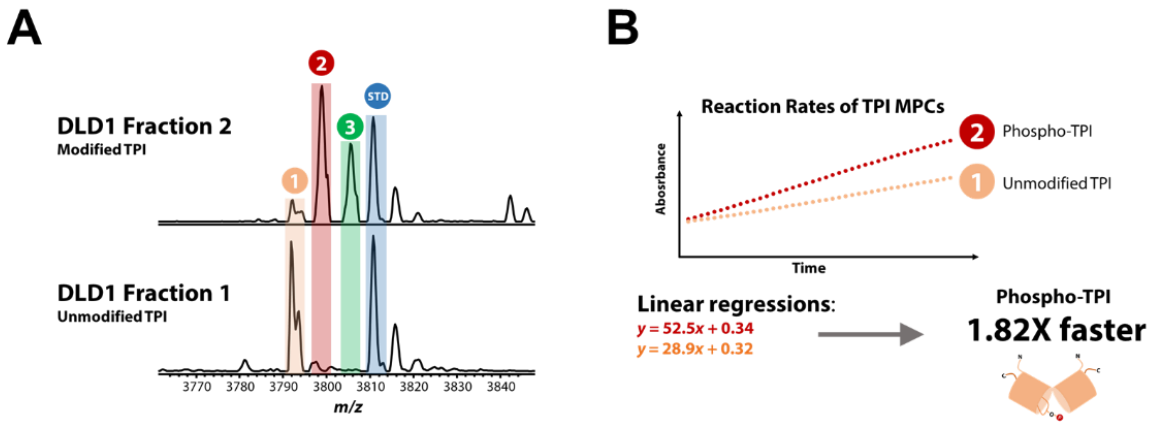


Figure 6.6. Kinetic analysis of endogenous TPI MPCs. (a) Fractionated DLD1 cell lysate using ion exchange chromatography resulted in two TPI-containing fractions, which were analysed using nTDMS to determine relative abundance of TPI MPCs. A TPI standard (STD) was spiked in at various concentrations and the relative intensities of the different MPCs were measured to determine ratio of MetOFF (MPC 1, orange), phospho-TPI (MPC 2, red), and MetON-NtAc (MPC 3, green). **(b)** With known concentrations, the TPI activity of each fraction was measured, yielding absorbance measurements that are reflective of TPI specific activity.

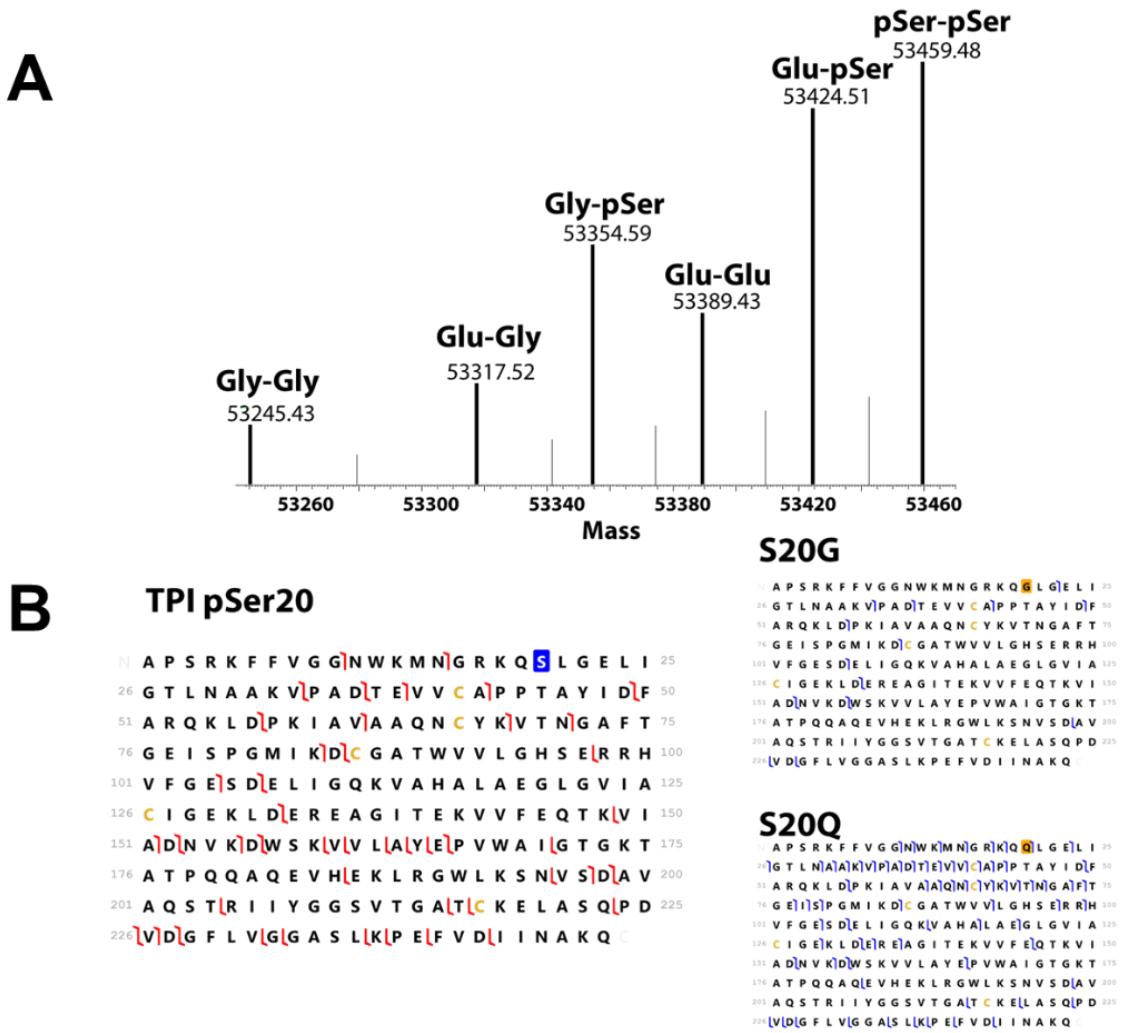
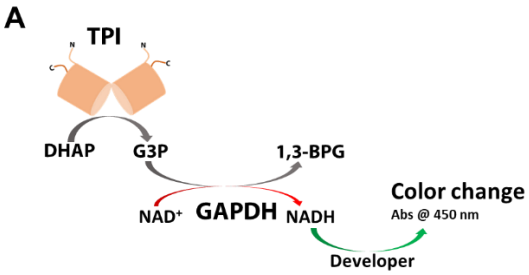


Figure 6.7. **nTDMS analysis of phospho-TPI generated by CFPS. (a)** Monoisotopic masses of TPI dimers generated by CFPS, which include TPI proteoforms with the misincorporation of glutamine and glycine at the site of the intended phosphorylation. The resultant spectrum is the dimerization of the various TPI proteoforms generated. (B) Fragmentation of each TPI proteoform indicating localization of phosphorylation and the misincorporation of various aminoacids at the intended phospho-site.

To validate this result, we next sought to obtain homogeneous samples of TPI featuring phosphorylation at serine 20. To achieve this, we applied approaches described earlier in this chapter, leveraging the use of OTSs in CFPS to directly incorporate Sep into TPI monomers at the indicated position. Lysates derived from rEcoli.759.Δ*serB* carrying the B40OTS plasmid were

directed to synthesize several proteoforms of TPI, namely the wildtype form lacking the initiator methionine (OFF) and wildtype lacking the initiator methionine with an amber codon at position 20 to encode Sep (phospho-TPI). Product was purified via affinity chromatography, after which nTDMS was applied to investigate the resulting dimers. Interestingly, nTDMS analysis of the CFPS-generated phospho-TPI shows that it can homodimerize in vitro **(Figure 6.7.A)**, even though the doubly phosphorylated dimer (*i.e.* one phosphorylation on each subunit) was not observed endogenously. Furthermore, nTDMS revealed that our use of the Sep OTS resulted in the misincorporation of glutamine and glycine **(Figure 6.7)** at the site of the intended phosphorylation and thus generated a complex population of TPI MPCs, which contained both homo- and hetero-dimers of the phosphorylated TPI proteoforms. Next, we performed kinetic assays of both the OFF and phospho-TPI samples, which revealed that TPI activity of the phospho-TPI is up to 4.44 times more active than the MetOFF TPI MPC, a remarkable observation in light of TPI's existing status as a "perfect enzyme" **(Figure 6.8)**.



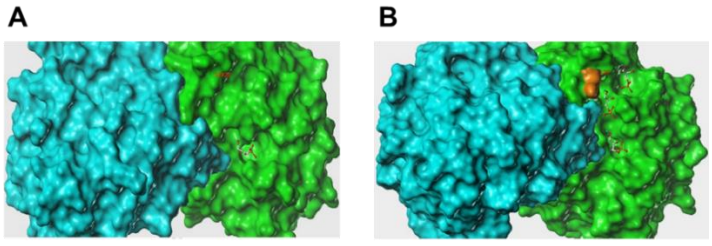
B

TPI Activity
(nmol NADH/min/mL)

TPI	Mean	St Error (95%)
OFF	186.4	±16.7
pSer20	827.8	±74.5

Figure 6.8 (previous page). Activity assay of TPI MPCs generated by CFPS. **(a)** Schematic of TPI activity assay, where DHAP (dihydroxyacetone phosphate) is converted into G3P (glyceraldehyde-3-phosphate). Next, GAPDH converts G3P into 1,3-BPG (1,3-bisphosphoglycerate), generating an equivalent of NADH, which reacts with a developer that elicits a color change (absorbance at 450 nm). **(b)** Table conveying the mean and standard error (95% confidence interval) for the activities of each sample tested at five different concentrations.

We next wanted to establish a structural basis for the increase in catalytic activity observed for phospho-TPI. To achieve this, we performed molecular dynamics simulations of the MetOFF TPI MPC, singly phosphorylated, and doubly phosphorylated TPI MPCs **(Figure 6.9)**. The simulation for the singly phosphorylated MPC suggested that the phosphorylation induced the formation of a channel that facilitates substrate entry into the active site, thereby reducing the strain energy associated with the substrate docking. Importantly, the simulation of the doubly-phosphorylated TPI did not show any significant structural changes relative to the MetOFF MPC, suggesting an explanation as to why the doubly-phosphorylated MPC is not observed endogenously.



E

	pSer20	Unmodified
ΔG	-13.52 kcal/mol	-4.28 kcal/mol
Strain Energy (Penalty)	1.56 kcal/mol	11.25 kcal/mol

Figure 6.9 (previous page). The molecular dynamics simulation of phospho-TPI reveals the formation of a substrate- guiding channel that facilitates substrate diffusion into the active site. (PDB: 4POC). Panels **(a)** and **(c)** correspond to the unmodified WT TPI and show different views of the region containing Ser20. Panels **(b)** and **(d)** correspond to different views of the simulated TPI structure upon phosphorylation at Ser20. In panel **(b)**, three stages of substrate diffusion through the phosphorylation-induced guiding channel can be visualized. Panel **(e)** tabulates the ΔG and strain energies associated with substrate binding to TPI when phosphorylated (pSer20) and when unmodified.

This investigation is still proceeding as of this writing. Recently, we have been trying to significantly scale up the synthesis of the various TPI proteoforms under investigation in the interest of obtaining enough protein to grow crystals for structural determinations.

6.5 Discussion

Phosphorylation is an incredibly important PTM implicated in the regulation of nearly all cellular processes[8, 10, 44-46]. Our ability to understand and characterize this key modification is currently limited by our ability to generate homogeneous protein samples featuring a specific set of phosphorylations. In this chapter, I describe my efforts to address this limitation by developing a highly-productive CFPS platform for synthesis of proteins featuring templated serine phosphorylations via the cotranslational incorporation of Sep via OTSs in amber suppression.

I first generated a novel CFPS chassis strain derived from a recoded organism lacking RF1 that had previously been optimized for CFPS (see Chapter 4), optimizing it for use with Sep by eliminating the activity of the Sep-specific phosphatase SerB from the system. My approach was to circumvent lysate productivity limitations imposed by outright inactivation of the enzyme by adding an affinity tag to it (for affinity chromatography-based removal) and installing mutations making it a more efficient substrate for the periplasmic protease OmpT. In this way, I reasoned that SerB activity could be preserved in growing cells but quickly eliminated by OmpT when

cellular compartments mix during lysate preparation. The final strain rEcoli.759.T7.Sep retains high productivity in CFPS, though its capacity for facilitating Sep incorporation has yet to be determined. The strain has not yet been tested using the original Sep OTS components, so this is an obvious next step in this effort. Other future directions include functional analysis of the mutant *serB* to establish whether or not the mutations made actually promote proteolysis of the enzyme by OmpT and, furthermore, whether or not Ompt-mediated cleavage impacts the activity of SerB.

I next created a novel plasmid, pSepOpt, encoding optimized Sep OTS components with ~20-fold higher catalytic efficiency than the original system[225]. This was ultimately achieved by employing a series of DNA manipulation and cloning tricks to overcome inefficient assembly of multiple parts to form a single, large plasmid. Eventually, the complete sequence-confirmed construct was obtained. The plasmid was utilized to generate several different CFPS lysates to assess its ability to drive cotranslation Sep incorporation, but unfortunately a complete assessment remains incomplete. Based on the preliminary results described in this chapter, I suspect that pSepOpt is not fully functional. My hypothesis is that sequence encoding the optimized tRNASepB4 lacks the proper flanking sequence elements to promote maturation into a functional tRNA following transcription. A rebuild of pSepOpt using Golden Gate Assembly to assemble the tandem tRNA cassette is a fruitful future direction in this project.

Parallel to the above efforts, I sought to encode each individual optimized Sep OTS components into an expression vector for overexpression *in vivo* followed by purification for direct supplementation to CFPS reactions. This approach, which is unique to CFPS, could be used in conjunction with or independently of in-strain expression of the OTS. I confirmed assembly of vectors for the optimized Sep tRNA (tRNASepB4) and elongation factor (EF-Sep), though a

functional expression vector encoding the optimized Sep synthetase (SepRS2) remains elusive. As a future direction, I propose moving the his-tag on the SepRS2 gene to the N-terminus to see if it improves expression and/or detection of the product.

Lastly, collaborators and I applied the techniques discussed here towards the synthesis of specific proteoforms of the glycolytic enzyme TPI. Analysis of endogenous samples from several cell lines revealed the presence of TPI MPCs featuring a phosphorylation at serine 20 on only one of the two monomers in the complex, and kinetics data suggested that the singly-phosphorylated MPC is twice as fast the unmodified complex. To validate this result, we used CFPS to synthesize both wild type and phosphorylated TPI proteoforms. nTDMS analysis of the CFPS-derived samples revealed that the Sep OTS regularly misincorporated glutamine and glycine instead of Sep at the desired position, highlighting the poor performance of the original Sep OTS components. Despite this limitation, we observed that singly-phosphorylated TPI MPCs derived from CFPS are over 4-fold faster than their wild type counterpart, in agreement with the results obtained from endogenous samples. Molecular dynamics modeling suggested a mechanistic basis for this, indicating that the presence of a single phosphorylation on the dimer opens a channel to the active site which could facilitate substrate entry. Going forward, we will need to devise a method by which to achieve preparative synthesis of the phosphorylated proteoform of TPI. Structural data is necessary to confirm the results of the molecular dynamics modeling, and crystal growth requires very large amounts of very pure protein.

Designer phosphoproteoform synthesis remains one of the defining challenges in studying this post-translational modification. Additionally, the ability to robustly and faithfully incorporate Sep at amber codons could enable templated synthesis of proteins featuring site-specific instances

of practically any side chain of interest via a recently reported chemical route[481]. This further motivates development of this technology. Though incomplete, I expect that the efforts described in this chapter could form the basis for such a synthesis platform.

6.6 Acknowledgments

Thanks to my collaborators Luis F. Schachner (L.F.S), Ashley N. Ives (A.N.I), Samuel R. Davidson (S.R.D.), and Rama K. Mishra (R.M.K.) for their contributions to the work described in this chapter. Professor Michael C. Jewett (M.C.J.) and I conceived of the Sep-optimized CFPS chassis strain and pSepOpt assembly projects. S.R.D. provided support and feedback in experimental design and was a second set of hands for many of the experiments described in this chapter. L.F.S. and Professor Neil L. Kelleher conceived of the TPI project. L.F.S. and A.N.I. performed native top-down mass spectrometry and TPI kinetics assays. R.M.K. performed the molecular dynamics modeling. I performed all other experiments and analysis. L.F.S. and I wrote this chapter.

CHAPTER 7

7 Summary and Future Directions

7.1 Summary

The establishment of mechanisms by which the addition of functional groups at specific positions on a protein regulates protein function has been a longstanding challenge in studying PTMs. This has been due in large part to a general inability to obtain preparative scale, homogeneous samples of protein featuring only a specific set of modifications for controlled structural and functional assays *in vitro*. To alleviate this limitation, my focus in this thesis was on developing synthesis platforms capable of robust synthesis of homogeneous protein populations featuring definable PTMs to enable fundamental studies into the effects of specific PTMs of interest on the structure and function of modified proteins. My approach was founded on incorporation of ncAAs via amber suppression, which could be leveraged to cotranslationally incorporate genetically-encoded modified residues into nascent proteins at specific positions. I selected the use of bacterial CFPS systems as a synthesis platform in order to circumvent the cytotoxicity effects associated with the OTSs used in amber suppression, as well as to benefit from the beneficial features of these systems.

As amber suppression is limited by competition with RF1, I hypothesized that the synthesis of proteins containing ncAAs could be improved by using lysates derived from organisms lacking RF1. To test this, I contributed to the development of a CFPS platform derived from a partially-

recoded strain of *E. coli* with an RF1deletion. By functionally inactivating several nucleases in the strain, we were able to dramatically increase the amount of protein synthesized by its lysates in CFPS, and the absence of RF1 in the system led to highly-efficient ncAA incorporation as hypothesized. Still, even in the absence of RF1 we observed the formation of errant truncated products, which we hypothesized were the result of competition with stalled ribosomal rescue factors (*e.g.* ssrA) upregulated in the partially-recoded organism to rescue translational termination at native amber codons. To address this, I aided in the development of a CFPS platform derived from a fully-recoded RF1-deficient *E. coli* strain lacking any endogenous use of the amber codon. The functional deactivation of candidate negative effectors of CFPS in the strain led to a significant 4.5-fold increase in the productivity of its lysates. Furthermore, the absence of RF1 (and, presumably, amber codon-associated ribosomal rescue factors) enabled highly-efficient incorporation of ncAAs using the system, in agreement with our hypotheses. These works are presented in greater detail in Chapters 2 and 3.

In my next aim, my goal was to simplify the use of our ncAA-optimized platform to increase its accessibility. To that end, I hypothesized that imbuing our best recoded RF1-deficient source strain with the ability to synthesize the viral T7 RNA polymerase would expand the capabilities of the platform to yield a one-pot system that was fully functional without supplementation with any purified biologics. As discussed in Chapter 4, after genomically incorporating a construct encoding the polymerase into the strain and subsequently installing mutations protecting the enzyme from proteolysis, the platform was capable of robust ncAA incorporation independent of purified polymerase supplementation, agreeing with my hypothesis. In particular, this effort revealed that exposure to the periplasmic protease OmpT during lysate

preparation could limit CFPS productivity by leading to the efficient cleavage of key cytoplasmic proteins involved in translation. This aim demonstrated a method for alleviating this by mutating the affected proteins to eliminate putative OmpT recognition and binding sites, which might be used in the future to improve the function of CFPS systems.

Next, I developed a novel CFPS system derived from the fast-growing non-model bacterium *Vibrio natriegens*. Because *V. natriegens* has the fastest known growth rate of any organism, I hypothesized that its lysates would be enriched with highly-active translational machinery, which in general could be beneficial for preparative-scale synthesis of protein. Elucidation of optimal cell growth, lysis, and reaction conditions culminated in a highly-productive CFPS platform comparable to the state-of-the-art, in agreement with my hypothesis. I also found that the system was uniquely capable of synthesizing short peptides, which have historically been difficult to produce recombinantly. Finally, the platform demonstrated an ability to facilitate incorporation of ncAAs into proteins, though this was severely limited likely by the presence of active RF1 in the lysate. The details of this work are presented in Chapter 5.

For my last aim, I sought a proof of principle that applied my methodology and synthesis platforms to the synthesis of proteins featuring site-specific PTMs for characterization. Specifically, I pursued the synthesis of proteins featuring definable serine phosphorylations. To that end, I set out to develop a CFPS platform optimized for direct incorporation of Sep. My approach was two-fold. First, I engineered my recoded, one-pot, RF1-deficient system to selectively inactivate the Sep-specific phosphatase SerB in lysates. Second, I assembled improved Sep OTS components into an expression vector for overexpression during chassis cell growth. Successful testing and validation of this system is incomplete as of this writing. Finally, I applied

existing Sep incorporation systems to the synthesis of specific phosphoproteoforms of the glycolytic enzyme triosephosphate isomerase (TPI) in CFPS. Analysis of the proteoforms via nTDMS and kinetics assays revealed that TPI dimers featuring a phosphorylation at serine 20 on only one monomer are more than four-fold faster than the wild type dimer. Molecular dynamics modeling proposed a structural mechanism by which the presence of the single phosphate group opens a channel to the enzyme active site, accelerating catalysis by facilitate substrate access. These works are described in Chapter 6. Together, these results highlight the transformative potential of templated proteoform synthesis in CFPS using amber suppression to directly incorporate genetically-encoded PTMs.

Taken together, my work will aid efforts to fully characterize and understanding PTMs by providing one approach whereby proteins featuring defined sets of modifications can be synthesized in pure, homogeneous samples for structural and functional evaluation.

7.2 Future Directions

Amber suppression in CFPS systems leads to many possibilities for fundamental biological studies into the roles played by specific PTMs in effecting changes in protein function. That said, there is still a good deal to learn about the CFPS production platform itself. While the "black box" of CFPS gets a little less opaque all the time, a systems understanding of the composition and workings of cell-free lysates has yet to be fully realized. This is perhaps best exemplified in Chapter 3 of this thesis, where the simultaneous inactivation of negative effectors that individually were very beneficial to CFPS frequently did *not* lead to further synergistic increases in system productivity. A more thorough understanding of the system operating in CFPS lysates could inform future efforts to rationally modify that system to increase production by identifying targets

for upregulation or inactivation. To that end, one fruitful future direction is to catalog and quantify all of the proteins that make it into our lysates, which to my knowledge has not been performed. It is admittedly a technically challenging goal, but recent advances in mass spectrometry-based proteomics could enable such an investigation.

Going forward, efforts to use OTSs in any environment to synthesize proteins featuring specific PTMs have a number of challenges to overcome. Orthogonal synthetases are typically 1000-fold less efficient than their native counterparts, and generally exhibit a high-degree of polysubstrate specificity with regards to other ncAAs. For these technologies to enable the synthesis of more complicated proteoforms (*i.e.* proteins featuring multiple different kinds of genetically-encoded PTMs), strategies to evolve more efficient OTSs *that are also orthogonal to other OTSs* will be key. A potential future direction here could be the continued development of *in vitro* evolution platforms that use microfluidics to rapidly screen phenotypes. Additionally, the cotranslational incorporation of multiple different kinds of PTMs as ncAAs into proteins will require access to additional coding channels. Realistically, the current state-of-the-art is incorporation of up to 2 different ncAA species, though incorporation of only a single kind of ncAA is significantly more efficient. A number of approaches, such as the development of a strain of *E. coli* that uses only 57 codons or the addition of synthetic nucleotides to the genetic code, seek to address this limitation. Success via either of these approaches would be transformative, not only for the synthesis of more complex proteoforms but also for the development of novel protein-based therapeutics and materials.

There is still much to be learned about the *V. natriegens* platform described in Chapter 5, and indeed I do not believe that we realized the maximal production potential of this organism.

This system broke from one of the tenets of bacterial CFPS in that its most productive lysates are derived from stationary phase cells rather than exponential phase cells, a feature that I am keenly interested to understand. A fruitful future direction for this system could be to quantify ribosome abundance in cell populations at different points in the growth cycle and observe how well they correspond with the productivity of lysates derived from cells at those points. This is important for establishing whether or not the trend in productivity can be simply attributed to ribosome abundance. My hypothesis is that ribosome abundance is actually highest in mid-exponential phase, but that most of the ribosomes are sequestered by native mRNAs in the lysate such that the pool of ribosomes available to participate in CFPS is lower than in stationary phase lysates. This sort of phenomenon is usually overcome in other systems by a runoff reaction, but as described in Chapter 5 that approach did not work here. The discovery that ribosome abundance is elevated in mid-exponential phase would support my hypothesis, and ought to spur investigations into why those lysates are less productive. The development of an alternative method for freeing ribosomes from native messages might be required. One future direction here could be to try dialyzing most of the Mg out of the lysate to force ribosomal subunits to dissociate to free them from bound mRNAs, after which Mg can be dialyzed back in to promote ribosome reassembly. Other future directions in development of this system include the exploration of alternative energy regeneration systems and screening a larger library of knockout strains. The latter could be informed by a course analysis of small molecules in *V. natriegens* CFPS reactions, with knockouts targeted towards metabolic pathways that siphon away key reaction substrates. Such an approach has been very fruitful in developing *E. coli* systems.

While I was able to detect synthesis of short AMPs using the *V. natriegens* platform, functional characterization of the products remains to be completed. When I first conceptualized the screen, I imagined synthesizing the peptides in CFPS and then adding crude CFPS reaction to small cultures to observe the killing effects of the AMPs. In practice, this did not really work. AMPs are generally cationic, and associate with their cellular targets via electrostatic interactions. These interactions are disrupted by salt, which is present at high levels in the *V. natriegens* CFPS mix. In other words, in order to actually use the AMPs synthesized using this platform, I believe it will first be necessary to purify the peptides out of the reaction mix or otherwise desalt them. This is an important future direction for this system, and could be done a number of ways. Peptides could be expressed fused to an affinity tag and a protease site such that the whole complex could be purified and then cleaved to yield the peptide. Dialysis using low MWCO membranes or C18 reversed phase columns could be used to desalt the entire crude CFPS mix.

Though *V. natriegens* CFPS was capable of facilitating ncAA incorporation, the efficiency of incorporation was quite low. I hypothesize that this is the result of truncation caused by the activities of *V. natriegens*' RF1 equivalent. As in *E. coli*, the removal of RF1 from *V. natriegens* will likely require at least partial recoding of essential/highly expressed genes that depend on the amber codon to terminate. Recoded *V. natriegens* would be a massive undertaking, but if the yields of the platform improve enough it may be worth it to enable preparative scale synthesis of proteins with ncAAs.

As for my efforts to optimize our recoded CFPS platforms for Sep incorporation summarized in Chapter 6, much remains incomplete. As a future direction, pSepOpt should be rebuilt to more faithfully mimic the plasmid architecture of B40OTS. The tandem tRNA cassette

will be tricky to assemble (which is why I did not build it that way the first time) but it is doable using type IIs restriction enzymes in a Golden Gate assembly scheme. The optimized Sep chassis strain rEcoli.759.T7.Sep should be characterized using an operational Sep OTS overexpression vector (*e.g.* B40OTS plasmid) to determine whether or not the changes made improve Sep incorporation. Some functional analysis of the mutant SerB in the strain should be performed to confirm or refute the hypothesis that the amino acid changes lead to proteolytic processing and subsequent loss of function of the phosphatase during lysate preparation. Though the exact outcome of my efforts here remain undetermined, I believe this chapter presents a generalizable approach for optimizing CFPS systems synthesis of proteins with specific PTMs via deactivation of genes whose products would be expected to interfere with or reverse a PTM of interest. For instance, a platform to investigate methylation patterns may improve upon inactivation of demethylases, a platform to investigate acetylation patterns may improve upon inactivation of deacetylases, *etc*.

The TPI investigation remains underway. Future directions will include attempts at scaling up the synthesis of the various TPI proteoforms to obtain enough protein to grow a crystal for determination of proteoform structures. Very precise kinetics assays using homogeneous proteoform samples will be necessary to convincingly demonstrate the effects of phosphorylation on the activity of this enzyme.

Elucidation of the mechanisms by which PTMs regulate protein function is the defining challenge in protein biology. Direct encoding of these modifications using methods such as amber suppression (in CFPS and elsewhere) will transform our ability to synthesize and characterize

proteins featuring PTMs of interest and eliminate a long-standing bottleneck limiting this field. I eagerly anticipate future developments in this space.

REFERENCES

1. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
2. Keshet, Y. and R. Seger, *The MAP kinase signaling cascades: a system of hundreds of components regulates a diverse array of physiological functions.* Methods Mol Biol, 2010. **661**: p. 3-38.
3. Knorre, D.G., N.V. Kudryashova, and T.S. Godovikova, *Chemical and functional aspects of posttranslational modification of proteins.* Acta Naturae, 2009. **1**(3): p. 29-51.
4. Doll, S. and A.L. Burlingame, *Mass spectrometry-based detection and assignment of protein posttranslational modifications.* ACS Chem Biol, 2015. **10**(1): p. 63-71.
5. Zhao, Y. and O.N. Jensen, *Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques.* Proteomics, 2009. **9**(20): p. 4632-41.
6. Park, H.S., et al., *Expanding the genetic code of Escherichia coli with phosphoserine.* Science, 2011. **333**(6046): p. 1151-1154.
7. Lee, S., et al., *A facile strategy for selective incorporation of phosphoserine into histones.* Angew. Chem. Int. Ed. Engl., 2013. **52**(22): p. 5771-5775.
8. Delom, F. and E. Chevet, *Phosphoprotein analysis: from proteins to proteomes.* Proteome Sci, 2006. **4**: p. 15.
9. Shechter, D., et al., *Extraction, purification and analysis of histones.* Nat Protoc, 2007. **2**(6): p. 1445-57.
10. Macek, B., M. Mann, and J.V. Olsen, *Global and site-specific quantitative phosphoproteomics: principles and applications.* Annu Rev Pharmacol Toxicol, 2009. **49**(1): p. 199-221.
11. McPherson, A. and J.A. Gavira, *Introduction to protein crystallization.* Acta Crystallogr F Struct Biol Commun, 2014. **70**(Pt 1): p. 2-20.
12. Chatterjee, A., et al., *A tryptophanyl-tRNA synthetase/tRNA pair for unnatural amino acid mutagenesis in E. coli.* Angew Chem Int Ed Engl, 2013. **52**(19): p. 5106-9.
13. Young, D.D., et al., *An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity.* Biochemistry, 2011. **50**(11): p. 1894-1900.
14. 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.
15. Young, T.S. and P.G. Schultz, *Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon.* Journal of Biological Chemistry, 2010. **285**(15): p. 11039-11044.

16. Wang, F., et al., *Unnatural amino acid mutagenesis of fluorescent proteins.* Angew. Chem. Int. Ed. Engl., 2012. **51**(40): p. 10132-10135.

17. Niu, W., P.G. Schultz, and J. Guo, *An expanded genetic code in mammalian cells with a functional quadruplet codon.* ACS Chem. Biol., 2013. **8**(7): p. 1640-1645.

18. Chatterjee, A., et al., *A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli.* Biochemistry, 2013. **52**(10): p. 1828-37.

19. 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. 255-263.

20. Zhang, M.S., et al., *Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing.* Nat Methods, 2017. **14**(7): p. 729-736.

21. Lajoie, M.J., et al., *Genomically recoded organisms expand biological functions.* Science, 2013. **342**(6156): p. 357-360.

22. 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.

23. 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.

24. 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.

25. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution.* Nature, 2009. **460**(7257): p. 894-8.

26. Isaacs, F.J., et al., *Precise manipulation of chromosomes in vivo enables genome-wide codon replacement.* Science, 2011. **333**(6040): p. 348-53.

27. 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.

28. Lee, H.H., et al., *Vibrio natriegens, a new genomic powerhouse*. bioRxiv, 2016. **058487**.

29. Weinstock, M.T., et al., *Vibrio natriegens as a fast-growing host for molecular biology.* Nat Methods, 2016. **13**(10): p. 849-51.

30. Pardee, K., et al., *Portable, On-Demand Biomolecular Manufacturing.* Cell, 2016. **167**(1): p. 248-259 e12.

31. Li, Y.F., *Recombinant production of antimicrobial peptides in Escherichia coli: A review (vol 80, pg 206, 2011).* Protein Expression and Purification, 2012. **82**(1): p. 252-252.

32. Rogers, L.D. and C.M. Overall, *Proteolytic Post-translational Modification of Proteins: Proteomic Tools and Methodology.* Molecular & Cellular Proteomics : MCP, 2013. **12**(12): p. 3532-3542.

33. James, M.N.G., *Handbook of proteolytic enzymes*. 1998, London: Academic Press.

34. Giglione, C., A. Boularot, and T. Meinnel, *Protein N-terminal methionine excision.* Cellular and Molecular Life Sciences CMLS, 2004. **61**(12): p. 1455-1474.

35. Huber, R. and W. Bode, *Structural basis of the activation and action of trypsin.* Accounts of Chemical Research, 1978. **11**(3): p. 114-122.

36. Kaufmann, J.E., J.C. Irminger, and P.A. Halban, *Sequence requirements for proinsulin processing at the B-chain/C-peptide junction.* Biochemical Journal, 1995. **310**(Pt 3): p. 869-874.

37. Wardlaw, S.L., *Hypothalamic proopiomelanocortin processing and the regulation of energy balance.* European journal of pharmacology, 2011. **660**(1): p. 213-219.

38. Pertea, M. and S.L. Salzberg, *Between a chicken and a grape: estimating the number of human genes.* Genome Biology, 2010. **11**(5): p. 206-206.

39. Blattner, F.R., et al., *The Complete Genome Sequence of Escherichia coli K-12.* Science, 1997. **277**(5331): p. 1453-1462.

40. Bonham-Carter, O., et al., *A study of bias and increasing organismal complexity from their post-translational modifications and reaction site interplays.* Briefings in Bioinformatics, 2017. **18**(1): p. 69-84.

41. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nature Reviews Molecular Cell Biology, 2012. **13**: p. 448.

42. Swatek, K.N. and D. Komander, *Ubiquitin modifications.* Cell Research, 2016. **26**: p. 399.

43. Zolnierowicz, S. and M. Bollen, *Protein phosphorylation and protein phosphatases De Panne, Belgium, September 19–24, 1999.* The EMBO Journal, 2000. **19**(4): p. 483-488.

44. Kennelly, P.J., *Protein Ser/Thr/Tyr Phosphorylation in the Archaea.* The Journal of Biological Chemistry, 2014. **289**(14): p. 9480-9487.

45. Macek, B., et al., *Phosphoproteome Analysis of E. coli Reveals Evolutionary Conservation of Bacterial Ser/Thr/Tyr Phosphorylation.* Molecular & Cellular Proteomics, 2008. **7**(2): p. 299.

46. Gu, Y., J. Rosenblatt, and D.O. Morgan, *Cell cycle regulation of CDK2 activity by phosphorylation of Thr160 and Tyr15.* The EMBO Journal, 1992. **11**(11): p. 3995-4005.

47. Graves, J.D. and E.G. Krebs, *Protein Phosphorylation and Signal Transduction.* Pharmacology & Therapeutics, 1999. **82**(2): p. 111-121.

48. Van Hoof, D., et al., *Phosphorylation Dynamics during Early Differentiation of Human Embryonic Stem Cells.* Cell Stem Cell, 2009. **5**(2): p. 214-226.

49. Shen, C.-H., et al., *Phosphorylation of BRAF by AMPK impairs BRAF-KSR1 association and cell proliferation.* Molecular cell, 2013. **52**(2): p. 161-172.

50. Deutscher, J., C. Francke, and P.W. Postma, *How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria.* Microbiology and Molecular Biology Reviews, 2006. **70**(4): p. 939-1031.

51. Radivojac, P., et al., *Gain and loss of phosphorylation sites in human cancer.* Bioinformatics, 2008. **24**(16): p. i241-i247.

52. Kumar, S., et al., *Extracellular phosphorylation of the amyloid β-peptide promotes formation of toxic aggregates during the pathogenesis of Alzheimer's disease.* The EMBO Journal, 2011. **30**(11): p. 2255-2265.

53. Rao, A.K. and J. Gabbeta, *Congenital disorders of platelet signal transduction.* Arterioscler Thromb Vasc Biol, 2000. **20**(2): p. 285-9.

54. Viatour, P., et al., *Phosphorylation of NF-κB and IκB proteins: implications in cancer and inflammation.* Trends in Biochemical Sciences, 2005. **30**(1): p. 43-52.

55. Plotnikov, A., et al., *The MAPK cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2011. **1813**(9): p. 1619-1633.

56. Shaul, Y.D. and R. Seger, *The MEK/ERK cascade: From signaling specificity to diverse functions.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2007. **1773**(8): p. 1213-1226.

57. Lane, D.P., *p53, guardian of the genome.* Nature, 1992. **358**: p. 15.

58. Ashcroft, M., M.H.G. Kubbutat, and K.H. Vousden, *Regulation of p53 Function and Stability by Phosphorylation.* Molecular and Cellular Biology, 1999. **19**(3): p. 1751.

59. Bates, S. and K.H. Vousden, *p53 in signaling checkpoint arrest or apoptosis.* Current Opinion in Genetics & Development, 1996. **6**(1): p. 12-18.

60. Loughery, J., et al., *Critical role for p53-serine 15 phosphorylation in stimulating transactivation at p53-responsive promoters.* Nucleic Acids Research, 2014. **42**(12): p. 7666-7680.

61. Maclaine, N.J. and T.R. Hupp, *The regulation of p53 by phosphorylation: a model for how distinct signals integrate into the p53 pathway.* Aging, 2009. **1**(5): p. 490-502.

62. Appella, E. and C.W. Anderson, *Post-translational modifications and activation of p53 by genotoxic stresses.* European Journal of Biochemistry, 2001. **268**(10): p. 2764-2772.

63. Kruse, J.-P. and W. Gu, *Modes of p53 Regulation.* Cell, 2009. **137**(4): p. 609-622.

64. Shieh, S.-Y., et al., *DNA Damage-Induced Phosphorylation of p53 Alleviates Inhibition by MDM2.* Cell, 1997. **91**(3): p. 325-334.

65. Blaydes, J.P. and T.R. Hupp, *DNA damage triggers DRB-resistant phosphorylation of human p53 at the CK2 site.* Oncogene, 1998. **17**: p. 1045.

66. Finlan, L.E., et al., *CK2-site Phosphorylation of p53 is Induced in ΔNp63 Expressing Basal Stem Cells in UVB Irradiated Human Skin.* Cell Cycle, 2006. **5**(21): p. 2489-2494.

67. Hupp, T.R., A. Sparks, and D.P. Lane, *Small peptides activate the latent sequence-specific DNA binding function of p53.* Cell, 1995. **83**(2): p. 237-245.

68. Wallace, M., et al., *Differential post-translational modification of the tumour suppressor proteins Rb and p53 modulate the rates of radiation-induced apoptosis in vivo.* Oncogene, 2001. **20**: p. 3597.

69. Thompson, T., et al., *Phosphorylation of p53 on Key Serines Is Dispensable for Transcriptional Activation and Apoptosis.* Journal of Biological Chemistry, 2004. **279**(51): p. 53015-53022.

70. Adeva-Andany, M.M., et al., *Glycogen metabolism in humans()().* BBA Clinical, 2016. **5**: p. 85-100.

71. Beurel, E., S.F. Grieco, and R.S. Jope, *Glycogen synthase kinase-3 (GSK3): Regulation, actions, and diseases.* Pharmacology & Therapeutics, 2015. **148**: p. 114-131.

72. Berg, J.M., et al., *Biochemistry*. Vol. 7. 2010, New York: W.H. Freeman.

73. Ji, J., et al., *Phosphorylated fraction of H2AX as a measurement for DNA damage in cancer cells and potential applications of a novel assay.* PLOS ONE, 2017. **12**(2): p. e0171582.

74. Podhorecka, M., A. Skladanowski, and P. Bozko, *H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy.* Journal of Nucleic Acids, 2010. **2010**: p. 9.

75. Sharma, A., K. Singh, and A. Almasan, *Histone H2AX Phosphorylation: A Marker for DNA Damage*, in *DNA Repair Protocols*, L. Bjergbæk, Editor. 2012, Humana Press: Totowa, NJ. p. 613-626.

76. Shiloh, Y. and Y. Ziv, *The ATM protein kinase: regulating the cellular response to genotoxic stress, and more.* Nature Reviews Molecular Cell Biology, 2013. **14**: p. 197.

77. Foster, E.R. and J.A. Downs, *Histone H2A phosphorylation in DNA double-strand break repair.* The FEBS Journal, 2005. **272**(13): p. 3231-3240.

78. Ambler, R.P. and M.W. Rees, *Epsilon-N-Methyl-lysine in bacterial flagellar protein.* Nature, 1959. **184**: p. 56-57.

79. Murn, J. and Y. Shi, *The winding path of protein methylation research: milestones and new frontiers.* Nature Reviews Molecular Cell Biology, 2017. **18**: p. 517.

80. Flanagan, J.F., et al., *Double chromodomains cooperate to recognize the methylated histone H3 tail.* Nature, 2005. **438**: p. 1181.

81. Li, H., et al., *Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF.* Nature, 2006. **442**(7098): p. 91-95.

82. Strahl, B.D., et al., *Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(26): p. 14967-14972.

83. Barski, A., et al., *High-Resolution Profiling of Histone Methylations in the Human Genome.* Cell, 2007. **129**(4): p. 823-837.

84. Fekner, T. and M.K. Chan, *The pyrrolysine translational machinery as a genetic-code expansion tool.* Curr Opin Chem Biol, 2011. **15**(3): p. 387-91.

85. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**: p. 593.

86. Zeng, W., A.R. Ball, and K. Yokomori, *HP1: Heterochromatin binding proteins working the genome.* Epigenetics : official journal of the DNA Methylation Society, 2010. **5**(4): p. 287-292.

87. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**: p. 41.

88. Chuikov, S., et al., *Regulation of p53 activity through lysine methylation.* Nature, 2004. **432**: p. 353.

89. Huang, J., et al., *G9a and Glp Methylate Lysine 373 in the Tumor Suppressor p53.* The Journal of Biological Chemistry, 2010. **285**(13): p. 9636-9641.

90. Huang, J., et al., *Repression of p53 activity by Smyd2-mediated methylation.* Nature, 2006. **444**: p. 629.

91. Jansson, M., et al., *Arginine methylation regulates the p53 response.* Nature Cell Biology, 2008. **10**: p. 1431.

92. Shi, X., et al., *Modulation of p53 function by SET8-mediated methylation at lysine 382.* Molecular cell, 2007. **27**(4): p. 636-646.

93. Metzger, E., et al., *LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription.* Nature, 2005. **437**: p. 436.

94. Brahms, H., et al., *Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein.* RNA, 2001. **7**(11): p. 1531-1542.

95. Friesen, W.J., et al., *SMN, the Product of the Spinal Muscular Atrophy Gene, Binds Preferentially to Dimethylarginine-Containing Protein Targets.* Molecular Cell, 2001. **7**(5): p. 1111-1117.

96. Yu, M.C., et al., *Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors.* Genes & Development, 2004. **18**(16): p. 2024-2035.

97. Li, H., et al., *Lipopolysaccharide-induced Methylation of HuR, an mRNA-stabilizing Protein, by CARM1.* Journal of Biological Chemistry, 2002. **277**(47): p. 44623-44630.

98. Jones, J.D. and C.D. O'Connor, *Protein acetylation in prokaryotes.* PROTEOMICS, 2011. **11**(15): p. 3012-3022.

99. Linster, E. and M. Wirtz, *N-terminal acetylation: an essential protein modification emerges as an important regulator of stress responses.* Journal of Experimental Botany, 2018: p. ery241-ery241.

100. Monda, J.K., et al., *Structural conservation of distinctive N-terminal acetylation-dependent interactions across a family of mammalian NEDD8 ligation enzymes.* Structure (London, England : 1993), 2013. **21**(1): p. 42-53.

101. Scott, D.C., et al., *N-Terminal Acetylation Acts as an Avidity Enhancer Within an Interconnected Multiprotein Complex.* Science (New York, N.Y.), 2011. **334**(6056): p. 674-678.

102. Yang, D., et al., *Nα-acetylated Sir3 stabilizes the conformation of a nucleosome-binding loop in the BAH domain.* Nature Structural &Amp; Molecular Biology, 2013. **20**: p. 1116.

103. Forte, G.M.A., M.R. Pool, and C.J. Stirling, *N-Terminal Acetylation Inhibits Protein Targeting to the Endoplasmic Reticulum.* PLoS Biology, 2011. **9**(5): p. e1001073.

104. Mercier, E., et al., *Signal recognition particle binds to translating ribosomes before emergence of a signal anchor sequence.* Nucleic Acids Research, 2017. **45**(20): p. 11858-11866.

105. Kim, H.-K., et al., *The N-Terminal Methionine of Cellular Proteins as a Degradation Signal.* Cell, 2014. **156**(0): p. 158-169.

106. Varshavsky, A., *The Ubiquitin System, an Immense Realm.* Annual Review of Biochemistry, 2012. **81**(1): p. 167-176.

107. Drazic, A., et al., *The world of protein acetylation.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2016. **1864**(10): p. 1372-1401.

108. Recht, J., et al., *Histone chaperone Asf1 is required for histone H3 lysine 56 acetylation, a modification associated with S phase in mitosis and meiosis.* Proceedings of the National Academy of Sciences, 2006. **103**(18): p. 6988-6993.

109. Shahbazian, M.D. and M. Grunstein, *Functions of Site-Specific Histone Acetylation and Deacetylation.* Annual Review of Biochemistry, 2007. **76**(1): p. 75-100.

110. Smith, S. and B. Stillman, *Stepwise assembly of chromatin during DNA replication in vitro.* EMBO Journal, 1991. **10**(4): p. 971-980.

111. Harbison, C.T., et al., *Transcriptional regulatory code of a eukaryotic genome.* Nature, 2004. **431**: p. 99.

112. Kurdistani, S.K., S. Tavazoie, and M. Grunstein, *Mapping Global Histone Acetylation Patterns to Gene Expression.* Cell, 2004. **117**(6): p. 721-733.

113. Tessarz, P. and T. Kouzarides, *Histone core modifications regulating nucleosome structure and dynamics.* Nature Reviews Molecular Cell Biology, 2014. **15**: p. 703.

114. Shogren-Knaak, M., et al., *Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions.* Science, 2006. **311**(5762): p. 844-847.

115. Ye, J., et al., *Histone H4 Lysine 91 Acetylation: A Core Domain ModificationAssociated with Chromatin Assembly.* Molecular Cell, 2005. **18**(1): p. 123-130.

116. Ito, A., et al., *p300/CBP-mediated p53 acetylation is commonly induced by p53-activating agents and inhibited by MDM2.* The EMBO Journal, 2001. **20**(6): p. 1331-1340.

117. Tang, Y., et al., *Acetylation Is Indispensable for p53 Activation.* Cell, 2008. **133**(4): p. 612-626.

118. Matsuzaki, H., et al., *Acetylation of Foxo1 alters its DNA-binding ability and sensitivity to phosphorylation.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(32): p. 11278-11283.

119. Bettelheim, F.R., *TYROSINE-O-SULFATE IN A PEPTIDE FROM FIBRINOGEN.* Journal of the American Chemical Society, 1954. **76**(10): p. 2838-2839.

120. Dong, J., C.Q. Li, and J.A. Lopez, *Tyrosine Sulfation of the Glycoprotein Ib-IX Complex: Identification of Sulfated Residues and Effect on Ligand Binding.* Biochemistry, 1994. **33**(46): p. 13946-13953.

121. Leyte, A., et al., *Sulfation of Tyr1680 of human blood coagulation factor VIII is essential for the interaction of factor VIII with von Willebrand factor.* Journal of Biological Chemistry, 1991. **266**(2): p. 740-746.

122. Kanan, Y., et al., *Complement Factor H, Vitronectin, and Opticin Are Tyrosine-Sulfated Proteins of the Retinal Pigment Epithelium.* PLOS ONE, 2014. **9**(8): p. e105409.

123. Sherry, D.M., et al., *Differential Developmental Deficits in Retinal Function in the Absence of either Protein Tyrosine Sulfotransferase-1 or -2.* PLOS ONE, 2012. **7**(6): p. e39702.

124. Sherry, D.M., et al., *Lack of protein-tyrosine sulfation disrupts photoreceptor outer segment morphogenesis, retinal function and retinal anatomy.* European Journal of Neuroscience, 2010. **32**(9): p. 1461-1472.

125. Farzan, M., et al., *Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry*. Cell, 1999. **96**(5): p. 667-676.

126. Kehoe, J.W. and C.R. Bertozzi, *Tyrosine sulfation: a modulator of extracellular protein–protein interactions.* Chemistry & Biology, 2000. **7**(3): p. R57-R61.

127. Clemetson, K.J., *Primary haemostasis: Sticky fingers cement the relationship.* Current Biology, 1999. **9**(3): p. R110-R112.

128. Savage, B., F. Almus-Jacobs, and Z.M. Ruggeri, *Specific synergy of multiple substrate-receptor interactions in platelet thrombus formation under flow*. Cell, 1998. **94**(5): p. 657-666.

129. Marchese, P., et al., *Identification of Three Tyrosine Residues of Glycoprotein Ib with Distinct Roles in von Willebrand Factor and -Thrombin Binding.* Journal of Biological Chemistry, 1995. **270**(16): p. 9571-9578.

130. Litman, G.W., et al., *Phylogenetic diversification of immunoglobulin genes and the antibody repertoire.* Molecular Biology and Evolution, 1993. **10**(1): p. 60-72.

131. Murphy, K. and C. Weaver, *Janeway's Immunobiology, 9th edition*. 2016: CRC Press.

132. Stanfield, C.L. and W.J. Germann, *Principles of human physiology*. 4th ed. 2011, San Francisco, CA: Pearson/Benjamin Cummings. xxviii, 767 p.

133. Kaboord, B. and M. Perr, *Isolation of Proteins and Protein Complexes by Immunoprecipitation*, in *2D PAGE: Sample Preparation and Fractionation*, A. Posch, Editor. 2008, Humana Press: Totowa, NJ. p. 349-364.

134. Wu, J.J., et al., *Identification of a High-Affinity Anti-Phosphoserine Antibody for the Development of a Homogeneous Fluorescence Polarization Assay of Protein Kinase C.* Journal of Biomolecular Screening, 2000. **5**(1): p. 23-30.

135. Liang, Z., et al., *Development of pan-specific antibody against trimethyllysine for protein research.* Proteome Science, 2008. **6**(1): p. 2.

136. Oza, J.P., et al., *Robust production of recombinant phosphoproteins using cell-free protein synthesis.* Nat Commun, 2015. **6**: p. 8168.

137. Sigismund, S., et al., *Threshold-controlled ubiquitination of the EGFR directs receptor fate.* The EMBO Journal, 2013. **32**(15): p. 2140-2157.

138. Furey, T.S., *ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions.* Nature Reviews Genetics, 2012. **13**: p. 840.

139. Guenther, M.G., et al., *A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells.* Cell, 2007. **130**(1): p. 77-88.

140. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nature genetics, 2007. **39**(3): p. 311.

141. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-560.

142. Wang, Z., et al., *Combinatorial patterns of histone acetylations and methylations in the human genome.* Nature genetics, 2008. **40**(7): p. 897.

143. Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes.* Nature Reviews Genetics, 2010. **12**: p. 7.

144. Mahmood, T. and P.-C. Yang, *Western Blot: Technique, Theory, and Trouble Shooting.* North American Journal of Medical Sciences, 2012. **4**(9): p. 429-434.

145. Li, C.L., et al., *SRF Phosphorylation by Glycogen Synthase Kinase-3 Promotes Axon Growth in Hippocampal Neurons.* The Journal of Neuroscience, 2014. **34**(11): p. 4027-4042.

146. Mertins, P., et al., *Integrated proteomic analysis of post-translational modifications by serial enrichment.* Nature Methods, 2013. **10**: p. 634.

147. Jaroentomeechai, T., et al., *Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery.* Nature Communications, 2018. **9**(1): p. 2686.

148. Hattori, T. and S. Koide, *Next-generation antibodies for post-translational modifications.* Current Opinion in Structural Biology, 2018. **51**: p. 141-148.

149. Fuchs, S.M., et al., *Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition.* Current biology : CB, 2011. **21**(1): p. 53-58.

150. Fuchs, S.M. and B.D. Strahl, *Antibody recognition of histone post-translational modifications: emerging issues and future prospects.* Epigenomics, 2011. **3**(3): p. 247-249.

151. Glish, G.L. and R.W. Vachet, *The basics of mass spectrometry in the twenty-first century.* Nature Reviews Drug Discovery, 2003. **2**: p. 140.

152. Gohlke, R.S., *Time-of-Flight Mass Spectrometry and Gas-Liquid Partition Chromatography.* Analytical Chemistry, 1959. **31**(4): p. 535-541.

153. Zubarev, R.A. and A. Makarov, *Orbitrap Mass Spectrometry.* Analytical Chemistry, 2013. **85**(11): p. 5288-5296.

154. Karas, M., et al., *Matrix-assisted ultraviolet laser desorption of non-volatile compounds.* International Journal of Mass Spectrometry and Ion Processes, 1987. **78**: p. 53-68.

155. Yamashita, M. and J.B. Fenn, *Electrospray ion source. Another variation on the free-jet theme.* The Journal of Physical Chemistry, 1984. **88**(20): p. 4451-4459.

156. Mann, M. and N.L. Kelleher, *Precision proteomics: The case for high resolution and high mass accuracy.* Proceedings of the National Academy of Sciences, 2008. **105**(47): p. 18132-18138.

157. Toby, T.K., L. Fornelli, and N.L. Kelleher, *Progress in Top-Down Proteomics and the Analysis of Proteoforms.* Annual review of analytical chemistry (Palo Alto, Calif.), 2016. **9**(1): p. 499-519.

158. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**: p. 198.

159. Zhang, Y., et al., *Protein Analysis by Shotgun/Bottom-up Proteomics.* Chemical reviews, 2013. **113**(4): p. 2343-2394.

160. Chen, Y., et al., *Integrated Approach for Manual Evaluation of Peptides Identified by Searching Protein Sequence Databases with Tandem Mass Spectra.* Journal of Proteome Research, 2005. **4**(3): p. 998-1005.

161. Olsen, J.V., S.-E. Ong, and M. Mann, *Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues.* Molecular & Cellular Proteomics, 2004. **3**(6): p. 608-614.

162. Gilmore, J.M., A.N. Kettenbach, and S.A. Gerber, *Increasing phosphoproteomic coverage through sequential digestion by complementary proteases.* Analytical and bioanalytical chemistry, 2012. **402**(2): p. 711-720.

163. Gillet, L.C., A. Leitner, and R. Aebersold, *Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing.* Annual Review of Analytical Chemistry, 2016. **9**(1): p. 449-472.

164. Liu, H., et al., *A method for systematic mapping of protein lysine methylation identifies new functions for HP1β in DNA damage repair.* Molecular cell, 2013. **50**(5): p. 723-735.

165. Svinkina, T., et al., *Deep, Quantitative Coverage of the Lysine Acetylome Using Novel Anti-acetyl-lysine Antibodies and an Optimized Proteomic Workflow.* Molecular & Cellular Proteomics : MCP, 2015. **14**(9): p. 2429-2440.

166. Zappacosta, F., et al., *An Optimized Platform for Hydrophilic Interaction Chromatography–Immobilized Metal Affinity Chromatography Enables Deep Coverage of the Rat Liver Phosphoproteome.* Journal of Proteome Research, 2015. **14**(2): p. 997-1009.

167. Catherman, A.D., O.S. Skinner, and N.L. Kelleher, *Top Down Proteomics: Facts and Perspectives.* Biochemical and biophysical research communications, 2014. **445**(4): p. 683-693.

168. Nesvizhskii, A.I. and R. Aebersold, *Interpretation of Shotgun Proteomic Data.* The Protein Inference Problem, 2005. **4**(10): p. 1419-1440.

169. Kelleher, N.L., et al., *Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry.* Journal of the American Chemical Society, 1999. **121**(4): p. 806-812.

170. Ge, Y., et al., *Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(31): p. 12658-12663.

171. Anderson, L.C., et al., *Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry.* Journal of proteome research, 2017. **16**(2): p. 1087-1096.

172. Skinner, O.S., et al., *Top-down characterization of endogenous protein complexes with native proteomics.* Nat Chem Biol, 2017.

173. Zhang, H. and Y. Ge, *Comprehensive Analysis of Protein Modifications by Top-down Mass Spectrometry.* Circulation. Cardiovascular Genetics, 2011. **4**(6): p. 711-711.

174. Steen, H. and M. Mann, *The abc's (and xyz's) of peptide sequencing.* Nature Reviews Molecular Cell Biology, 2004. **5**: p. 699.

175. Han, X., et al., *Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons.* Science, 2006. **314**(5796): p. 109-12.

176. Barber, K.W. and J. Rinehart, *The ABCs of PTMs.* Nature chemical biology, 2018. **14**(3): p. 188-192.

177. Abbott, A., *How to spot a protein in a crowd.* Nature, 1999. **402**: p. 716.

178. Larsen, M.R., et al., *Analysis of posttranslational modifications of proteins by tandem mass spectrometry.* BioTechniques, 2006. **40**(6): p. 790-798.

179. Mann, M., et al., *Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome.* Trends in Biotechnology, 2002. **20**(6): p. 261-268.

180. Seo, J. and K.J. Lee, *Post-translational modifications and their biological functions: proteomic analysis and systematic approaches.* J Biochem Mol Biol, 2004. **37**(1): p. 35-44.

181. Kim, H.-J., E.J. Song, and K.-J. Lee, *Proteomic Analysis of Protein Phosphorylations in Heat Shock Response and Thermotolerance.* Journal of Biological Chemistry, 2002. **277**(26): p. 23193-23207.

182. Melo-Braga, M.N., et al., *Modulation of Protein Phosphorylation, N-Glycosylation and Lys-Acetylation in Grape (Vitis vinifera) Mesocarp and Exocarp Owing to Lobesia botrana Infection.* Molecular & Cellular Proteomics, 2012. **11**(10): p. 945-956.

183. Graves, H.K., et al., *Mutations that prevent or mimic persistent post-translational modifications of the histone H3 globular domain cause lethality and growth defects in Drosophila.* 2016. **9**(1): p. 9.

184. Yang, Y., et al., *Role of acetylation and extra-cellular location of heat shock protein 90α in tumor cell invasion.* Cancer research, 2008. **68**(12): p. 4833-4842.

185. White, R.H., M. Keberlein, and V. Jackson, *A Mutational Mimic Analysis of Histone H3 Post-Translational Modifications: Specific Sites Influence the Conformational State of H3/H4, Causing either Positive or Negative Supercoiling of DNA.* Biochemistry, 2012. **51**(41): p. 8173-8188.

186. Liu, N., et al., *Proteomic Profiling and Functional Characterization of Multiple Post-Translational Modifications of Tubulin.* Journal of Proteome Research, 2015. **14**(8): p. 3292-3304.

187. Kashiwagi, K., et al., *A Ser75-to-Asp phospho-mimicking mutation in Src accelerates ageing-related loss of retinal ganglion cells in mice.* Scientific Reports, 2017. **7**(1): p. 16779.

188. Paleologou, K.E., et al., *Phosphorylation at Ser-129 but Not the Phosphomimics S129E/D Inhibits the Fibrillation of α-Synuclein.* 2008. **283**(24): p. 16895-16905.

189. Nagata, K.-i., I. Izawa, and M. Inagaki, *A decade of site- and phosphorylation state-specific antibodies: recent advances in studies of spatiotemporal protein phosphorylation.* 2001. **6**(8): p. 653-664.

190. Russo, A.A., P.D. Jeffrey, and N.P. Pavletich, *Structural basis of cyclin-dependent kinase activation by phosphorylation.* Nature Structural Biology, 1996. **3**: p. 696.

191. Gu, W. and R.G. Roeder, *Activation of p53 Sequence-Specific DNA Binding by Acetylation of the p53 C-Terminal Domain.* Cell, 1997. **90**(4): p. 595-606.

192. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Research, 2012. **40**(Database issue): p. D261-D270.

193. Xue, L., et al., *Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(15): p. 5615-5620.

194. Chatterjee, C. and T.W. Muir, *Chemical Approaches for Studying Histone Modifications.* 2010. **285**(15): p. 11045-11050.

195. Chen, Z. and P.A. Cole, *Synthetic Approaches to Protein Phosphorylation.* Current opinion in chemical biology, 2015. **28**: p. 115-122.

196. Haj-Yahya, M. and H.A. Lashuel, *Protein Semisynthesis Provides Access to Tau Disease-Associated Post-translational Modifications (PTMs) and Paves the Way to Deciphering*

the Tau PTM Code in Health and Diseased States. Journal of the American Chemical Society, 2018. **140**(21): p. 6611-6621.

197. Wang, K., et al., *Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion.* Nat Biotechnol, 2007. **25**(7): p. 770-7.

198. Terasaka, N., et al., *An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center.* Nat Chem Biol, 2014. **10**(7): p. 555-7.

199. van Hest, J.C.M., K.L. Kiick, and D.A. Tirrell, *Efficient incorporation of unsaturated methionine analogues into proteins in vivo.* Journal of the American Chemical Society, 2000. **122**(7): p. 1282-1288.

200. Ngo, J.T. and D.A. Tirrell, *Noncanonical amino acids in the interrogation of cellular protein synthesis.* Acc Chem Res, 2011. **44**(9): p. 677-85.

201. Chin, J.W., *Expanding and reprogramming the genetic code of cells and animals.* Annu Rev Biochem, 2014. **83**: p. 379-408.

202. Dumas, A., et al., *Designing logical codon reassignment - Expanding the chemistry in biology.* Chemical Science, 2015. **6**(1): p. 50-69.

203. Chin, J.W., et al., *Addition of p-azido-L-phenylaianine to the genetic code of Escherichia coli.* Journal of the American Chemical Society, 2002. **124**(31): p. 9026-9027.

204. Yang, H., et al., *A General Method for Artificial Metalloenzyme Formation through Strain-Promoted Azide-Alkyne Cycloaddition.* Chembiochem, 2014. **15**(2): p. 223-227.

205. 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.

206. Tada, S., et al., *Genetic PEGylation.* PLoS One, 2012. **7**(11): p. e49235.

207. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding Nε-acetyllysine in recombinant proteins.* Nat Chem Biol, 2008. **4**(4): p. 232-234.

208. Ai, H.W., J.W. Lee, and P.G. Schultz, *A method to site-specifically introduce methyllysine into proteins in E. coli.* Chem Commun (Camb), 2010. **46**(30): p. 5506-8.

209. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

210. Wang, L., J. Xie, and P.G. Schultz, *Expanding the genetic code.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 225-49.

211. Santoro, S.W., et al., *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity.* Nat Biotechnol, 2002. **20**(10): p. 1044-8.

212. Nozawa, K., et al., *Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality.* Nature, 2009. **457**(7233): p. 1163-7.

213. 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.

214. Li, W.T., et al., *Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs.* J Mol Biol, 2009. **385**(4): p. 1156-64.

215. Ellefson, J.W., et al., *Directed evolution of genetic parts and circuits by compartmentalized partnered replication.* Nat Biotechnol, 2014. **32**(1): p. 97-101.

216. Chin, J.W., et al., *An expanded eukaryotic genetic code.* Science, 2003. **301**(5635): p. 964-7.

217. Ai, H.W., et al., *Genetically encoded alkenes in yeast.* Angew Chem Int Ed Engl, 2010. **49**(5): p. 935-7.

218. Zhang, Y., et al., *Crystal structures of apo wild-type M. jannaschii tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine.* Protein Sci, 2005. **14**(5): p. 1340-9.

219. Haruna, K., et al., *Engineering the elongation factor Tu for efficient selenoprotein synthesis.* Nucleic Acids Res, 2014. **42**(15): p. 9976-83.

220. O'Donoghue, P., et al., *Upgrading protein synthesis for synthetic biology.* Nat Chem Biol, 2013. **9**(10): p. 594-8.

221. Rackham, O. and J.W. Chin, *A network of orthogonal ribosome x mRNA pairs.* Nat Chem Biol, 2005. **1**(3): p. 159-66.

222. Jewett, M.C., et al., *In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation.* Mol Syst Biol, 2013. **9**: p. 678.

223. 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.

224. Liu, Y., et al., *Characterizing and alleviating substrate limitations for improved in vitro ribosome construction.* ACS Synth Biol, 2015. **4**(4): p. 454-462.

225. Rogerson, D.T., et al., *Efficient genetic encoding of phosphoserine and its non-hydrolyzable analog.* Nature chemical biology, 2015. **11**(7): p. 496-503.

226. Groff, D., et al., *A Genetically Encoded ε-N-methyl Lysine in Mammalian Cells.* Chembiochem : a European journal of chemical biology, 2010. **11**(8): p. 1066-1068.

227. Fan, C., K. Ip, and D. Söll, *Expanding the Genetic Code of Escherichia coli with Phosphotyrosine.* FEBS letters, 2016. **590**(17): p. 3040-3047.

228. Ostrov, N., et al., *Design, synthesis, and testing toward a 57-codon genome.* Science, 2016. **353**(6301): p. 819-822.

229. Johnson, D.B., et al., *RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites.* Nat. Chem. Biol., 2011. **7**(11): p. 779-786.

230. Mukai, T., et al., *Codon reassignment in the Escherichia coli genetic code.* Nucleic Acids Res., 2010. **38**(22): p. 8188-8195.

231. 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. 3716-3722.

232. 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. 844-853.

233. Rovner, A.J., et al., *Recoded organisms engineered to depend on synthetic amino acids.* Nature, 2015. **518**(7537): p. 89-93.

234. Mandell, D.J., et al., *Biocontainment of genetically modified organisms by synthetic protein design.* Nature, 2015. **518**(7537): p. 55-60.

235. Wan, W., et al., *A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli.* Angew. Chem. Int. Ed. Engl., 2010. **49**(18): p. 3211-3214.

236. Odoi, K.A., et al., *Nonsense and sense suppression abilities of original and derivative Methanosarcina mazei pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the Escherichia coli BL21(DE3) cell strain.* PLoS One, 2013. **8**(3): p. e57035.

237. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon.* Proc Natl Acad Sci U S A, 2004. **101**(20): p. 7566-71.

238. Chatterjee, A., et al., *A bacterial strain with a unique quadruplet codon specifying non-native amino acids.* Chembiochem, 2014. **15**(12): p. 1782-6.

239. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome.* Nature, 2010. **464**(7287): p. 441-4.

240. Wang, K., et al., *Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET.* Nat Chem, 2014. **6**(5): p. 393-403.

241. Sachdeva, A., et al., *Concerted, rapid, quantitative, and site-specific dual labeling of proteins.* J Am Chem Soc, 2014. **136**(22): p. 7785-8.

242. Gautier, A., A. Deiters, and J.W. Chin, *Light-activated kinases enable temporal dissection of signaling networks in living cells.* J Am Chem Soc, 2011. **133**(7): p. 2124-7.

243. Lemke, E.A., et al., *Control of protein phosphorylation with a genetically encoded photocaged amino acid.* Nat Chem Biol, 2007. **3**(12): p. 769-72.

244. Wilkins, B.J., et al., *A cascade of histone modifications induces chromatin condensation in mitosis.* Science, 2014. **343**(6166): p. 77-80.

245. Uttamapinant, C., et al., *Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins.* J Am Chem Soc, 2015. **137**(14): p. 4602-5.

246. Greiss, S. and J.W. Chin, *Expanding the genetic code of an animal.* J Am Chem Soc, 2011. **133**(36): p. 14196-9.

247. Bianco, A., et al., *Expanding the genetic code of Drosophila melanogaster.* Nat. Chem. Biol., 2012. **8**: p. 748-750.

248. Annaluru, N., et al., *Total synthesis of a functional designer eukaryotic chromosome.* Science, 2014. **344**(6179): p. 55-8.

249. Longman, D., et al., *Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway.* Genes Dev, 2007. **21**(9): p. 1075-85.

250. Hancock, S.M., et al., *Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair.* J Am Chem Soc, 2010. **132**(42): p. 14819-24.

251. Ma, H., J. Zhang, and H. Wu, *Designing Ago2-specific siRNA/shRNA to Avoid Competition with Endogenous miRNAs.* Mol Ther Nucleic Acids, 2014. **3**: p. e176.

252. Meiler J., B.D., *ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility.* Proteins, 2006.

253. Malyshev D., D.K., Lavergne T., Chen T., Dai N., Foster J., Corrêa I., Romesberg F., *A semi-synthetic organism with an expanded genetic alphabet.* Nature, 2014.

254. Neumann H. , S.A.L., Chin J.W., *De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs.* J. Am. Chem. Soc., 2010.

255. 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.

256. 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.

257. Takahashi, M.K., et al., *Characterizing and prototyping genetic networks with cell-free transcription-translation reactions.* Methods, 2015. **86**: p. 60-72.

258. Karim, A.S. and M.C. Jewett, *A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery.* Metab Eng, 2016. **36**: p. 116-26.

259. Dudley, Q.M., K.C. Anderson, and M.C. Jewett, *Cell-Free Mixing of Escherichia coli Crude Extracts to Prototype and Rationally Engineer High-Titer Mevalonate Synthesis.* ACS Synth Biol, 2016. **5**(12): p. 1578-1588.

260. Watanabe, M., et al., *Cell-free protein synthesis for structure determination by X-ray crystallography.* Methods Mol Biol, 2010. **607**: p. 149-60.

261. Martemyanov, K.A., et al., *Cell-free production of biologically active polypeptides: application to the synthesis of antibacterial peptide cecropin.* Protein Expr Purif, 2001. **21**(3): p. 456-61.

262. Renesto, P. and D. Raoult, *From genes to proteins: in vitro expression of rickettsial proteins.* Ann N Y Acad Sci, 2003. **990**: p. 642-52.

263. Xu, Z., et al., *High-level expression of soluble human beta-defensin-2 fused with green fluorescent protein in Escherichia coli cell-free system.* Appl Biochem Biotechnol, 2005. **127**(1): p. 53-62.

264. 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.

265. Sullivan, C.J., et al., *A cell-free expression and purification process for rapid production of protein biologics.* Biotechnol J, 2016. **11**(2): p. 238-48.

266. Li, J., et al., *Cell-free protein synthesis enables high yielding synthesis of an active multicopper oxidase.* Biotechnol J, 2016. **11**(2): p. 212-8.

267. Heinzelman, P., J.A. Schoborg, and M.C. Jewett, *pH responsive granulocyte colony-stimulating factor variants with implications for treating Alzheimer's disease and other central nervous system disorders.* Protein Eng Des Sel, 2015. **28**(10): p. 481-9.

268. Kightlinger, W., et al., *Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases.* Nat Chem Biol, 2018.

269. Schoborg, J.A., et al., *A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases.* Biotechnol Bioeng, 2018. **115**(3): p. 739-750.

270. Gurramkonda, C., et al., *Improving the recombinant human erythropoietin glycosylation using microsome supplementation in CHO cell-free system.* Biotechnol Bioeng, 2018. **115**(5): p. 1253-1264.

271. Slomovic, S., K. Pardee, and J.J. Collins, *Synthetic biology devices for in vitro and in vivo diagnostics.* Proc Natl Acad Sci U S A, 2015. **112**(47): p. 14429-35.

272. Gootenberg, J.S., et al., *Nucleic acid detection with CRISPR-Cas13a/C2c2.* Science, 2017. **356**(6336): p. 438-442.

273. Karig, D.K., et al., *Preservation of protein expression systems at elevated temperatures for portable therapeutic production.* J R Soc Interface, 2017. **14**(129).

274. Smith, M.T., et al., *Lyophilized Escherichia coli-based cell-free systems for robust, high-density, long-term storage.* Biotechniques, 2014. **56**(4): p. 186-93.

275. Hunt, J.P., et al., *The growing impact of lyophilized cell-free protein expression systems.* Bioengineered, 2017. **8**(4): p. 325-330.

276. Adiga, R., et al., *Point-of-care production of therapeutic proteins of good-manufacturing-practice quality.* Nature Biomedical Engineering, 2018.

277. Stark, J.C., et al., *BioBits Bright: a fluorescent synthetic biology education kit.* Science Advances, 2018.

278. Huang, A., et al., *BioBits Explorer: a modular synthetic biology education kit.* Science Advances, 2018.

279. Schoborg, J.A. and M.C. Jewett, *Cell-Free Protein Synthesis: An Emerging Technology for Understanding, Harnessing, and Expanding the Capabilities of Biological Systems*, in *Synthetic Biology*, C. Smolke, et al., Editors. 2018.

280. Carlson, E.D., et al., *Cell-free protein synthesis: applications come of age.* Biotechnol Adv, 2012. **30**(5): p. 1185-94.

281. Hodgman, C.E. and M.C. Jewett, *Cell-free synthetic biology: thinking outside the cell.* Metab. Eng., 2012. **14**(3): p. 261-269.

282. Perez, J.G., J.C. Stark, and M.C. Jewett, *Cell-Free Synthetic Biology: Engineering Beyond the Cell.* Cold Spring Harb Perspect Biol, 2016. **8**(12).

283. Swartz, J.J.J.o.I.M. and Biotechnology, *Developing cell-free biology for industrial applications.* 2006. **33**(7): p. 476-485.

284. 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.

285. Calhoun, K.A. and J.R. Swartz, *Energizing cell-free protein synthesis with glucose metabolism.* Biotechnol Bioeng, 2005. **90**(5): p. 606-13.

286. Kim, D.M. and J.R. Swartz, *Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis.* Biotechnol Bioeng, 2001. **74**(4): p. 309-16.

287. Jewett, M.C. and J.R. Swartz, *Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis.* Biotechnol Bioeng, 2004. **86**(1): p. 19-26.

288. Michel-Reydellet, N., K. Calhoun, and J. Swartz, *Amino acid stabilization for cell-free protein synthesis by modification of the Escherichia coli genome.* Metab. Eng., 2004. **6**(3): p. 197-203.

289. Michel-Reydellet, N., K. Woodrow, and J. Swartz, *Increasing PCR fragment stability and protein yields in a cell-free system with genetically modified Escherichia coli extracts.* J. Mol. Microbiol. Biotechnol., 2005. **9**(1): p. 26-34.

290. 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.* 2012. **53**(3): p. 163-174.

291. 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.

292. Siller, E., et al., *Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency.* Journal of Molecular Biology, 2010. **396**(5): p. 1310-1318.

293. 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.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(2): p. 559-564.

294. Chang, H.-C., et al., *De novo Folding of GFP Fusion Proteins: High Efficiency in Eukaryotes but Not in Bacteria.* Journal of Molecular Biology, 2005. **353**(2): p. 397-409.

295. Goshima, N., et al., *Human protein factory for converting the transcriptome into an in vitro–expressed proteome.* Nature Methods, 2008. **5**: p. 1011.

296. Endo, Y. and T. Sawasaki, *Cell-free expression systems for eukaryotic protein production.* Current Opinion in Biotechnology, 2006. **17**(4): p. 373-380.

297. Iizuka, N., et al., *Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae.* Molecular and Cellular Biology, 1994. **14**(11): p. 7322-7330.

298. Wang, X., et al., *An optimized yeast cell-free system: Sufficient for translation of human papillomavirus 58 L1 mRNA and assembly of virus-like particles.* Journal of Bioscience and Bioengineering, 2008. **106**(1): p. 8-15.

299. 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.

300. 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.

301. Schoborg, J.A., et al., *Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis.* 2014. **9**(5): p. 630-640.

302. Ezure, T., et al., *A Cell-Free Protein Synthesis System from Insect Cells*, in *Cell-Free Protein Production: Methods and Protocols*, Y. Endo, K. Takai, and T. Ueda, Editors. 2010, Humana Press: Totowa, NJ. p. 31-42.

303. Suzuki, T., et al., *N-Terminal protein modifications in an insect cell-free protein synthesis system and their identification by mass spectrometry.* 2006. **6**(16): p. 4486-4495.

304. Brödel, A.K., A. Sonnabend, and S. Kubick, *Cell-free protein expression based on extracts from CHO cells.* 2014. **111**(1): p. 25-36.

305. Stech, M., et al., *Cell-Free Systems: Functional Modules for Synthetic and Chemical Biology*, in *Fundamentals and Application of New Bioproduction Systems*, A.-P. Zeng, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 67-102.

306. Thoring, L., et al., *High-yield production of "difficult-to-express" proteins in a continuous exchange cell-free system based on CHO cell lysates.* Sci Rep, 2017. **7**(1): p. 11710.

307. Martin, R.W., et al., *Development of a CHO-Based Cell-Free Platform for Synthesis of Active Monoclonal Antibodies.* ACS Synth Biol, 2017. **6**(7): p. 1370-1379.

308. Ahn, J.-H., J.-W. Keum, and D.-M. Kim, *Expression Screening of Fusion Partners from an E. coli Genome for Soluble Expression of Recombinant Proteins in a Cell-Free Protein Synthesis System.* PLOS ONE, 2011. **6**(11): p. e26875.

309. Yang, W.C., et al., *Cell-free production of transducible transcription factors for nuclear reprogramming.* Biotechnol Bioeng, 2009. **104**(6): p. 1047-58.

310. Bundy, B.C. and J.R. Swartz, *Efficient disulfide bond formation in virus-like particles.* Journal of Biotechnology, 2011. **154**(4): p. 230-239.

311. Moore, S.J., et al., *Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria.* Proc Natl Acad Sci U S A, 2018. **115**(19): p. E4340-E4349.

312. Marshall, R., et al., *Rapid and Scalable Characterization of CRISPR Technologies Using an E. coli Cell-Free Transcription-Translation System.* Molecular Cell, 2018. **69**(1): p. 146-157.e3.

313. Smith, M.T., et al., *Creating a completely "cell-free" system for protein synthesis.* Biotechnol Prog, 2015. **31**(6): p. 1716-9.

314. Pardee, K., et al., *Paper-based Synthetic Gene Networks.* Cell, 2014. **159**(4): p. 940-954.

315. Shin, J. and V. Noireaux, *Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70.* J Biol Eng, 2010. **4**: p. 8.

316. 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.

317. 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., 2013.

318. Karig, D.K., et al., *Probing Cell-Free Gene Expression Noise in Femtoliter Volumes.* ACS Synthetic Biology, 2013. **2**(9): p. 497-505.

319. Siuti, P., S.T. Retterer, and M.J. Doktycz, *Continuous protein production in nanoporous, picolitre volume containers.* Lab on a Chip, 2011. **11**(20): p. 3523-3529.

320. Ohuchi, M., H. Murakami, and H. Suga, *The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus.* Curr Opin Chem Biol, 2007. **11**(5): p. 537-42.

321. 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. 400-416.

322. 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. 5949-5963.

323. Smith, M.T., et al., *Alternative fermentation conditions for improved Escherichia coli-based cell-free protein synthesis for proteins requiring supplemental components for proper synthesis.* Process Biochemistry, 2014. **49**(2): p. 217-222.

324. Shimizu, Y., et al., *Cell-free translation reconstituted with purified components.* Nat Biotechnol, 2001. **19**(8): p. 751-5.

325. Ahn, J.-H., et al., *Preparation Method for Escherichia coli S30 Extracts Completely Dependent upon tRNA Addition to Catalyze Cell-free Protein Synthesis.* Biotechnology and Bioprocess Engineering, 2006. **11**(5): p. 420-424.

326. 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.

327. Cui, Z., et al., *Semisynthetic tRNA Complement Mediates in Vitro Protein Synthesis.* J Am Chem Soc, 2015. **137**(13): p. 4404-13.

328. 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.

329. Gerrits, M., et al., *Cell-free synthesis of defined protein conjugates by site-directed cotranslational labeling*, in *Madame Curie Bioscience Database [Internet]*. 2013, Landes Bioscience.

330. Yin, G., et al., *RF1 attenuation enables efficient non-natural amino acid incorporation for production of homogeneous antibody drug conjugates.* Scientific Reports, 2017. **7**: p. 3026.

331. Swartz, J.R., *Transforming biochemical engineering with cell-free biology.* AIChE J., 2012. **58**(1): p. 5-13.

332. Hayashi, Y., J. Morimoto, and H. Suga, *In vitro selection of anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors.* ACS Chem. Biol., 2012. **7**(3): p. 607-613.

333. 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.

334. 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. 16101-6.

335. Albayrak, C. and J.R. Swartz, *Direct polymerization of proteins.* ACS Synth. Biol., 2014: p. on line.

336. Ugwumba, I.N., et al., *Improving a natural enzyme activity through incorporation of unnatural amino acids.* J. Am. Chem. Soc., 2010. **133**(2): p. 326-333.

337. Li, X. and C.C. Liu, *Biological Applications of Expanded Genetic Codes.* ChemBioChem, 2014. **15**(16): p. 2335-2341.

338. Liu, C.C. and P.G. Schultz, *Adding New Chemistries to the Genetic Code.* Annual Review of Biochemistry, 2010. **79**(1): p. 413-444.

339. 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.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(19): p. 10092-10097.

340. Hoesl, M.G. and N. Budisa, *Recent advances in genetic code engineering in Escherichia coli.* Curr. Opin. Biotechnol., 2012. **23**(5): p. 751-757.

341. Hughes, R.A. and A.D. Ellington, *Rational design of an orthogonal tryptophanyl nonsense suppressor tRNA.* Nucleic Acids Res., 2010. **38**(19): p. 6813-6830.

342. Wang, Y.-S., et al., *A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum.* J. Am. Chem. Soc., 2012. **134**(6): p. 2950-2953.

343. Ko, J.-h., et al., *Suppression of amber codons in Caulobacter crescentus by the orthogonal Escherichia coli histidyl-tRNA synthetase/tRNA^{His} Pair.* PLoS ONE, 2013. **8**(12): p. e83630.

344. Bröcker, M.J., et al., *Recoding the genetic code with selenocysteine.* Angew. Chem. Int. Ed. Engl., 2014. **53**(1): p. 319-323.

345. Ma, Y., et al., *Coupling bioorthogonal chemistries with artificial metabolism: intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins.* Molecules 2014. **19**(1): p. 1004-1022.

346. Johnson, D.B.F., et al., *Release factor one is nonessential in Escherichia coli.* ACS Chem. Biol., 2012. **7**(8): p. 1337-1344.

347. 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.

348. Sando, S., et al., *In vitro selection of RNA aptamer against Escherichia coli release factor 1.* Bioorg. Med. Chem. Lett., 2007. **17**(5): p. 1216-1220.

349. Agafonov, D.E., et al., *Efficient suppression of the amber codon in E. coli in vitro translation system.* FEBS Lett., 2005. **579**(10): p. 2156-2160.

350. Gerrits, M., et al., *Cell-free synthesis of defined protein conjugates by site-directed cotranslational labeling*, in *Cell-Free Protein Expression*, W. Kudlicki, F. Katzen, and R. Bennett, Editors. 2007, Landes Bioscience: Austin, TX. p. 166–180.

351. Wang, H.H. and G.M. Church, *Multiplexed genome engineering and genotyping methods: applications for synthetic biology and metabolic engineering*, in *Methods Enzymol.*, V. Christopher, Editor. 2011, Academic Press. p. 409-426.

352. Jiang, X.P., et al., *Reduction of protein degradation by use of protease-deficient mutants in cell-free protein synthesis system of Escherichia coli.* J. Biosci. Bioeng., 2002. **93**(2): p. 151-156.

353. Sambrook, J. and D.W. Russell, *Molecular Cloning, A Laboratory Manual*. 3rd ed. 2001, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

354. Swartz, J.R., M.C. Jewett, and K.A. Woodrow, *Cell-free protein synthesis with prokaryotic combined transcription-translation.* Methods Mol Biol, 2004. **267**: p. 169-182.

355. Fritz, B.R., O.K. Jamil, and M.C. Jewett, *Implications of macromolecular crowding and reducing conditions for in vitro ribosome construction.* Nucleic Acids Res, 2015. **43**(9): p. 4774-84.

356. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.

357. Airen, I.O., *Genome-wide functional genomic analysis for physiological investigation and improvement of cell-free protein synthesis*, in *Dept. of Chemical, Engineering*. 2011, Stanford University.

358. Raines, R.T., *Ribonuclease A.* Chem. Rev., 1998. **98**(3): p. 1045-1066.

359. Kushner, S.R., *mRNA decay in Escherichia coli comes of age.* J. Bacteriol., 2002. **184**(17): p. 4658-4665.

360. Zhang, Y., et al., *MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli.* Mol. Cell, 2003. **12**(4): p. 913-923.

361. Prud'homme-Genereux, A., et al., *Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'.* Mol. Microbiol., 2004. **54**(5): p. 1409-1421.

362. Bermardo, G. and C. Cordonnier, *Mechanism of Degradation of DNA by Endonuclease I from Escherichia Coli.* J Mol Biol, 1965. **11**: p. 141-3.

363. Calhoun, K.A. and J.R. Swartz, *Total amino acid stabilization during cell-free protein synthesis reactions.* J. Biotechnol., 2006. **123**(2): p. 193-203.

364. Borja, G.M., et al., *Engineering Escherichia coli to increase plasmid DNA production in high cell-density cultivations in batch mode.* Microb. Cell Fact, 2012. **11**(1): p. 132.

365. Lehman, I.R., G.G. Roussos, and E.A. Pratt, *The deoxyribo-nucleases of Escherichia coli. III. Studies on the nature of the inhibition of endonuclease by ribonucleic acid.* J Biol Chem, 1962. **237**: p. 829-33.

366. 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.

367. Zawada, J. and J. Swartz, *Maintaining rapid growth in moderate-density Escherichia coli fermentations.* Biotechnol Bioeng, 2005. **89**(4): p. 407-15.

368. Voloshin, A.M. and J.R. Swartz, *Efficient and scalable method for scaling up cell free protein synthesis in batch mode.* Biotechnology and Bioengineering, 2005. **91**(4): p. 516-521.

369. Kim, D.M. and C.Y. Choi, *A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane.* Biotechnol Prog, 1996. **12**(5): p. 645-9.

370. Kigawa, T., et al., *Cell-free production and stable-isotope labeling of milligram quantities of proteins.* FEBS Lett, 1999. **442**(1): p. 15-9.

371. 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.

372. 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.

373. 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, 2015. **4**(5): p. 503-15.

374. Kay, J.E. and M.C. Jewett, *Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol.* Metab Eng, 2015. **32**: p. 133-42.

375. Pardee, K., et al., *Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components.* Cell, 2016. **165**(5): p. 1255-66.

376. Salehi, A.S., et al., *Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system.* Biotechnol J, 2016. **11**(2): p. 274-81.

377. 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.

378. Iwane, Y., et al., *Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes.* Nat Chem, 2016. **8**(4): p. 317-25.

379. 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.

380. Currier, N.V., et al., *Targeted Drug Delivery with an Integrin-Binding Knottin-Fc-MMAF Conjugate Produced by Cell-Free Protein Synthesis.* Molecular Cancer Therapeutics, 2016. **15**(6): p. 1291-1300.

381. Chin, J.W., *Reprogramming the genetic code.* Embo Journal, 2011. **30**(12): p. 2312-2324.

382. Pirman, N.L., et al., *A flexible codon in genomically recoded Escherichia coli permits programmable protein phosphorylation.* Nat Commun, 2015. **6**: p. 8130.

383. Josephson, K., M.C. Hartman, and J.W. Szostak, *Ribosomal synthesis of unnatural peptides.* J Am Chem Soc, 2005. **127**(33): p. 11727-35.

384. Sambrook, J., Fritsch, E.F., and Maniatis, T., *Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press. ColdSS Spring Harbor, NY*. 1989.

385. Gallagher, R.R., et al., *Multilayered genetic safeguards limit growth of microorganisms to defined environments.* Nucleic Acids Res, 2015. **43**(3): p. 1945-54.

386. Henderson, J.W., Brooks, A., *Improved Amino Acid Methods using Agilent ZORBAX Eclipse Plus C18 Columns for a Variety of Agilent LC Instrumentation and Separation Goals*, A. Technologies, Editor. 2010: USA.

387. Yang, J., et al., *Expression of active murine granulocyte-macrophage colony-stimulating factor in an Escherichia coli cell-free system.* Biotechnol Prog, 2004. **20**(6): p. 1689-96.

388. Goerke, A.R., K.G. Knapp, and J.R. Swartz, *BIOT 4-Developing a cell-free protein synthesis platform for producing proteins requiring disulfide bonds.* Abstracts of Papers of the American Chemical Society, 2006. **232**.

389. Goerke, A.R., et al., *Cell-free metabolic engineering promotes high-level production of bioactive Gaussia princeps luciferase.* Metab Eng, 2008. **10**(3-4): p. 187-200.

390. Grabowska, A., M. Nowicki, and J. Kwinta, *Glutamate dehydrogenase of the germinating triticale seeds: gene expression, activity distribution and kinetic characteristics.* Acta Physiologiae Plantarum, 2011. **33**(5): p. 1981-1990.

391. Lin, E.C., *Glycerol dissimilation and its regulation in bacteria.* Annu Rev Microbiol, 1976. **30**: p. 535-578.

392. Rittmann, D., S.N. Lindner, and V.F. Wendisch, *Engineering of a glycerol utilization pathway for amino acid production by Corynebacterium glutamicum.* Applied and Environmental Microbiology, 2008. **74**(20): p. 6216-6222.

393. Atkinson, D.E., *The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers.* Biochemistry, 1968. **7**(11): p. 4030-4.

394. Des Soye, B.J., et al., *Repurposing the translation apparatus for synthetic biology.* Curr Opin Chem Biol, 2015. **28**: p. 83-90.

395. Raucher, D. and J.S. Ryu, *Cell-penetrating peptides: strategies for anticancer treatment.* Trends Mol Med, 2015. **21**(9): p. 560-70.

396. Catherine, C., et al., *Engineering thermal properties of elastin-like polypeptides by incorporation of unnatural amino acids in a cell-free protein synthesis system.* Biotechnology and Bioprocess Engineering, 2015. **20**(3): p. 417-422.

397. Wu, I.L., et al., *Multiple site-selective insertions of noncanonical amino acids into sequence-repetitive polypeptides.* Chembiochem, 2013. **14**(8): p. 968-78.

398. Kuznetsov, G., et al., *Optimizing complex phenotypes through model-guided multiplex genome engineering.* Genome Biol, 2017. **18**(1): p. 100.

399. Richardson, S.M., et al., *Design of a synthetic yeast genome.* Science, 2017. **355**(6329): p. 1040-1044.

400. Wang, K., et al., *Defining synonymous codon compression schemes by genome recoding.* Nature, 2016. **539**(7627): p. 59-64.

401. Zhang, Y., et al., *A semi-synthetic organism that stores and retrieves increased genetic information.* Nature, 2017. **551**(7682): p. 644-647.

402. Dudley, Q.M., A.S. Karim, and M.C. Jewett, *Cell-free metabolic engineering: biomanufacturing beyond the cell.* Biotechnol J, 2015. **10**(1): p. 69-82.

403. Garamella, J., et al., *The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology.* ACS Synth Biol, 2016. **5**(4): p. 344-55.

404. Karim, A.S., Q.M. Dudley, and M.C. Jewett, *Cell-free synthetic systems for metabolic engineering and biosynthetic pathway prototyping*, in *Industrial Biotechnology: Microorganisms*, J.C. Liao and C. Wittmann, Editors. 2015, Wiley-VCH: Weinheim.

405. Martin, R.W., et al., *Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids.* Nat Commun, 2018. **9**(1): p. 1203.

406. d'Aquino, A.E., D.S. Kim, and M.C. Jewett, *Engineered Ribosomes for Basic Science and Synthetic Biology.* Annual Review of Chemical and Biomolecular Engineering, 2018. **9**(1): p. 311-340.

407. Liu, C.C., et al., *Toward an orthogonal central dogma.* Nat Chem Biol, 2018. **14**(2): p. 103-106.

408. Liu, Y., D.S. Kim, and M.C. Jewett, *Repurposing ribosomes for synthetic biology.* Curr Opin Chem Biol, 2017. **40**: p. 87-94.

409. 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.

410. 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.

411. Mosberg, J.A., M.J. Lajoie, and G.M. Church, *Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate.* Genetics, 2010. **186**(3): p. 791-9.

412. 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.

413. Horton, R.M., *PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes.* Mol Biotechnol, 1995. **3**(2): p. 93-9.

414. Jewett, M.C., et al., *An integrated cell-free metabolic platform for protein production and synthetic biology.* Mol Syst Biol, 2008. **4**: p. 220.

415. Lederberg, J. and E.M. Lederberg, *Replica plating and indirect selection of bacterial mutants.* J Bacteriol, 1952. **63**(3): p. 399-406.

416. Davanloo, P., et al., *Cloning and expression of the gene for bacteriophage T7 RNA polymerase.* Proc Natl Acad Sci U S A, 1984. **81**(7): p. 2035-9.

417. 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.

418. Inouye, S. and M. Inouye, *Up-promoter mutations in the lpp gene of Escherichia coli.* Nucleic Acids Res, 1985. **13**(9): p. 3101-10.

419. Gan, R. and M.C. Jewett, *Translation system engineering in Escherichia coli enhances non-canonical amino acid incorporation into proteins.* 2017.

420. McAllister, W.T., et al., *Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection.* J Mol Biol, 1981. **153**(3): p. 527-44.

421. Zawada, J. and J. Swartz, *Effects of growth rate on cell extract performance in cell-free protein synthesis.* Biotechnol Bioeng, 2006. **94**(4): p. 618-24.

422. Bremer, H. and P. Dennis, *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2 ed. Vol. 1. 1996, Washington, D.C.: ASM Press.

423. Stefano, J.E. and J. Gralla, *Lac UV5 transcription in vitro. Rate limitation subsequent to formation of an RNA polymerase-DNA complex.* Biochemistry, 1979. **18**(6): p. 1063-7.

424. Espah Borujeni, A., A.S. Channarasappa, and H.M. Salis, *Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites.* Nucleic Acids Res, 2014. **42**(4): p. 2646-59.

425. Salis, H.M., E.A. Mirsky, and C.A. Voigt, *Automated design of synthetic ribosome binding sites to control protein expression.* Nat Biotechnol, 2009. **27**(10): p. 946-50.

426. Ellinger, T. and R. Ehricht, *Single-step purification of T7 RNA polymerase with a 6-histidine tag.* Biotechniques, 1998. **24**(5): p. 718-20.

427. 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.

428. Tabor, S. and C.C. Richardson, *A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes.* Proc Natl Acad Sci U S A, 1985. **82**(4): p. 1074-8.

429. 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.

430. Muller, D.K., C.T. Martin, and J.E. Coleman, *Processivity of proteolytically modified forms of T7 RNA polymerase.* Biochemistry, 1988. **27**(15): p. 5763-71.

431. Ikeda, R.A. and C.C. Richardson, *Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7.* J Biol Chem, 1987. **262**(8): p. 3790-9.

432. Ikeda, R.A. and C.C. Richardson, *Interactions of a proteolytically nicked RNA polymerase of bacteriophage T7 with its promoter.* J Biol Chem, 1987. **262**(8): p. 3800-8.

433. Gottesman, S., *Proteases and their targets in Escherichia coli.* Annu Rev Genet, 1996. **30**: p. 465-506.

434. Hwang, B.Y., et al., *Substrate specificity of the Escherichia coli outer membrane protease OmpP.* J Bacteriol, 2007. **189**(2): p. 522-30.

435. Tunitskaya, V.L. and S.N. Kochetkov, *Structural-functional analysis of bacteriophage T7 RNA polymerase.* Biochemistry (Mosc), 2002. **67**(10): p. 1124-35.

436. Petrov, A.S., et al., *RNA-magnesium-protein interactions in large ribosomal subunit.* J Phys Chem B, 2012. **116**(28): p. 8113-20.

437. Sousa, R., *T7 RNA Polymerase*, in *Encyclopedia of Biological Chemistry*, W.J. Lennarz and M.D. Lane, Editors. 2004, Elsevier.

438. Wang, L., et al., *Addition of the keto functional group to the genetic code of Escherichia coli.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 56-61.

439. Despanie, J., et al., *Elastin-like polypeptides: Therapeutic applications for an emerging class of nanomedicines.* J Control Release, 2016. **240**: p. 93-108.

440. Ledent, P., et al., *Unexpected influence of a C-terminal-fused His-tag on the processing of an enzyme and on the kinetic and folding parameters.* FEBS Letters, 1997. **413**(2): p. 194-196.

441. Sabaty, M., et al., *Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis.* BMC Biochemistry, 2013. **14**: p. 28-28.

442. Lu, Y., *Cell-free synthetic biology: Engineering in an open world.* Synth Syst Biotechnol, 2017. **2**(1): p. 23-27.

443. Buntru, M., et al., *Tobacco BY-2 cell-free lysate: an alternative and highly-productive plant-based in vitro translation system.* BMC Biotechnol, 2014. **14**: p. 37.

444. Buntru, M., et al., *A versatile coupled cell-free transcription-translation system based on tobacco BY-2 cell lysates.* Biotechnol Bioeng, 2015. **112**(5): p. 867-78.

445. Penalber-Johnstone, C., et al., *Optimizing cell-free protein expression in CHO: Assessing small molecule mass transfer effects in various reactor configurations.* Biotechnol Bioeng, 2017. **114**(7): p. 1478-1486.

446. Zemella, A., et al., *Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems.* Chembiochem, 2015. **16**(17): p. 2420-31.

447. Li, J., H. Wang, and M.C. Jewett, *Expanding the palette of Streptomyces-based cell-free protein synthesis systems with enhanced yields.* Biochemical Engineering Journal, 2018. **130**: p. 29-33.

448. Moore, S.J., et al., *Streptomyces venezuelae TX-TL - a next generation cell-free synthetic biology tool.* Biotechnol J, 2017. **12**(4).

449. Kelwick, R., et al., *Development of a Bacillus subtilis cell-free transcription-translation system for prototyping regulatory elements.* Metab Eng, 2016. **38**: p. 370-381.

450. Payne, W.J., *Studies on bacterial utilization of uronic acids. III. Induction of oxidative enzymes in a marine isolate.* J Bacteriol, 1958. **76**(3): p. 301-7.

451. Eagon, R.G., *Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes.* J Bacteriol, 1962. **83**: p. 736-7.

452. Aiyar, S.E., T. Gaal, and R.L. Gourse, *rRNA promoter activity in the fast-growing bacterium Vibrio natriegens.* J Bacteriol, 2002. **184**(5): p. 1349-58.

453. Dalia, T.N., et al., *Multiplex Genome Editing by Natural Transformation (MuGENT) for Synthetic Biology in Vibrio natriegens.* ACS Synth Biol, 2017. **6**(9): p. 1650-1655.

454. Wang, H., J. Li, and M.C. Jewett, *Development of a Pseudomonas putida cell-free protein synthesis platform for rapid screening of gene regulatory elements.* Synthetic Biology, 2018. **3**(1): p. ysy003-ysy003.

455. Failmezger, J., et al., *Cell-free protein synthesis from non-growing, stressed Escherichia coli.* Sci Rep, 2017. **7**(1): p. 16524.

456. Liu, D.V., J.F. Zawada, and J.R. Swartz, *Streamlining Escherichia coli S30 extract preparation for economical cell-free protein synthesis.* Biotechnol Prog, 2005. **21**(2): p. 460-5.

457. Nirenberg, M.W., *Cell-Free Protein Synthesis Directed by Messenger Rna.* Methods in Enzymology, 1963. **6**: p. 17-23.

458. Fagerbakke, K.M., S. Norland, and M. Heldal, *The inorganic ion content of native aquatic bacteria.* Can J Microbiol, 1999. **45**(4): p. 304-11.

459. Voloshin, A.M. and J. Swartz, *Large-Scale Batch Reactions for Cell-Free Protein Synthesis*, in *Cell-Free Protein Synthesis: Methods and Protocols*, A.S. Spirin and J. Swartz, Editors. 2008, Wiley-VCH Verlab GmbH & Co. KGaA: Weinheim, Germany. p. 207-235.

460. Gross, A., et al., *Synthetic Peptides as Protein Mimics.* Front Bioeng Biotechnol, 2015. **3**: p. 211.

461. Haussner, C., J. Lach, and J. Eichler, *Synthetic antibody mimics for the inhibition of protein-ligand interactions.* Curr Opin Chem Biol, 2017. **40**: p. 72-77.

462. Devkota, A.K., et al., *Fluorescent peptide assays for protein kinases.* Curr Protoc Mol Biol, 2010. **Chapter 18**: p. Unit 18 17.

463. Bengtsson-Palme, J., E. Kristiansson, and D.G.J. Larsson, *Environmental factors influencing the development and spread of antibiotic resistance.* FEMS Microbiol Rev, 2018. **42**(1).

464. Zhao, Q., et al., *Recombinant production of medium- to large-sized peptides in Escherichia coli using a cleavable self-aggregating tag.* Microb Cell Fact, 2016. **15**(1): p. 136.

465. Silvestro, L., J.N. Weiser, and P.H. Axelsen, *Antibacterial and antimembrane activities of cecropin A in Escherichia coli.* Antimicrob Agents Chemother, 2000. **44**(3): p. 602-7.

466. Failmezger, J., et al., *Cell-free protein synthesis from fast-growing Vibrio natrigens.* Front Microbiol, 2018.

467. Belov, M.E., et al., *From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry.* Anal Chem, 2013. **85**(23): p. 11163-73.

468. Zhang, Z. and A.G. Marshall, *A universal algorithm for fast and automated charge state deconvolution of electrospray mass-to-charge ratio spectra.* Journal of the American Society for Mass Spectrometry, 1998. **9**(3): p. 225-233.

469. Fellers, R.T., et al., *ProSight Lite: graphical software to analyze top-down mass spectrometry data.* Proteomics, 2015. **15**(7): p. 1235-8.

470. Plaut, B. and J.R. Knowles, *pH-dependence of the triose phosphate isomerase reaction.* The Biochemical journal, 1972. **129**(2): p. 311-320.

471. Mukai, T., et al., *Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon.* Scientific Reports, 2015. **5**: p. 9699.

472. Ronda, C., et al., *CRMAGE: CRISPR Optimized MAGE Recombineering.* Scientific Reports, 2016. **6**: p. 19452.

473. Torella, J.P., et al., *Unique nucleotide sequence (UNS)-guided assembly of repetitive DNA parts for synthetic biology applications.* Nature protocols, 2014. **9**(9): p. 2075-2089.

474. Knowles, J.R. and W.J. Albery, *Perfection in enzyme catalysis: the energetics of triosephosphate isomerase.* Accounts of Chemical Research, 1977. **10**(4): p. 105-111.

475. Sharma, P. and P. Guptasarma, *'Super-perfect' enzymes: Structural stabilities and activities of recombinant triose phosphate isomerases from Pyrococcus furiosus and Thermococcus onnurineus produced in Escherichia coli.* Biochemical and Biophysical Research Communications, 2015. **460**(3): p. 753-758.

476. Blacklow, S.C., et al., *Triosephosphate isomerase catalysis is diffusion controlled.* Biochemistry, 1988. **27**(4): p. 1158-1165.

477. Stroppolo, M.E., et al., *Superefficient enzymes.* Cell Mol Life Sci, 2001. **58**(10): p. 1451-60.

478. Davidi, D., et al., *A Bird's-Eye View of Enzyme Evolution: Chemical, Physicochemical, and Physiological Considerations.* Chemical Reviews, 2018.

479. Kenney, G.E., et al., *The biosynthesis of methanobactin.* Science, 2018. **359**(6382): p. 1411-1416.

480. Park, Y.J., et al., *Repurposed HisC Aminotransferases Complete the Biosynthesis of Some Methanobactins.* Biochemistry, 2018. **57**(25): p. 3515-3523.

481. Yang, A., et al., *A chemical biology route to site-specific authentic protein modifications.* Science (New York, N.Y.), 2016. **354**(6312): p. 623-626.