2016-03-01

Aminoacyl-tRNA Synthetase Production for Unnatural Amino Acid Incorporation and Preservation of Linear Expression Templates in Cell-Free Protein Synthesis Reactions

Andrew Broadbent
Brigham Young University - Provo

Aminoacyl-tRNA Synthetase Production for Unnatural Amino Acid Incorporation and Preservation of Linear Expression Templates in Cell-Free Protein Synthesis Reactions

Andrew Broadbent

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Bradley C. Bundy, Chair
Morris D. Argyle
Richard K. Watt

Department of Chemical Engineering

Brigham Young University

March 2016

ABSTRACT

Aminoacyl-tRNA Synthetase Production for Unnatural Amino Acid Incorporation and Preservation of Linear Expression Templates in Cell-Free Protein Synthesis Reactions

Andrew Broadbent
Department of Chemical Engineering, BYU
Master of Science

Proteins—polymers of amino acids—are a major class of biomolecules whose myriad functions facilitate many crucial biological processes. Accordingly, human control over these biological processes depends upon the ability to study, produce, and modify proteins. One innovative tool for accomplishing these aims is cell-free protein synthesis (CFPS). This technique, rather than using living cells to make protein, simply extracts the cells' natural protein-making machinery and then uses it to produce protein *in vitro*. Because living cells are no longer involved, scientists can freely adapt the protein production environment in ways not otherwise possible. However, improved versatility and yield of CFPS protein production is still the subject of considerable research. This work focuses on two ideas for furthering that research.

The first idea is the adaptation of CFPS to make proteins containing unnatural amino acids. Unnatural amino acids are not found in natural biological proteins; they are synthesized artificially to possess useful properties which are then conferred upon any protein made with them. However, current methods for incorporating unnatural amino acids do not allow incorporation of more than one type of unnatural amino acid into a single protein. This work helps lay the groundwork for the incorporation of different unnatural amino acid types into proteins. It does this by using modified aminoacyl-tRNA synthetases (aaRSs), which are key components in CFPS, to be compatible with unnatural amino acids.

The second idea is the preservation of DNA templates from enzyme degradation in CFPS. Among the advantages of CFPS is the option of using linear expression templates (LETs) in place of plasmids as the DNA template for protein production. Because LETs can be produced more quickly than plasmids can, using LETs greatly reduces the time required to obtain a DNA template for protein production. This renders CFPS a better candidate for high-throughput testing of proteins. However, LETs are more susceptible to enzyme-mediated degradation than plasmids are, which means that LET-based CFPS protein yields are lower than plasmid-based CFPS yields. This work explores the possibility of increasing the protein yield of LET-based CFPS by addition of sacrificial DNA, DNA which is not used as a protein-making template but which is degraded by the enzymes in place of the LETs.

Keywords: Andrew Broadbent, cell-free protein synthesis (CFPS), unnatural amino acid incorporation, aminoacyl-tRNA synthetase, linear expression template (LET), aminoacylation assay, sacrificial DNA

ACKNOWLEDGEMENTS

This work would not have been possible without the kindness and generosity of my mentor, Dr. Brad Bundy. His great patience, insight, cheerfulness, and encouragement helped me to keep going and not give up. I am grateful as well for the generous stipend and payment of tuition with which he provided me during my time in the lab. Dr. Morris Argyle and Dr. Richard Watt provided valuable support as teachers of my graduate courses and also by reviewing my prospectus and thesis.My fellow graduate students, Mark Smith, Jeff Wu, Amin Salehi, Matt Schinn, Kristen Wilding, and Seung-Ook Yang provided much-appreciated technical help and camaraderie that improved the quality of my work and made my time at the lab more enjoyable.

The undergraduate students with whom I had the pleasure of working gave me many of my most cherished memories and lessons in the lab. These students included Ashtyn Groesbeck, Lauren Greer, Conner Earl, Anthony Bennett, Jeremy Hunt, Hayley Ford, Andrea Cutler, Christina Muhlestein, Garrett Kraus, Brooke Linney, Kelly Arrington, Amanda Peterson, Aaron Escamilla, Mark Lindsay, Yunju Lee, Chris Werner, and especially Steven Stanley, who spent many hours working side-by-side with me on the aminoacylation assay (see Chapter 3) in particular. These students inspired me with their curiosity, ability, good spirits, and testimonies.

Lastly, I am grateful for the support which my parents, siblings, friends, and church leaders have lent me. The love of God and of those around me has been as a beacon to me during these past two difficult years of graduate work.

TABLE OF CONTENTS

LIST OF TABLES ..... **vi**

LIST OF FIGURES ..... **vii**

1 Introduction to Cell-free Protein Synthesis (CFPS) **1**

1.1 Cell-free Protein Synthesis (CFPS) 1

1.1.1 Advantages of Cell-free Protein Synthesis 4

1.1.2 Disadvantages of Cell-free Protein Synthesis 4

1.1.3 Method of Crude Extract Preparation 4

1.2 Scope of the Project 6

1.3 Outline 6

2 Background Information about CFPS Improvement Methods **8**

2.1 Incorporation of Different Unnatural Amino Acid Types 8

2.1.1 What Is Unnatural Amino Acid Incorporation? 8

2.1.2 Applications of Unnatural Amino Acid Incorporation 12

2.2 Using Sacrificial DNA to Improve LET-Based CFPS Protein Yields 13

2.2.1 The Role of DNA Templates in CFPS 13

2.2.2 LETs as DNA Templates in CFPS 14

2.2.3 PCR (Polymerase Chain Reaction) 15

2.2.4 The Challenge of LET Preservation 17

3 Production and Evaluation of aaRS for Incorporation of Unnatural Amino Acids **18**

3.1 Introduction 18

3.2 Materials and Methods 21

3.2.1 Design of Orthogonal tRNA/aaRS Combinations 21

3.2.2 Protein Preparation 23

3.2.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) 25
3.2.4 Aminoacylation Assay 25
3.3 Results and Discussion 27
3.4 Conclusion 38
4 Using Sacrificial DNA to Improve LET-Based CFPS Protein Yields 39
4.1 Introduction 39
4.2 Materials and Methods 39
4.3 Results and Discussion 43
4.4 Economics of Sacrificial DNA Use in LET-based CFPS 50
4.5 Conclusion 54
5 Conclusions and Future Work 55
5.1 Conclusions 55
5.2 Future Work 56
6 References 59
7 Appendix 61
7.1 Sequences of Selected tRNA/aaRS Pairs (List Courtesy of Mark Smith) 61
7.2 Sequence of sfGFP (Superfolder Green Fluorescent Protein) [17] 62
7.3 Nucleic Acid Sequences of Primer Pairs Used to Make Sacrificial DNA [29] 62

LIST OF TABLES

Table 1: tRNA/aaRS Orthogonal Pair Selections for Incorporation of Different Unnatural Amino Acids. 19

Table 2: Reagents Needed for the Aminoacylation Assay. 27

Table 3: Reagents Needed for Cell-free Protein Synthesis Reactions. 40

Table 4: Temperature Cycle Used to Hybridize Complimentary PCR Primers for Use as Sacrificial DNA. 42

Table 5: Properties of Sacrificial DNA Primer Pairs. 43

Table 6: Low T_m Values and Yields Displayed from Least to Greatest T_m 48

LIST OF FIGURES

Figure 1: A comparison of the *in vivo* and CFPS methods of protein production. 3

Figure 2: The cell-free extract production process. 7

Figure 3: General amino acid structure. 9

Figure 4: Three natural amino acids, separate and linked. 10

Figure 5: Protein synthesis using cellular machinery. 11

Figure 6: Comparative time requirements for *in vivo*, plasmid-based CFPS, and LET-based CFPS methods of protein production. 14

Figure 7: The PCR Process. 16

Figure 8: Three Kinds of Cross-talk Among Cellular Machinery. 20

Figure 9: Plasmid map of the MbRS plasmid. 22

Figure 10: Polyacrylamide Gel on Purified Solutions of PhRS and ScRS. Ladder legend on the left obtained from [2]. 28

Figure 11: Polyacrylamide Gel on Purified Solutions of MjRS. 29

Figure 12: Polyacrylamide Gel on Purified Cell Lysate Supernatant from Cells Containing the MbRS Plasmid. 30

Figure 13: Polyacrylamide Gel on the Supernatant of Centrifuged Lysates of Cells Containing the MbRS Plasmid. 31

Figure 14: Polyacrylamide Gel on Noncentrifuged Lysates of Cells Containing the MbRS Plasmid. 32

Figure 15: Image of a Clear, Flat-bottomed 96-well Plate Used for an Aminoacylation Assay. 34

Figure 16: Image of a Clear, Flat-bottomed 96-well Plate Used for an Aminoacylation Assay, with 50 uL of 8% PVA Added to Each Well. 35

Figure 17: Standard Curve Data Points and Second-order Polynomial Regression from the Plate Shown in Figure 15. 36

Figure 18: Standard Curve Data Points and Second-order Polynomial Regression from the Plate Shown in Figure 16. 37

Figure 19: RFU-to-ug/mL Standard Curve. 41

Figure 20: Magnesium Optimization of CFPS with Pair A Added. 44

Figure 21: Magnesium Optimization of CFPS with Pair A Added in the Amount of Seven Times the Mass of LET Present. 45

Figure 22: Yields of CFPS Reactions with Various Sacrificial DNA Pairs and 15 mM $Mg(Glu)_2$. All reactions were performed in triplicate, and error bars extend one standard deviation up and down from the average result. 46

Figure 23: Magnesium Optimization of CFPS Reactions Containing Sacrificial DNA. 47

Figure 24: Magnesium Optimization of CFPS Reactions with Increased Amounts of LET but No Sacrificial DNA. 49

Figure 25: Costs of LET-based CPFS without Sacrificial DNA. Costs shown are for enough CFPS to make 1 mg of sfGFP. Costs of materials assumed the same as in [4]. 52

Figure 26: Costs of LET-based CFPS with Sacrificial DNA (Pair D, for this chart). 53

Figure 27: Costs of LET-based CFPS with 3x the Normal Amount of LET. Costs shown are for enough CPFS to make 1 mg of sfGFP. Costs of materials assumed the same as in [4]. 53

1 INTRODUCTION TO CELL-FREE PROTEIN SYNTHESIS (CFPS)

This thesis presents preliminary work on two ways to improve cell-free protein synthesis (CFPS): (1) incorporation of different unnatural amino acid types into protein and (2) the use of sacrificial DNA to improve LET-based CFPS protein yields.

Prior to discussing the author's work in these areas, a brief introduction to cell-free protein synthesis itself will be given in the sections below.

1.1 Cell-free Protein Synthesis (CFPS)

Protein, one of the major classes of biomolecules, includes myriad important biological compounds, from antibodies and transporters to structural elements and enzymes. Without proteins, life would not exist. In recent years, scientists have leveraged the vast body of extant genomic data to produce, *in vitro*, both natural and modified versions of proteins. This ability to make proteins has potential applications as broad as proteins' roles are diverse; both in studies and in actual production, proteins are already claiming a prominent place on the world stage. In 2013, biopharmaceuticals (many of which are proteins) garnered $140 billion in sales worldwide [6]. They form the backbone of virus-like particles—a new vaccine and drug delivery technology [7]. Even laundry detergent contains engineered proteins, enzymes which enhance the cleaning ability of the detergent [8]. It is within this exciting and expanding context of applications that cell-free protein synthesis, one important developing method of producing protein, is making its mark.

Cell-free protein synthesis (CFPS) is a method for making protein which extracts the protein-making machinery from living cells and uses this machinery to produce the protein in an open, test-tube environment (i.e. *in vitro*). It is used as an alternative to the *in vivo* method, which produces protein inside cells that are still alive. CFPS is not yet used in industrial production except in a very few instances, but the technology is developing toward industrial use [9]. Startups like Sutro Biopharma are already using cell-free protein synthesis [10]. Figure 1 compares the *in vivo* method to CFPS.

The *in vivo* and CFPS methods of protein production have similar steps. Both require growing cells, lysing the cells, and purifying out the protein product. The difference lies in when protein production is started. *In vivo* production of protein happens inside the cells while they are still alive. CFPS production of protein waits until the cells have been lysed, then salvages the cellular protein-making machinery and uses it to make the protein without living cells, since in this way there is no cell wall to hinder manipulation of the production process [11].

A cell-free protein synthesis system needs several components: ribosomes, elongation factors, protein folding chaperones, transfer RNA/aminoacyl-tRNA synthetase (tRNA/aaRS) pairs, amino acids, magnesium ions, catabolism enzymes, ATP, etc. There are two major approaches to creating such a system. The first, called the PURE (protein synthesis using recombinant elements) system [12], does so by combining individual components in known quantities. This method has the advantage of known component amounts, but suffers from the associated increased production cost and reduced productivity. The second approach, crude cell-free extract preparation, uses the soluble fraction of cell lysate: in other words, the liquid removed after breaking open bacterial cells. This liquid extract already contains ribosomes, protein folding chaperones, tRNA/aaRS pairs, amino acids, magnesium ions, catabolism enzymes, and ATP. This

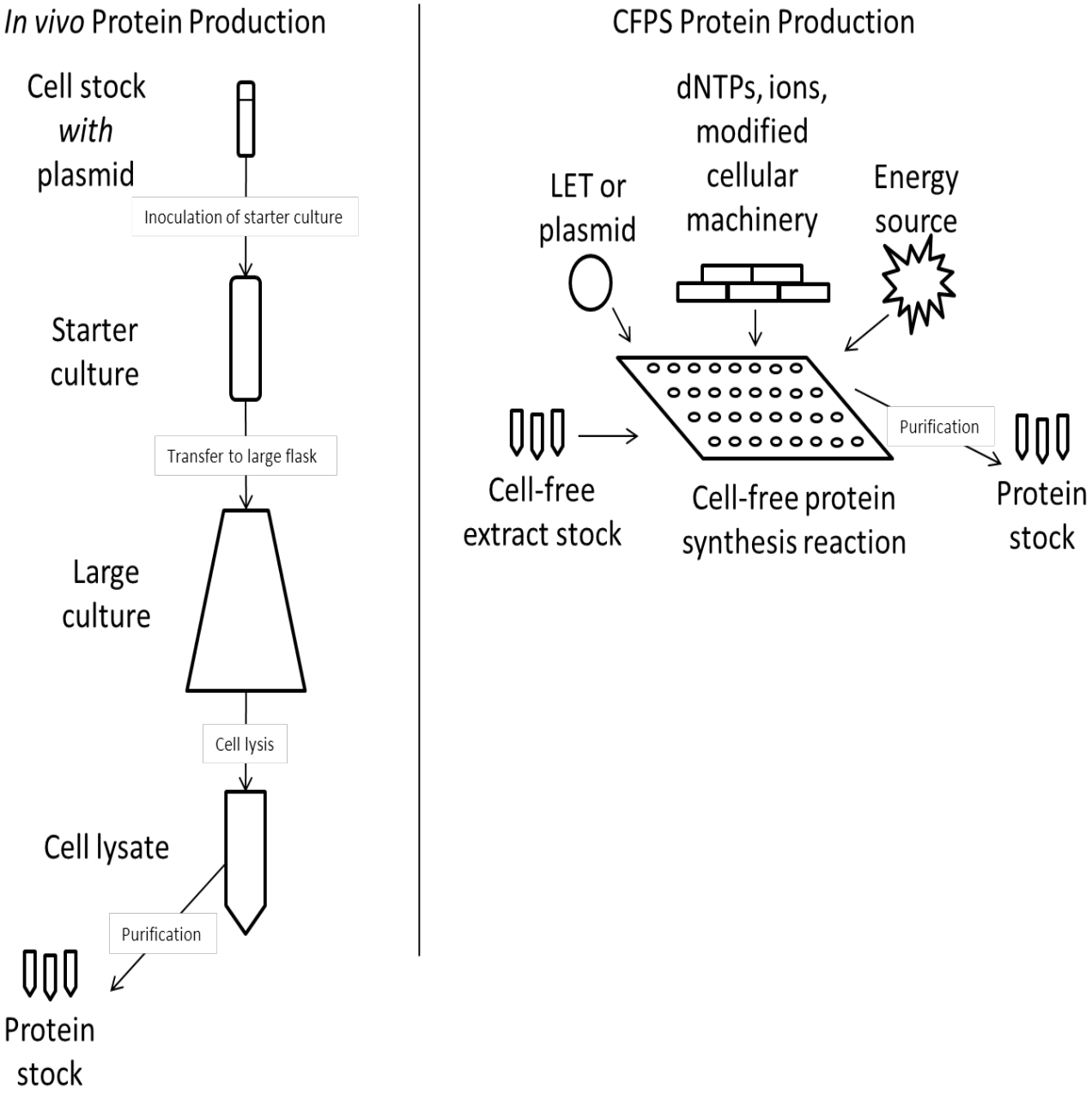

Figure 1: A comparison of the *in vivo* and CFPS methods of protein production. Using cell-free extract can be a much less laborious route than is growing cells and lysing them.

work uses crude cell-free extract [9], which is 100 times less expensive than the PURE system and produces proteins at much higher yields.

1.1.1 Advantages of Cell-free Protein Synthesis

Cell-free protein synthesis offers several advantages over *in vivo* protein synthesis. First, because there are no living cells, energy normally used to keep cells alive can instead be used for making the product protein. Second, because there is no cell wall, reagents can be added, measurements can be made, and product can be purified out more easily; for example, this direct access to the cellular machinery greatly facilitates the production of protein microarrays, or broad sets of related proteins which are compared and contrasted to investigate their structure and function. Third, cell-free extract can be stored in a much greater range of temperatures without major damage to the system, and it can even be lyophilized (made into freeze-dried powder), which makes storage, transportation, and use easier [3, 9].

1.1.2 Disadvantages of Cell-free Protein Synthesis

At present, cell-free protein synthesis systems, especially the PURE variety, are too expensive to be economical in industry. In addition, the crude cell-free extract, which is much less expensive than the PURE system, is still too expensive for some applications [9], and there can be variations from batch to batch of prepared components. However, the Bundy lab and others continue to improve the crude system [13].

1.1.3 Method of Crude Extract Preparation

The following is a general description of a method for crude cell-free extract preparation which has been optimized by previous members of the Bundy lab (see Figure 2 on page 6 for a

pictorial representation of the process) [14]. Cell-free extracts begin with the growth of an *E. coli* (cell line BL21 (DE3) Star™) cell culture. After inoculation of a plate with cells from a glycerol stock, bacterial colonies grow in the plate and are used to inoculate a test tube containing 5 mL of autoclaved LB (Luria broth) growth medium. Alternatively, the 5 mL of medium may be inoculated directly from the glycerol stock. The resulting 5 mL culture is incubated overnight at 37º C with 280 rpm shaking in an incubator shaker, after which the contents are transferred to a flask containing 100 mL of autoclaved LB growth medium. This flask is incubated at 37º C with 280 rpm shaking, and then its contents are transferred to a container containing 1 L of autoclaved LB growth medium. The growth of bacteria in the 1 L container is monitored via optical density (OD) readings on samples taken hourly. When the growth begins to exit the exponential phase, the culture is removed from incubation. Next, the culture is centrifuged at 11,270 RCF (relative centrifugal force; equivalent to "times g") in a supercentrifuge for 30 minutes to separate out the cells from the growth medium.

After the medium is removed, the cells are washed with Buffer A (10 mM Tris base, 14 mM anhydrous magnesium acetate, 60 mM potassium glutamate, 1 mM dithiothreitol; pH adjusted to 8.2 using acetic acid, sterilized in an autoclave), resuspended in the same kind of buffer, and lysed with a French press homogenizer. The lysed cells are then centrifuged at 12,000 RCF for 10 minutes to separate out the cell wall and other unwanted cell parts. The supernatant, which contains ribosomes, mRNA, tRNA, and aaRS, is then removed, leaving the unwanted cell parts to be discarded. The mRNA is degraded through a run-off reaction at 37º C and 280 rpm using endogenous RNAses. At this point, the cell-free extract is almost complete. The last remaining component, magnesium ion, is adjusted for each new extract made, and so several aliquots are set aside for testing with various concentrations of magnesium ion. Whichever yields the most protein

in a cell-free protein synthesis reaction has the best concentration, and that concentration is then applied to the rest of the batch, thus completing the preparation of the cell-free extract [15].

1.2 **Scope of the Project**

The objectives of this work were the following:

1. Produce a set of aminoacyl-tRNA synthetases for a CFPS system capable of incorporating different unnatural amino acid types into a protein.
2. Evaluate the use of sacrificial DNA as a means to increase protein yield of LET-based CFPS.

1.3 **Outline**

Chapter 2: Background Information about CFPS Improvement Methods

This chapter presents information relevant to the work on CFPS improvement.

Chapter 3: Production and Evaluation of aaRS for Incorporation of Unnatural Amino Acids

This chapter details methods and results concerning the production of a set of aminoacyl-tRNA synthetases for a CFPS system capable of incorporating different unnatural amino acid types into a protein.

Chapter 4: Using Sacrificial DNA to Improve LET-based CFPS Protein Yields

This chapter details methods and results concerning the use of sacrificial DNA as a means to increase protein yield of LET-based CFPS.

Chapter 5: Conclusions and Future Work

This chapter summarizes the conclusions of work on the two CFPS improvement methods and suggests a direction for future work in these areas.

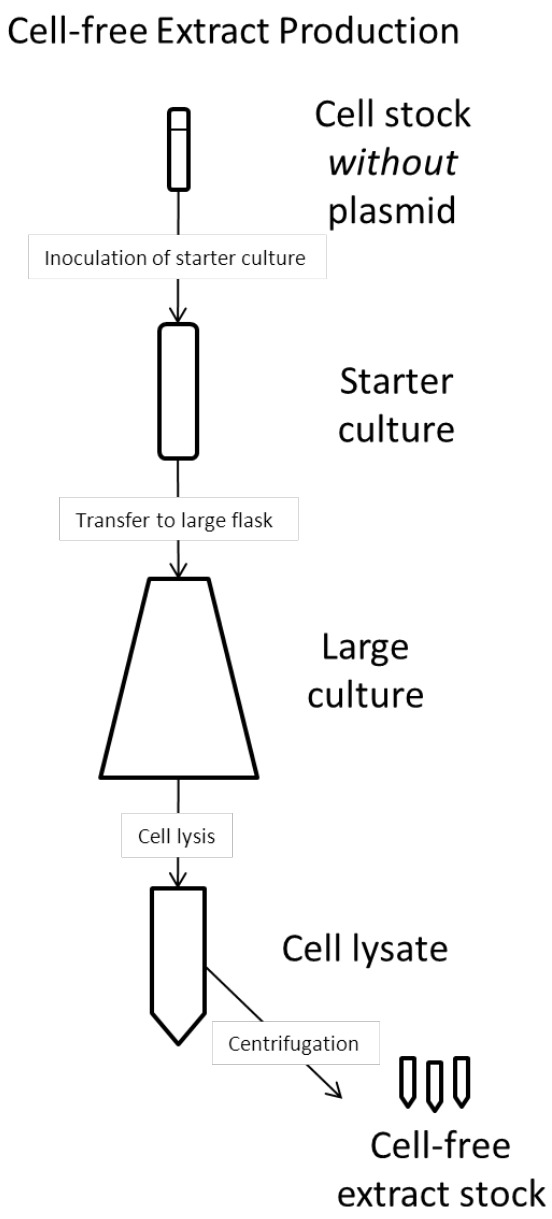

Figure 2: The cell-free extract production process.

Chapter 6: Appendix

This chapter contains supplementary material too large or ancillary to be placed in the text, such as full sequences of tRNA and aminoacyl-tRNA synthetases and the sequence of sfGFP (superfolder green fluorescing protein).

2 BACKGROUND INFORMATION ABOUT CFPS IMPROVEMENT METHODS

This section of the thesis presents background information on two ideas for improving CFPS: (1) incorporation of different unnatural amino acid types into protein and (2) the use of sacrificial DNA to improve LET-based CFPS protein yields.

2.1 Incorporation of Different Unnatural Amino Acid Types

This subsection provides background information about the first idea for improving CFPS, incorporation of unnatural amino acid types, including what it is and its applications.

2.1.1 What Is Unnatural Amino Acid Incorporation?

An amino acid is a molecule which has the structure shown in Figure 3. The group symbolized by the letter "R" in the figure is a molecular branch whose identity is different for each individual amino acid. In fact, the differences among the various amino acids amount to nothing more than having different R groups. In nature, a set of 20 amino acids is used by all organisms to produce proteins, which are simply amino acids linked by a bond between the amine and carboxyl group (see Figure 4).

All protein synthesis done using cellular machinery involves the same basic requirements. First, there must be a DNA template. It is referred to as a template because it contains the information needed to make the protein product. Just as protein is a polymer of amino acids, DNA is a polymer of nucleic acids, and it is the order of the nucleic acids in DNA that dictates the order of amino acids in the protein product. These nucleic acids come in triplets called codons.

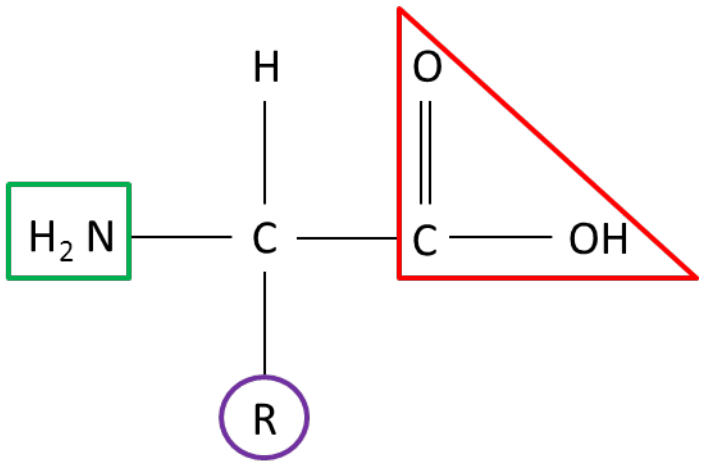

Figure 3: General amino acid structure. The name "amino acid" refers to the amine group (also called the N-terminus) surrounded by the rectangle and the carboxylic acid group (also called the C-terminus) surrounded by the triangle. Amino acids link up N-terminus-to-C-terminus to form long chains, or polymers, known as proteins. The R group surrounded by the circle signifies a molecular branch which varies by amino acid; it is the feature that distinguishes one amino acid from another.

Each codon codes for a specific amino acid, and taken together, the entire DNA template codes for all those amino acids linked together: a protein. Next, the information needs to be brought to the site where the amino acids are linked—the ribosome. A process called transcription makes mRNA (also called messenger RNA), another polymer of nucleic acids, based on the DNA template. The resulting mRNA contains the information from the DNA template in its sequence. The mRNA travels to the ribosome and dictates the sequence of amino acids in the protein made there. As the ribosome reads the mRNA, molecules called tRNAs (transfer RNAs) bring amino acids to the ribosome so they can be added to the growing protein chain.

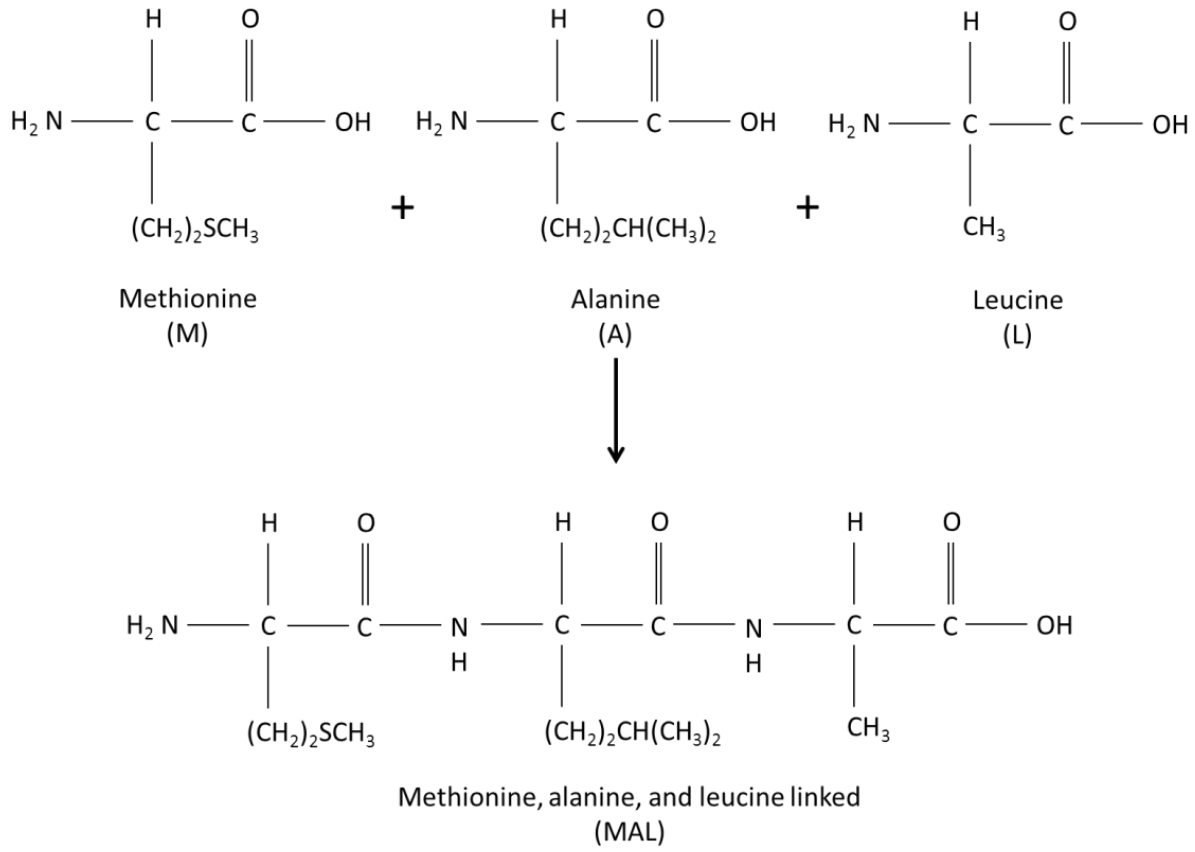

Figure 4: Three natural amino acids, separate and linked. These linked amino acids (MAL) are the first three in the protein insulin, which is 110 amino acids long in total.

In order for the mRNA information to be translated into an amino acid sequence, each tRNA matches up with an mRNA codon via a complimentary triplet of nucleic acids called an anticodon. The anticodon end of the tRNA binds to the appropriate codon, and the other end, which carries the amino acid, positions the amino acid for binding to the end of the growing protein chain. The amino acid is attached to the end of the protein chain. Then the next codon of the mRNA is read, and a new tRNA takes the old one's place. After a tRNA delivers its amino acid to the ribosome and is released from the ribosome, it must be reloaded with amino acid. An aminoacyl-tRNA synthetase (aaRS) specific to a tRNA and an amino acid loads the tRNA with the amino acid [16] and then releases it to again deliver its payload at the ribosome. See Figure 5.

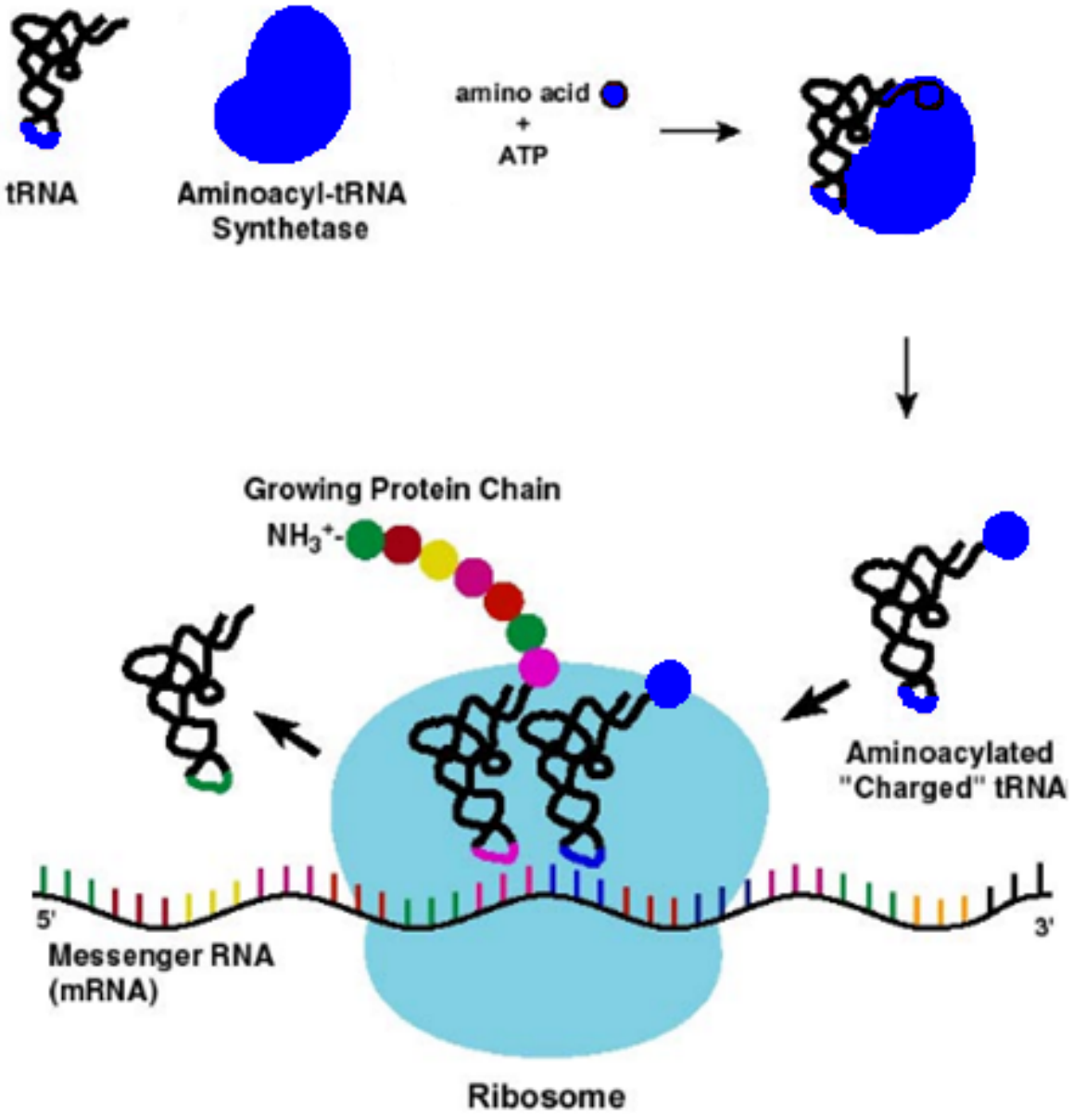

Figure 5: Protein synthesis using cellular machinery. This process occurs both *in vivo* and in CFPS. DNA template not depicted. Figure modified from [2].

This work adapts this cellular system of protein synthesis to include unnatural amino acids by using modified aaRSs that load unnatural amino acids onto tRNAs, which then bring them to the ribosome for addition to the protein product (see Chapter 3 for more on this topic).

2.1.2 Applications of Unnatural Amino Acid Incorporation

One application for unnatural amino acids is protein immobilization, or attachment of a protein to a surface. Enzymes, which are proteins that catalyze chemical reactions, are often lost when used in industrial processes because they are difficult to recover from the chemicals with which they are mixed. By attaching enzymes to a surface, one can ensure that they do not become lost in the solution. However, the enzymes' performance is dependent on the site of attachment. Unnatural amino acids would provide attachment sites whose chemical uniqueness in the protein would ensure a consistent attachment site. For instance, some unnatural amino acids can participate in protein-protein linking reactions such as the copper(I)-catalyzed azide−alkyne [3 + 2] cycloaddition reaction [17].

Many other applications of unnatural amino acid incorporation exist. In fact, they are as numerous as are the properties of the unnatural amino acids themselves. Peter G. Schultz, a prominent researcher in this area, lists many of the useful properties of unnatural amino acids in the following excerpt from his lab webpage:

> "[They] include heavy atom containing amino acids to facilitate x-ray crystallographic studies; amino acids with novel steric/packing and electronic properties; photocrosslinking amino acids which can be used to probe protein-protein interactions *in vitro* or *in vivo*; keto, acetylene, azide, and boronate containing amino acids which can be used to selectively introduce a large number of biophysical probes, tags, and novel chemical functional groups into proteins *in vitro* or *in vivo*; redox active amino acids to probe and modulate electron transfer; photocaged and photoisomerizable amino acids to photoregulate biological processes; metal binding amino acids for catalysis and metal ion sensing; amino acids that contain fluorescent or IR active side chains to probe protein

structure and dynamics; α-hydroxy acids and D-amino acids as probes of backbone conformation and hydrogen bonding interactions; and sulfated amino acids and mimetics of phosphorylated amino acids as probes of posttranslational modifications." [18]

2.2 Using Sacrificial DNA to Improve LET-Based CFPS Protein Yields

This subsection provides background information about the second idea for improving CFPS: the use of sacrificial DNA to improve LET-based CFPS protein yields, including the role of DNA templates in CFPS, the special advantages and disadvantages of LETs as DNA templates, how LETs are made, and the preservation of LETs in CFPS using sacrificial DNA.

2.2.1 The Role of DNA Templates in CFPS

In nature, DNA provides the instructions for making proteins. The information contained in the DNA sequence is transcribed into messenger RNA, which takes the information to the ribosome. The ribosome can then make protein with the sequence dictated by the messenger RNA. This protein production process is the same in CFPS, since CFPS uses extracted bacterial ribosomes and other cellular machinery to make protein. Thus, DNA has the same function in CFPS as it does in living cells: to provide instructions or information that will eventually determine the sequence of the protein product.

In living cells, such as the bacteria from which cell-free extract is made, protein synthesis and other cellular processes are regulated in part using enzymes called nucleases, which degrade DNA and RNA. Nucleases are divided into two major categories: endonucleases and exonucleases. Endonucleases begin degradation somewhere in the middle of the DNA or RNA molecule, whereas exonucleases start degradation at the ends of the molecule. When degradation lowers the levels of DNA and RNA in the cell, the proteins for which they code are not produced as much. In

this way, nucleases act as a set of biological brakes for the cell. In CFPS, however, where the only goal is to produce as much of a single protein as possible, nucleases only get in the way of making the desired product.

2.2.2 LETs as DNA Templates in CFPS

Traditionally, the DNA template used in CFPS is plasmid DNA, meaning circularized DNA. Because plasmids are circularized, there are no free ends for exonucleases to degrade. The exonucleases must wait for the endonucleases to cut the plasmid and expose free ends in order to begin their degradation. This lends plasmids more time to help make protein. However, in order to obtain plasmids to use in CFPS, cells must be cultured and harvested and the plasmid must be purified out from them, a process that takes days to complete. In order to reduce the time needed to obtain a DNA template for CFPS, one can use LETs instead of plasmids (see Figure 6 below). LETs can be made by PCR (polymerase chain reaction) and then purified, a process which can be completed in a matter of hours.

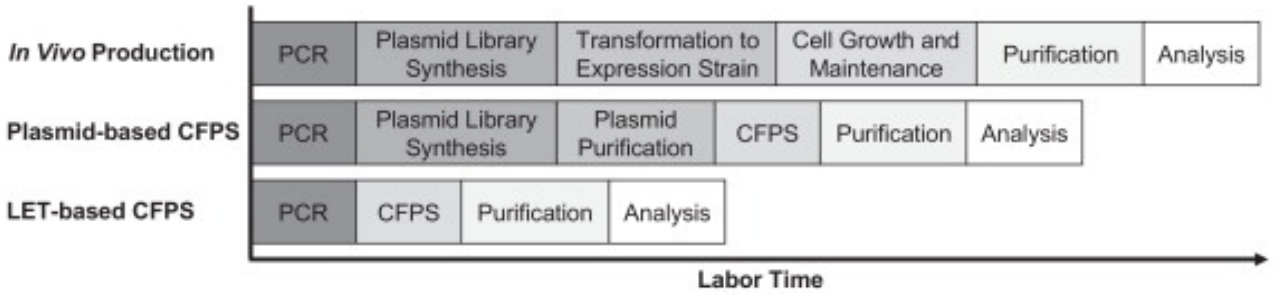

Figure 6: Comparative time requirements for *in vivo*, plasmid-based CFPS, and LET-based CFPS methods of protein production. Using LETs in CFPS requires the least time of the three methods. Figure reproduced from [13].

2.2.3 PCR (Polymerase Chain Reaction)

PCR, or polymerase chain reaction, is a commonly-used method for synthesizing large amounts of DNA. Figure 7 is a representation of how PCR works.

During the first step in PCR, denaturation, the thermocycler heats the DNA template to 90-100º C in order to separate its two strands. In the next step, annealing, the thermocycler cools the DNA just enough to allow short, single strands of DNA called primers to bind to complimentary spots on the template strands (annealing temperatures vary by primer). These flank the area in the template that is to be replicated. Then, in a step called extension, the thermocycler heats the DNA to about 72º C, activating an enzyme called DNA polymerase to construct new complimentary DNA strands to match the separated single strands, starting at where the primers annealed and extending toward the 5' ends of the DNA strands [5]. These newly formed double-stranded DNA molecules are heated in another denaturation step, and the cycle is repeated. Typically a PCR protocol will contain about 30 such cycles. After completion of the last cycle, the DNA is kept at 4º C until needed.

Completed PCR reactions contain more than just the desired product DNA, however; they contain leftover raw nucleic acid units for generating new DNA strands, DNA polymerase enzyme, salt, the template DNA, and leftover primers. Because of this, PCR products need to be purified. One way of doing this, which is used in this work, is alcohol precipitation. In this method, the PCR product solution is mixed with ammonium acetate and isopropyl alcohol, cooled, centrifuged (supernatant is discarded), mixed with cold ethanol, and centrifuged (supernatant is again discarded). This cycle is performed twice, and the final pellet of DNA is air-dried and resuspended in ddH_2O. The resulting DNA solution's concentration is measured via spectrophotometer using an absorbance assay.

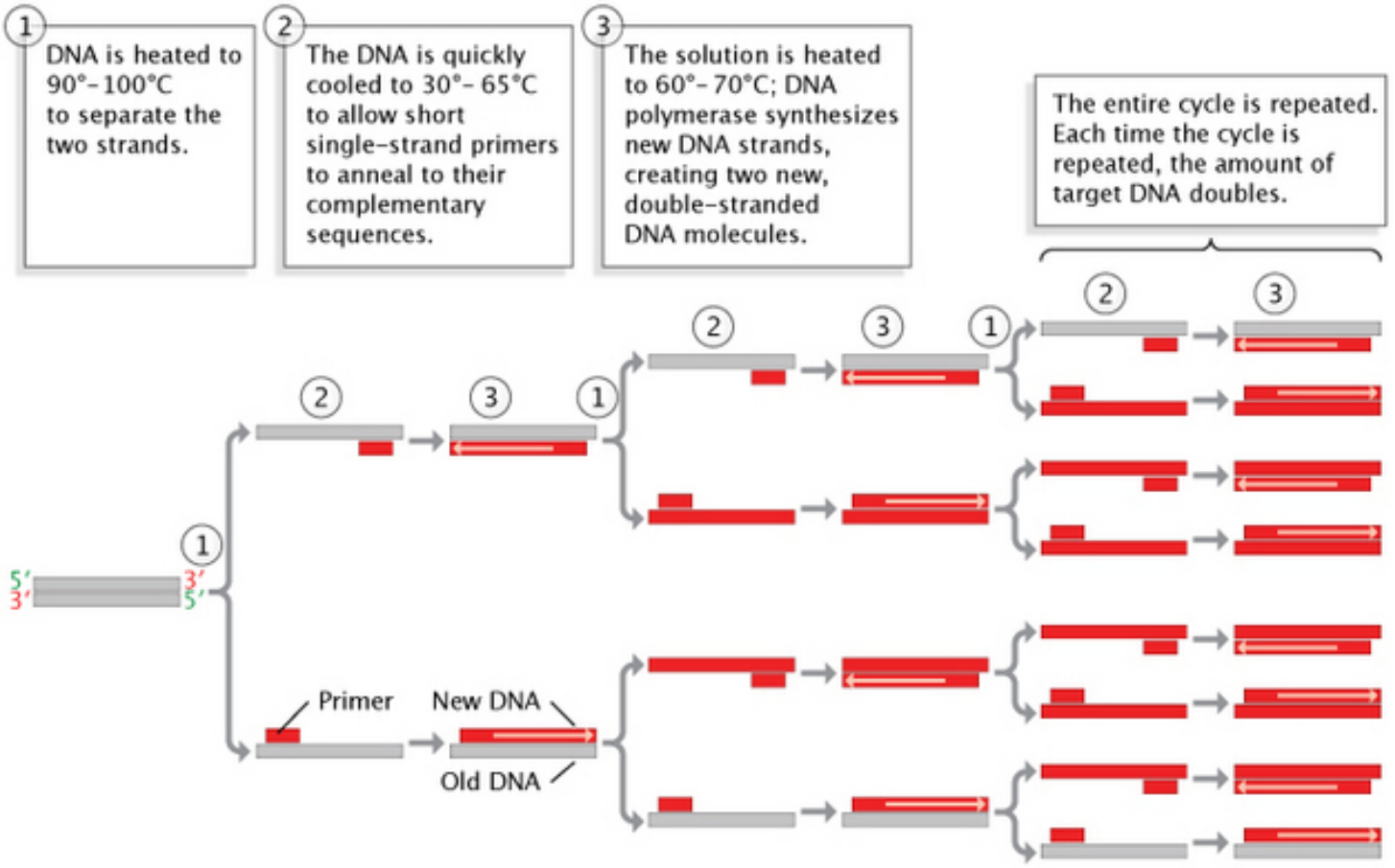

Figure 7: The PCR Process. Template DNA (shown leftmost in the figure), together with raw nucleotides, DNA polymerase enzyme, and primers, is subjected to a heating/cooling cycle in a thermocycler, each temperature of which triggers a step in the synthesis of more DNA. By the end of a typical PCR protocol, the copies made from the template number in the billions. Figure reproduced from [5].

2.2.4 The Challenge of LET Preservation

Unlike plasmids, LETs are susceptible to degradation by both endonucleases and exonucleases, since they are not circularized and have free ends. The combined assault of endonucleases and exonucleases degrades LETs more quickly than plasmids, and so protein production is less with LETs than with plasmids.

If LETs could last longer in CFPS without being degraded, then the protein yield of LET-based CFPS could be increased. There are a number of ways in which LETs could be preserved in CFPS, including inhibition of nucleases, removal of nucleases, and addition of sacrificial DNA. The first two methods, while they would be effective, are difficult and costly in comparison to the third, which is the method that this work explores. Sacrificial DNA preserves an LET in much the same way that a sacrificial anode preserves a valuable metal: it is degraded in place of the LET, either because of its relative availability, its ease of degradation, or both. Further details about sacrificial DNA are presented in Chapter 4.

3 PRODUCTION AND EVALUATION OF AARS FOR INCORPORATION OF UNNATURAL AMINO ACIDS

3.1 Introduction

Unnatural amino acid incorporation involves the same kind of cellular machinery as that used in natural amino acid incorporation—namely, a codon in the mRNA which signals for insertion of the unnatural amino acid; a tRNA bearing the anticodon matching the unnatural amino acid signal codon; and an aminoacyl-tRNA synthetase that attaches the unnatural amino acid to the tRNA. Each amino acid has a set of these three components dedicated uniquely to it. In this way, cells are able to incorporate amino acids with extremely high specificity. In order for amino acids to be incorporated in the locations specified by the mRNA, there must be no cross-talk among cellular protein synthesis machinery. There are three conceivable pairings within which cross-talk can occur: (1) tRNA/aaRS; (2) aaRS/amino acid; and (3) tRNA/codon. Corresponding examples of each kind of cross-talk follow: (1) If an alanine aaRS binds a leucine tRNA, then the leucine tRNA will be charged with the amino acid alanine, and so it will insert alanine where leucine should have been; (2) If a leucine aaRS binds to methionine, then leucine tRNA will be charged with methionine and will be inserted where leucine should have been; (3) If a leucine codon pairs with tRNA other than leucine tRNA, then amino acids other than leucine could be inserted where leucine should have gone. See Figure 8 below for a visualization of cross-talk.

tRNA/aaRS pairs which avoid all three kinds of cross-talk are called orthogonal. They insert amino acids only in those places where their codon signals them to be. One orthogonal tRNA/aaRS pair for unnatural amino acids which has already been successfully used in *E. coli* cell-free protein synthesis was taken from the archaeal species *Methanococcus jannaschii* [17]. The cellular machinery of archaea is sufficiently different from bacterial machinery for it not to experience cross-talk with the bacterial machinery. Other orthogonal pairs have also been engineered [17, 19-25]. However, to this point, work in this area has focused on incorporating only one type of unnatural amino acid per protein. The first goal of this work was to design a system capable of inserting not only one, but several different types of unnatural amino acid into a protein. Together with Mark Smith, the author identified a number of tRNA/aaRS pairs which seemed promising candidates for incorporation of different unnatural amino acids (Table 1).

Table 1: tRNA/aaRS Orthogonal Pair Selections for Incorporation of Different Unnatural Amino Acids.

tRNA/aaRS Pair	Organism of Origin	Reason Chosen	Unnatural Amino Acid(s)	Reference
Pyrrolysine (Pyl)	*M. barkeri*	Pyl machinery naturally absent from *E. coli*	N6-[(2-propynyloxy)carbonyl]-L-lysine	[19]
Phenylalanine (Phe)	*S. cerevisiae*	*S. cerevisiae* different domain than *M. barkeri*, *E. coli* and used orthogonally in *E. coli*	3-(1-naphthyl)-alanine, 1-methyl-tryptophan, 3-benzothienyl-alanine, and 6-methyl-tryptophan	[20]
Lysine (Lys)	*P. horikoshii*	*P. horikoshii* different domain than *S. cerevisiae* and *E. coli*	L-homoglutamine (Lys) and none (Lys): merely outcompetes release factor	[21] (Lys) [22] (Lys)
Tyrosine (Tyr)	*M. jannaschii*	Used orthogonally in *E. coli*	*p*-propargyloxyphenylalanine, *p*-azidophenylalanine	[17]

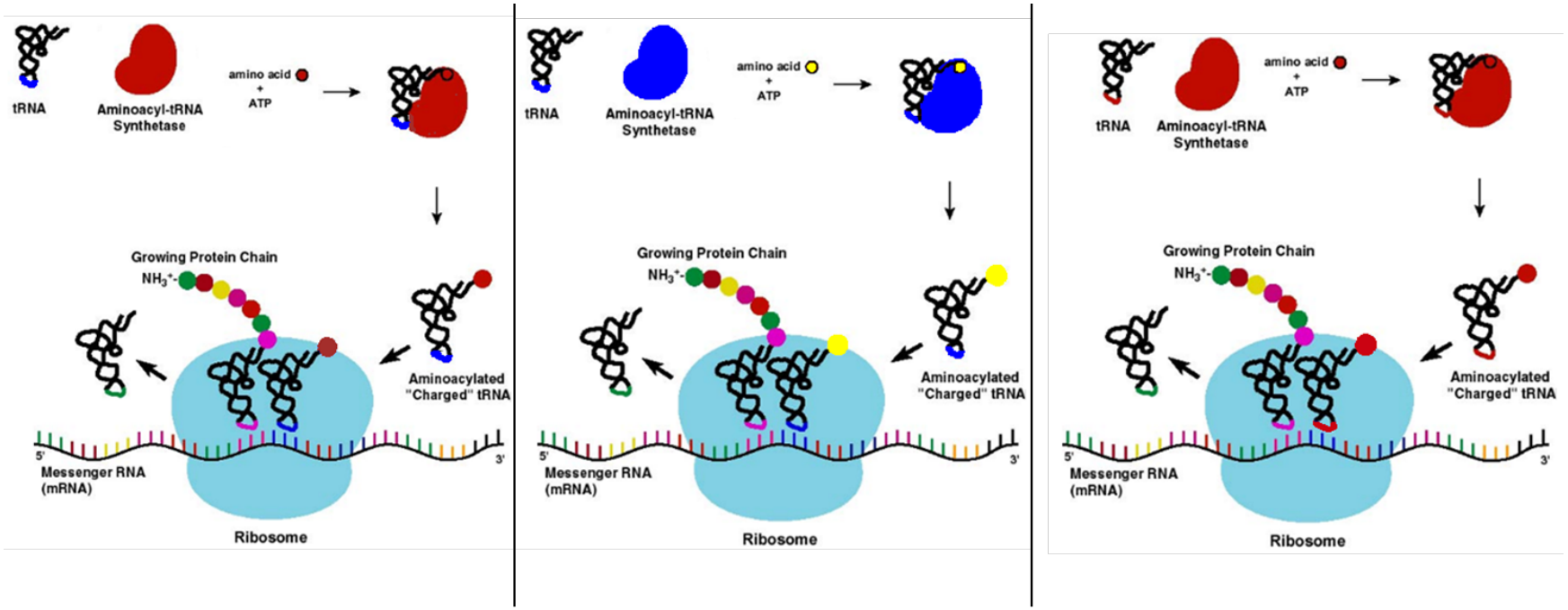

Figure 8: Three Kinds of Cross-talk Among Cellular Machinery. The numbers (1), (2), and (3) correspond to cross-talk types (1), (2), and (3), respectively, as described in the text. Color coding: Red—alanine; blue—methionine; yellow—leucine. Figure modified from [2].

Incorporation of different unnatural amino acids means that each of the unnatural tRNA/aaRS pairs must not cross-talk with natural pairs or amino acids.

3.2 Materials and Methods

3.2.1 Design of Orthogonal tRNA/aaRS Combinations

Table 1 summarizes the choices and criteria for choosing the four tRNA/aaRS pairs. Using the literature cited in Table 1 and online gene databases such as GenBank, four tRNA/aaRS pairs were selected for their potential to act orthogonally in an *E. coli*-derived cell extract environment to incorporate unnatural amino acids into protein products of cell-free protein synthesis. They were also selected with the functionality of their respective unnatural amino acids in view. For example, the unnatural amino acid N6-[(2-propynyloxy)carbonyl]-L-lysine is click-chemistry compatible [26].

The next step was to actually make these aaRS. Since aaRS are proteins, they could be made by *E. coli in vivo*. To do this, however, plasmids coding for the aaRS were required. Plasmids, which are circularized DNA, can be divided into various parts or regions, each of which has its own function. For example, the gene of interest is the region of the plasmid which codes for the desired protein product. The antibiotic resistance region enables the bacterium to survive antibiotic. The origin of replication (Ori) directs the cells to replicate the plasmid and pass it on to progeny. The *lac* operator region prevents the regions downstream of it from being used, until the operator is unlocked by the presence of isopropyl β-D-1-thiogalactopyranoside (IPTG). Plasmids are commonly referred to as consisting of two major parts: the gene of interest and the plasmid vector (everything other than the gene of interest). Figure 9 is a map of the plasmid containing the gene for *M. barkeri* aaRS.

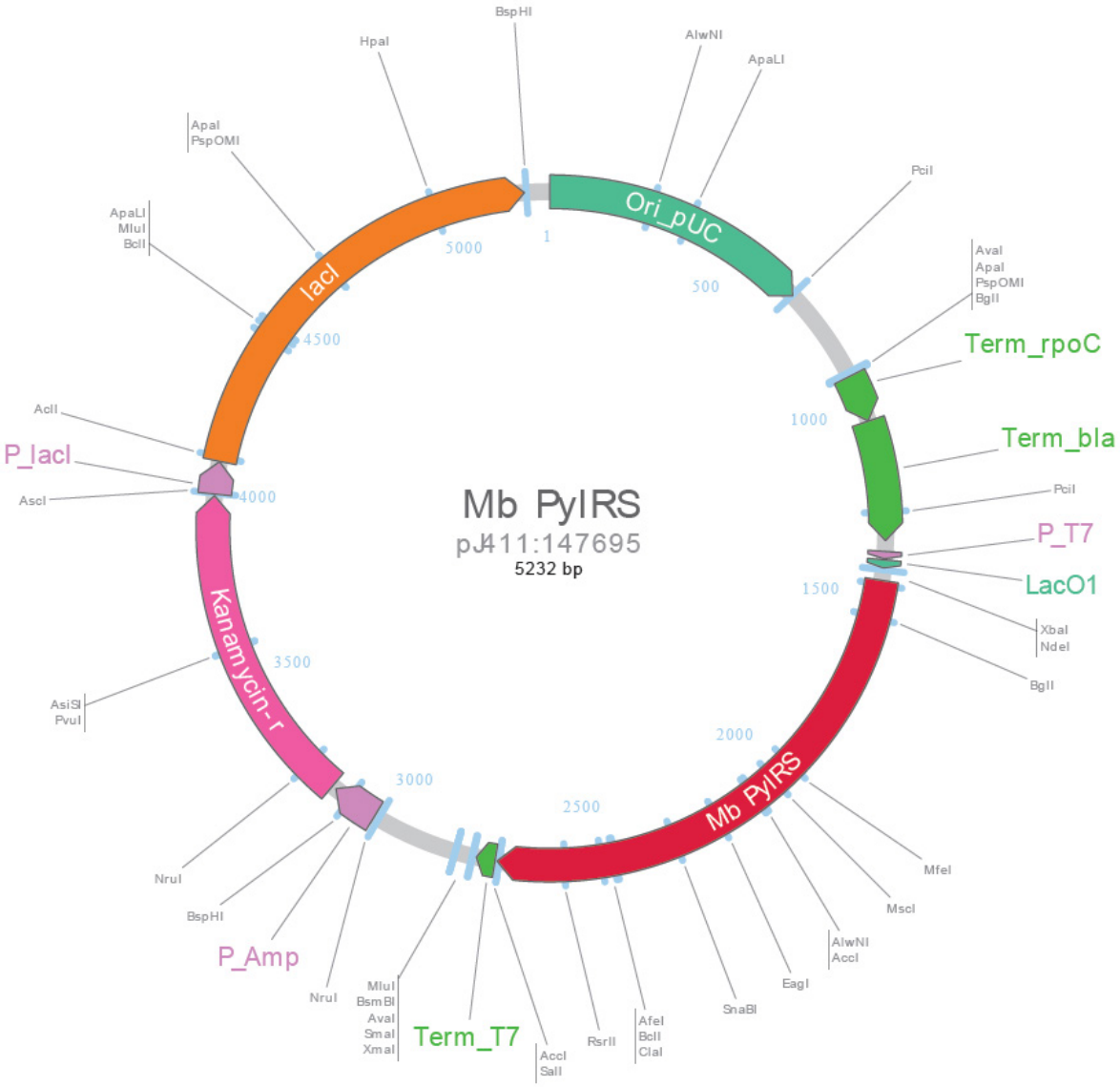

Figure 9: Plasmid map of the MbRS plasmid. Obtained from DNA 2.0, the company which prepared and shipped the plasmid [3].

Using information obtained from GenBank and the literature, gene sequences for the *M. barkeri*, *S. cerevisiae*, and *P. horikoshii* aaRS (abbreviated hereafter MbRS, ScRS, and PhRS, respectively) were identified (see the appendix for these sequences in full). Literature review, aaRS/tRNA pair selection, and plasmid design were all done together with Mark Smith, a Bundy Lab doctoral candidate. These gene sequences were submitted to a third-party company, DNA 2.0, to be inserted into plasmid vectors and shipped to the Bundy laboratory.

3.2.2 Protein Preparation

The protocol used in this work for preparing stocks of the four aaRS proteins begins with transformation (insertion into bacteria) of the appropriate plasmids into BL21 (DE3) Star™ cells (Life Technologies). These cells had been genetically modified to produce a truncated form of RNase E, which results in less mRNA degradation and thus enhanced protein production [27], and they were made chemically competent in the Bundy lab. The plasmids were added to a cell suspension, which was then heat treated for one minute, allowing the plasmids to enter the cells.

After transformation of the plasmids, protein production follows a procedure similar to that described in 1.1.3 above for preparing cell-free extract. While cell-free extract preparation uses cells which were not transformed with plasmid, protein production uses cells which have been transformed with a plasmid that codes for the protein of interest. After inoculation of a plate with cells from the appropriate glycerol stock, bacterial colonies grow in the plate and are used to inoculate a test tube containing 5 mL of autoclaved LB growth medium and supplemented with 100 ug/mL kanamycin. The kanamycin is added to ensure that only those bacteria that contain the aaRS plasmid (and, thus, the kanamycin resistance) survive. Alternatively, the 5 mL of medium may be inoculated directly from the glycerol stock. The resulting 5 mL culture is incubated

overnight at 37º C with 280 rpm shaking in an incubator shaker, after which the contents are transferred to a flask containing 100 mL of autoclaved LB growth medium supplemented with 100 ug/mL kanamycin. This flask is incubated at 37º C with 280 rpm shaking until cell growth reaches the log phase (OD of ~2), and then its contents are transferred to a container containing 1 L of autoclaved LB growth medium supplemented with 100 ug/mL kanamycin. The growth of bacteria in the 1 L container is monitored via optical density (OD) readings using 600 nm wavelength light on samples taken at least hourly. When the OD is 0.4-0.6, protein production is induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), which unlocks the *lac* operator mentioned in section 3.2.1. When the growth begins to exit the exponential phase, the culture is removed from incubation. Next, the culture is centrifuged at 11,270 RCF in a supercentrifuge for 15 minutes to separate out the cells from the growth medium.

After the medium is removed, the cells are washed with Buffer A (10mM Tris base, 14 mM anhydrous magnesium acetate, 60 mM potassium glutamate, 1 mM dithiothreitol; pH adjusted to 8.2 using acetic acid, sterilized in an autoclave), resuspended in the same kind of buffer, and lysed with three passes at 21,000 psi through a French press homogenizer. The lysed cells are then centrifuged at 12,000 RCF for 10 minutes to separate out the cell wall and other unwanted cell parts. The supernatant, which contains the protein of interest, is then removed, leaving the unwanted cell debris to be discarded. Next, the supernatant is treated with DNase in order to reduce the viscosity of the supernatant. After this, the supernatant is run through a nickel affinity column (GE Medical, Uppsala, Sweden) which binds the hexahistidine-tagged product proteins and holds them until elution. The eluent fractions are run on a 10% polyacrylamide gel to determine purity and molecular weight, and fractions with the desired purity and correct molecular weight are pooled. The protein concentration of the pooled fractions is determined using a BioRad™ DC

Assay, and if the solution needs to be more concentrated, then it is dialyzed against sucrose or poly-(ethylene glycol) using a cellulose membrane with a MW cutoff of 6-8 kD. The concentration post-dialysis is measured via DC Assay.

3.2.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

10% polyacrylamide gels (Life Technologies) were used to verify the products of protein preparations. Samples of the proteins were prepared according to the following protocol, which is based on the instructions included with the Life Technologies gels. 5 uL of protein solution was mixed in a microcentrifuge tube with 2.5 uL of NuPAGE® LDS sample buffer (4x), 1 uL of 500 mM dithiothreitol (DTT) in ddH_2O, and 1.5 uL of ddH_2O. The resulting sample solution was incubated at 70º C for ten minutes in a water bath. After incubation, 7 uL of the sample solution was mixed with 2 uL of purple loading dye, and 7 uL of the resulting mixture was loaded into a well of the polyacrylamide gel. The gel was immersed in 200 mL of 1x MES buffer with 500 uL of 1M sodium disulfite in ddH_2O added. The gel was connected to an electrophoresis power unit, and a constant voltage of 200 V was applied for one hour or until the dye line reached the bottom of the gel. The gel was then removed, placed on a visible-light illuminator, and imaged using a digital camera.

3.2.4 Aminoacylation Assay

In order to monitor the aaRS's attachment of amino acids to tRNA (a reaction known as aminoacylation), the author searched out and attempted to replicate an aminoacylation assay from the literature [28] in an effort to measure aminoacylation in a cell-free protein synthesis environment. This assay would indicate whether the aaRSs work correctly.

The strategy behind this aminoacylation assay is to split a byproduct of the aminoacylation reaction, pyrophosphate, into two molecules of inorganic phosphate, which can be complexed with a dye molecule to produce a measurable color change (Equation 1, Equation 2). The degree of color change implies the amount of inorganic phosphate produced, from which the number of tRNA molecules aminoacylated can be calculated.

Equation 1: An Example Aminoacylation Reaction. Here, the amino acid glutamine (abbreviated Gln) is attached to its corresponding tRNA (abbreviated $tRNA^{Gln}$), as facilitated by its corresponding aaRS (glutaminyl-tRNA synthetase), using ATP as an energy source. AMP and pyrophosphate (Pp_i), both parts of the spent ATP molecule, are byproducts of the reaction [27].

$$Gln + tRNA^{Gln} + ATP \xrightarrow{\textit{glutaminyl-tRNA synthetase}} Gln - tRNA^{Gln} + AMP + Pp_i$$

Equation 2: The Splitting of Pyrophosphate into Two Inorganic Phosphate Molecules, as Facilitated by Pyrophosphatase [28].

$$Pp_i \xrightarrow{\textit{pyrophosphatase}} 2\ P_i$$

Since the point of this assay was to measure the amount of aminoacylation carried out by an aaRS, the aaRS is produced as described in 3.2.2 above, and the resulting purified solution becomes the sample for the assay to test. Once the aaRS production was complete, it was incorporated into the aminoacylation assay. The reagents needed for the assay are listed in Table 2 below. These reagents were combined in a microcentrifuge tube, with the aaRS being added last, since the aminoacylation reaction will not begin without it. The final mix was briefly vortexed and then aliquotted into a clear, flat-bottomed 96-well plate. The plate was sealed with tape and incubated for 30 minutes at 37º C, during which time the aminoacylation reaction proceeded. After incubation, the reaction was quenched by the addition of twice the sample

volume of malachite green solution to each well. This solution contained the dye molecule that causes a color change (to green) in the presence of inorganic phosphate.

Table 2: Reagents Needed for the Aminoacylation Assay.

Reagent	Final Concentration in Reaction
ddH_2O	N/A; used as diluent
Aminoacylation Buffer (5x stock)	1x
DTT	1 mM
ATP	200 uM
aaRS of Interest	1 uM
Amino Acid of Interest	1 mM
tRNA of Interest	10 uM
Pyrophosphatase	2.5 ug/mL

The plate is then shaken at room temperature for 30 minutes to allow the color change to develop. Directly following this development period, the plate is read for absorbance at 620 nm. Using a standard curve made with known amounts of phosphate salt as the inorganic phosphate source, the absorbance values may be converted into concentrations of phosphate ions, which in turn may be divided by two to yield concentrations of aminoacylated tRNA.

3.3 Results and Discussion

Polyacrylamide gels on the eluent from nickel affinity columns indicated the presence of PhRS, ScRS, and MjRS in their respective eluents (Figure 10, Figure 11). Polyacrylamide gels on the eluent from nickel affinity columns did not indicate the presence of MbRS (Figure 12). Neither did gels on the supernatant from centrifuged lysate of cells containing the MbRS plasmid (Figure 13). However, gels on the lysate before centrifugation indicated that MbRS was present and overexpressed (Figure 14). This led to the conclusion that the MbRS protein was aggregating inside the bacteria during growth and then was centrifuged down into a pellet with the cell debris. MbRS was observed to be not very soluble in the *E. coli* cellular environment.

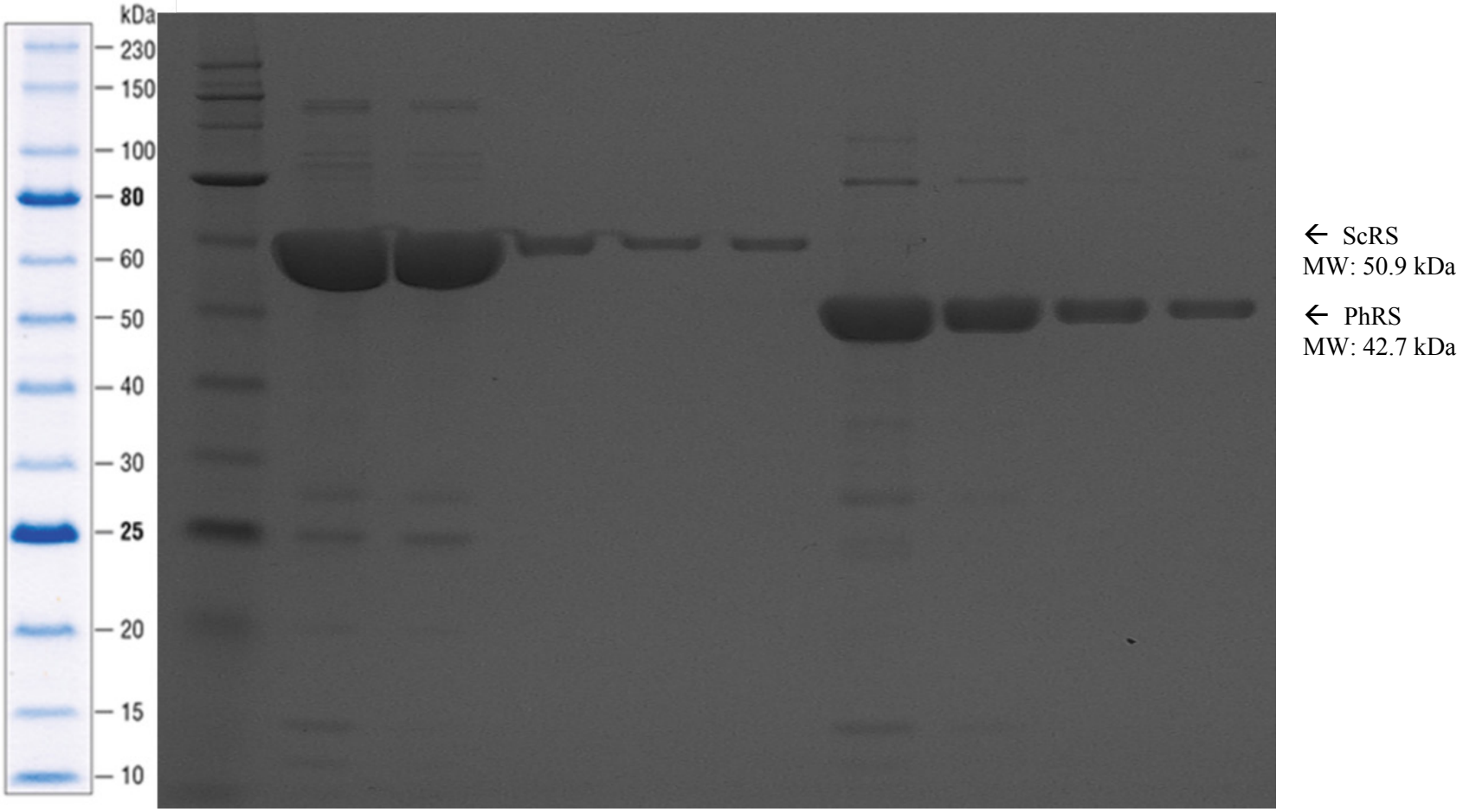

Figure 10: Polyacrylamide Gel on Purified Solutions of PhRS and ScRS. Ladder legend on the left obtained from [1].Lane contents, left to right: (1) ladder; (2)-(6) increasing dilutions of purified ScRS; (7)-(10) increasing dilutions of PhRS. Expected location of protein bands shown by arrows on the right, based on proteins' molecular weight.

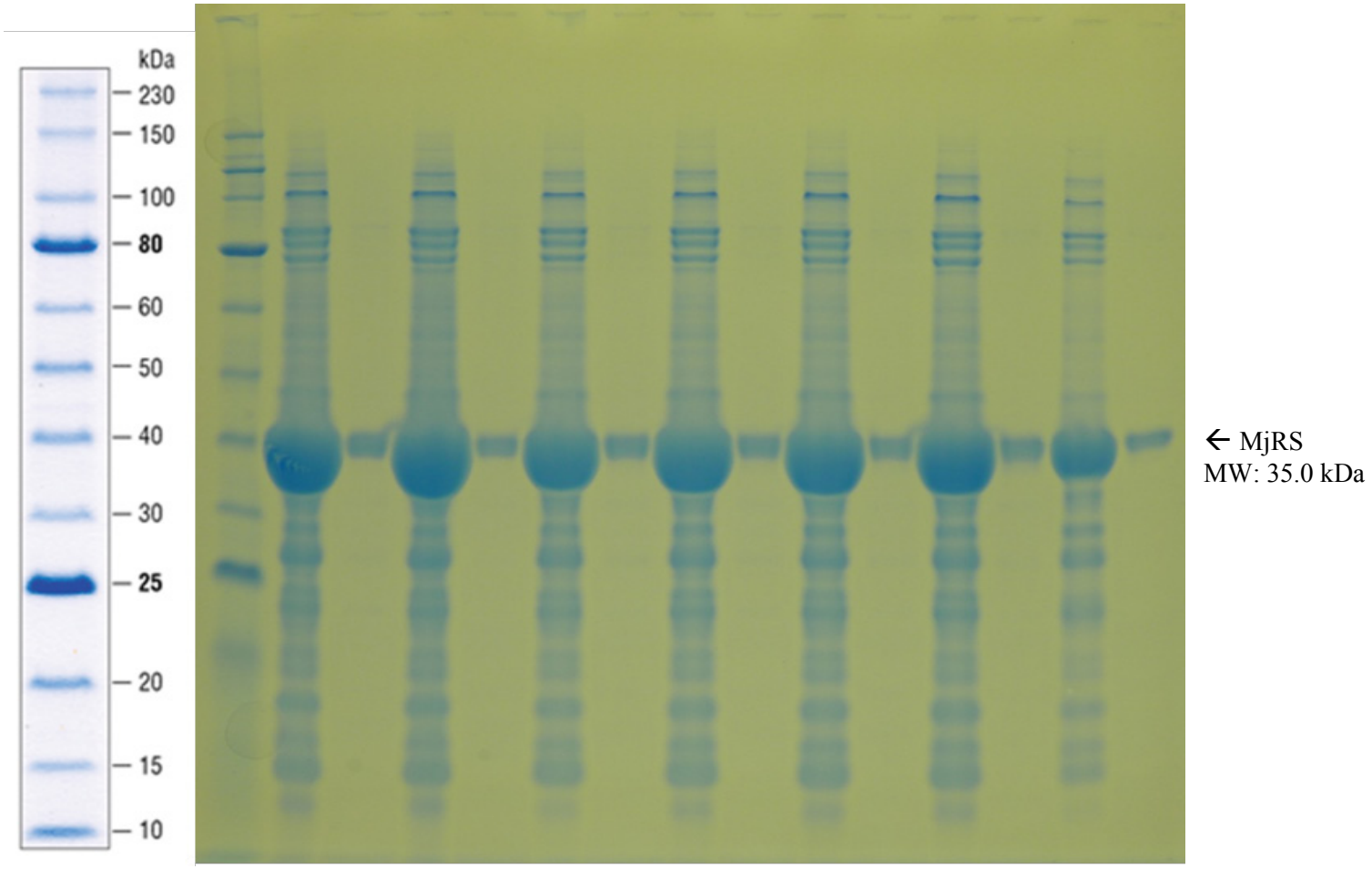

Figure 11: Polyacrylamide Gel on Purified Solutions of MjRS. Lane contents, left to right: (1) protein ladder; (2)-(15) alternating sixth and twelfth eluent fractions from seven nickel-affinity chromatography cycles. Ladder legend on the left obtained from [1]. Expected location of protein bands shown by arrows on the right, based on protein's molecular weight.

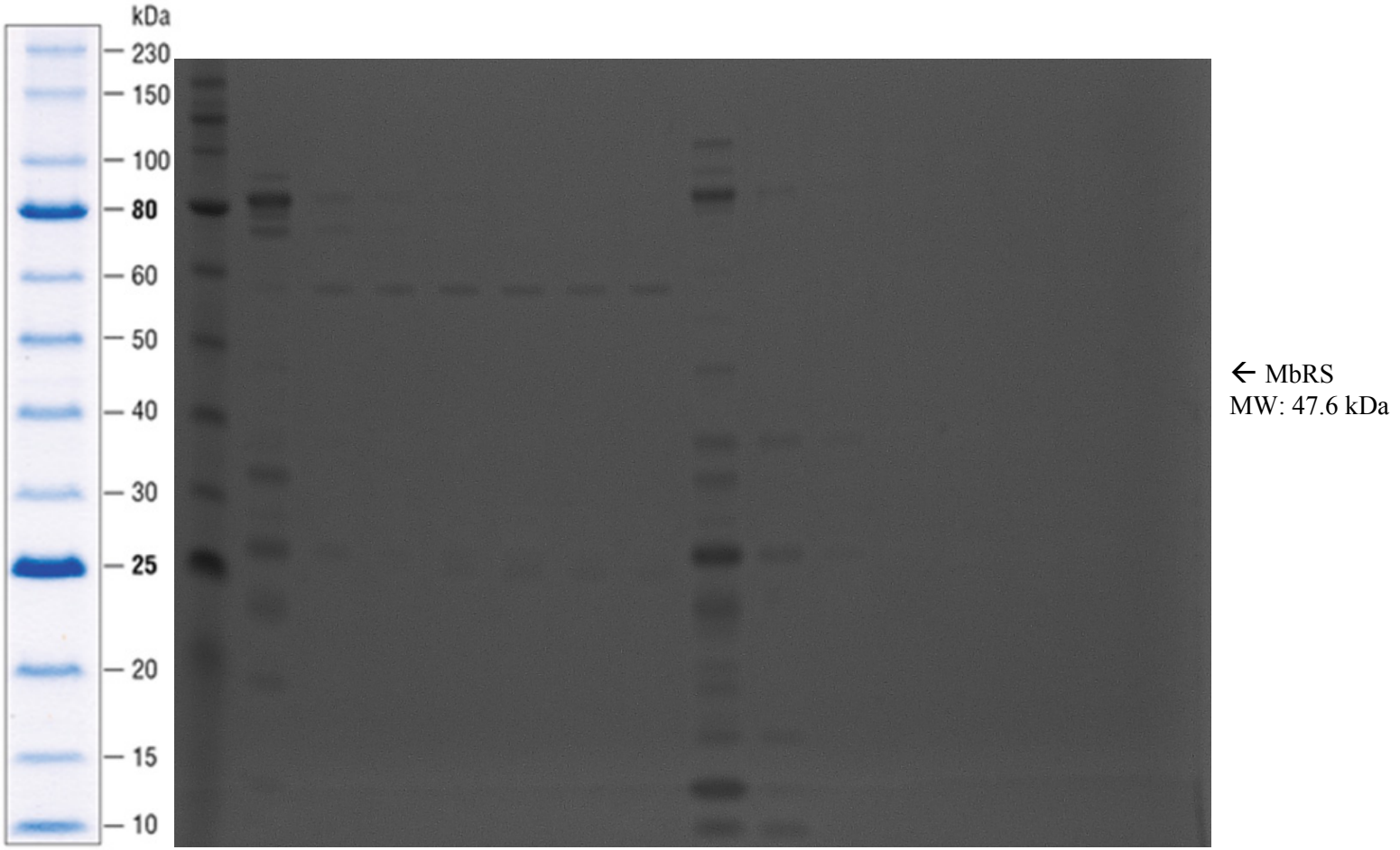

Figure 12: Polyacrylamide Gel on Purified Cell Lysate Supernatant from Cells Containing the MbRS Plasmid. Lane contents, left to right: (1) ladder; (2)-(8) increasing dilutions of purified supernatant from *induced* cells; (9)-(15) increasing dilutions of purified supernatant from *non-induced* cells. Expected location of protein band indicated by arrow at right, based on protein's molecular weight. No MbRS was observed.

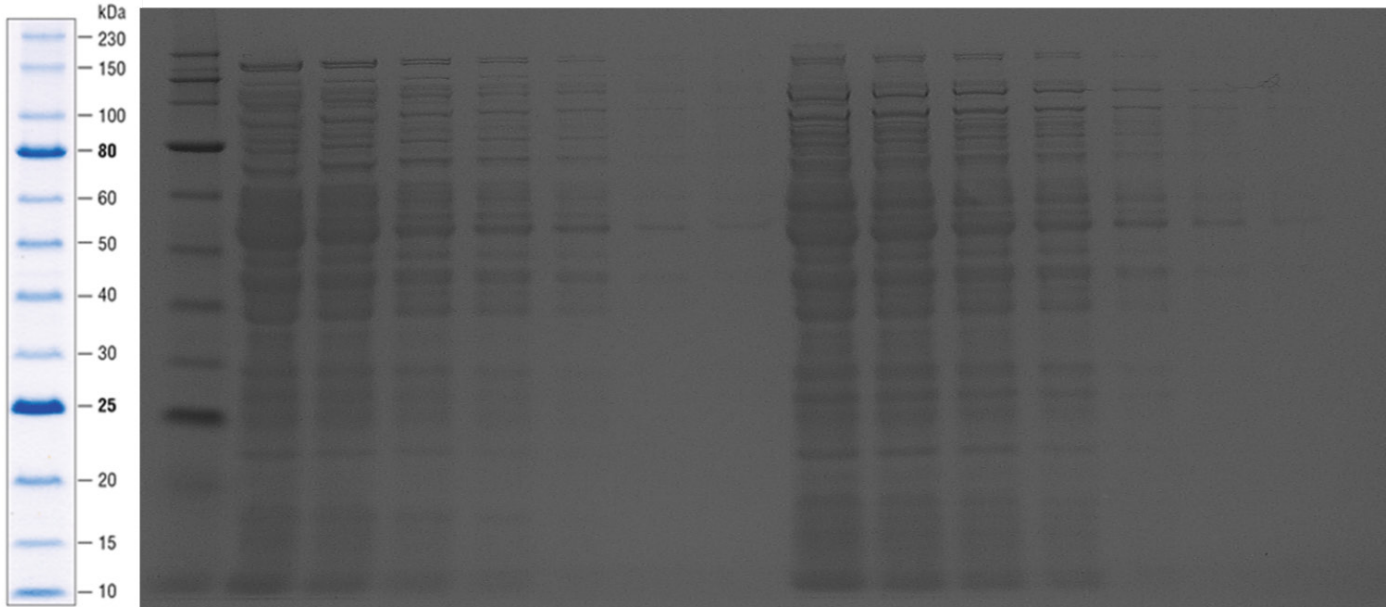

Figure 13: Polyacrylamide Gel on the Supernatant of Centrifuged Lysates of Cells Containing the MbRS Plasmid. Lane contents: (1) ladder; (2)-(8) increasing dilutions of lysate supernatant from *non-induced* cells; (9)-(15) increasing dilutions of lysate supernatant from *induced* cells. Supernatant from non-induced cells was included as a negative control. Expected location of protein band indicated by arrow at right, based on protein's molecular weight. No MbRS was observed.

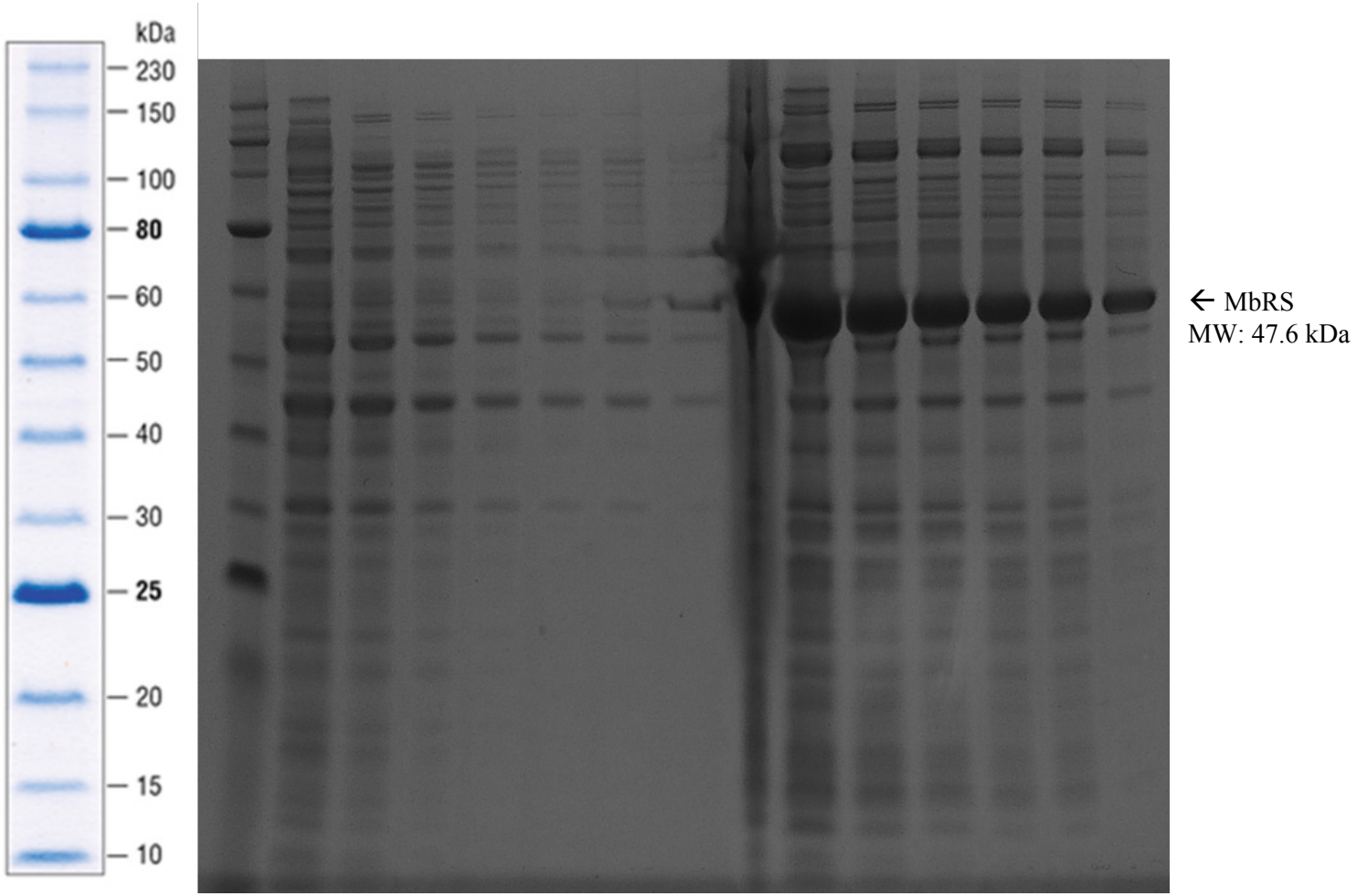

Figure 14: Polyacrylamide Gel on Noncentrifuged Lysates of Cells Containing the MbRS Plasmid. Lane contents, left to right: (1) ladder; (2)-(8) increasing dilutions of *noninduced* cells; (9)-(15) increasing dilutions of *induced* cells. Lysates from noninduced cells were included as a negative control. Expected location of protein bands is indicated by the arrow at right, based on protein's molecular weight. This time, an overexpressed protein band near MbRS's expected location was observed. Viscosity of the sample may account for the difference in actual and expected locations.

The aminoacylation assay did not turn out as expected; the formation of a green flaky precipitate skewed the absorbance readings, resulting in unreasonable data. Figure 15 below is an image of a 96-well plate containing such precipitate after an aminoacylation assay had been carried out. In order to counteract the precipitation, an 8% solution in ddH_2O of poly-(vinyl alcohol), also called PVA, was added to the assay, and samples were diluted 4x. This succeeded in greatly reducing or even preventing precipitation in the aminoacylation mixtures (Figure 16). The first four rows of wells in Figure 15 and the first two rows of wells in Figure 16 were both PO_4^- standard curves covering the range of 0-2,000 pmol PO_4^-. Figure 17 and Figure 18 are the data points and regressed standard curves for the assay performed with the plates in Figure 15 and Figure 16, respectively.

Further experiments established that PVA can prevent precipitate not only in standard curve wells, but also in reaction wells, though this required a 16% wt./vol. solution of PVA in ddH_2O to be mixed in 1:5 volumetric ratio with malachite green solution before the malachite solution was added to the plate. However, when precipitate was successfully eliminated, the phosphate reading for the negative control (a reaction lacking tRNA) was equivalent to that of the full reaction, at least within the ability of this assay to distinguish. This may be because PVA somehow prevents one or more components of the aminoacylation reaction from working properly. Thus, in the end, obtaining reliable results from this assay was prevented, if not by the formation of precipitate that could skew absorbance readings, then perhaps by the very agent that solved the precipitate problem.

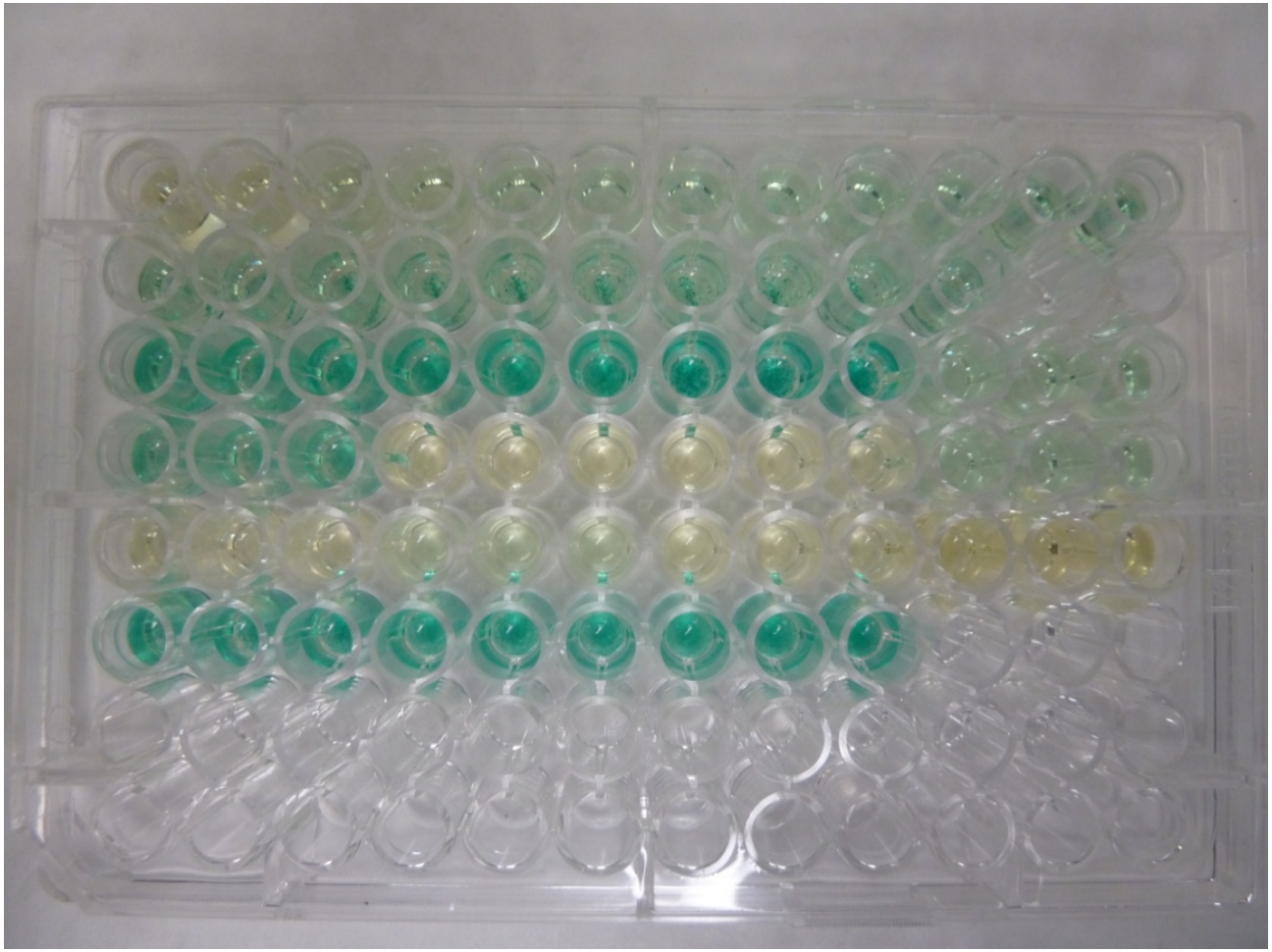

Figure 15: Image of a Clear, Flat-bottomed 96-well Plate Used for an Aminoacylation Assay. The image was taken just after the plate was read for absorbance at 620 nm. Note the green precipitate, which is especially visible in the second column of wells from the left. Image taken about 25 minutes after plate was read.

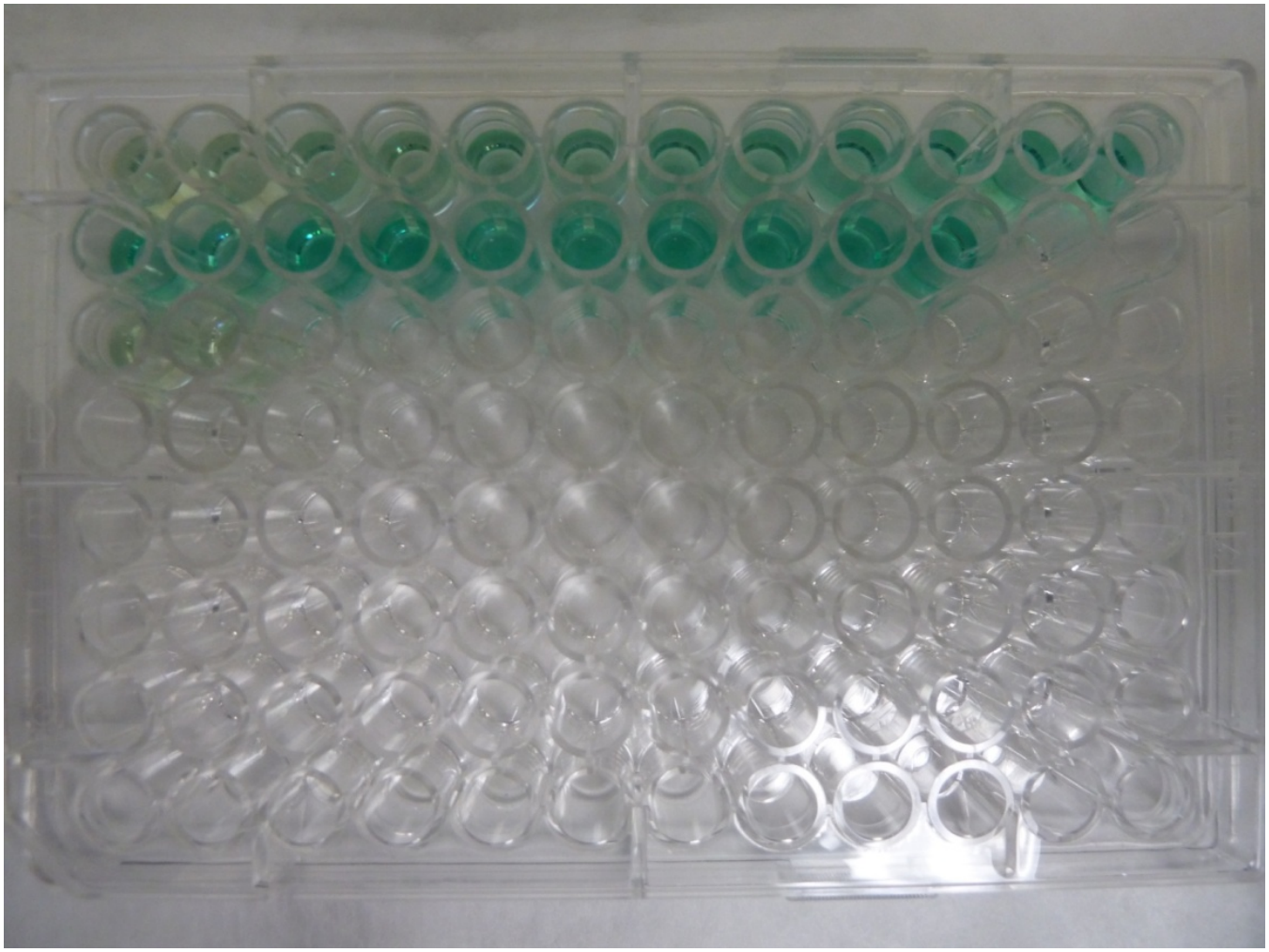

Figure 16: Image of a Clear, Flat-bottomed 96-well Plate Used for an Aminoacylation Assay, with 50 uL of 8% PVA Added to Each Well.The image was taken about two hours, thirty minutes after the plate was read. Significantly, no precipitate was visible even after that much time. Compare Figure 15.

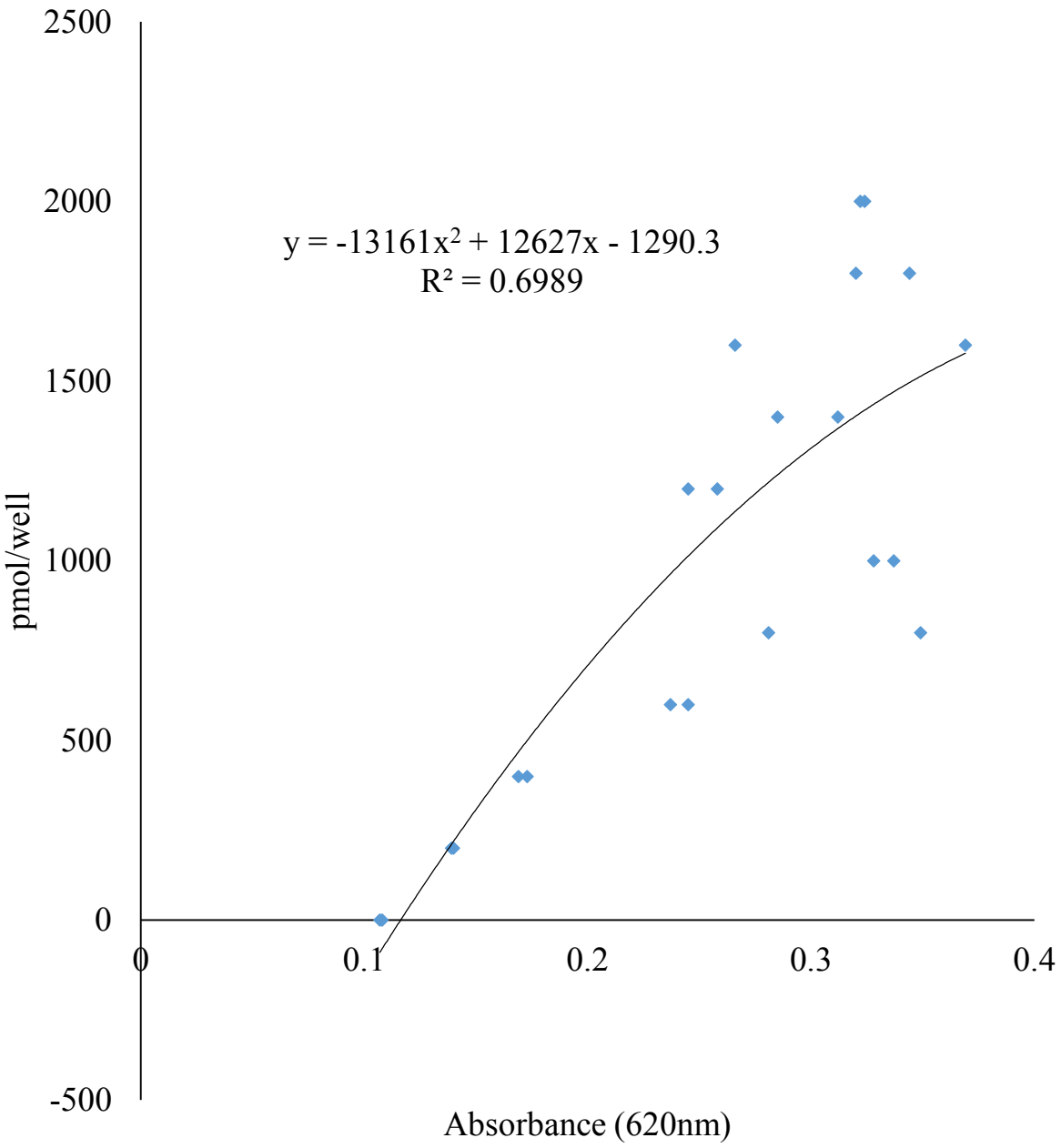

Figure 17: Standard Curve Data Points and Second-order Polynomial Regression from the Plate Shown in Figure 15. Note the large scatter of the rightmost points about the regression, likely caused by aggregation during the assay.

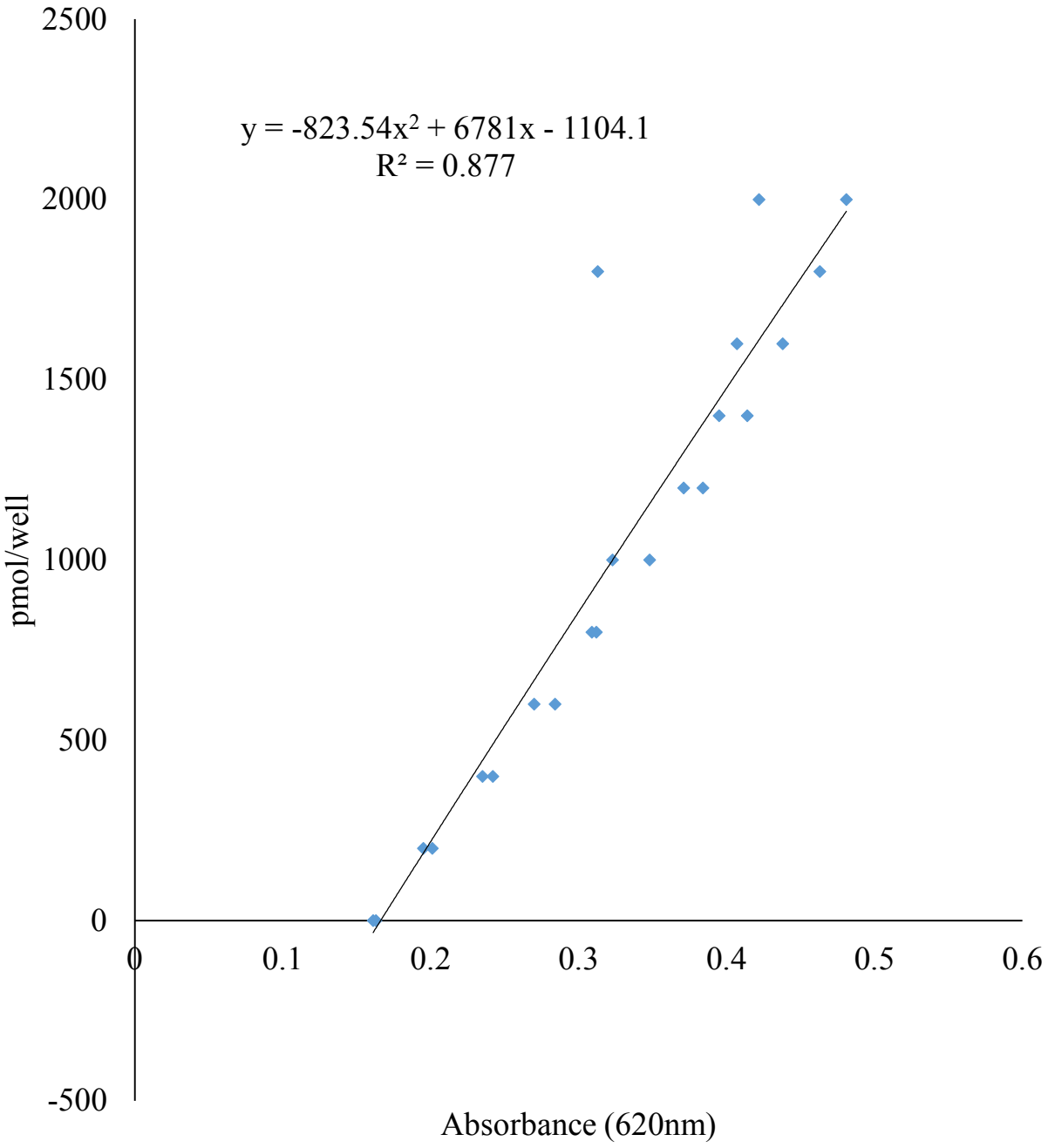

Figure 18: Standard Curve Data Points and Second-order Polynomial Regression from the Plate Shown in Figure 16. Note the tighter clustering of the points about the regression as compared with that of the points in Figure 17.

3.4 **Conclusion**

A set of four independent aaRS/tRNA pairs for incorporation of different unnatural amino acids was designed, and all four aaRSs were successfully produced *in vivo* as verified by polyacrylamide gel analysis. MbRS appears not to be very soluble in the *E. coli* cellular environment, since it was found with the cellular debris and not in the supernatant following centrifugation of cell lysate. This problem will need to be addressed before MbRS can be used in cell-free protein synthesis. The aminoacylation assay tested in this work does not seem to be reliable, at least for this particular application, since the negative control and the full reaction had nearly identical results. The author attributes this to PVA's possible inactivation of one or more components involved in the aminoacylation reaction.

4 USING SACRIFICIAL DNA TO IMPROVE LET-BASED CFPS PROTEIN YIELDS

4.1 Introduction

Cell-free protein synthesis uses the same method to make protein as do living cells, and so it is important to ensure that all needed biological components are present and are not degraded. For example, DNA, which contains the information that determines the sequence of the protein product, is susceptible to degradation by DNases, enzymes naturally present in *E. coli* (and in cell-free extract derived from *E. coli*). If DNA is degraded, then mRNA coding for the correct protein product will no longer be made, and if less correct mRNA reaches the ribosomes, then not as much of the desired protein product will be made. Accordingly, it is reasonable to expect that preventing the degradation of DNA would increase the yield of protein in a cell-free protein synthesis reaction. This work explored the possibility of accomplishing greater CFPS protein yields through the addition of extra DNA which would be degraded in place of the template DNA that codes for the product protein. This extra DNA is called, in this work, sacrificial DNA.

4.2 Materials and Methods

A summary of the procedure for a generic cell-free protein synthesis reaction follows. All reagents are kept on ice, allowing those that are frozen to thaw on ice. Table 3 below is a list of the needed reagents and their required final concentrations. Once thawed, these reagents are

combined in the concentrations specified in Table 3. The DNA template codes for sfGFP (superfolder green fluorescent protein) since this product is detectable by fluorescence reading. The cell-free extract is added last, since once this is added, the cell-free protein synthesis reaction can begin. The final mix is briefly vortexed and aliquotted into a black, flat-bottomed 96-well plate, with 20 uL mix per well. The plate is then sealed with tape and incubated at 37º C for three hours. After the incubation, 45 uL of ddH_2O is added to each well, and the plate is briefly shaken and then read for fluorescence (excitation: 485 nm, emission: 510 nm).

Table 3: Reagents Needed for Cell-free Protein Synthesis Reactions.

Reagent	Final Concentration in Reaction
ddH_2O	N/A; acts as diluent
$Mg(Glu)_2$ (up to 1000 mM stock; dilute if hard to pipette tiny vol.)	varies; see "$Mg(Glu)_2$ Optimization" below
DNA template (plasmid or LET)	12 nM if plasmid or 120 nM if LET
PANOx-SP (4x concentrated); see [13] for a list of components and concentrations for PANOx-SP	4x diluted
Extract	4x diluted

The fluorescence data from the cell-free protein synthesis reactions is converted to protein yield data using a standard curve which links known amounts of ^{14}C-labeled sfGFP (ug/mL, determined by scintillation count) to their corresponding fluorescence readings. The sfGFP was made in CFPS reactions, and the reaction product solution was measured without purification, just like it would be in any other CFPS reaction. In this way, any background fluorescence native to the CFPS product solution would be shared by the standard curve and subsequent CFPS reactions, thus allowing the standard curve to apply to those reactions' data. The standard curve used in this work is shown below as Figure 19.

With sacrificial DNA, the procedure changes only in that the sacrificial DNA is added to the mix at some point before the extract. Complimentary PCR primers were used as sacrificial DNA, since they were already available in the lab and contained neither T7 RNA polymerase

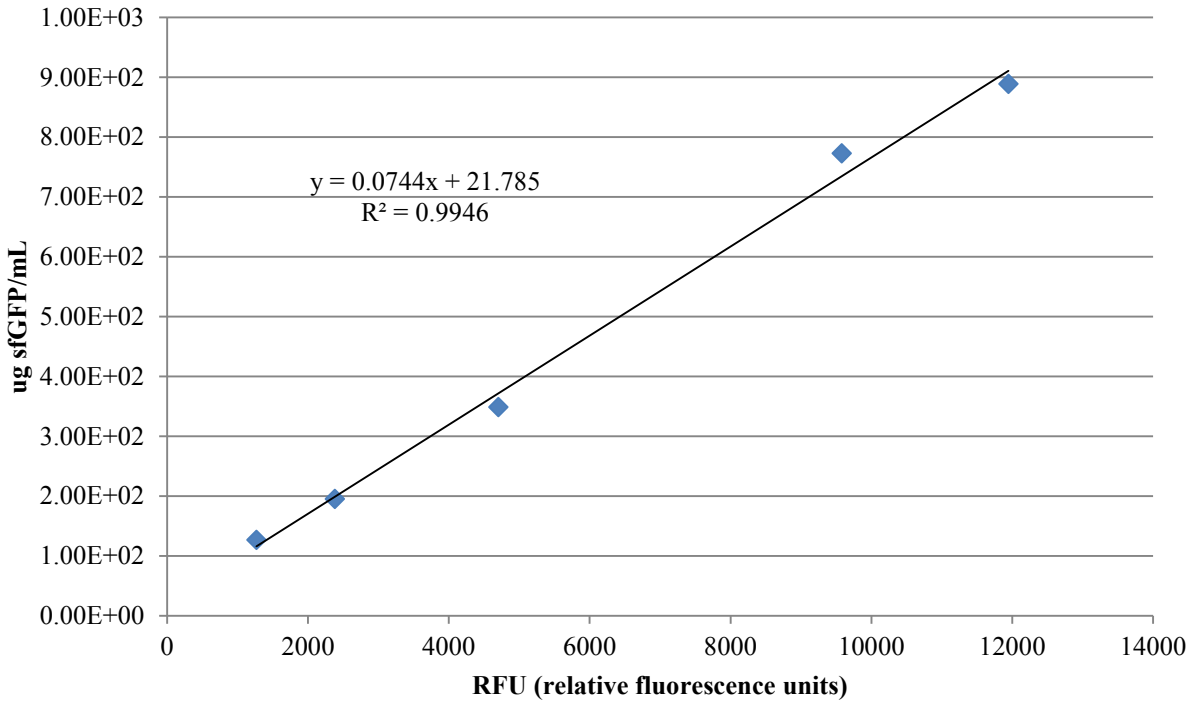

Figure 19: RFU-to-ug/mL Standard Curve. This curve correlates the concentration of sfGFP present in a solution to the fluorescence reading of that solution. sfGFP concentration was determined beforehand by scintillation count; the sfGFP contained ^{14}C-labeled amino acids in a known quantity, allowing calculation of protein mass from scintillation count. Each point on the graph represents the average result of a cell-free protein synthesis (CFPS) reaction performed in triplicate.

promoter sequence (which would allow transcription of the sacrificial DNA into RNA) nor ribosome binding site (which would allow the translation of protein from the above-mentioned RNA). These complimentary primers were about 50 base pairs in length and were originally used for QuikChange™ mutagenesis in PCR. Because these primers were single-stranded, they were hybridized (matched up to form double-stranded DNA) using a thermocycler before use as sacrificial DNA.

To hybridize the primers, 25 uL of the forward primer and 25 uL of the reverse primer were added to each of the desired number of PCR tubes, and these tubes were subjected to a temperature cycle like that shown in Table 4 below. Table 5, also below, contains information about all the sacrificial DNA primer pairs tested in this work.

Table 4: Temperature Cycle Used to Hybridize Complimentary PCR Primers for Use as Sacrificial DNA. The specific temperatures shown are for Pair D; for other pairs, the 95-degree step was the same regardless of the pair being hybridized, but the second step's temperature was set one or two degrees above the primers' melting temperature, and it descended by one degree in each of the four subsequent steps until, at the last step, the temperature was lowered to 4° C.

Temperature (degrees Celsius)	Duration (minutes:seconds)
95	2:30
68	1:00
67	1:00
66	1:00
65	1:00
64	1:00
4	(until tubes are removed)

Table 5: Properties of Sacrificial DNA Primer Pairs.# base pairs is the length in base pairs of the hybridized primer pair. Overall T_m is the melting temperature of the entire hybridized primer sequence. An end T_m is the melting temperature of the four codons (12 base pairs) at one end of the hybridized primer pair. Low T_m is the lesser of the two end T_ms, and high T_m is the greater of the two. Average T_m is the arithmetic average of the low T_m and the high T_m. G/C content is the percentage of bases in the primer pair that are either G or C (i.e. not A or T).

	Pair A	Pair B	Pair C	Pair D
# base pairs	43	58	34	53
Overall T_m (entire primer)	56.8	70.0	66.1	67.3
Low end T_m (4 codons)	26.5	36	36	20.3
HIgh end T_m (4 codons)	27	43.5	43.5	47.6
Average end T_m	26.75	39.75	39.75	33.95
G/C content	27.9%	48.3%	52.9%	43.4%

4.3 **Results and Discussion**

At first, the mass of sacrificial DNA added was equal to the mass of the template DNA (LET, in this work). It seemed reasonable, since protection of the template DNA was the object of adding the sacrificial DNA, that the amount of sacrificial DNA added should be proportional to the amount of template DNA present. A mass-based proportion was chosen rather than a moles-based proportion. An early experiment tested the effect of adding sacrificial DNA Pair A (see Figure 20 below). This experiment included a magnesium optimization, in case the added sacrificial DNA altered the amount of magnesium required for the reaction.

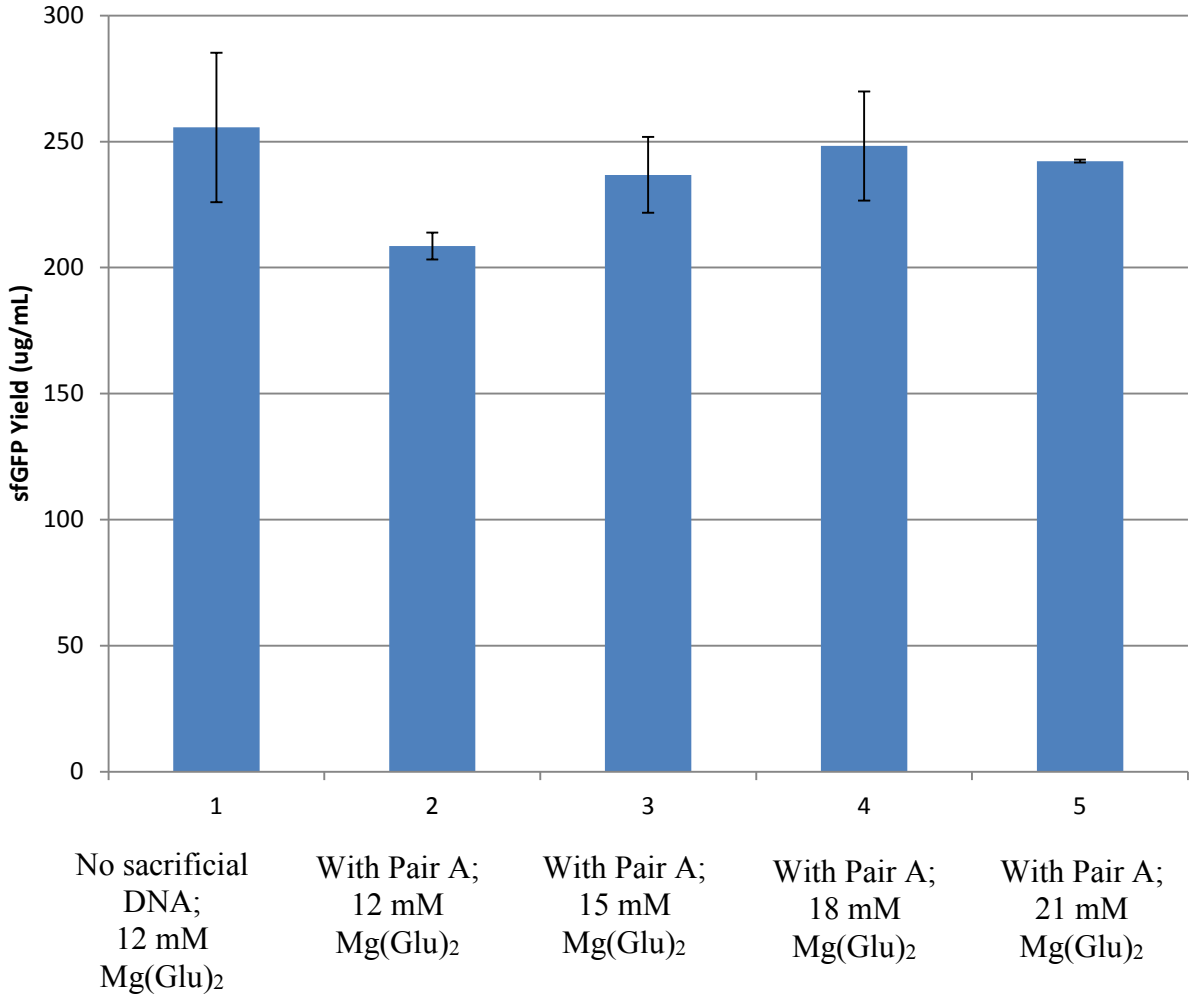

Figure 20: Magnesium Optimization of CFPS with Pair A Added. Each reaction was performed in duplicate, and error bars extend one standard deviation up and down from the average result.

Magnesium optimization is standard practice even when conducting CFPS reactions without sacrificial DNA; thus, magnesium optimization was important not only because of possible changes in magnesium requirements due to addition of sacrificial DNA (e.g. chelation of magnesium ions), but also because of variation in optimal magnesium levels of the extract itself.

The experiment revealed little difference among the various reactions; in fact, none of the Pair A-supplemented reactions' yields surpassed that of the reaction lacking Pair A, no matter what the concentration of magnesium. However, it did seem that more magnesium was indeed needed for the Pair-A supplemented reactions. Since the sacrificial DNA did not have the expected effect on yield, the experiment was repeated with a mass of sacrificial DNA seven times greater than before. The results of this experiment are shown in Figure 21 below.

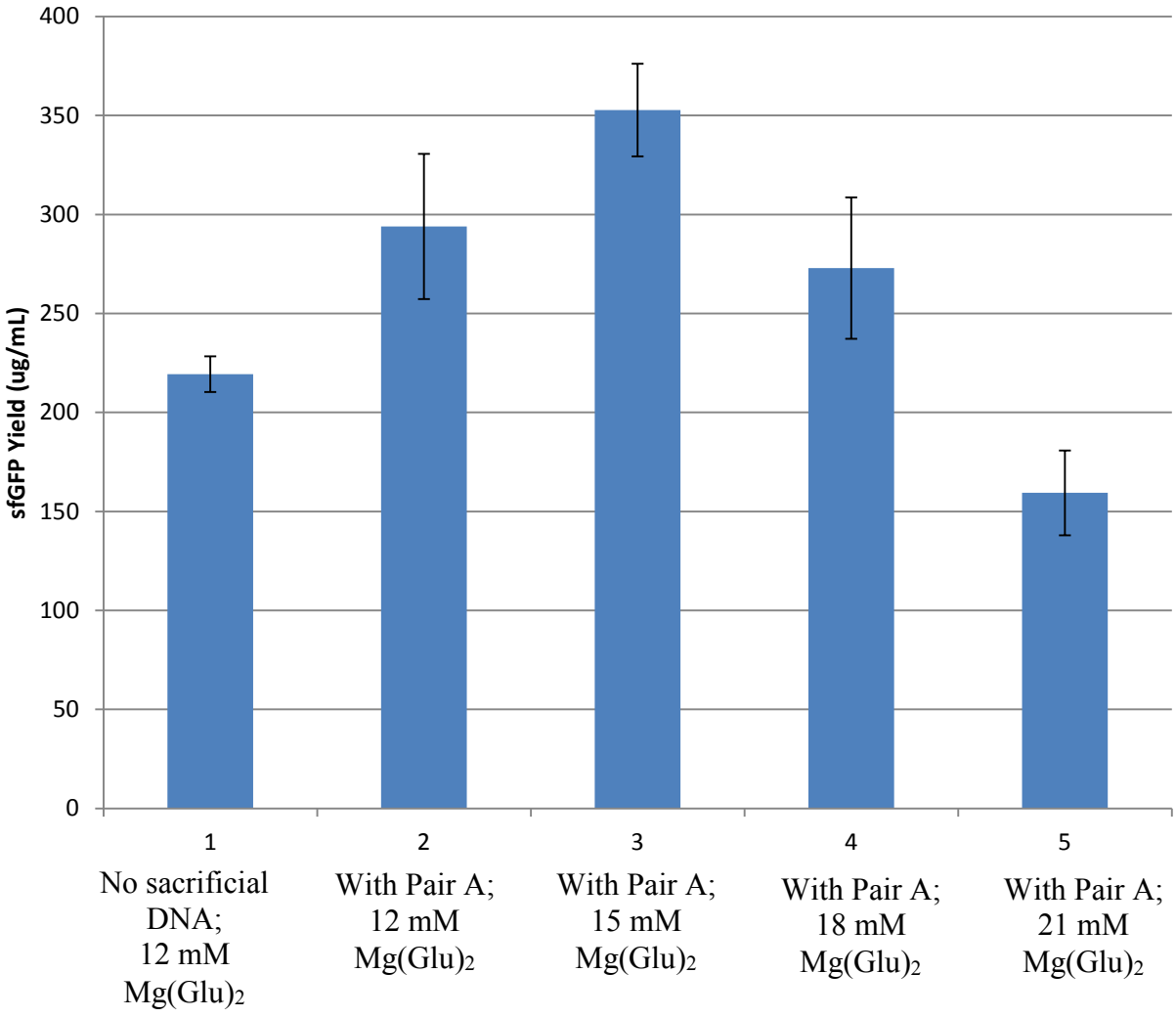

Figure 21: Magnesium Optimization of CFPS with Pair A Added in the Amount of Seven Times the Mass of LET Present. All reactions were performed in triplicate, and error bars extend one standard deviation up and down from the average result.

This time, multiple Pair-A supplemented reactions' yields surpassed that of the control. In the case of the 15 mM $Mg(Glu)_2$ reaction, the average yield was 67% greater than that of the control. In order to confirm these results and extend the test to other sacrificial DNA primer pairs, CFPS reactions were carried out with Pairs A, B , C, and D (see Table 5 above for descriptions of these pairs), one pairs per reaction. The resulting sfGFP yields were compared to that of a reaction lacking sacrificial DNA. Since 15 mM $Mg(Glu)_2$ was the best concentration for Pair A-supplemented reactions, it was also used for the other pairs' reactions. The results of this experiment are shown below in.

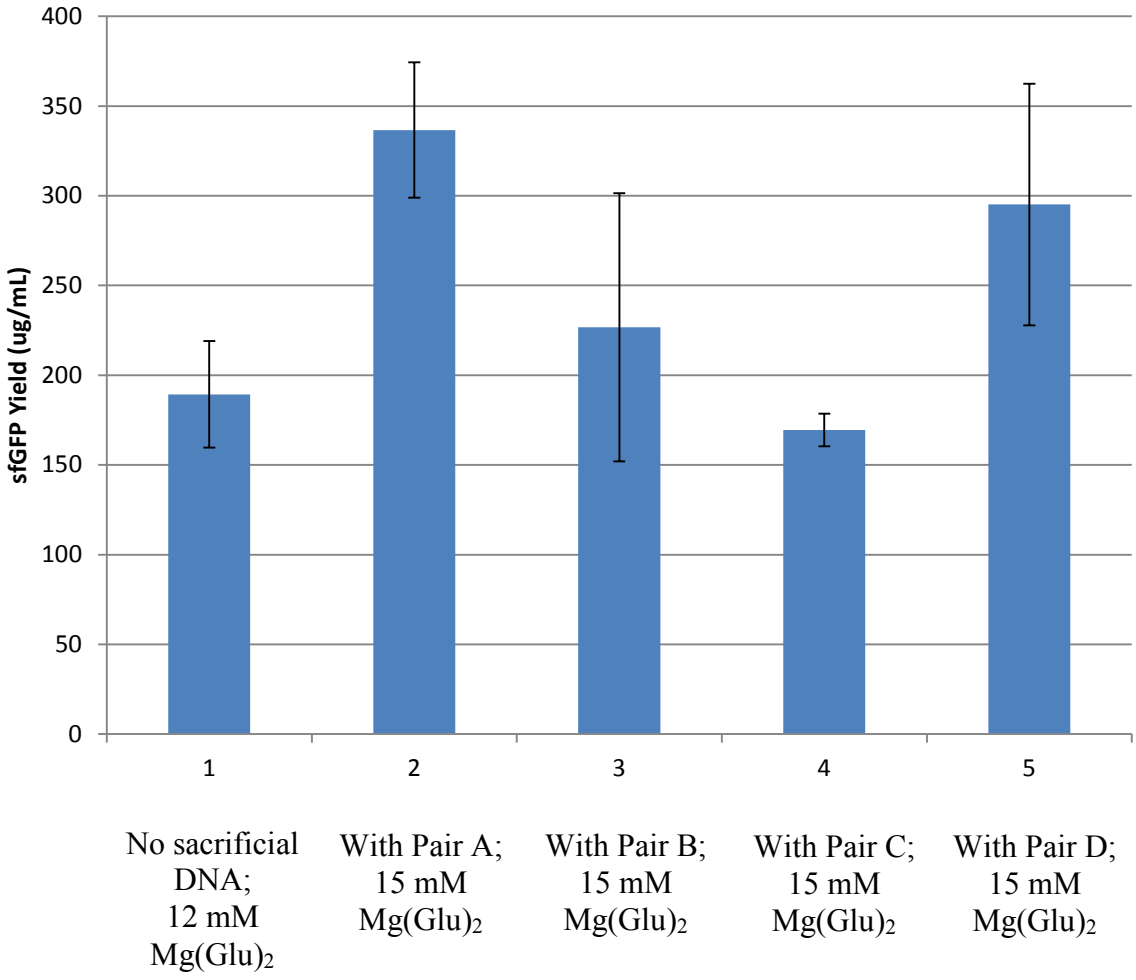

Figure 22: Yields of CFPS Reactions with Various Sacrificial DNA Pairs and 15 mM $Mg(Glu)_2$. All reactions were performed in triplicate, and error bars extend one standard deviation up and down from the average result.

The results displayed in Figure 22 confirmed those displayed in Figure 21: Pair A-supplemented CFPS with 15 mM $Mg(Glu)_2$ has sfGFP yields that are substantially greater than those of a CFPS reaction without Pair A—in this case, 88% greater. They also suggested that not all sacrificial DNA pairs perform the same way in CFPS, though large error values called for further tests on the subject. Accordingly, another experiment was performed wherein all reactions (including those with Pairs B-D) were Mg-optimized. The results of this experiment are shown in Figure 23 below.

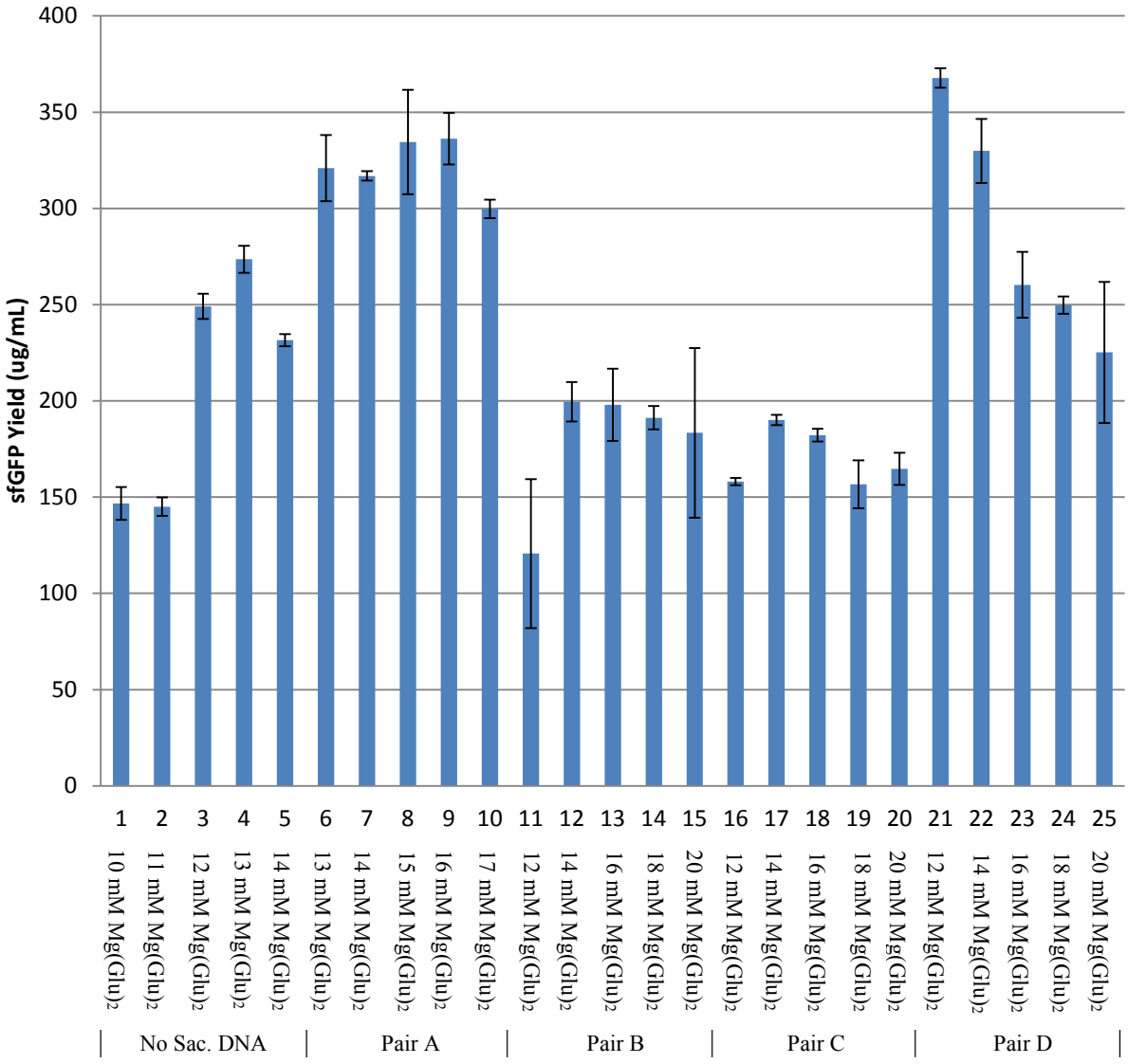

Figure 23: Magnesium Optimization of CFPS Reactions Containing Sacrificial DNA. Reactions 1-14 were performed in triplicate, while reactions 15-25 were performed in duplicate. Error bars extend one standard deviation above and below the average result.

Once again, Pair A's efficacy was verified, with the 15 mM and 16 mM $Mg(Glu)_2$ reactions producing substantially more sfGFP than the best reaction lacking sacrificial DNA (13 mM $Mg(Glu)_2$). In addition, the magnesium optimization of the other pairs' reactions revealed that yields for Pairs B and C both peak when $Mg(Glu)_2$ is at about 14 mM but never achieve levels comparable to those of the Pair A yields. Pair D, however, exceeds even the greatest Pair A reaction yield, and it does so with levels of $Mg(Glu)_2$ as low as those in the reactions without any sacrificial DNA. This was not intuitive because of the expectation that additional DNA would chelate Mg^{2+} ions and thus require more $Mg(Glu)_2$ to be added to the reaction. Among the properties of the four sacrificial DNA primer pairs (see Table 5), only one followed the same trend as did yield: low end T_m. One likely reason for this is that the end melting temperature of a DNA molecule is connected to how easily a DNase can degrade it. Since the melting temperature is one measure of how easily the two strands of DNA come apart, a low end T_m could mean less difficulty for a DNase during the degradation process, in particular an exonuclease, which is a DNase that degrades starting at the ends of linear DNA. Table 6, shown below, highlights the correlation of low end T_m to yield.

Table 6: Low T_m Values and Yields Displayed from Least to Greatest T_m. Note the negative correlation between low end T_m and yield.

Sacrificial DNA Primer Pair	Low End T_m (degrees Celsius)	Mg-optimized Yield (ug/mL)
D	20.3	368
A	26.5	336
B	36	200
C	36	190

Another question touching the use of sacrificial DNA in CFPS was whether simply changing LET concentration would have a superior effect to that of adding sacrificial DNA. To

investigate this, CFPS reactions were carried out with one, three, and eight times the normal mass of LET (the normal mass of LET plus seven times more, like the seven times' worth of sacrificial DNA) and no sacrificial DNA. The results of these reactions are shown in Figure 24.

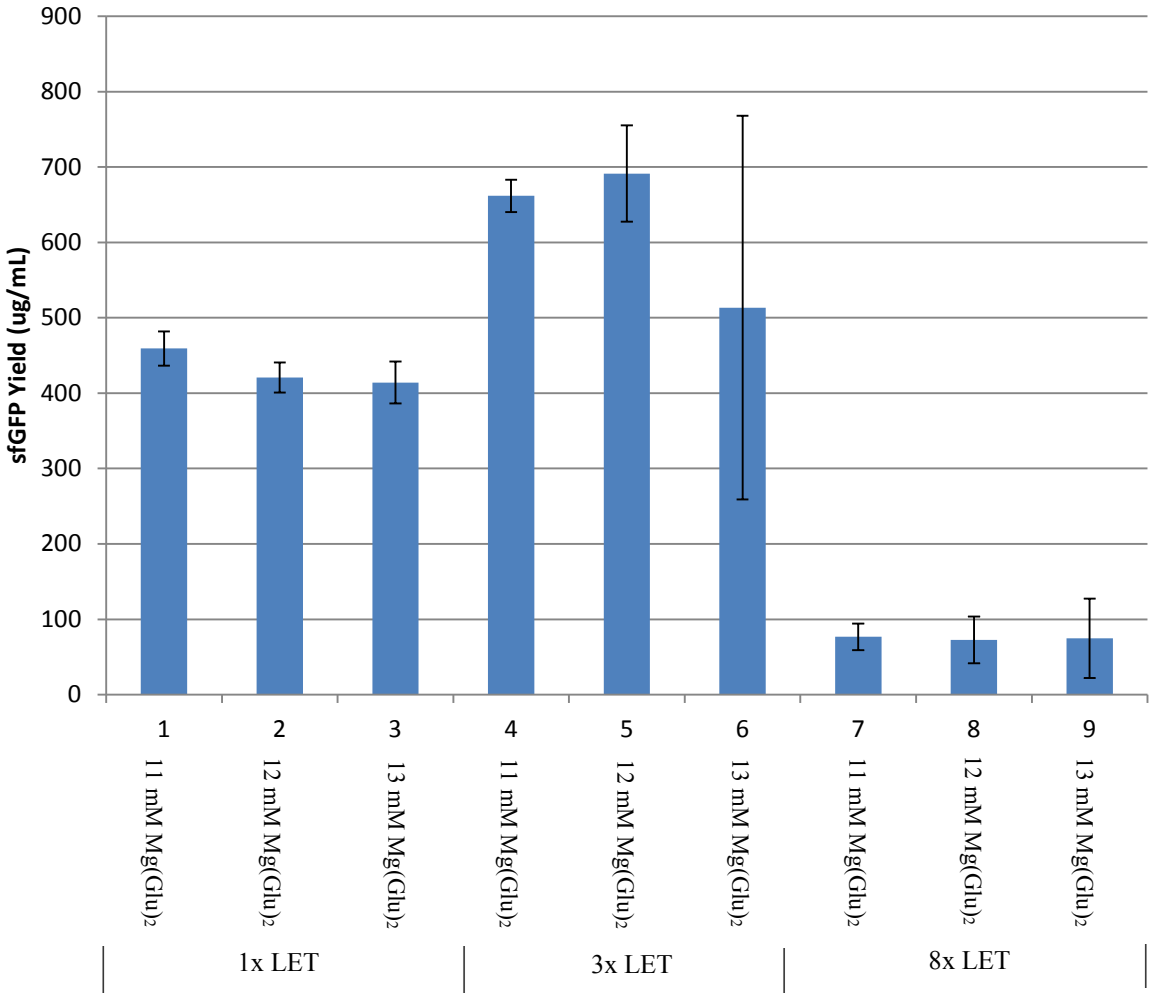

Figure 24: Magnesium Optimization of CFPS Reactions with Increased Amounts of LET but No Sacrificial DNA. All reactions were performed in triplicate, and error bars extend up and down one standard deviation from the average result.

As the data in Figure 24 show, the addition of more LET increases yield in the 3x case, but it decreases yield in the 8x case, at least for the levels of $Mg(Glu)_2$ tested. The next step was to test CFPS reactions with both elevated levels of LET *and* the addition of sacrificial DNA.

As before, the 8x LET reaction yielded much less protein than did the 1x LET reaction. Adding more LET provides more template for mRNA transcription and so could be beneficial; however, it is important to match transcription and translation rates such that ribosomes are constantly translating and protecting mRNA from RNAses, enzymes which degrade RNA. If too little LET is present, then not enough mRNA will be present for the ribosomes to be constantly producing protein; if too much is present, then energy will be wasted on making mRNA that cannot be used by the already-occupied ribosomes.The 3x LET reactions did not perform as well as the 3x LET reactions did in the previous experiment, although the addition of Pair D to the 3x LET reactions improved yield somewhat over that of the 3x LET reaction alone. Oddly, the addition of Pair D to the 1x LET reaction yielded less protein than did the 1x LET reaction alone. This contradicts the results from previous reactions with Pair D and 1x LET. Here it should be noted that the ubiquitous presence of DNAses makes the use of LET for CFPS reactions challenging. Meticulous adherance to protocols, the use of sterile technique, and care to prevent the unintentional transfer of DNAses and RNAses to LET stocks and CFPS reactions is essential. The author attributes the occasional lower yields of 3x LET and Pair D CFPS reactions to variance in procedure and to contamination issues.

4.4 **Economics of Sacrificial DNA Use in LET-based CFPS**

This chapter has discussed the yield increases in LET-based CFPS that are possible when either sacrificial DNA or more LET is added to the CFPS reaction. However, these yield increases come with the added costs of making sacrificial DNA or additional LET. This section examines these costs in time, labor, and materials.

There are multiple potential ways of obtaining sacrificial DNA, such as plasmid digestion and PCR with a plasmid as a template. This work hybridized complimentary, commercially-

obtained primers in a thermocycler to obtain sacrificial DNA, and it is this method whose costs will be analyzed here. The primers hybridized for this work's sacrificial DNA were lab stocks left over from previous projects. Because these primers had already been purchased, no money needed to be spent on them. This highlights one advantage of using sacrificial DNA: if leftover complimentary primers with suitable sequences (e.g. low end T_m) are already present in laboratory stocks, they can be put to use as sacrificial DNA. If only a single leftover stock is available, it can still be used if a complimentary primer is purchased. This is still less expensive than purchasing two entirely new primers. Only two ingredients, the forward and the reverse primer, were required for each sacrificial DNA hybrid pair, and the thermocyler hybridization cycle, shown in Table 4, required less than ten minutes. Furthermore, no purification was required, since neither DNA polymerase nor any other of the PCR ingredients was needed. All in all, the preparation time for a batch of sacrificial DNA (i.e. up to 4.8 mL with a full 96-tube thermocycler) was less than one hour.

Cost calculations were based on average results from experiments done on different days. Some of these days, as mentioned already, may have included procedural aberrations and/or contamination of reactions which altered the yields of the CFPS reactions. As more consistent data become available in this area, cost calculations may be made with more certainty. The present calculations indicate that 3x LET decreases the cost per ug of protein product made in CFPS. 3x LET also resulted in the greatest average yield—571 ug/mL, with a standard deviation of 153 ug/mL. Depending on the source of sacrificial DNA, sacrificial DNA can either be very expensive, no expense at all (as in this study, which used leftover primers for sacrificial DNA), or somewhere in between. Even in the case of no sacrificial DNA expense, however, 3x LET reactions were still the least expensive method of making sfGFP. Sacrificial DNA CFPS's average yield was 324

ug/mL with a standard deviation of 51.0 ug/mL, and LET-based CPFS alone had an average yield of 296 ug/mL with a standard deviation of 89.2 ug/mL.

Figure 25-27 display the costs of LET-based CFPS: without sacrificial DNA, with sacrificial DNA, and with 3x LET, respectively.

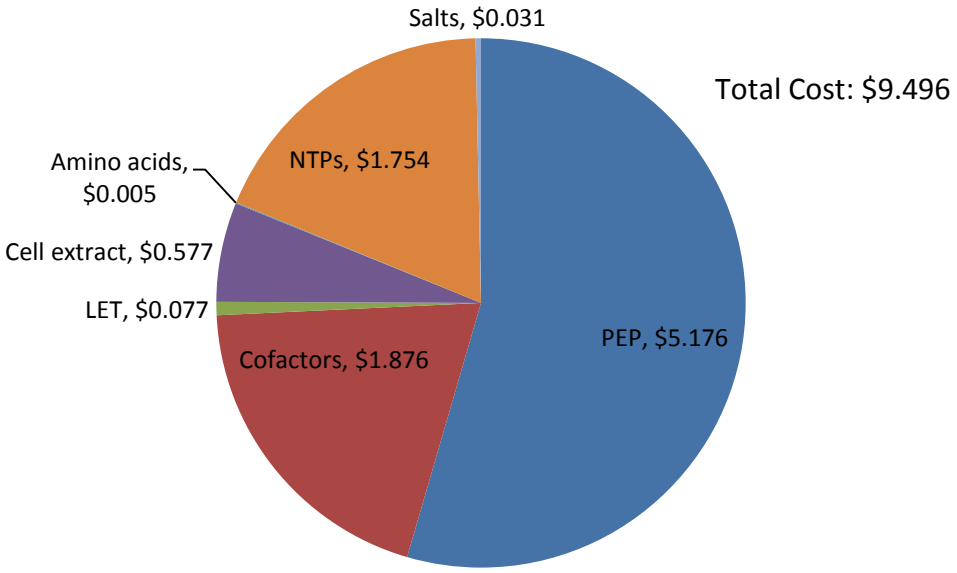

Figure 25: Costs of LET-based CPFS without Sacrificial DNA. Costs shown are for enough CFPS to make 1 mg of sfGFP. Costs of materials assumed the same as in [4].

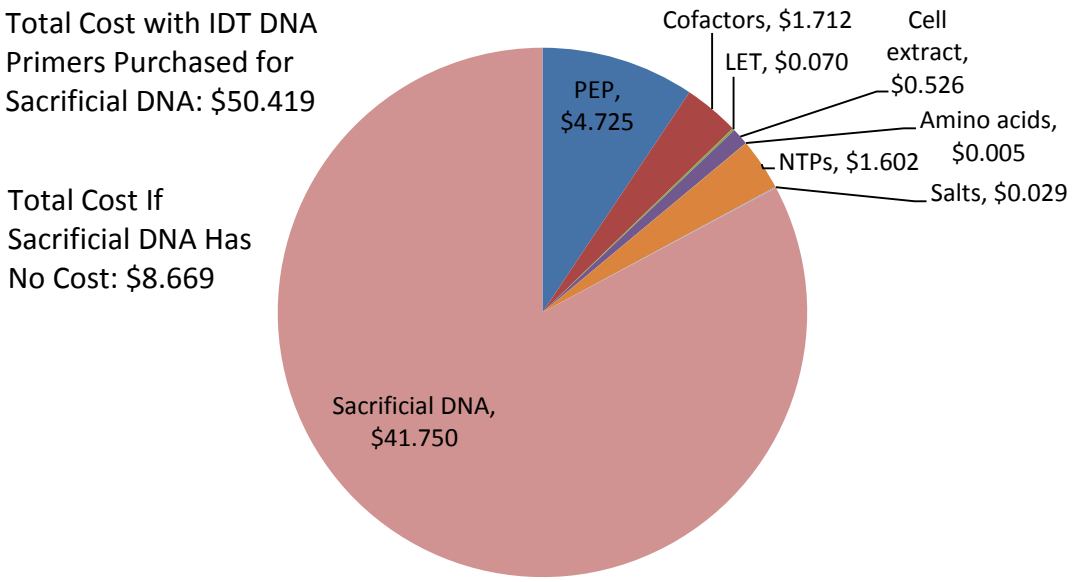

Figure 26: Costs of LET-based CFPS with Sacrificial DNA (Pair D, for this chart).Costs shown are for enough CFPS to make 1 mg of sfGFP. Price of sacrificial DNA based on purchase price of Pair D primers from IDT DNA. Costs of other materials assumed the same as in [4].

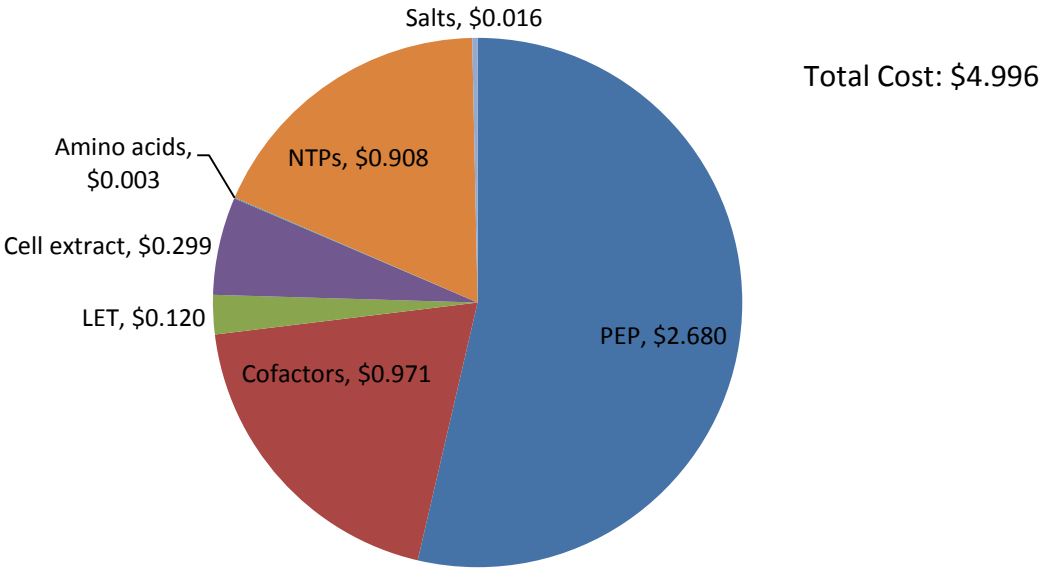

Figure 27: Costs of LET-based CFPS with 3x the Normal Amount of LET. Costs shown are for enough CPFS to make 1 mg of sfGFP. Costs of materials assumed the same as in [4].

4.5 **Conclusion**

LET-based CFPS yields have been shown to increase when either certain sacrificial DNA is added (e.g. either Pair A or Pair D) or 3x as much LET as is normal is used. 8x LET reactions, on the other hand, result in lower yields. It should be noted that some lower yields were occasionally observed; these were attributed to protocol inconsistencies and the ubiquitous nature of DNases and RNases. Further work adding sacrificial DNA Pair D to the 3x LET CFPS reaction showed either marginal or no increase in yield over the 3x LET CFPS reaction's yield. Based on the data in this work, 3x LET CFPS reactions are more cost-effective and have higher yield per reaction than either sacrificial DNA CFPS reactions or LET-based CFPS alone.

5 CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

This work centered on two ways of enlarging the capabilities and effectiveness of cell-free protein synthesis: producing aminoacyl-tRNA synthetases (aaRSs) capable of accepting unnatural amino acids, and preserving linear expression template DNA (LET) during CFPS reactions in order to increase yield. Chapter 3 described the selection and production of four aaRSs, as well as the testing of an aminoacylation assay from the literature for application to one of the aaRSs in this work. Chapter 4 discussed the effect on yield of adding more DNA to CFPS reactions, be it LET or sacrificial DNA.

As described in Chapter 3, a set of four aaRS/tRNA pairs for incorporation of unnatural acids in CFPS was designed, and plasmids containing the genes for the aaRSs were obtained. The *in vivo* production of these four aaRSs was carried out and confirmed by polyacrylamide gel electrophoresis (PAGE). PAGE analysis at various points in the protein purification process indicated that the aaRS from *M. barkeri* (MbRS for short, in this work) was being pelleted with the cellular debris during centrifugation of the cell lysate. This led to the conclusion that MbRS was not soluble in the *E. coli* cellular environment at the production concentrations and perhaps was forming inclusion bodies. The aminoacylation assay in this work proved unreliable for monitoring aminoacylation of *M. jannaschii* tRNA with the unnatural amino acid *p*-

propargyloxyphenylalanine. In particular, even when precipitate formation, which initially plagued the assay, was successfully prevented by addition of poly-(vinyl alcohol) PVA, allowing reliable absorption readings, the no-tRNA control's phosphate production reading (i.e. color change) in the assay about equaled that of the full reaction. This undermined confidence that the assay would measure only aminoacylation activity, or even measure it at all.

As detailed in Chapter 4, the yield of LET-based CFPS reactions can be improved—as much as 88%—by either using 3x the normal mass of LET or supplementing the reactions with Pair A or Pair D sacrificial DNA primer pairs in the amount of seven times the normal mass of LET. The only property of the sacrificial DNA which correlated consistently with amount of yield improvement was low end T_m. One possible explanation for this is that a weaker affinity of one DNA strand for the other, in particular at the ends, makes it easier for DNases to degrade the hybridized DNA molecule as a whole. Using 8x the normal mass of LET, however, drastically reduced yields compared to the 1x LET control. Also, some day-to-day variation in CFPS yields was observed and somewhat expected because of possible effect of using different LET stocks and because of the ubiquitous presence of DNAses and RNAses.

5.2 **Future Work**

Although this work took some preliminary steps in the investigation of unnatural amino acid-compatible aaRSs and of how to increase yield in LET-based CFPS, much remains to be done in these areas.

First of all, the aaRSs discussed in this work must not only be made and purified; their function must also be tested. One possible plan for such testing follows. The first stage of this testing would focus on assessing each aaRS/tRNA pair's orthogonality individually in the cell-free protein synthesis system. In order to test the orthogonality of individual unnatural tRNA/aaRS

pairs in the cell-free environment, pairs would be included in the system separately. All unnatural-pair tRNA would bear the anticodon corresponding to the amber stop codon AUG. Protein production would be evaluated for the system both with and without the pair's unnatural amino acid. For the pair to be verified as orthogonal, the tests (CFPS reactions) should give the following results: (1) with the unnatural amino acid present, the full protein product, e.g. some model protein green fluorescent protein (sfGFP), should be translated, since the unnatural amino acid will be inserted at the amber stop codon site; (2) with the unnatural amino acid absent, the protein product will be truncated at the amber stop codon, and no active GFP will be detected. Spectrophotometry data would be obtained showing the translation of the full model protein-GFP product, indicating that the stop codon was successfully suppressed by the tRNA.

The next stage of this testing would seek to establish the orthogonality of multiple unnatural aaRS/tRNA pairs relative to one another in the same CFPS reaction. This would advance research toward incorporating more than one type of unnatural amino acid into a single protein, rather than only incorporating one type per protein, as in this work. In a set-up similar to that for the previous stage, an mRNA encoding for GFP (which includes an amber stop codon insertion) would be translated in the cell-free system. However, all four unnatural tRNA/aaRS pairs would be present in the system at once. Five separate protein synthesis experiments would be conducted, one without any of the unnatural amino acids, and then four additional reactions, each of which would lack one of the tRNA/aaRS pairs and have only that pair's corresponding unnatural amino acid present (i.e. the other three unnatural amino acids will not be present). If the pairs are all mutually orthogonal, then all of these reactions will yield product which is truncated at the amber stop codon site. In addition, mass spectrometry would be used to verify correct incorporation of the unnatural amino acids will all components present. GFP action would not be detected in the absence of an

unnatural amino acid and its corresponding tRNA/aaRS pair. Correct unnatural amino acid insertion would be verified by trypsin digest followed by HPLC-mass spectrometry analysis. However, insufficient solubility of the MbRS in an *E. coli*-based CFPS environment will likely present a major challenge to such testing, since the aaRSs need to be present in sufficient concentration during the reaction to work properly.

For improvement of LET-based CFPS yields, it would be valuable to achieve greater consistency from experiment to experiment in yield improvement as compared with the control. In other words, it would be worthwhile to learn how to avoid the occasional poor performance of LET-based CFPS reactions. Also, it would be valuable to consider in more depth the question of what characteristics make one sort of sacrificial DNA effective over another. For example, the hypothesis that low end T_m is the primary determinant of sacrificial DNA effectiveness, which is corroborated by the data in this work, could be tested further by running CFPS reactions with a greater variety of sacrificial DNA: one group of DNA which would have two or three levels of low end T_m while keeping length and internal sequence constant, and another group which would have low end T_m constant with length and internal sequence varying. By comparing the yields of these reactions, it should become clearer which properties are actually important for increasing yield. One other idea for further study is to find a way to accurately measure the amount of intact LET present before and after a CFPS reaction. Such a measurement would lend insight into whether increased yield is really tied to better LET preservation or is, instead, not related. Such insight would direct future yield improvement efforts either toward or away from LET preservation depending on the extent to which it turns out to be important. Finally, in the same way as LET (DNA) preservation has been studied in this work, mRNA preservation could be studied in future work.

6 References

1. *Prestained Protein Ladder, Broad Range (10-230 kDa)*. 2015; Available from: https://www.neb.com/products/p7710-prestained-protein-ladder-broad-range-10-230-kda.
2. DNA Damage Recognition & Tolerance Laboratory, N.U. *Translation of the Genetic Code*. Available from: http://www.dna.neu.edu/research/projects/.
3. Mark Thomas Smith, S.D.B., Christopher J. Werner, and Bradley C. Bundy, *Lyophilized Escherichia coli-based cell-free systems for robust, high-density, long-term storage.* BioTechniques, 2014. **56**: p. 186-193.
4. Shrestha, P., *Streamlined Extract Preparation for E. coli-Based Cell-Free Protein Synthesis and Rapid Site-Specific Incorporation of Unnatural Amino Acids in Proteins.* All Theses and Dissertations, 2012.
5. *The polymerase chain reaction (PCR).* Nature Education, 2014.
6. Walsh, G., *Biopharmaceutical benchmarks 2014.* Nat Biotech, 2014. **32**(10): p. 992-1000.
7. Carlson, E.D., et al., *Cell-free protein synthesis: Applications come of age.* Biotechnology Advances, 2012. **30**(5): p. 1185-1194.
8. Banerjee, U.C., et al., *Thermostable alkaline protease from Bacillus brevis and its characterization as a laundry detergent additive.* Process Biochemistry, 1999. **35**(1–2): p. 213-219.
9. Ranji, A., et al., *Chapter 15 - Transforming Synthetic Biology with Cell-Free Systems*, in *Synthetic Biology*. 2013, Academic Press: Boston. p. 277-301.
10. *Our Scalable Biochemical Protein Synthesis Platform: Designing and Manufacturing a New Generation of Biotherapeutics.* 2014.
11. Ahn, J.H., T.J. Kang, and D.M. Kim, *Tuning the expression level of recombinant proteins by modulating mRNA stability in a cell-free protein synthesis system.* Biotechnol Bioeng, 2008. **101**(2): p. 422-427.
12. Shimizu, Y., T. Kanamori, and T. Ueda, *Protein synthesis by pure translation systems.* Methods, 2005. **36**(3): p. 299-304.
13. Shrestha, P., M.T. Smith, and B.C. Bundy, *Cell-free unnatural amino acid incorporation with alternative energy systems and linear expression templates.* New Biotechnology, 2014. **31**(1): p. 28-34.
14. Shrestha P, H.T., Bundy BC, *Economical Cell Lysis Technique Optimized for Cell-free Protein Synthesis.* Emerging Ideas in Biomedical Research, 2012.
15. Bundy, B.C.e.a., *Standard Operating Procedure: BL21 Star DE3 Cell Culture Protocol.* 2011.
16. Söll, M.I.a.D., *Aminoacyl-tRNAs: setting the limits of the genetic code.* Genes & Dev., 2004. **18**: p. 731-738.
17. Bundy, B.C. and J.R. Swartz, *Site-Specific Incorporation of p-Propargyloxyphenylalanine in a Cell-Free Environment for Direct Protein-Protein Click Conjugation.* Bioconjugate Chemistry, 2010. **21**(2): p. 255-263.
18. Schultz, P.G., *Schultz Lab (website).*

19. Polycarpo, C., et al., *Activation of the Pyrrolysine Suppressor tRNA Requires Formation of a Ternary Complex with Class I and Class II Lysyl-tRNA Synthetases.* Molecular Cell, 2003. **12**(2): p. 287-294.
20. Chatterjee, A., et al., *A Tryptophanyl-tRNA Synthetase/tRNA Pair for Unnatural Amino Acid Mutagenesis in E.coli.* Angewandte Chemie-International Edition, 2013. **52**(19): p. 5106-5109.
21. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(20): p. 7566-7571.
22. Santoro, S.W., et al., *An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli.* Nucleic Acids Research, 2003. **31**(23): p. 6700-6709.
23. Chatterjee, A., H. Xiao, and P.G. Schultz, *Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli.* Proceedings of the National Academy of Sciences, 2012. **109**(37): p. 14841-14846.
24. Liu, C.C. and P.G. Schultz, *Adding New Chemistries to the Genetic Code.* Annual Review of Biochemistry, Vol 79, 2010. **79**: p. 413-444.
25. Iraha F, e.a., *Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNATyr pair by the archaeal tyrosine pair in Escherichia coli for genetic code expansion.* Nucleic Acids Res., 2010. **38**: p. 3682-3691.
26. Nguyen, D.P., et al., *Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-tRNA Synthetase/tRNACUA Pair and Click Chemistry.* Journal of the American Chemical Society, 2009. **131**(25): p. 8720-8721.
27. Woese, C.R., et al., *Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process.* Microbiology and Molecular Biology Reviews, 2000. **64**(1): p. 202-236.
28. Cestari, I. and K. Stuart, *A Spectrophotometric Assay for Quantitative Measurement of Aminoacyl-tRNA Synthetase Activity.* Journal of Biomolecular Screening, 2013. **18**(4): p. 490-497.
29. *Oligonucleotide Specification Sheet*. 2015; Available from: http://www.idtdna.com.

7 APPENDIX

7.1 Sequences of Selected tRNA/aaRS Pairs (List Courtesy of Mark Smith)

***Methanosarcina barkeri*: pyrrolysine**

tRNA nucleotide sequence
TAATACGACTCACTATAGGGAGATTCCCCTGATGAGTCCGTGAGGACGAAACGGTA
CCCGGTACCGTCGGGAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTTAGC
CGGGTTAGATTCCCGGGGTTTCCGCCAGGAAGCT

aaRS amino acid sequence
MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTAR
AFRHHKYRKTCKRCRVSDEDINNFLTRSTESKNSVKVRVVSAPKVKKAMPKSVSRAPK
PLENSVSAKASTNTSRSVPSPAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISLNM
AKPFRELEPELVTRRKNDFQRLYTNDREDYLGKLERDITKFFVDRGFLEIKSPILIPAEYV
ERMGINNDTELSKQIFRVDKNLCLRPMLAPTLYNYLRKLDRILPGPIKIFEVGPCYRKESD
GKEHLEEFTMVNFCQMGSGCTRENLEALIKEFLDYLEIDFEIVGDSCMVYGDTLDIMHG
DLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRSESYYNGIST
NL

***Saccharomyces cerevisiae*: tryptophan**

tRNA nucleotide sequence
TAATACGACTCACTATAGGGAGAAGTTCCTGATGAGTCCGTGAGGACGAAACGGTA
CCCGGTACCGTCGAACTGGTGGCTCAATGGTAGAGCTCCCGCCTCTAAAGCGGGGG
GTTGCAGGTTCAATTCCTGTCCAGTTCACCAGGAAGCT

aaRS amino acid sequence
MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARA
LRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAP
KPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITS
MSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIYAEE
RENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPM
LAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLE
SIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGA
GFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

***Pyrococcus horikoshii*: lysine**

tRNA nucleotide sequence
TAATACGACTCACTATAGGGAGAGCCCCTGATGAGTCCGTGAGGACGAAACGGTAC
CCGGTACCGTCGGGCCCGTAGCTCAGCCTGGTAGAGCGGCGGGCTCTAAACCCGCA
GGTCGCGGGTTCAAATCCCGCCGGGCCCGCCAGGAAGCT

aaRS amino acid sequence
MVHWADYIADKIIRERGEKEKYVVESGITPSGYVHVGNFRILFTAYIVGHALRDKGYEV
RHIHMWDDYDRFRKVPRNVPQEWKDYLGMPISEVPDPWGCHESYAEHFMRKFEEEVE
KLGIEVDFLYASELYKRGEYSEEIRLAFEKRDKIMEILNKYREIAKQPPLPENWWPAMVY
CPEHRREAEIIEWDGGWKVKYKCPEGHEGWVDIRSGNVKLRWRVDWPMRWSHFGVD
FEPAGKDHLVAGSSYDTGKEIIKEVYGKEAPLSLMSEFVGIKGQKGKMSGSKGNVILLS
DLYEVLEPGLVRFIYARHRPNKEIKIDLGLGILNLYDEFDKVERIYFGVEGGKGDDEELR
RTYELSVMLPTY

***Pyrococcus horikoshii*: lysine (four-base codon)**

tRNA nucleotide sequence
TAATACGACTCACTATAGGGAGAGACCACTGATGAGTCCGTGAGGACGAAACGGTA
CCCGGTACCGTCTGGTCCGTAGCTCAGCCTGGTAGAGCGGCGGGCTTCCTCACCCGC
AGGTCGCGGGTTCAAATCCCGCCGGACTAGCCAGGAAGCT

aaRS amino acid sequence
(same as for regular lysine above)

7.2 Sequence of sfGFP (Superfolder Green Fluorescent Protein) [17]

MSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATIGKLTLKFICTTGKLPV
PWPTLVTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGKYKTRAVVK
FEGDTLVNRIELKGTDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKGTRDHMVLHEYVNAA
GITWSHPQFEK

7.3 Nucleic Acid Sequences of Primer Pairs Used to Make Sacrificial DNA [29]

Pair A Primers:

5- GGA GAT ATA CAT ATG TAG ATA TTT GAA ATG TTA CGT ATA GAT G -3
5- GGA GAT ATA CAT ATG TAG ATATTT GAA ATG TTA CGT ATA GAT G -3

Pair B Primers:

5- CCG AAC GAA AAA GGC ACATGGTCT CAT CCG CAGTTT GAA AAATAA CGG GAC CAC ATG G -3
5- CCATGT GGT CCC GTT ATT TTT CAA ACT GCG GAT GAG ACC ATG TGC CTT TTT CGT TCG G -3

Pair C Primers:

5- CCG AAC GAA AAA GGC ACATAA CGG GAC CAC ATG G -3
5- CCATGT GGT CCC GTT ATG TGC CTT TTT CGT TCG G -3

Pair D Primers:

5- CTA AAT TAA AAC CGGTTT ATG ATT CTT AGG ATG CGGTTC GTC GCG CTG CAT TG -3
5- CAATGC AGC GCG ACG AAC CGC ATC CTA AGA ATC ATA AAC CGGTTT TAATTT AG -3