THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF FOOD SCIENCE

CHARACTERIZATION OF THE SELECTION OF MULTIPLE START CODONS BASED ON VARIATIONS IN RIBOSOME BINDING SITE TRANSLATION INITIATION STRENGTH

HANNAH JEPSEN-BURGER
SPRING 2014

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Food Science
with honors in Food Science

Reviewed and approved* by the following:

Thomas L. Richard
Professor of Agricultural and Biological Engineering
Thesis Supervisor

Edward G. Dudley
Associate Professor of Food Science
Thesis Adviser

*Signatures are on file in the Schreyer Honors College

ABSTRACT

Translation initiation is the process by which a ribosome selects an initiation point from within an mRNA strand and initiates the process of converting genetic code into protein for the cell. Although alternative methods of eukaryotic translation initiation have been extensively cataloged, analogous alternative prokaryotic translation initiation mechanisms still remain largely unstudied. This research addresses the effects of multiple start codons within seven nucleotides of each other on protein expression using common cloning techniques. Plasmid DNA sequences were manipulated to observe these effects. By varying the ribosome binding site translation initiation rate of each start codon and measuring the fluorescent output of two reporters, the research determined that these two start codons may individually be selected by the ribosome for translation, resulting in multiple discrete protein species.

Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I would like to thank Dr. Tom Richard, for his support and guidance as this project was completed, and his assistance in the writing of this thesis.

I would like to acknowledge Dr. Howard Salis for his extensive contributions to the conception of the fundamental research idea, as well as his instrumental assistance in designing laboratory protocols suitable for the completion of the project.

I would also like to thank my laboratory's graduate student, Jyotsna Lavanya, for her help in ordering laboratory supplies and oligonucleotides.

Lastly, I would like to thank Dr. Edward Dudley for his assistance in the writing and review of this thesis.

Chapter 1

Introduction

1.1 Problem

The formation of a polypeptide chain through mRNA translation is the final major step in protein production. To begin protein synthesis, a ribosome must first select a codon from which to initiate translation. The process of translation initiation is an important regulatory mechanism in cellular function as well. Through it, cells may maintain and adjust metabolic activities and process kinetics [24]. While biologists have understood the basic principles of translation initiation for decades, the precise mechanisms of these processes have proven difficult to study, until recently.

With the advent of molecular biology, vast advances in our understanding of protein translation initiation have been made. An initial flurry of activity investigating the precise mechanisms of both prokaryotic and eukaryotic translation initiation coincided with technological improvements in the study of cell kinetics in the late 1980's, but interest stagnated as many scientists turned their attention toward the Human Genome Project and other major molecular biology undertakings [1].

Interest in these poorly understood translational mechanisms has been rekindled in the last 15 years due to the observation of several non-canonical phenomena in prokaryotic and eukaryotic translation initiation and subsequent protein expression. Modern knowledge of DNA replication, transcription, and translation dictates that these processes should, and usually do, follow highly-conserved, predictable processes, yet recent research investigating the finer mechanisms of translation reveal a startling number of translational apparatuses that do not behave as anticipated.

While research has elucidated some of the mechanisms by which cells govern alternative methods of translation initiation, especially in eukaryotes, the function of these non-canonical processes remains unknown. Continued study of these phenomena in cellular processes may one day show that these mechanisms are yet another way in which organisms evolve through genetic diversity.

Standard translation initiation models have been utilized by molecular and synthetic biologists for decades, and the emerging power of biological computational and analytical software has tremendously improved scientists' abilities to target the correct translation initiation codon in an unfamiliar DNA sequence [15].

With these new abilities, new questions have been raised about the validity of identifying a single start codon in a prokaryotic polycistronic mRNA strand which begins initiation of a desired protein. Recent studies in yeast genes and other higher order plants indicate that a single functional protein may indeed have several translation initiation sites, one of which is selected based on its upstream and downstream sequences and other factors

for translation in a single mRNA strand. These studies, however, are almost exclusively reactive rather than proactive, and are most often designed to investigate unusual observations in other primary research.

Most studies in this emerging area of research characterize multiple points of translation initiation in eukaryotes, or in virus genetic material which replicates via eukaryotic cellular machinery. Few studies specifically address this phenomenon in prokaryotic species, and none, to the author's knowledge, have studied it extensively.

As synthetic biology and cloning techniques become increasingly mainstream in raw agricultural practices and food modification and processing (consider genetically modified organisms), it is important to expand our knowledge of alternative translation initiation sites in both eukaryotes and prokaryotes. Scientists may use this information to more accurately assess unconventional research results, or they may use the knowledge as a tool to develop better agriculturally-useful organisms.

1.2 Hypothesis

Prior research indicates that multiple points of translation initiation corresponding to a single protein sequence are possible, and are in fact frequently observed in nature. This research tested whether multiple start codons in close proximity to each other in an mRNA strand will produce discrete protein species due to varying ribosomal selection during translation initiation. In addition, the research tested whether the preferential selection of one or the other start codons could be controlled by altering the ribosome

binding site (RBS) translation initiation strength of each AUG triplet.

The testing of plasmid cassettes containing sequences that reflected the project design were expected to show that multiple start codons within seven nucleotides of each other can indeed be individually utilized as translation initiation points, and that altering the ribosome binding site strength results in corresponding alterations in translation initiation at each codon.

Efficiency in the translation of these multiple protein species is expected to rise or fall with the predicted strength of the site directly preceding the AUG codons; that is, the ribosome binding site.

1.3 Research Objective

The objective of this research was to construct a plasmid cassette capable of testing the effects of multiple start codons on protein expression in *Escherichia coli*. DNA constructs were tested to firstly observe if it is possible for a ribosome to select mulitple start codons for multiple points of protein translation initiation, and secondly to observe if this start codon selection may be controlled through adjustment of the preceding RBS sequence.

The strategy used to accomplish these goals included the assembly of a plasmid using synthetic biology techniques such as Gibson Assembly, polymerase chain reaction (PCR), restriction enzyme digestion, and the insertion of annealed oligonucleotides into a vector plasmid sequence. To test the validity of the hypothesis, fluorescent protein reporters were measured as an indicator of the translation initiation rate of each start codon .

Chapter 2

Literature Review

2.1 Translation Initiation in Prokaryotes and Eukaryotes

2.1.1 Translation Initiation in Eukaryotes

The central dogma of molecular biology dictates that DNA makes RNA makes protein [13]. Inherent in this mantra, which biologists have doggedly characterized for decades, is the understanding that a single start codon initiates translation of a single functional protein. While the amino acid sequence that will ultimately form a protein may lie adjacent to several other translatable protein sequences (as is the case in the polycistronic mRNA strands of prokaryotes), or may specifically spliced together from a string of codons representing a single protein (as in the monocistronic mRNA strands of eukaryotes), a single start codon has traditionally been believed to initiate the synthesis of a single protein [13, 24, 28].

Although both eukaryotic and prokaryotic translation initiation traditionally begin with a single start codon, the mechanisms of the selection of that start codon differ between the two types of organisms [24]. Eukary-

otic cells begin protein translation with a 43S preinitiation complex, which consists of the 40S ribosomal subunit, GTP, Met-tRNAi$^{\mathrm{Met}}$ and a series of proteins known as eukaryotic initiation factors, specifically eIF2, eIF3, eIF1, and eIF1A, which act to regulate the binding of the preinitiation complex to the capped 5' end of the mRNA strand. [1].

After the preinitiation complex has successfully bound to the capped mRNA, it begins to "scan" in the 5' to 3' direction for the translation initiation codon, AUG, which complements the attached tRNA in what is referred to as the 5' UTR, or untranslated region. In most events, the first AUG codon in the sequence is found suitable for translation initiation, and the complex ends its scanning of the mRNA strand. The larger 60S ribosomal subunit binds the smaller 43S preinitiation complex in the formation of the full ribosome, and translation begins in the open reading frame, or ORF, determined by the placement of the first AUG codon [1, 24].

The binding of the preinitiation complex, and the subsequent scanning mechanism of the 5' UTR in eukaryotic translation initiation is highly dependent on the primary and secondary structures of the targeted mRNA strand. In most cases, interaction between the 5' cap and the various regulator proteins required for initiation, but recent revelations in the field have demonstrated alternative methods of translation initiation [30, 36, 27, 22].

2.1.2 Translation Initiation in Prokaryotes

Prokaryotic translation initiation differs significantly from eukaryotic translation through the former's lack of a comparable ribosomal scanning mechanism for start codons [24, 18, 25, 28]. Translation initiation sites

on polycistronic prokaryotic mRNA are additionally characterized by the Shine-Dalgarno (SD) sequence, a highly conserved set of nucleotides that bind preferentially to the 16S rRNA portion of the 30S ribosomal subunit several nucleotides upstream from the intended AUG codon [25, 16]. As the prokaryotic ribosome does not scan an mRNA strand, searching for the first start codon avilable, as eukaryotic systems do, the Shine-Dalgarno sequence-AUG codon relationship must satisfy certain requirements to properly initiate protein production.

The DNA sequence immediately preceding the SD section must be free of secondary structure, allowing the smaller ribosomal subunit involved in prokaryotic translation initiation to loosely bind tp the area and prevent any further mRNA secondary structure from forming and obscuring access to the AUG start codon [25]. The SD sequence must end between four and nine nucleotides prior to the start codon for the highest chance of binding and ribosome recognition. About 15 nucleotides upstream and downstream of the start codon are sheltered from nuclease degradation when the full 70S unit prepares for translation, effectively further preventing further mRNA secondary structure [24].

With the assistance of several prokaryotic initiation factors and GTP, the smaller ribosomal subunit begins translation at the start codon [25, 13]. Thus, while translation initiation may seem simple in prokaryotes, requiring only a start codon and a nearby SD sequence, the exact structure and placement of the upstream mRNA nucleotides plays a vital role in ribosomal recognition of an AUG codon. Just as in eukaryotic translation initiation, prokaryotic translation relies on a number of crucial factors, both interactive

and structural, to locate the appropriate AUG triplet and initiate protein production. Some of the proposed mechanisms of unconventional prokaryotic and eukaryotic translation initiation are described below.

2.2 Multiple Sites for Translation Initiation

Despite the central dogma's traditionally rigid view concerning protein translation, investigations testing the conformity of nature to man's theories continue. Recently, certain translational phenomena have been observed in human viruses [33, 20, 8, 7, 6, 5] several yeast genes [4, 2], insects [3], higher-order plants [39], mammals [37], and even the human genome [21, 34, 19, 15], that challenge the conventional understanding of protein translation, espcially in eukaryotic organisms.

Various theories delineating the mechanisms underlying alternative translation initiation for both prokaryotes and eukaryotes were first proposed by Marilyn Kozak in the late 1970's. Her hypotheses describing the leaky scanning ribosome, the importance of a set of nucleotides now known as the Kozak Consensus Sequence in eukaryotic mRNA, and the effect of mRNA secondary structure on prokaryotic ribosomal placement were at first met with skepticism from molecular biologists [1]. Through many studies and highly-publicized results, her theories are now the basis of our understanding of alternative translation initiation in both prokaryotes and eukaryotes.

2.2.1 Alternative Translation Initiation in Eukaryotic Systems

As previously noted, eukaryotic translation is primarily reliant on 43S preinitiation complex - 5' cap interactions and the ability of the subunit to identify with fidelity the target AUG triplet for normal translation initiation. Although both were once considered absolutely necessary for initiation, it is from these two mechanisms that several translational anomalies have arisen to be studied and quantified [1].

A phenomenon referred to as IRES, or internal ribosome entry site, results in the by-passing of normal 5' capped mRNA - 40S subunit complex interactions. The ribosome complex binds directly instead to a start codon located between nucleotide sequences with favorable secondary structure for its binding, effectively removing the need for a scanning mechanism [25, 12]. This particular alternative form of translation may occur in both capped and uncapped mRNA strands, and requires only eIF3, eIF1, and eIF1A to initiate 48S initiation complex formation [30, 12, 23, 31, 20]. Gunnery et. al demonstrated cap-independent translation of the HIV-1 Tat protein in HeLa cells [17]

Another form of cap-independent translation initiation may be attributed to cap-independent translation enhancers, or CITEs. Although significantly less studied than IRES, CITEs are believed to be utilized primarily by viral genetic material with uncapped mRNA to propagate translation initiation. A CITE sequence is likely a leader sequence near the 5' end of the mRNA strand that ensures the binding of initiation factors and the 43S preinitia-

tion complex. It was shown by analyzing naturally long 5' UTR sequences that the translational efficiency of capped mRNA sequences was significantly different depending on the leader sequence of the cap, indicating that some additional sequence in the cap structure was assisting translation initiation [9]. The sequences appear to bear some resemblance to the function of the Shine-Dalgarno sequence in prokaryotes [36, 12].

Independent of 5' capped or uncapped translation is a more common and thus better studied form of alternative initiation known most commonly as "leaky scanning". While the canonical model of the scanning ribosome in eukaryotes indicates that the 40S complex should cease its movement from the 5' to 3' end of the mRNA strand after binding to the first and correct AUG codon it encounters [23], the leaky scanning ribosome may skip over the first canonical start codon, settling instead on another AUG triplet that better supports its binding [1, 24].

This phenomenon is commonly observed in mRNA sequences in which two possible start codons are located within close proximity to each other, not exceeding 15 nucleotides apart. Ribosomes may even slip backwards several nucleotides to recover initiation of a nearby start codon, as was shown in influenza B virus NA and NB glycoproteins [41], and later in turnip yellow mosaic virus (TYMV) [26] If the secondary structure of a second start codon allows for more entropically favorable conditions, that codon will initiate translation, reducing the efficiency of initiation from the original AUG codon [14, 32].

2.2.2 Alternative Translation Initiation in Prokaryotic Systems

Prokaryotic translation initiation is not as well characterized as in eukaryotic systems, and thus it is more difficult to discern the mechanisms behind alternative start codon selection in these organisms. A reliance on the secondary structure of the sequence surrounding a potential start codon however, as well as the polycistronic nature of the mRNA strands, allows for a high degree of variability in start codon selection.

Proteins produced from the selection of multiple start codons in prokaryotic mRNA strands, especially in those that code for regulatory proteins transcription factors, kinases, and growth factors, may have the ability to up- or down-regulate protein expression [38]. These multiple protein species may also work in tandem to improve the overall function of each other. It was shown that multiple start codons produced multiple forms of the SafA protein integral to the formation of *Bacillus subtilis* spore coating [29]. When point mutations of one or the other suspected translation initiation sites were introduced, the spore coat was not as robust, indicating that the cell needs both translation initiation sites to function properly [29].

The vast majority of information collected on the effect of multiple start codons on protein expression are from reactive rather than proactive studies, stemming from the observation of abnormal results of an unrelated experiment. Given the greater interest in eukaryotic studies by NIH and other agencies, it is unsurprising that the mechanisms for this translation initiation phenomenon in eukaryotes are in fact far better characterized than their

prokaryotic counterparts. This disproportionate amount of prior research is evidenced by the lack of literature available on the subject of alternative mechanisms of prokaryotic translation initiation.

While the standard steps of translation initiation are well-characterized in prokaryotes, mechanisms that are analogous to ribosomal start codon slippage as described above for eukaryotes are still not well understood in prokaryotes. Start codons in prokaryotic systems are notoriously difficult to identify and predict with software, traditionally pinpointing the correct AUG codon for a certain protein with 60-70 percent accuracy in high-GC sequences [35]. Much of the information gathered in identifying correct start codons results from mutating and deleting various possible translation initiation sites and observing the effects on translation of the desired protein.

Although the mechanisms of cellular translation initiation have engaged scientific attention for decades, the once-rigid foundations of the translational canon are slowly being chiseled away. They yield instead to the idea that the abilities of both prokaryotic and eukaryotic cells to produce multiple and diverse functional proteins through the initiation of several start codons from a single mRNA strand are much more complex than previously hypothesized. While eukaryotic organisms compensate for the many intricacies and requirements of translation initiation through ribosomal scanning with IRESs, CITEs, and the leaky scanning mechanism, prokaryotic systems rely on altered secondary structure and multiple non-canonical start codon triplets to increase the number of functional proteins available for translation from a single mRNA strand [1, 28].

Although the potential for multiple start codons in the same protein

sequence is a very real, and indeed frequently observed, phenomenon, it is a problem many researchers may encounter as a secondary hurdle in their primary research. Characterizing the effects of multiple start codons on protein expression through ribosome binding site strength control provides information useful in elucidating the possible influences and impacts of this issue in prokaryotes and, with further study, its mechanism.

Chapter 3

Materials and Methods

Common synthetic biology techniques were utilized to construct the plasmid cassette that was used to test ribosomal selection of multiple start codons in protein translation. Plasmid editing software was used to design a sequence with the appropriate placement of start codons and reporter proteins (see Chapter 4). Long oligonucleotides, known as gBlocks, were used to build a plasmid cassette using the Gibson Assembly method. The testing of this first cassette encouraged further study of the phenomenon through RBS manipulation.

The Salis Laboratory Ribosome Binding Site Calculator provided RBS sequences fitting the needs of the experiment, and these sequences were obtained, then inserted into the plasmid vector via oligonucleotide (oligo) annealing, restriction enzyme digestion, and ligation. Cassettes were tested for fluorescence in a monochomator-based microplate reader capable of measuring fluorescence intensity. Materials and methods for this design are detailed in this chapter, while the experimental strategy and design are covered in

Chapter 4.

3.1 Bacterial strains and growth media

Escherichia coli strain DH10B (referred to as *E. coli*) was used in all experiments for cloning and characterization of multiple translation initiation sites. Strains were cultured in LB media at 37C, shaking at 250rpm, or grown on LB-agar plates. Cultured wild-type strains were supplemented with streptomycin (25 µ/mL), while strains containing plasmid cassettes were grown with chloramphenicol (25 µ/mL).

3.2 Plasmid and primer design

A plasmid Editor (ApE) is a free software application available through the University of Utah. This program was utilized for all plasmid and primer design and editing. In addition to ApE, a SciTools application, available on Integrated DNA Technologies's (IDT) website, entitled *Oligo Analyzer*, provided necessary information regarding primer GC content, melting temperatures, potential secondary structure interference, as well as dimer probabilities. Refer to Table A.1 for primer sequences used in PCR and sequencing. All *E. coli* stocks and plasmid vectors were obtained from the Salis Synthetic Biology and Metabolic Engineering Laboratory at Penn State. All oligos were commercially synthesized by IDT.

3.3 Gibson Assembly® of vector plasmid

The Gibson Assembly® method was used to assemble the initial plasmid vector, pR0sfG0. The method included the annealing of multiple several-hundred-basepair double-stranded oligos with adjacent complementary ends together [11]. IDT synthesizes oligos to these specifications, called "gBlocks", which are to be used in conjunction with New England Biolab's (NEB) Gibson Assembly®. Master Mix. Refer to Table B.1 for the complete sequences of these gBlocks. 3.1 demonstrates the reaction volumes used in the Gibson Assembly®. method. After mixing, the reaction tube was incubated in a thermocycler for 60 mins at 50 °C for a 4-fragment assembly.

Table 3.1: Gibson Assembly® Master Mix reaction specifications

fragment concentration (pmoles)	0.2-1.0 (X µL)
deionized water	10-X µL
Gibson Assembly® Master Mix	10 µL
total volume	20 µL

3.4 Gel electrophoresis and DNA extraction

Electrophoresis and subsequent DNA extraction from the agarose gel was performed in order to separate excised DNA fragments and inactivate heat-stable endonucleases after enzymatic restriction digestion. Low molecular weight, short basepair fragments, typically <100bb, were run on a 2.5% agarose gel at 80V for 120 minutes. High molecular weight, long basepair fragments were run on a 1.5% agarose gel at 100V for 100 minutes.

DNA fragments were observed under a transilluminator, and were ex-

cised for purification. The E.Z.N.A Gel Extraction Kit supplied by Omega Bio-Tek was utilized for this protocol, and purification was conducted according to the manufacturer's instructions.

3.5 Salis Lab Ribosome Binding Site Calculator

The Salis Lab Ribosome Binding Site Calculator (https://salis.psu.edu/software) was integral in the design of the plasmid cassettes, namely in the adjustment of RBS translation initiation strength. The Calculator generates RBS sequences at a desired strength by allowing the user to input an mRNA sequence, beginning with a start codon triplet, along with a 20 basepair pre-sequence upstream from the desired RBS site placement, and a translation initiation rate on a proportional scale. Using this information, the Calculator produces a RBS sequence with the prescribed translation initiation rate, as demonstrated in Figure 3.1.

In designing the RBS sequences for the various test cassettes, the translation initiation rate target for the RBS of the start codon corresponding to RFP was increased by an order of magnitude between each cassette (i.e. around 100 in the first cassette, to approximately 1000 in the second, to 10,000 in the third, to 100,000 in the fourth), while the sfGFP start codon remained at a constant rate (around 5,000 au). The "RBS With Constraints" option in the Calculator was used to ensure that the RBS sequence that provided a certain translation initiation rate for one codon would not interfere with the desired expression of the the other start codon. Thus, RBS

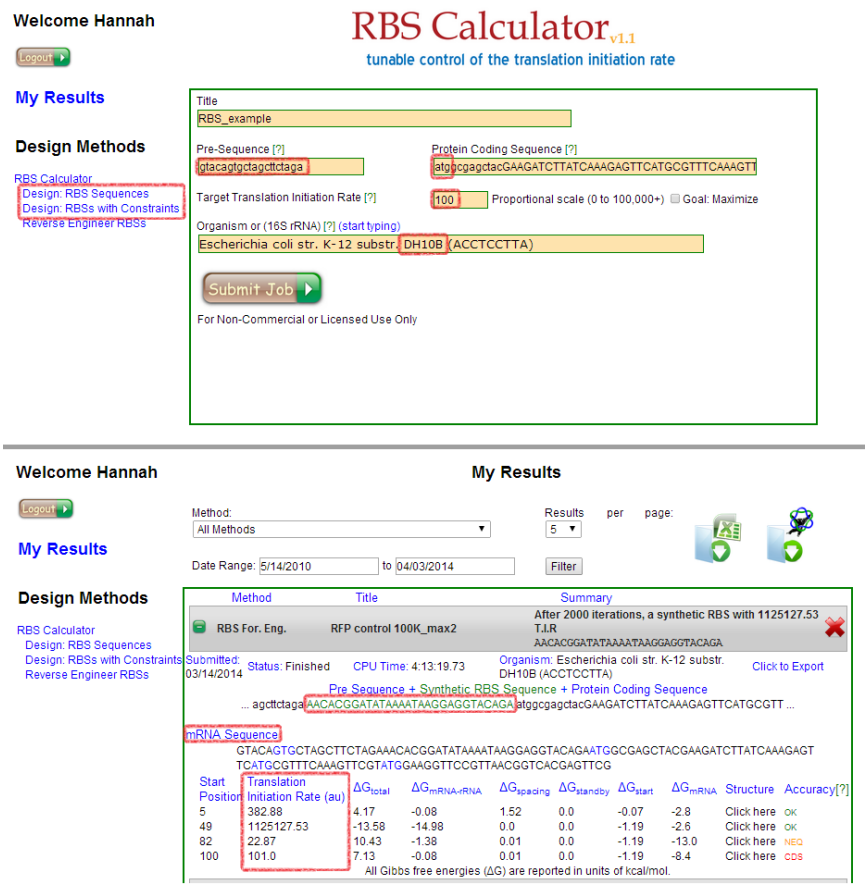

Figure 3.1: Salis Lab RBS Calculator example. A pre-sequence, protein coding sequence, organism type, and translation initiation rate target are inserted by the user, and the RBS sequence output may be used to design plasmid cassettes.

sequences which suited the needs of the experiment were produced by this application. See page 26 for complete details concerning the alteration of RBS translation initiation rates in cassettes.

3.6 Ligation of annealed oligonocleotides

Single-stranded oligos were designed and annealed to contain "sticky ends" complementary to the 3 nucleotide overhangs left by restriction enzyme digestion of the plasmid vector. Annealed oligos were ligated using T4 ligase manufactured by NEB, as shown in Table 3.2. The ligation reaction was incubated at room temperature for 10 minutes.

Table 3.2: Ligation using T4 ligase reaction specifications

10X T4 DNA Ligase Buffer	2 µL
vector DNA	50 ng
insert DNA	50 ng
nuclease-free water	to 20 µL
T4 DNA ligase	1 µL

3.7 Restriction enzyme digestion

The vector plasmids were designed to contain two restriction enzyme sites flanking the RBS + AUG + AUG sequence of the plasmid. Restriction enzyme digestions were used to linearize and remove RBS and AUG codon fragments from the vector plasmid, leading to the formation of nucleotide overhangs. These sticky ends allowed for the insertion of annealed oligos with complementary overhangs. Table 3.3 demonstrates a typical restriction enzyme digest reaction. Restriction enzyme digests were incubated at 37 °C for 6 hours before being separated on an agarose gel, as previously described.

Table 3.3: Restriction enzyme digest specifications

BSA	1 µL
NEBuffer 2	5 µL
vector DNA	1 µg
XbaI	1 µL
BglII	1 µL
nuclease-free water	to 50 µL

3.8 Microplate reader fluorescence measurement

A monochomator-based microplate reader(Infinite® M1000 PRO, TECAN, Männedorf Switzerland) capable of measuring fluorescence intensity was used to test each construct for fluorescence reporter protein production. Fluorescent molecules, such as the reporters used to visualize start codon selection in this experiment, emit a fluorescent photon at a specific wavelength when exposed to a light of shorter wavelength. In this assay, the amount of fluorescence emitted from the two fluorescent reporters contained within the construct was quantified.

The four test clones were shaken in LB media with the appropriate antibiotic at 200 RPM at 37 °C. Wild-type DH10B *E. coli* grown under the same conditions was used as negative control, while LB media was used as a blank.

The following morning, the clones and negative control were diluted to 5 µL in 190 µL LB media in a 96-well microplate and loaded into the microplate reader. The plate was shaken for 3.5 hours at 37 °C; optical density (OD600) of cell cultures, as well as RFP and sfGFP fluorescence was measured in 10 minute intervals until corrected OD600 reached approximately

2.5.

The samples from the microplate were diluted a second time in a 1:40 ratio in LB media and measured until corrected OD600 reached 2.5. The plate was diluted a third time in the same manner and measured again.

Chapter 4

Chapter 4. Plasmid Cassette Design, Construction, and Analysis

4.1 Plasmid Cassette Construction

In order to explore the effects of alternative translation initiation sites on protein expression and translation, a plasmid containing multiple start codons separated by a seven nucleotide sequence was devised. An existing plasmid sequence, available from the Salis Laboratory in the Department of Agricultural and Biological Engineering, was used as the basis of the design, providing a DNA backbone that allowed for manipulation in the ApE software, and subsequent physical manipulation via cloning techniques.

The plasmid included a promoter sequence, a ribosome binding site,an

origin of replication (ColE1), a chloramphenicol resistance gene, and two start codons within a short distance of the RBS. The first start codon was a single AUG codon in ORF[0], or open reading frame 0, which initiated translation of red fluorescent protein (RFP) in the same ORF[0].

In order to easily measure the effects of multiple start codons on protein expression in prokaryotes, the original plasmid sequence was modified to include a second start codon in a different open reading frame than the first. The second start codon was positioned seven nucleotides downstream from the first start codon, in ORF[+1]. This AUG triplet initiated translation of a super-folder green fluorescent protein (sfGFP) reporter protein in the same +1 ORF as the second start codon.

Since the object of most studies investigating multiple start codons is to discover whether proteins produced by initiation at a non-canonical start codon are similarly functional to their traditionally-translated counterparts, only multiple start codons within the *same ORF* are studied in those experiments. In this project design, however, multiple start codons in *two different ORFs* are used to determine start codon selection. By simply altering the ORF of the second start codon to force it out of frame with the first, and attaching separate reporters, each in frame with a single start codon, the degree of ribosomal selection of each codon may be measured by the fluorescence of each reporter. Figure 4.1 demonstrates the placement of each start codon, and its corresponding ORF and subsequent reporter protein contained within that distinct ORF.

For example, if the first AUG triplet is selected, RFP will be expressed

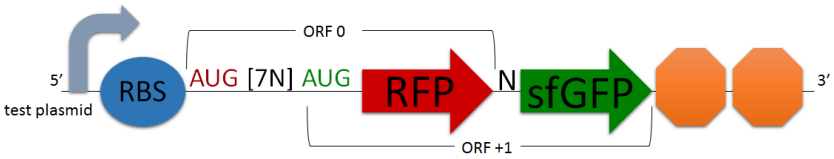

Figure 4.1: Cartoon of initial test cassette design. The red AUG triplet in ORF[0] represents the start codon which initiates translation of RFP, while the green AUG triplet in ORF[+1] represents the start codon which initiates translation of sfGFP. The double-octagon symbol represents a double-terminator sequence.

but sfGFP will not be expressed. If the second start codon is selected for translation initiation, sfGFP will be expressed but RFP will not. If both start codons are used for translation initiation in different mRNA strands, both reporters will fluoresce. The amount of each reporter present, measured by the fluorescence per cell, indicates to what degree each start codon was chosen for initiation by the ribosome.

Thus, the first several codon triplets of the RFP protein sequence were altered in the plasmid design to include a second start codon in ORF[+1], seven nucleotides downstream from the original RFP start codon. This second point of translation initiation, out of frame with the first ATG sequence by a single nucleotide frameshift, and subsequently the out of frame with the RFP reporter sequence due to the seven nucleotide difference between start codons, was instead designed to be in frame for translation of the sfGFP.

Super-folder GFP was used as the second reporter because of it ability to fold correctly even in the presence of an attached "junk" DNA sequence. Thus, should the ribosome select the second AUG codon for translation, the series of out-of-frame codons comprising the mis-translated RFP sequence

would not affect the protein's ability to fold and fluoresce.

With these start codons in the desired open reading frames, the plasmid was then rationally designed to modify and neutralize any potential stop codons in the preceding RFP sequence that might terminate protein synthesis of sfGFP upon the event of its codon's selection by the ribosome for translation initiation.

After fluorescence measurements of the first plasmid cassette, the design for which is pictured in Figure 4.1, demonstrated that a prokaryotic translation system can indeed promote the translation of one or another start codons in close proximity to each other, additional plasmid cassettes were designed. These constructs were designed with the intent to control the effects of multiple start codons on protein translation by varying the RBS sequence preceding the AUG triplets, and thus the translation initiation strength of the start codons. The heavy reliance of prokaryotic ribosomes on mRNA secondary structure for AUG codon selection suggests that altering the sequence directly preceding the first start codon (the RBS site), as well as the seven nucleotides separating the first start codon from the second may have an impact on the selection of a start codon from translation initiation.

In total, four additional cassettes were designed, constructed, and tested. The design of these cassettes was identical to that shown in Figure 4.1, with the exception of the RBS sequence in each, which was altered in order to measure the effects of different RBS translation initiation rates on start codon selection. The design and construction of these test cassettes is discussed in Section 4.2.

Although this research may be expanded in the future to include multiple permutations of high and low RFP and sfGFP translation initiation, as well as varied lengths between start codons, this thesis involves only the study of a single set of altered RBS translation initiation strengths, discussed in the following section.

4.2 Varying RBS Translation Initiation Rates of Reporter Proteins

Four additional cassettes were designed and constructed to manipulate the translation initiation rates of the two start codons in the original plasmid design. RBS sequences were designed in the Salis Lab Ribosome Binding Site Calculator. In order to reduce variability between cassettes, sequences were designed to increase translation initiation rates of the first start codon by one order of magnitude between cassettes, while a constant level of sfGFP was maintained.

In the first cassette, pR100G5K, the first start codon was intended to initiate translation at a low level, around 100au, while the second start codon should have been initiated at around 5000au. The second cassette contained an RBS sequence that was predicted to initiate the RFP start codon at near 1000au, while the sfGFP start codon remained constant at around 5000au, and so on, so that each plasmid contained an RBS site which initiated translation of the RFP start codon at increasing rates, while sfGFP was expressed at relatively constant levels between the cassettes.

In this way, the project aimed to better understand the nature of multiple start codons by controlling the probability that one start codon would be selected over another, and to observe if the experimental results conformed to the predicted translation initiation rates. Table 4.1 shows the RBS and AUG codon sequences used in the test cassettes, as well as their predicted translation initiation rates for each codon.

Table 4.1: RBS Sequences and their predicted translation initiation rates in test cassettes

Name	Predicted RFP translation initiation rate (au)	Predicted sfGFP translation initiation rate (au)	RBS + start codon sequences
pR100G5K	408	5341	TATTACCGTCACTACT CCGGAATACCATTCAA GGATGTTTTTCAATG
pR1KG5K	1657	5341	TAAACTATAAAAGAGA AGTAAGCAAGACAAAG GATGTACAACTATG
pR10KG5K	18687	5079	TAGTCTCTAAGGGG CTAGAGAGGAGTAGGTA AGGATGGAAGACGATG
pR100KG5K	108355	5557	AGAAAGAATCAAAAAT ACATAAGGAGAGTAAG GTATGGTAGGCGATG

RBS and AUG sequences and their projected translation initiation rates for test constructs according to Salis Lab RBS Calculator estimates. Blue text indicate the RBS, red text the RFP start codon, and green text the sfGFP start codon.

4.3 Control Cassette Design and Construction

Designing control constructs for the experiment entailed eliminating a start codon from the original plasmid sequence. After construction of these

control plasmids, clones containing the control plasmids would then be tested for fluorescence in the same manner as the test clones in order to observe the effects on translation of both proteins. Since the dynamic reporter protein in this set of cassettes was RFP, the AUG codon corresponding to sfGFP translation was altered so that it no longer represented a start codon. Figure 4.2 demonstrates the removal of the second start codon to prevent sfGFP expression.

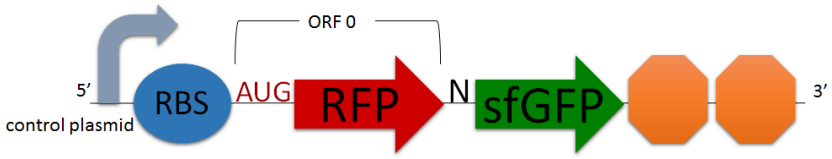

Figure 4.2: Control plasmid cassette, in which the start codon corresponding to sfGFP reporter protein was altered to a non-AUG sequence to evaluate the expression of sfGFP at increasing translation rates of RFP, despite the lack of a sfGFP reporter start codon. The red start codon initiates translation of RFP in ORF 0, while the green start codon initiates translation of sfGFP in ORF +1.

With this alteration, the expression of RFP might reflect the level of expression of the protein in the test cassettes, while sfGFP expression was, in theory, eliminated. Table 4.2 below represents altered RBS sequences mirroring translation initiation rates for the RFP start codon in the test cassettes. Clones containing these control plasmids were not successfully tested at the time of this writing.

Oligonucleotides with sticky ends were commercially synthesized, annealed in the laboratory, and ligated into the plasmid vector in a protocol

Table 4.2: RBS sequences and their predicted translation initation rates in control cassettes

Name	Predicted RFP translation initiation rate (au)	Predicted sfGFP translation initiation rate (au)	RBS + start codon sequences
pR100G0	106	0	TACCTAAACGAAAGACA GCGCAGCGCAAGGCG TATGG CGAGCTACG
pR1KG0	923	0	CTAGAATACGAAAATCACTT TAAAATAGTCCAAACGT TAATGG CGAGCTACG
pR10KG0	10410	0	AAGGCGTTCTAATAAAT ATAAGAAGGGCAACATG GCGAGCTACG
pR100KG0	104148	0	ATACATACAGCGAAAAAA TAAGGAGTAATAT TAATGGCGAGCTAC GATGGCGAGCTACG

RBS and AUG sequences and their projected translation initiation rates for control constructs according to Salis Lab RBS Calculator estimates. Light blue text indicate the RBS, red text the RFP start codon, and dark blue text the altered, non-initiating sfGFP start codon.

identical to that used to create the test cassettes (see Section 3.1 for details about materials and methods).

4.4 Fluorescence Measurement

A measurement of total fluorescent output over several cell growth cycles was measured using the TECAN monochomator. The instrument shakes a microplate, allowing cells to grow, while aiming an excitatory wavelength that causes fluorescent proteins present in the microplate wells to emit photons at a different wavelength used for reporting. The RFP excitatory wavelength is 584 nm, and the protein emits fluorescence at 607 nm, while sfGFP is excitable at 475 nm and emits at 508 nm.

In this measurement, serial dilutions of the four test cassette clones and wild-type DH10B *E. coli* were performed in triplicate (as described in Section 3.8, with each dilution representing the beginning of a new cell growth cycle. These serial dilutions were measured for total RFP and sfGFP fluorescence, as well as cell density in each microplate well.

The equation

$$FLPC = \frac{FL_{sample} - FL_{LB}}{OD_{sample} - OD_{LB}} - \frac{FL_{DH10B} - FL_{LB}}{OD_{DH10B} - OD_{LB}} \tag{4.1}$$

where LB is the growth media, FL is the fluorescence measured in the well at a given time, and DH10B is the wild-type *E. coli* strain acting as the negative control , describes the amount of RFP and sfGFP expressed as comparable fluorescence per cell (FLPC) measurements based on well fluorescent measurements and background fluorescence from wild-type *E. coli*

strain DH10B cells. Although the equation does not account for normalization between the two fluorescent proteins, it does permit a ratio comparison between the reporters in each clone.

While this lumped data is acceptable for making very generalized comparisons, it is more meaningful when manipulated to express the fluorescence per indiviual *E. coli* cell. To analyze this data, corrected RFP, sfGFP, and OD600 values were calculated. Since clones were measured in triplicate, the average corrected RFP and sfGFP values between clones were used to determine total fluorescence per cell. As the data represented the lag to peak log growth phase of three growth cycles for each clone, the linear portion of the three growth curves, and corresponding increase in fluorescence, was averaged for the graphical representation in Figure 4.3. Based on Equation 4.1, fluorescence per cell of RFP and sfGFP reporters was compared and analyzed. Results are discussed in Chapter 5.

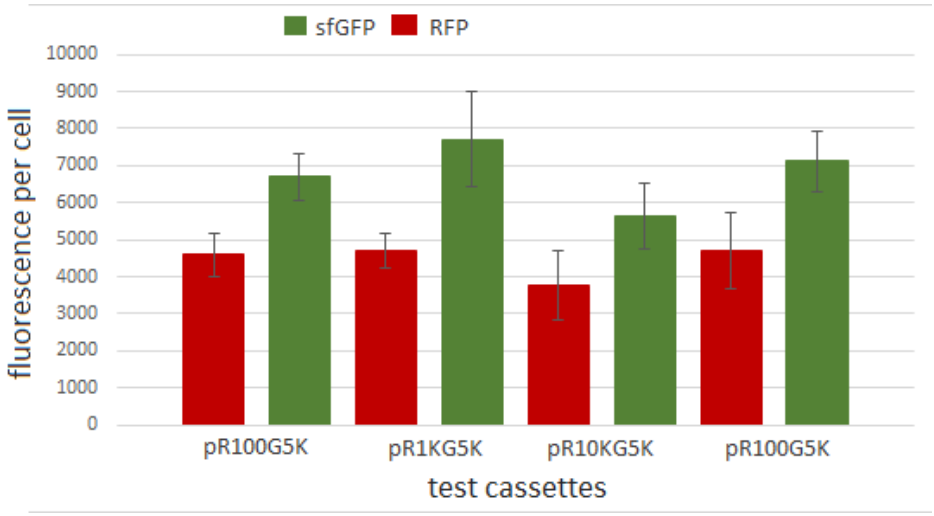

Figure 4.3: Fluorescence per cell (FLPC) of plasmid test cassettes. Red bars represent RFP per cell, while green bars represent sfGFP per cell, averaged across log growth cycles. Statistical analyses using one-way ANOVA and Tukey's HSD test indicate no significant difference between the means of each group.

Chapter 5

Discussion

The purpose of this project was to observe and, if positive observations were made, to both quantify and attempt to control the expression of proteins produced through the phenomenon of alternative translation initiation sites. The use of non-canonical start codons and ribosomal leaky scanning has been extensively mapped in eukaryotes, but few mechanisms have been proposed, and fewer supported with rigorous research, to explain the incidence of this anomaly in prokaryotes.

This research aimed to discover if two start codons within seven nucleotides of each other would result in the translation of multiple distinct protein species, and if so, could these multiple points of translation initiation be controlled through the alteration of a ribosome binding site sequence.

After the first clone containing multiple start codons was constructed, it successfully demonstrated that multiple start codons within seven nucleotides did indeed express multiple proteins, as evidenced by the expression of both RFP and sfGFP reporters. To evaluate how accurately this

phenomenon could be controlled, a series of plasmids were constructed with increasing predicted levels of RBS translation initiation strength. These cassettes were tested by measuring total fluorescence over a series of serial dilutions and regrowth cycles.

In order to better quantify and compare the data collected from these fluorescence intensity measurements, the fluorescence per cell of both RFP and sfGFP were calculated (Figure 4.3) While the predictive model in the Salis Lab Ribosome Binding Site Calculator anticipated that each clone would differ in its RFP start codon translation initiation rate by a factor of 10, the experimental results show unexpectedly similar rates of RFP and sfGFP expression in each *E. coli* clone. A one-way ANOVA statistical test, coupled with Tukey's HSD post-hoc test demonstrated that there were no statistically significant differences between the means of each FLPC group.

To facilitate comprehension of the desired outcome of the TECANM1000 fluorescence measurement compared to the actual experimental results, Figure 5.1 has been included in this section. It serves as a visual reference only, and its graphical representation is based on fabricated data. Figure 5.1 demonstrates the *expected* trends in RFP and sfGFP expression between each test clone. RFP levels increase by an order of magnitude between each clone, while sfGFP levels remain constant. In the *experimental* data (Figure 4.3), RFP and sfGFP levels were not statistically different between clones, as had been expected.

While the experimental FLPC values attributed to each clone in Figure 4.3 are all part of the same statistical group, the values were expected to look similar to those pictured in Figure 5.1. According to the RBS Calcula-

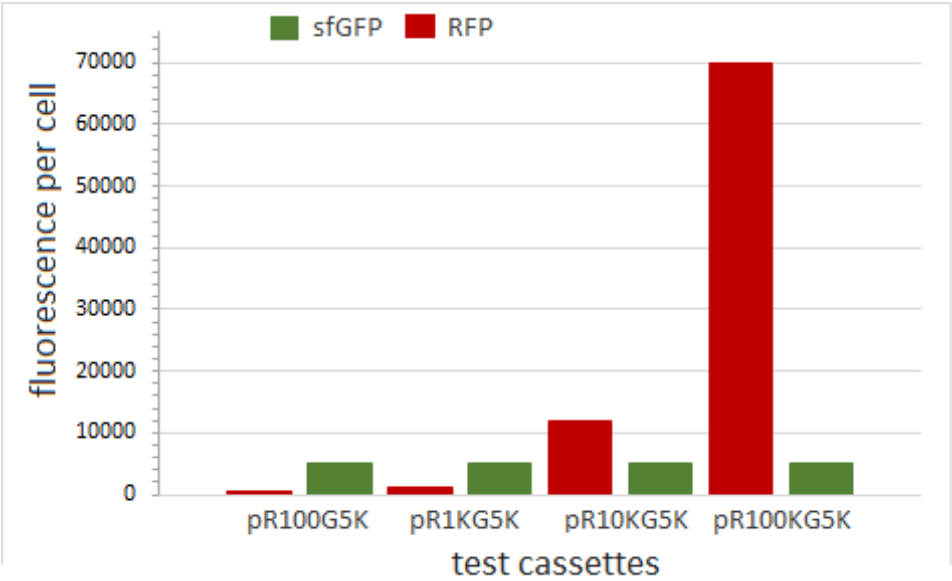

Figure 5.1: Expected RFP and sfGFP results from TECAN measurements of test cassettes. sfGFP was expected to remain constant in each clone while RFP expression was expected to increase approximately one order of magnitude in each cassette from left to right, according to increasing RBS translation initiation strength.

tor, the predicted translation initiation rates of the RBS sequences ligated into each test cassette dictated that the RFP expression increase by an order of magnitude between each clone, while sfGFP remain relatively constant. Instead of seeing this steady increase in RFP expression, the test cassettes showed similar amounts of RFP and sfGFP expression across the set.

Without statistically significant differences between group means, its is impossible to correctly judge trends between the clones. Still, the data is not devoid of meaning. From this experiment, it is clear that there is more to controlling the selection of potential start codons than altering the RBS strength. The Salis Lab RBS Calculator is moderately well-equipped to

anticipate the possibility of multiple start codons; in the reviewing of an mRNA strand for AUG sequences, it even notifies the user if overlapping start codons exist within the sequence of interest. However, based on the failed predictions of the application in this experiment, there are certainly other variables which must be considered.

Even without the expected differences in RFP and sfGFP expression in the clones, this experiment demonstrates that multiple start codons do indeed result in the production of multiple protein species. In fact, the lack of differentiation between clones when differentiation was expected indicates that secondary structure, here manipulated by altering the RBS sequence, may not be as vitally important in the selection of start codons in prokaryotic species as is currently believed. Perhaps a downstream sequence alters ribosome selection, or an unidentified internal regulator prevents the overexpression of highly-initiate start codons. In any case, this data suggests that start codons within seven nucleotides of each other compete for translation initiation by the ribosome, at least somewhat independently from the upstream sequence.

In addition to these considerations, we must also keep in mind that the translation initiation rates supplied for the RBS sequences inserted into these plasmid cassettes do not necessarily imply similar protein expression level in the cell. Especially in the clones with higher RFP translation initiation rates associated with their start codons, the cells may be under too much biological stress to produce a protein at such high rates, resulting in skewed data. Other factors, such as the obscuring of sfGFP fluorescence by the connected "junk" DNA sequence which remains attached to the sfGFP

molecule after translation termination, as well as so-called codon "bottlenecking" may contribute to varied levels of sfGFP.

An anomalous growth pattern was observed in pR1KG5K during plating, which reinforces the unusually high amount of sfGFP seen in this clone. While all other clones grew comparably on agar plates over an 18 hour period, colonies of this clone would not appear for one to two days after plating. When they did appear, they were very small, and fluoresced a much brighter green than the other constructs under transilluminator light.

This high expression of sfGFP is also apparent in Figure 4.3. The slow growth of this clone my be attributed to its apparently high expression of sfGFP, as the cell likely shunted more of its energy to producing the reporter instead of growing. Slow growth in comparison to the other clones tested under the same conditions and for the same amount of time may also have impacted the overall validity of the FLPC calculations, as the pR1KG5K clone did not reach the same OD600 as the other cells during the TECAN measurements.

Unfortunately, due to the simple nature of this experiment, it is very difficult to draw any reliable conclusions from the data described here. The small scale of the project, and the absence of additional cassettes or mRNA data makes it unwise to speculate on the mechanism of alternative translation initiation in prokaryotes.

We may say with certainty, however, that alternative translation initiation is, in fact, possible to induce in biologically-engineered systems. Based solely on this criteria, the research is, if nothing else, novel. The lack of reliable information regarding ribosomal selection of start codons in prokary-

otic systems may prove a more noticeable gap in knowledge as interest in genetically-modified prokaryotic organisms, especially in fermented dairy products, begins to take off in food science applications.

In the larger scope of molecular and synthetic biology, prokaryotic organisms are often the first choice of beginning researchers, as they proliferate quickly and are often hardier and less sensitive to environmental stressors than their eukaryotic counterparts. Many of the research studies cited here stumbled upon alternative translation initiation sites in the course of their primary research. By proactively attempting to look beyond canonical perspectives of prokaryotic translation initiation, not only will researchers better understand small-scale applications in their current research, but they can also use the information to engineer organisms which defy dogmatic theories, opening an entirely new world of protein translation.

Chapter 6

Further Research

6.1 Additional RBS Translation Initiation Rate Variations

Although no definitive mechanisms can be proposed based on the results of this research, continued attempts to control start codon selection by the ribosome in prokaryotic organisms is an integral part of understanding how the system operates. As previously mentioned in this thesis, the continued assembly of clones containing varied RBS sequences is an obvious continuation of this research. If time and resources were readily available, additional RBS sequences would be designed which promoted sfGFP at the same rates currently presented by RFP (that is, increasing by an order of magnitude), while repressing current RFP rates and maintaining them at a constant expression rate.

Additionally, varying the distance between the start codons may help

indicate at which point the ribosome no longer overlaps both start codons. When the protein expression of the second start codon stops, the distance between the two start codons is too great to warrant ribosomal slippage between the two, and a definitive codon may be continuously selected for initiation.

Lastly, the start codon sequences themselves may be altered to include non-canonical sequences such as GUG or UUG, which are less frequently observed in prokaryotic translation initiation, but are nonetheless viable codons from which to begin translation [13, 25]. the use of these alternative start codons may help indicate exactly how strong their ability to initiate translation truly is, and the manipulation of the RBS site and downstream codons may also determine the secondary structure which is necessary for their propagation or inhibition. In short, further characterization of multiple start codons is both intriguing and useful in a number of applications.

6.2 Food Science Applications

Beyond the proof of concept demonstrated in this research, the potential of multiple start codons, if they can be accurately controlled and their behavior correctly predicted, reaches far beyond the realm of synthetic and molecular biology. As primarily foundation work, the research can be extended into the field of food science as well, specifically in the study of human pathogens. Several viruses have been found to exhibit multiple start codons which code for several discrete protein species.

Interestingly, however, is the evidence that some of these non-canonical

start codons code for proteins which can up- or down-regulate different cellular systems, or which may be required for full functionality of what was once believed to be a single protein [40, 29, 8]. A single mutation leading to the addition or substitution of another start codon may heavily impact a virus's ability to disrupt a host cell's homeostasis.

The investigation of these incidences of attenuation or positive feedback mechanisms in bacteria, coupled with knowledge allowing for the control of the mechanism, may lead to interesting applications in food microbiology and the medical world.

Pathogens are not the only bacteria frequently cited in food science and technology. Fermentation bacteria can be used as vectors to confer additional benefits beside flavor to the food products they create. Dr. Ryan Elias's research [42, 10] concerning the biologically-active regions of antioxidant enzymes and other proteins may be coupled with an cloned organism that has the ability to both produce both a sequence rich in radical-scavenging amino acids, and the enzyme to cleave it and degrade secondary structure so the amino acids are available for antioxidant activity. Using multiple start codons to control the expression of one or the other sequence, simply by manipulating the surrounding mRNA structure, may open up a world of possibilities in conferring non-native benefits or enhancing existing ones in fermented food products.

Bibliography

[1] Alekhina, O. M. and Vassilenko, K. S. "Translation initiation in eukaryotes: versatility of the scanning model." *Biochemistry.* 77.13 (Dec. 2012), pp. 1465–77. DOI: 10.1134/S0006297912130056.

[2] Antúnez de Mayolo, A. et al. "Multiple start codons and phosphorylation result in discrete Rad52 protein species." *Nucleic acids research* 34.9 (Jan. 2006), pp. 2587–97. DOI: 10.1093/nar/gkl280.

[3] Beerman, R. W. and Jongens, T. a. "A non-canonical start codon in the Drosophila fragile X gene yields two functional isoforms." *Neuroscience* 181 (May 2011), pp. 48–66. DOI: 10.1016/j.neuroscience.2011.02.029.

[4] Ben-Yehezkel, T. et al. "Mapping the translation initiation landscape of an S. cerevisiae gene using fluorescent proteins." *Genomics* 102.4 (Oct. 2013), pp. 419–29. DOI: 10.1016/j.ygeno.2013.05.003.

[5] Brondke, H., Schmitz, B., and Doerfler, W. "Nucleotide sequence comparisons between several strains and isolates of human cytomegalovirus reveal alternate start codon usage." *Archives of virology* 152.11 (Jan. 2007), pp. 2035–46. DOI: 10.1007/s00705-007-1026-x.

[6] Byrd, M. P., Zamora, M., and Lloyd, R. E. "Generation of Multiple Isoforms of Eukaryotic Translation Initiation Factor 4GI by Use of Alternate Translation Initiation Codons". 22.13 (2002), pp. 4499–4511. DOI: 10.1128/MCB.22.13.4499.

[7] Cao, X. et al. "Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus." *Journal of virology* 69.1 (Jan. 1995), pp. 560–3.

[8] Chenik, M., Chebli, K., and Blondel, D. "Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism." *Journal of virology* 69.2 (Feb. 1995), pp. 707–12.

[9] Dmitriev, S. E. et al. "Efficient cap-dependent translation of mammalian mRNAs with long and highly structured 5-untranslated regions in vitro and in vivo". *Molecular Biology* 43.1 (Feb. 2009), pp. 108–113. DOI: 10.1134/S0026893309010154.

[10] Elias, R. J., Kellerby, S. S., and Decker, E. a. "Antioxidant activity of proteins and peptides." *Critical reviews in food science and nutrition* 48.5 (May 2008), pp. 430–41. DOI: 10.1080/10408390701425615.

[11] Gibson, D. G. "Enzymatic assembly of overlapping DNA fragments." *Methods in enzymology* 498 (Jan. 2011), pp. 349–61. DOI: 10.1016/B978-0-12-385120-8.00015-2.

[12] Gilbert, W. V. "Alternative ways to think about cellular internal ribosome entry." *The Journal of biological chemistry* 285.38 (Sept. 2010), pp. 29033–8. DOI: 10.1074/jbc.R110.150532.

[13] Gold, L. et al. "TRANSLATIONAL INITIATION". *Annual Review of Microbiology* 35 (1981), pp. 365–403.

[14] Gu, S. et al. "Alternative Translation Initiation of Human Regulators of G-Protein Signaling-2 Yields a Set of Functionally" (2008). DOI: 10.1124/mol.107.036285..

[15] Gu, W., Zhou, T., and Wilke, C. O. "A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes." *PLoS computational biology* 6.2 (Feb. 2010), e1000664. DOI: 10.1371/journal.pcbi.1000664.

[16] Gualerzi, C. O. and Pon, C. L. "Initiation of mRNA translation in prokaryotes." *Biochemistry* 29.25 (June 1990), pp. 5881–9.

[17] Gunnery, S. et al. "NUCLEIC ACIDS , PROTEIN SYNTHESIS , AND MOLECULAR GENETICS : Translation of an Uncapped mRNA Involves Scanning Translation of an Uncapped mRNA Involves Scanning *" (1997).

[18] Hauryliuk, V. V. "GTPases of the prokaryotic translation apparatus". *Molecular Biology* 40.5 (Oct. 2006), pp. 688–701. DOI: 10.1134/S0026893306050025.

[19] Honoré, E. "Alternative translation initiation further increases the molecular and functional diversity of ion channels." *The Journal of physiology* 586.Pt 23 (Dec. 2008), pp. 5605–6. DOI: 10.1113/jphysiol.2008.165019.

[20] Jackson, R. J. and Kaminski, A. “Internal initiation of translation in eukaryote: The picornavirus paradigm and beyond”. *RNA* 1.10 (1995), pp. 985–1000.

[21] Juanes, M. E. et al. “Biosynthesis of prion protein nucleocytoplasmic isoforms by alternative initiation of translation.” *The Journal of biological chemistry* 284.5 (Jan. 2009), pp. 2787–94. DOI: 10.1074/jbc.M804051200.

[22] Kochetov, a. V. “Alternative translation start sites and their significance for eukaryotic proteomes”. *Molecular Biology* 40.5 (Oct. 2006), pp. 705–712. DOI: 10.1134/S0026893306050049.

[23] Kozak, M. “The scanning model for translation: an update.” *The Journal of cell biology* 108.2 (Feb. 1989), pp. 229–41.

[24] Kozak, M. “Initiation of translation in prokaryotes and eukaryotes”. *Gene* 234.2 (July 1999), pp. 187–208. DOI: 10.1016/S0378-1119(99)00210-3.

[25] Kozak, M. “Regulation of translation via mRNA structure in prokaryotes and eukaryotes.” *Gene* 361 (Nov. 2005), pp. 13–37. DOI: 10.1016/j.gene.2005.06.037.

[26] Matsuda, D. and Dreher, T. W. “Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA”. c (2006), pp. 1338–1349. DOI: 10.1261/rna.67906.In.

[27] Matsuda, D. and Dreher, T. W. “Cap- and initiator tRNA-dependent initiation of TYMV polyprotein synthesis by ribosomes : Evaluation

of the Trojan horse model for TYMV RNA translation". 6 (2007), pp. 129–137. DOI: 10.1261/rna.244407.mRNAs.

[28] McCarthy, J. E. and Brimacombe, R. "Prokaryotic translation: the interactive pathway leading to initiation." *Trends in genetics : TIG* 10.11 (Nov. 1994), pp. 402–7.

[29] Ozin, A. J. et al. "Alternative Translation Initiation Produces a Short Form of a Spore Coat Protein in Bacillus subtilis". 183.6 (2001), pp. 2032–2040. DOI: 10.1128/JB.183.6.2032.

[30] Pestova, T. V. and Kolupaeva, V. G. "The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection". 3 (2002), pp. 2906–2922. DOI: 10.1101/gad.1020902.5.

[31] Racine, T. et al. "Leaky scanning and scanning-independent ribosome migration on the tricistronic S1 mRNA of avian reovirus." *The Journal of biological chemistry* 282.35 (Aug. 2007), pp. 25613–22. DOI: 10.1074/jbc.M703708200.

[32] Seo, S. W. et al. "Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency." *Metabolic engineering* 15 (Jan. 2013), pp. 67–74. DOI: 10.1016/j.ymben.2012.10.006.

[33] Sonenberg, N. "Picornavirus RNA translation continues to surprise." *Trends in genetics : TIG* 7.4 (Apr. 1991), pp. 105–6.

[34] Song, K. Y. et al. "Differential use of an in-frame translation initiation codon regulates human mu opioid receptor (OPRM1)." *Cellular and*

molecular life sciences : CMLS 66.17 (Sept. 2009), pp. 2933–42. DOI: 10.1007/s00018-009-0082-7.

[35] Suzek, B. E. et al. "A probabilistic method for identifying start codons in bacterial genomes." *Bioinformatics (Oxford, England)* 17.12 (Dec. 2001), pp. 1123–30.

[36] Terenin, I. M. et al. "Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2." *Nature structural & molecular biology* 15.8 (Aug. 2008), pp. 836–41. DOI: 10.1038/nsmb.1445.

[37] Thomas, D. et al. "Alternative translation initiation in rat brain yields K2P2.1 potassium channels permeable to sodium." *Neuron* 58.6 (June 2008), pp. 859–70. DOI: 10.1016/j.neuron.2008.04.016.

[38] Touriol, C. et al. "Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons". *Biology of the Cell* 95.3-4 (May 2003), pp. 169–178. DOI: 10.1016/S0248-4900(03)00033-9.

[39] Wamboldt, Y. et al. "Participation of leaky ribosome scanning in protein dual targeting by alternative translation initiation in higher plants." *The Plant cell* 21.1 (Jan. 2009), pp. 157–67. DOI: 10.1105/tpc.108.063644.

[40] Weekes, D. B. et al. "Attenuation of Disease Phenotype Through Alternative Translation Initiation in Low-Penetrance Retinoblastoma". 28.September 2006 (2007), pp. 159–167. DOI: 10.1002/humu.

[41] Williams, M. a. and Lamb, R. a. "Effect of mutations and deletions in a bicistronic mRNA on the synthesis of influenza B virus NB and NA glycoproteins." *Journal of virology* 63.1 (Jan. 1989), pp. 28–35.

[42] Zhou, L. and Elias, R. J. "Factors influencing the antioxidant and pro-oxidant activity of polyphenols in oil-in-water emulsions." *Journal of agricultural and food chemistry* 60.11 (Mar. 2012), pp. 2906–15. DOI: 10.1021/jf204939h.

Appendix A

Oligonucleotides

Table A.1: Primer sequences

Name	Sequence
Sequencing Primers	
JCFTV F2	GCTGAACGGTCTGGTTATAGG
JCFTV R	CTGATACCGCTCGCCGCAGCCGAAC
JC311 SGFP SEQR	ACCATAGGTCAGGGTGGTAACC
Vector PCR Primers	
vec1 R	CTAGGACTGAGCTAGCCGTC
vec4 F	GCTACTAGAGCCAGGCATCAAATAA
gBlock Primers	
gBlock1 F	GACGGCTAGCTCAGTCC
gBlock1 R	GGTGGTGTTGTTACCGTT
gBlock2 F	ACGTGTTCTTAACTTCGA
gBlock2 R	GGTGATGTTAATGGTCATAAATTCA
gBlock3 F	CTGGTGGAACTGGATGGTGATG
gBlock3 R	CTGGAATATAATTTTAACAGCCATAACGTG
gBlock4 F	GGGCCATAAACTGGAATATAATTTTAACAG
gBlock4 R	CTACTAGAGCCAGGCATCAAATAA

Table A.2: Oligonucleotide sequences

Name	Sequence
Test Cassette Oligos	
7N R100G5K F	CTAGATATTACCGTCACTACTCCGGAATACCA TTCAAGGATGTTTTTCAATGAA
7N R100G5K R	GATCTTCATTGAAAAACATCCTTGAATGGTAT TCCGGAGTAGTGACGGTAATAT
7N R1KG5K F	CTAGATAAACTATAAAAGAGAAGTAAGCAAGA CAAAGGATGTACAACTATGAA
7N R1KG5K R	GATCTTCATAGTTGTACATCCTTTGTCTTGCT TACTTCTCTTTTATAGTTTAT
7N R10KG5K F	CTAGATAGTCTCTAAGGGGCTAGAGAGGAGTA GGTAAGGATGGAAGACGATGAA
7N R10KG5K R	GATCTTCATCGTCTTCCATCCTTACCTACTCC TCTCTAGCCCCTTAGAGACTAT
7N R100KG5K F	CTAGAAGAAAGAATCAAAAATACATAAGGAGA GTAAGGTATGGTAGGCGATGAA
7N R100KG5K R	GATCTTCATCGCCTACCATACCTTACTCTCCT TATGTATTTTTGATTCTTTCTT
Control Cassette Oligos	
7N R100G0 F	CTAGATACCTAAACGAAAGACAGCGCAGCGCA AGGCGTATGGCGAGCTACGAA
7N R100G0 R	GATCTTCGTAGCTCGCCATACGCCTTGCGCTG CGCTGTCTTTCGTTTAGGTAT
7N R1KG0 F	CTAGAATACGAAAATCACTTTAAAATAGTCC AAACGTTAATGGCGAGCTACGAA
7N R1KG0 R	GATCTTCGTAGCTCGCCATTAACGTTTGGAC TATTTTAAAGTGATTTTCGTATT
7N R10KG0 F	CTAGAAAGGCGTTCTAATAAATATAAGAAGGG CAACATGGCGAGCTACGAA
7N R10KG0 R	GATCTTCGTAGCTCGCCATGTTGCCCTTCTTA TATTTATTAGAACGCCTTT
7N R100KG0 F	CTAGAATACATACAGCGAAAAAATAAGGAGTAA TATTAATGGCGAGCTACGAA
7N R100KG0 R	GATCTTCGTAGCTCGCCATTAATATTACTCCTT ATTTTTTCGCTGTATGTATT

Appendix B

Plasmid Construction

Table B.1: gBlock sequences

Name	Sequence
gBlock Sequences	
gBlock 1	GAGGAGATATATATATGACGGCTAGCTCAGTCCT
	AGGTACAGTGCTAGCTTCTAGAAAAAAAATAAGG
	AGGTAAATGGCGAGCTATGAAGATCTTATCAAAG
	AGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTC
	CGTTAACGGTCACGAGTTCGAAATCGAAGGTGAA
	GGTGAAGGTCGTCCGTACGAAGGTACCCAGACCG
	CTAAACTCAAAGTTACCAAAGGTGGTCCGCTGCC
	GTTCGCTTGGGACATCCTGTCCCCGCAGTTCCAG
	TACGGTTCCAAAGCTTACGTTAAACACCCGGCTG
	ACATCCCGGACTACCTCAAACTGTCCTTCCCGGA

	AGGTTTCAAATGGGAACGTGTTCTTAACTTCGAA
	GACGGTGGTGTTGTTACCGTTATATATATATCTCCTC
gBlock 2	GAGGAGATATATATATACGTGTTCTTAACTTCGA
	AGACGGTGGTGTTGTTACCGTTACCCAGGACTCT
	TCCCTGCAAGACGGTGAGTTCATCTACAAAGTTA
	AACTGCGTGGTACCAACTTCCCGTCCGACGGTCC
	GGTTATGCAGAAAAAAACCATGGGTTGGGAAGCT
	TCCACCGAACGTATGTACCCGGAAGACGGTGCTC
	TCAAAGGTGAAATCAAAATGCGTCTCAAACTTAA
	AGACGGTGGTCACTACGACGCTGAAGTTAAAACC
	ACCTACATGGCTAAAAAACCGGTTCAGCTGCCGG
	GTGCTTACAAAACCGACATCAAACTGGACATCAC
	CTCCCACAACGAAGACTACACCATCGTTGAACAG
	TACGAACGTGCTGAAGGTCGTCACTCCACCGGTG
	CTTAATAAACGTAAAGGCGAAGAACTGTTTACCG
	GTGTGGTTCCGATTCTGGTGGAACTGGATGGTGA
	TGTTAATGGTCATAAATTCAATATATATATCTCCTC
gBlock 3	GAGGAGATATATATATCTGGTGGAACTGGATGGT
	GATGTTAATGGTCATAAATTCAGCGTTCGTGGTG
	AAGGCGAAGGTGATGCCACGAATGGTAAACTGAC
	CCTGAAATTTATCTGCACCACAGGTAAACTGCCG
	GTTCCGTGGCCGACCCTGGTTACCACCCTGACCT
	ATGGTGTTCAGTGTTTCGCACGTTATCCGGATCA

	TATGAAACAGCACGATTTCTTTAAAAGCGCCATG
	CCGGAAGGTTATGTTCAGGAACGTACCATTAGCT
	TTAAAGATGACGGCACCTATAAAACCCGTGCCGA
	AGTTAAATTCGAAGGCGATACCCTGGTGAATCGT
	ATCGAACTGAAAGGCATCGATTTTAAAGAGGATG
	GTAATATCCTGGGCCATAAACTGGAATATAATTT
	TAACAGCCATAACGTGATATATATATCTCCTC
gBlock 4	GAGGAGATATATATATGGGCCATAAACTGGAATA
	TAATTTTAACAGCCATAACGTGTATATCACCGCA
	GATAAACAGAAAAACGGCATTAAAGCGAACTTTA
	AAATCCGCCATAATGTGGAAGATGGTAGCGTTCA
	GCTGGCAGATCATTATCAGCAGAATACGCCGATC
	GGTGATGGTCCGGTTCTGCTGCCGGATAATCATT
	ATCTGAGCACCCAGAGCGTTCTGAGTAAAGATCC
	GAATGAAAAACGTGATCACATGGTGCTGTTAGAG
	TTCGTTACCGCAGCAGGTATTACACATGGTATGG
	ATGAACTGTATAAATGATAAGGCCGGCCGTGCTA
	GTGTAGATCGCTACTAGAGCCAGGCATCAAATAA
	TATATATATCTCCTC

Table B.2: pR0sfG0 plasmid sequence

TACGTGCCGATCAACGTCTCATTTTCGCCAGATATCCTCGAGCGCGGAATTCCCTAGGGGATCCGTTGACGGCTAG

CTCAGTCCTAGGTACAGTGCTAGCTTCTAGAAAAAAAATAAGGAGGTAAATGGCGAGCTATGAAGATCTTATCAAA

GAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTTCCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAG

GTCGTCCGTACGAAGGTACCCAGACCGCTAAACTCAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTTGGGACAT

CCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCGGCTGACATCCCGGACTACCTCAAACTG

TCCTTCCCGGAAGGTTTCAAATGGGAACGTGTTCTTAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACT

CTTCCCTGCAAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGACGGTCCGGTTAT

GCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTACCCGGAAGACGGTGCTCTCAAAGGTGAAATC

AAAATGCGTCTCAAACTTAAAGACGGTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGG

TTCAGCTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGACTACACCATCGTTGA

ACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTTAATAAACGTAAAGGCGAAGAACTGTTTACCGGTGT

GGTTCCGATTCTGGTGGAACTGGATGGTGATGTTAATGGTCATAAATTCAGCGTTCGTGGTGAAGGCGAAGGTGAT

GCCACGAATGGTAAACTGACCCTGAAATTTATCTGCACCACAGGTAAACTGCCGGTTCCGTGGCCGACCCTGGTTA

CCACCCTGACCTATGGTGTTCAGTGTTTCGCACGTTATCCGGATCATATGAAACAGCACGATTTCTTTAAAAGCGC

CATGCCGGAAGGTTATGTTCAGGAACGTACCATTAGCTTTAAAGATGACGGCACCTATAAAACCCGTGCCGAAGTT

AAATTCGAAGGCGATACCCTGGTGAATCGTATCGAACTGAAAGGCATCGATTTTAAAGAGGATGGTAATATCCTGG

GCCATAAACTGGAATATAATTTTAACAGCCATAACGTGTATATCACCGCAGATAAACAGAAAAACGGCATTAAAGC

GAACTTTAAAATCCGCCATAATGTGGAAGATGGTAGCGTTCAGCTGGCAGATCATTATCAGCAGAATACGCCGATC

GGTGATGGTCCGGTTCTGCTGCCGGATAATCATTATCTGAGCACCCAGAGCGTTCTGAGTAAAGATCCGAATGAAA

AACGTGATCACATGGTGCTGTTAGAGTTCGTTACCGCAGCAGGTATTACACATGGTATGGATGAACTGTATAAATG

ATAAGGCCGGCCGTGCTAGTGTAGATCGCTACTAGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACT

GGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCC

TTTCTGCGTTTATACAATTGACTAGTCTGCAGAAAAGCGGCCGCGGCATCAAATAAAACGAAAGGCTCAGTCGAAA

GACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGCCCTAGAGCA

TGCGCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAA

CGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTC

CATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTAT

AAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT

GTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTC

GTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTC

TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA

TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGC

GCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG

GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTAC

GGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGCCCGGGTGCTTGGATTCTCACCAA

TAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTATCAA

CAGGAGTCCAAGCGGTACCGATATCAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAG

CATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCT

TGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGG

TGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTC

ACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGAT

GAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTT

TCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTG

CTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACT

GACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCT

CCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATG

GTGAAAGTTGGAACCTCT

Magenta text indicates restriction enzyme sites, left to right, XbaI and BglII. Blue text represents the RBS. Red and green ATG sequences represent RFP and sfGFP start codons, respectively.

ACADEMIC VITA

Hannah R. Jepsen-Burger

hrj5014@psu.edu

EDUCATION

B.S. in Food Science — Expected May 2014
Minor in **Agribusiness Management**
Concentration in Microbiology and Quality Assurance
Schreyer Honors College
The Pennsylvania State University, University Park, PA

Oley Valley High School — Graduated Spring 2010

AWARDS

Schreyer Honors College Undergraduate Research Grant
Department of Food Science Speizer Undergraduate Research Award
College of Agriculture Academic Excellence Scholarship

EXPERIENCE

ConAgra Mills Quality Assurance Associate — Starting Summer 2014
Laurel Weilage – Quality Assurance Manager

Penn State Dept. of Biological Engineering Undergraduate Researcher — Summer 2012-present
Dr. Thomas Richard-Supervisor
- conduct synthetic biology research on protein manipulation using multiple start codon selection

CERTIFICATIONS

HACCP certified
ServSafe certified

PROFESSIONAL PRESENTATIONS

Institute of Biological Engineering Annual Conference Presenter — March 2013
"The Effect of Multiple Start Codons on Protein Translation in Prokaryotes"

ACTIVITIES

International Genetically Engineered Machines Competition (iGEM) — Summer 2012-present
- Penn State iGEM Team 2012 captain
- conducted synthetic biology research
- current undergraduate advisor

Penn State Food Science Club — Fall 2011-present
Gamma Sigma Delta Agricultural Honor Society — Spring 2013-present
Springfield THON — Fall 2011-present
Community Food Security Student Organization — Fall 2010-2011