The Dissertation Committee for Paige Sanders Spencer Certifies that this is the approved version of the following dissertation:

TRANSLATION RATE IS GENETICALLY ENCODED AND INFLUENCES PROTEIN FOLDING

Committee:

José M. Barral, M.D., Ph.D., Supervisor

Andrés F. Oberhauser, Ph.D., Chair

Darren F. Boehning, Ph.D.

Henry F. Epstein, M.D.

Vincent J. Hilser, Ph.D.

Muge Kuyumcu-Martinez, Ph.D.

Dean, Graduate School

TRANSLATION RATE IS GENETICALLY ENCODED AND INFLUENCES PROTEIN FOLDING

by
Paige Sanders Spencer, B.S.

Dissertation
Presented to the Faculty of The University of Texas Graduate School of Biomedical Sciences at Galveston
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Approved by the Supervisory Committee

José M. Barral
Andrés F. Oberhauser
Darren F. Boehning
Henry F. Epstein
Vincent J. Hilser
Muge Kuyumcu-Martinez

August, 2012
Galveston, Texas

Key words: protein folding, silent mutations, translation rates, molecular chaperones

DEDICATION

I dedicate it all to my wonderfully supportive husband, Bobby, and our beautiful baby boy, Dale. You have enriched my life and this work more than you know.
My cup overflows.

ACKNOWLEDGEMENTS

I offer my heartfelt gratitude and admiration to Dr. José M. Barral who has been a true mentor to me. From the beginning, you saw something in me that I had not yet realized was there. Your support, guidance, wisdom, and patience have greatly contributed to the scientist and individual that I am today. I chose you as my mentor because I thought I would learn more from you than anyone else, and I certainly have.

I also thank my committee members for their time and contributions to my research. Thank you, Dr. Andrés Oberhauser, for chairing my committee, for diligently keeping me on track, and for helpful advising. Thank you, Dr. Darren Boehning, for the motivational chats, energetic support, and suggestions. Thank you, Dr. Henry Epstein, for providing my project with a unique perspective that could only come from years of scientific inquisition and experience. Thank you, Dr. Muge Martinez, for providing fresh insight into my project. Thank you, Dr. Vincent Hilser for a brief official and continuing unofficial mentorship; you have helped to create a very cool project, and your youthful enthusiasm for our work is contagious.

I would like to thank Drs. John F. Anderson and Efrain Siller for technical assistance, including helping me reach items stored high in the lab. Thank you, Dr. James Wrabl, for providing the bioinformatics and thermodynamics portions of this work and for useful discussions. I would like to thank Dr. Darrell Carney for his support of my project. I would also like to thank the Welch Foundation and Sealy Center for Structural and Molecular Biology for financial support. Thank you, Dr. Coppenhaver and the

graduate school staff for ushering me along in this journey. I am also grateful to Dr. Tracy Toliver-Kinsky, my Biochemistry and Molecular Biology advisor, for meaningful advice and guidance.

A special note of gratitude goes to Carol Kerr for helping me find balance between school and family life.

I owe a lifetime gratitude to my parents, Jeff and Cheryl Sanders, for instilling drive and ambition in me and for providing the essentials that have gotten me this far.

Finally, I would like to thank God, whose actions we call science. Thank you for providing us with so many biological questions and the curiosity and ability to answer some of them.

TRANSLATION RATE IS GENETICALLY ENCODED AND INFLUENCES PROTEIN FOLDING

Publication No._____

Paige S. Spencer, PhD
The University of Texas Graduate School of Biomedical Sciences at Galveston, 2012
Supervisor: José M. Barral

The degeneracy of the genetic code allows most amino acids to be encoded by multiple codons. The distribution of these so-called synonymous codons among protein coding sequences is not random and multiple theories have arisen to explain the biological significance of such non-uniform codon selection. Many ideas revolve around the notion that certain codons allow for faster or more efficient translation, whereas the presence of others result in slower translation rates. The presence of these different types of codons along a message is postulated in turn to confer variable rates of emergence of the nascent polypeptide from the ribosome, which may influence its capacity to fold towards the native state, among other properties. Previous studies have reported conflicting results with regards to whether certain kinds of codons correlate or not with particular structural or folding properties of the encoded protein. We believe this has arisen, in part, because different criteria have been traditionally used for predicting whether a codon will be translated quickly or slowly in a given organism, including its frequency of occurrence among highly expressed genes and the concentration of tRNA species capable of decoding it, which do not always correlate. We have developed a metric to predict organism-specific polypeptide elongation rates of any mRNA based on whether each codon is decoded by tRNAs capable of Watson-Crick, non-Watson-Crick or both types of interactions. We demonstrate by pulse-chase analyses in living *E. coli* cells that sequence engineering based on these concepts predictably modulates translation rates due to changes in polypeptide elongation and show that such alterations significantly impact the folding of proteins of eukaryotic origin. We also demonstrate that sequence harmonization based on expression-host tRNA content designed to mimic ribosome movement of the original organism can significantly increase the folding of the encoded polypeptide. Additionally, we show that the rate at which a polypeptide emerges from the ribosome can affect co-translational chaperone binding, which may explain some of the observed changes in folding efficiencies. We have also begun to identify certain folding regions that may be more sensitive than others to translation speed modulation.

This body of work could provide insight into how synonymous nucleotide substitutions result in altered protein function and disease.[1]

[1] Portions of this abstract are reproduced from Journal of Molecular Biology, 2012, doi:10.1016/j.jmb.2012.06.010. Silent Substitutions Predictably Alter Translation Elongation Rates and Protein Folding Efficiencies. Spencer, P.S. *et al.*. with permission.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xii

LIST OF ABBREVIATIONS xiv

CHAPTER 1: INTRODUCTION 1

Protein synthesis and the redundancy of the genetic code 1

Polypeptide elongation rate determinants 4

Codon bias does not necessarily determine polypeptide elongation rate 9

Polypeptide elongation rates and protein folding 13

CHAPTER 2: SILENT SUBSTITUTIONS PREDICTABLY ALTER TRANSLATION ELONGATION RATES AND PROTEIN FOLDING EFFIENCIES 20

2.1 INTRODUCTION 20

2.2 RESULTS 22

2.3 DISCUSSION 35

2.4 MATERIALS AND METHODS 35

Prediction of relative translation elongation rates 35

Coding sequence engineering 37

Strains and growth conditions 38

Pulse-chase analysis and calculation of polypeptide elongation rates 39

Quantitative PCR assays 41

Recombinant protein production 41

Determination of protein solubility 42

Determination of luciferase activity 42

CHAPTER 3: POLYPEPTIDE ELONGATION RATES INFLUENCE CHAPERONE RECRUITMENT TO THE NASCENT POLYPEPTIDE 43

3.1 INTRODUCTION 43

3.2 RESULTS 45

3.3 DISCUSSION 52
3.4 MATERIALS AND METHODS 54
Coding sequence engineering 54
Strains and growth conditions 54
Recombinant protein production 54
Gel filtration chromatography 55
Pull down experiments 55
Translation inhibition experiments 55

CHAPTER 4: ARE CERTAIN REGIONS OF A NASCENT POLYPEPTIDE CHAIN MORE MISFOLDING SENSITIVE TO ALTERATIONS IN TRANSLATION RATE? 57
4.1 INTRODUCTION 57
4.2 RESULTS 61
Sequence manipulation predictably modulates GFP folding 63
Testable relationship exists between translation rate and DTdSconf 64
Design strategy to test translation rate and DTdSconf relationship 65
Global and local rate decelerations increase GFP folding efficiency 68
4.3 DISCUSSION 69
4.4 MATERIALS AND METHODS 71
Strain and growth conditions 71
Recombinant protein production 71
Recoding GFP mRNA 72
Determination of green fluorescence 73
Determination of GFP accumulation 73
Average Pearson correlation coefficient for codability and thermodynamic parameters 73

CHAPTER 5: SUMMARY AND PERSPECTIVES 75

Appendix 84
Luciferase DNA and protein sequences 84

VITA 98

HONORS 100

ABSTRACTS 101

PUBLICATIONS.....103

LIST OF TABLES

Table 1.1 Differences in tRNA gene content for *E. coli* and *H. sapiens* 11

Table 1.2 Overview of studies linking mRNA codon composition with protein folding. 19

Table 2.6 tRNA gene content and biased codon usage frequencies for *E. coli* and *D. melanogaster*. 27

LIST OF FIGURES

Figure 1.1 The nature of the codon:anticodon interaction influences translation elongation 8

Figure 2.1 The distribution of genes encoding tRNAs of different decoding capacities vary among archaea, bacteria, and eukarya. 21

Figure 2.2 Utilization of tRNA gene information and nature of codon:anticodon base pairing allows the prediction of relative translation elongation rates. 23

Figure 2.3 Avoidance of wobble-based interactions during mRNA decoding results in acceleration of translation elongation rates *in vivo*. 24

Figure 2.4 Steady-state accumulation of mRNA synthesized from the wild type and sequence-engineered Luc constructs..... 25

Figure 2.5 Comparison of fits between constant and variable theoretical elongation rates for proteins synthesized from wild type and sequence-engineered Luc constructs. 26

Figure 2.7 Autoradiogram of an SDS-PAGE from a pulse-chase experiment with the Luc_{slow} construct. 29

Figure 2.8 Effects of silent substitutions on observed translation rates are not due to changes in translation initiation. 30

Figure 2.9 Synonymous sequence-based acceleration influences the folding of the encoded polypeptide. 31

Figure 2.10 Mimicking eukaryotic tRNA population via synonymous sequence engineering of mRNA enhances folding efficiency of recombinant proteins in a bacterial host. 33

Figure 3.1 Synonymous sequence based deceleration increases the folding efficiency of the encoded polypeptide. 47

Figure 3.2 Elongation rate determines luciferase migration pattern. 48

Figure 3.3 Test of molecular chaperone co-elution pattern with Luc synthesized from various constructs..... 49

Figure 3.4 Luc pulldown reveals increased TF, DnaJ, and DnaK binding to Luc_{slow}..... 50

Figure 3.5 Luc_{slow} post-translational folding regime is enhanced relative to faster constructs. 52

Figure 3.6 Proposed mechanism explaining chaperone-assisted increased folding efficiency..... 53

Figure 4.1 Average Pearson correlation coefficient (r) between codability and each thermodynamic parameter for a large set of human proteins. 62

Figure 4.2 Western blot of GFP accumulation produced by each recoded construct..... 63

Figure 4.3 Relative fluorescence emission of GFP variants. 64

Figure 4.4 Plot displaying DTdSconf and codability of GFP protein and relative elongation rate of GFP mRNA expressed in *E. coli*. 66

Figure 4.5 Construction strategy for GFP chimeras to assess the impact of speed of folding of specific regions. 67

Figure 4.6 Translation speed profiles for GFP chimeras. 68

Figure 4.7 Recoded regions highlighted in GFP structure..... 69

Figure 4.8 Recoding translation speed affects folding of GFP chimeras. 70

LIST OF ABBREVIATIONS

~	approximately
º C	degrees Celsius
µg	microgram
µl	microliter
µM	micromolar
Å	Angstrom
A	adenosine
A_{600}	Absorbance at 600 nanometer
aa-tRNA	aminoacyl-tRNA
aa/s	amino acids per second
A site	aminoacyl-tRNA site
Arg	arginine
Asn	asparagine
B. subtilus	*Bacillus subtilis*
BSA	bovine serum albumin
C	cytosine
C. elegans	*Caenorhabditis elegans*
Ch	chimera
dG	Gibbs free energy
dH	enthalpy
D. melanogaster	*Drosophila melanogaster*
DdG	free energy of protein in the denatured state
DdHap	apolar enthalpy protein in the denatured state
DdHpol	polar enthalpy of protein in the denatured state
DTdSconf	conformational entropy of protein in the denatured state
E site	exit site
EF-tu	elongation factor tu
E. coli	*Escherichia coli*
G	guanine
GFP	green fluorescent protein
HeLa	Henrietta Lack's immortal cervical carcinoma cells
I	inosine
kDa	kiloDaltons
LB	Luria broth
Luc	Luciferase from *Photinus pyralis*
Luc_{cbf}	Luc recoded with most frequent *E. coli* codons
Luc_{re}	Luc recoded to recapitulate *D. melanogaster* tRNA population
min	minutes
Met	methionine
mg	milligram

ml	milliliter
mM	millimolar
mRNA	messenger ribonucleic acid
N	nucleic acid residue
NdG	free energy of protein in the native state
NdHap	apolar enthalpy of protein in the native state
NdHpol	polar enthalpy protein in the native state
NMR	nuclear magnetic resonance
NNN	nucleic acid residues composing a codon/anticodon
NTdSconf	conformational entropy of protein in the native state
PBS	phosphate-buffered saline
P site	peptidyl transfer site
S. cerevisiae	*Saccharomyces cerevisiae*
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
TdS	entropy
T_m	melting temperature
tRNA	transfer ribonucleic acid
Trp	Tryptophan
U	Uracil
WT	wild type
wt	weight (grams)
vol	volume (ml)

CHAPTER 1: INTRODUCTION[2]

PROTEIN SYNTHESIS AND THE REDUNDANCY OF THE GENETIC CODE

The transfer of genetic information into protein products is termed translation (Figure 1; for detailed reviews on the mechanisms of translation, please see (Bashan & Yonath, 2008; Schmeing & Ramakrishnan, 2009; Steitz, 2008)). Messenger RNA (mRNA), transcribed from DNA, is translated into protein by a template driven process. The template is composed of a specific combination of 61 trinucleotide codons which encode 20 amino acids. This genetic code is common to most organisms and is referred to as redundant because all amino acids, with the exception of tryptophan and methionine, are encoded by more than one codon (termed synonymous codons). Codons are read by adaptor molecules called transfer RNA (tRNA) that bear complementary (cognate) trinucleotide sequences, or anticodons. This reading or decoding of the codon occurs by recognition through base pairing, where at least two hydrogen bonds are formed between each of the nucleotides pairs that make up the codon:anticodon minihelix. Only one position of the codon:anticodon minihelix allows pairing that can deviate from standard Watson-Crick (G:C and A:U) interactions. In the third nucleotide of the codon and the first nucleotide of the anticodon, the so-called wobble position, nonstandard base pairing can occur and results in altered base stacking conformations

[2]This chapter is reproduced from Computational and Structural Biotechnology Journal, 2012, 1 (1): e201204006. doi: http://dx.doi.org/10.5936/csbj.201204006. Genetic code redundancy and its influence on the encoded polypeptides. Spencer, P.S. and Barral, J.M. with permission.

that are different from that of Watson-Crick pairing yet remain within the conformational constraints of the glycosidic bonds (Crick, 1966).

Interestingly, there are three conserved nucleotides in the bacterial 70S ribosome which maintain decoding fidelity by monitoring the conformation of the bases in the codon:anticodon minihelix (Schmeing & Ramakrishnan, 2009). The monitoring of base conformations is much more stringent in the first two nucleotide positions of the minihelix than in the wobble position, allowing for flexibility or wobble in the decoding of this position (Schmeing & Ramakrishnan, 2009). For example, nonstandard pairing of G:U and U:G, in which one less hydrogen bond is formed compared to standard G:C and C:G pairing, is allowed only in this position. Furthermore, post-transcriptional deamination of adenosine to inosine in the first anticodon position (INN) expands the decoding capacity from strictly Watson-Crick (A:U) to other allowed "wobble" base pairing (I:U, I:C, I:A) (Crick, 1966). Adenosine deamination occurs in many eukaryotic ANN anticodons; however, in bacteria, this modification is exclusive to the ACG anticodon of $tRNA^{Arg}$ (Grosjean et al, 2010). There are many other base modifications throughout the tRNA molecule, but these are more variable and will not be considered here. Upon decoding, peptide bond formation is catalyzed in the peptidyl-transferase center of the ribosome and is followed by translocation of the ribosome to the next codon. While diversity exists across evolution in the complexity of the ribosome (Ban et al, 2000; Schmeing & Ramakrishnan, 2009), translation regulation factors (Ban et al, 2000; Schmeing & Ramakrishnan, 2009), and tRNA gene composition (Chan & Lowe, 2009),

the core processes of translation are remarkably conserved and consist of three general steps: initiation, elongation, and termination.

Translation rates are not uniform along an mRNA and vary with the codon composition of the message, since the individual translation rates of codons have been shown to vary by as much as 25-fold (Curran & Yarus, 1989; Sorensen & Pedersen, 1991; Varenne et al, 1984). The non-uniformity of rates has been proposed to depend on tRNA concentration, the nature of base pairing, and/or mRNA secondary structure (Ikemura, 1985; Kudla et al, 2009; Varenne et al, 1984). The former two will be discussed later in this review. A logical assumption is that a stable mRNA secondary structure may hinder or slow translation by either preventing the ribosome from binding or by acting as a speed bump during ribosomal progression. Indeed, the presence of stable mRNA secondary structures in the ribosomal binding site have been shown to largely affect expression levels as a result of interference with translation initiation (Kudla et al, 2009). However, the role of mRNA secondary structure in determining polypeptide elongation rates has been disputed (Sorensen et al, 1989; Stadler & Fire, 2011; Varenne et al, 1984). Once the ribosome has initiated translation, it displays powerful helicase activity capable of disrupting very stable mRNA secondary structures (Tm = 70°C) (Takyar et al, 2005). This suggests that mRNA secondary structure plays an insignificant role in the rate of translation elongation, which is the main process addressed in this review. mRNA secondary structure likely plays a much more significant role in translation initiation and termination rates, which will not be discussed here.

Additionally, most of the material presented in this review pertains to the bacterial ribosome.

POLYPEPTIDE ELONGATION RATE DETERMINANTS

The process of polypeptide elongation occurs by the sequential addition to the growing polypeptide chain of a single amino acid brought to the ribosome by a molecular complex with three constituents: aminoacyl tRNA (aa-tRNA), elongation factor Tu (EF-Tu), and GTP (a so-called ternary complex) bearing the correct or cognate anticodon for the mRNA codon in the ribosomal A site (Figure 1). There are three general steps to the elongation cycle: tRNA selection, peptidyl transfer, and translocation. tRNA selection, or decoding, consists of an initial binding of the ternary complex to the ribosome followed by codon recognition. Then, the GTPase activity of EF-Tu is activated, which subsequently causes GTP hydrolysis, EF-Tu dissociation, and accommodation (Gromadski & Rodnina, 2004). Accommodation is the movement of the amino acid portion of the aa-tRNA in the A site closer to the peptidyl tRNA in the P site for peptidyl transfer to occur (Schmeing & Ramakrishnan, 2009). Following peptidyl transfer, binding of elongation factor G (EF-G) and GTP hydrolysis catalyze the translocation of the ribosome one codon forward, so that the tRNAs now reside in the E and P sites, respectively (Schmeing & Ramakrishnan, 2009). The elongation cycle continues as the codon in newly vacant ribosomal A site awaits the next tRNA arrival. Interestingly, the ribosomal A site is likely seldom vacant and is instead sampled by cognate, near-cognate, and non-cognate tRNAs (Fluitt et al, 2007). These terms, near-cognate and non-cognate, have conventionally been assigned to tRNAs which have single or multiple base

mismatches with a given codon, respectively. However, Plant *et al.* have challenged that a functional definition, namely the ability to form a minihelix with the codon in the ribosomal A site, better distinguishes a near- from a non-cognate (Plant et al, 2007). It is important to note, that as peptidyl transfer and translocation occur much faster, tRNA selection appears to be the rate limiting step of ribosomal progression along the mRNA during polypeptide elongation (Johansson et al, 2008; Uemura et al, 2010; Varenne et al, 1984). Independently, two groups have observed large rate differences in the steps of polypeptide elongation by performing high resolution kinetic studies of the bacterial ribosome *in vitro.* They have determined that the rate of ternary complex GTPase activation in response to codon recognition is the rate limiting step of peptidyl transfer. They found that GTP hydrolysis of the cognate ternary complex occurs 650-fold (Gromadski & Rodnina, 2004) or approximately 116-fold (Lee et al, 2007) faster than the near-cognate one (base mismatch in 1st codon position in these studies). The other measurable rates were similar between cognate and near-cognate tRNAs, with the exception of a faster dissociation of the near-cognate during codon recognition (Gromadski & Rodnina, 2004). Modeling of these kinetic data agrees with a competition for the A site whereby the binding and rejection of a number of near-cognate tRNAs, prior to the binding and accommodation of the cognate tRNA, delays the rate of translation (Chu et al, 2011; Fluitt et al, 2007). The faster rate of cognate anticodon recognition combined with the rapid rejection of the near-cognate anticodon emphasize the role of tRNA selection in determining the rate of polypeptide elongation.

Since the binding of the aa-tRNA-containing ternary complex to the ribosome is essentially a binding reaction, concentration of the cognate tRNA for a particular codon should influence the rate at which the ribosome translates that codon. This has indeed been shown by a examining the correlation between codon translation rates and cognate tRNA concentrations (Varenne et al, 1984). Increasing the concentration of $tRNA^{Trp}$ four-fold by overexpression results in a three-fold increase in translation rate of the corresponding codon, UGG (Curran & Yarus, 1989) (tryptophan is one of only two amino acids which are encoded by a single codon). Most codons can be read by more than one isoacceptor tRNA due to wobble pairing in the third position of the codon and first position of the anticodon (Crick, 1966). Conversely, a single tRNA anticodon can decode various synonymous codons, and these can vary in translation rates. For example, the only two codons encoding glutamate, GAA and GAG, are decoded by a single aa-tRNA species at differing rates of 21.6 and 6.4 codons/second, respectively (Sorensen & Pedersen, 1991) (Figure 1). Similar to GAA and GAG, other *in vivo* measured translation rates of synonymous codons read by identical aa-tRNAs show that those with Watson-Crick pairing in the wobble position are translated faster than those with wobble pairing in every instance (Curran & Yarus, 1989; Sorensen & Pedersen, 1991). When more than one codon is translated by a single tRNA, the only difference is the nature of the base pairing and base stacking between the third codon position and the first anticodon position. The different rates observed clearly demonstrate that base pairing in the wobble position, in addition to tRNA concentration, determines codon translation rate. Recent ribosomal profiling has solidly corroborated this effect on *in vivo* rates in *C. elegans* and

HeLa cells by showing genome wide that ribosomes occupy wobble read codons for 50% longer than Watson-Crick read codons (Stadler & Fire, 2011). Furthermore, out of all NNC and NNU codons, the former are translated faster in *C. elegans* and HeLa cells. This result agrees well with what has been reported previously in *E. coli* (Curran & Yarus, 1989). Interestingly, all NNC/NNU codon pairs are synonymous and are decoded by identical tRNAs by near-Watson-Crick (I:C) or wobble pairing (I:U) anticodons. Where comparison was possible, the difference in ribosomal occupancy was greater between Watson-Crick and wobble than near-Watson-Crick and wobble (Stadler & Fire, 2011), implying that rate of codon recognition can be ranked as follows: Watson-Crick > near-Watson-Crick > wobble.

What might be the advantages that organisms derive from being capable of modulating their translation elongation rates? In addition to enhancing the ability of individual segments of a polypeptide to fold (or avoid misfolding) during translation (please see below), *global* regulation of these rates might be greatly beneficial to cells whose growth is generally regulated by protein synthesis rates according to the "growth optimization model"(Ehrenberg & Kurland, 1984). It is well known that the process of translation is not absolutely accurate (Kurland et al, 1996). Yet, various mutations in the bacterial translational apparatus can result in so-called hyperaccurate protein synthesis, where significantly fewer mistakes are made during translation. However, these mutations result in considerably slower rates of polypeptide elongation. In other words, in these mutants, accuracy is achieved at the expense of speed. Thus, it can be concluded that wild type polypeptide elongation rates are a compromise between accuracy and

speed. In circumstances where nutrient availability is limited (and growth is restricted), the cell might need to decrease the production of proteins, yet ensure that those that are synthesized are relatively error free. In opposite circumstances, cells might take advantage of ample nutrients and not be particularly concerned about incorporating the wrong amino acids into their proteins, as these will be "diluted out" as cells grow and divide.

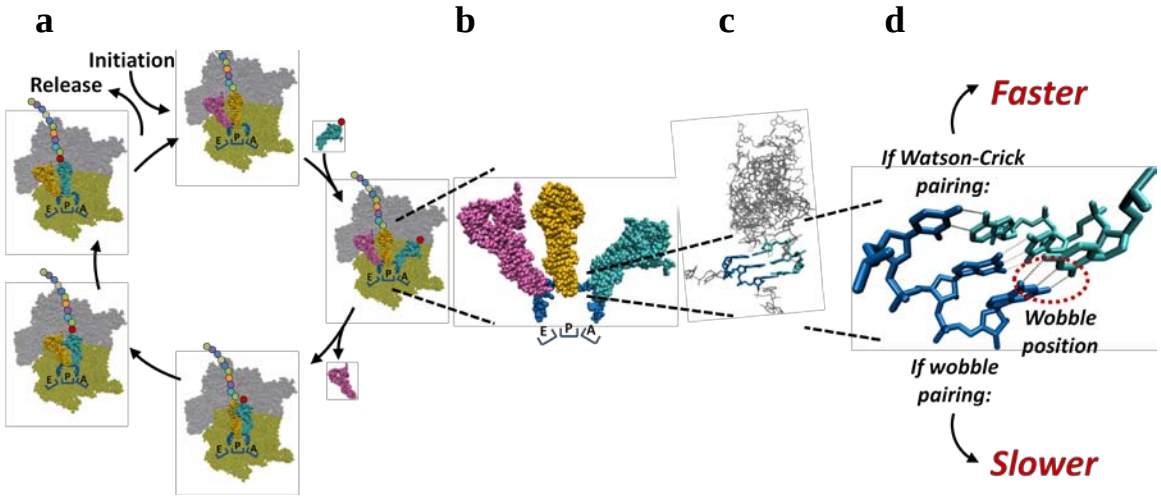


FIGURE 1.1 THE NATURE OF THE CODON:ANTICODON INTERACTION INFLUENCES TRANSLATION ELONGATION

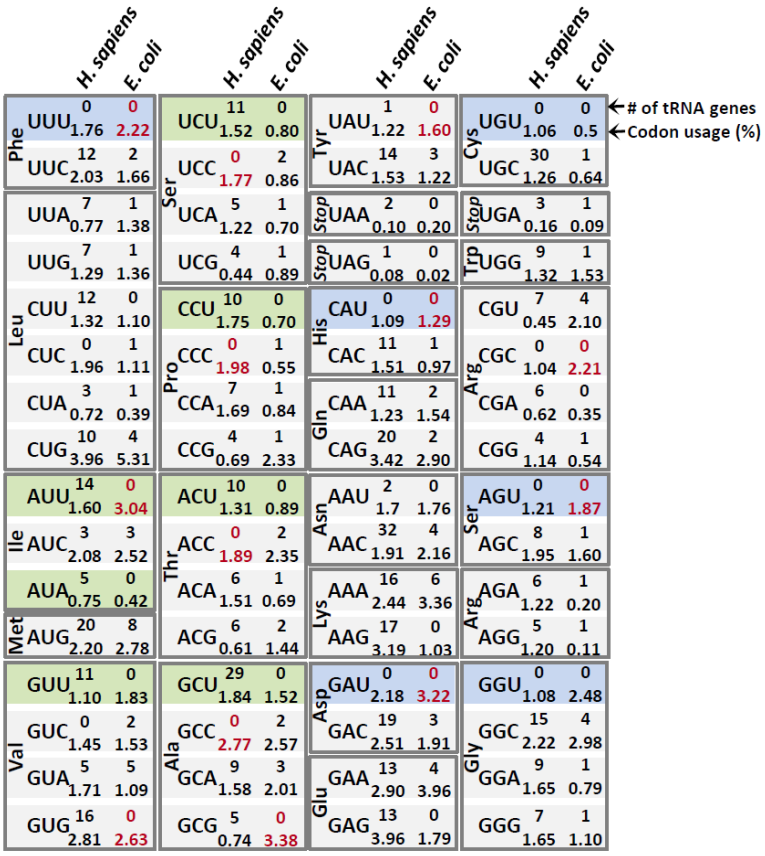
(**a**) Summary of salient steps during bacterial translation elongation. After initiation, a ternary complex of tRNA (cyan) charged with an amino acid (red dot) and EF-Tu:GTP (not shown) binds to the A site of the 70S complex (gray/green) (1). GTP is then hydrolyzed, which results in incoming tRNA accommodation and release of EF-Tu and deacylated tRNA from the E site (2). The nascent polypeptide (chain of colored dots) is then transferred from the peptidyl tRNA in the P site to the incoming tRNA (3). EF-G binding and subsequent GTP hydrolysis (not shown) results in the critical translocation step, by which the now empty tRNA in the P site is transferred to the E site and the new peptidyl-tRNA is placed in the P site (4). EF-G release now renders the complex competent for a new round of elongation (5) or release and termination, if a stop codon is now encountered in the A site. (**b**) Space filling representation depicting an actual complex of mRNA and tRNAs in the E, P and A sites (PDB file 2Y18, from [76]. (**c**) Stick representation displaying the details of the codon (blue):anticodon (cyan) interaction in the A site shown in b (from [same as above]). (**d**) Enlarged view of actual codon:anticodon complexes with Watson-crick-based interactions (above; PDB file 2Y18) and wobble-based interactions (below; PDB file 2Y0Y from [76]), which result in faster and slower rates of polypeptide elongation, respectively.

CODON BIAS DOES NOT NECESSARILY DETERMINE POLYPEPTIDE ELONGATION RATE

As discussed in the above section, it is likely that polypeptide elongation rates depend both on the nature of the anticodon-codon interaction as well as actual aa-tRNA concentrations. The concentrations of tRNA molecules have been experimentally determined for several organisms and cell types, although these measurements do not distinguish between charged and un-charged tRNAs. Regardless, the concentration of particular sets of tRNAs has been shown to correlate relatively well with corresponding tRNA gene numbers. For example, in *E. coli*, the r-values (numerical value describing the linear dependence of datasets such that r = 1.0 indicates a perfect, positive linear relationship) have been reported to vary between 0.74 and 0.9 while in *B. subtillis* r = 0.86 (Ikemura, 1981; Kanaya et al, 1999). In the eukaryote *S. cerevisiae*, the correlations reveal a similar dependency, with an: r = 0.91 (Percudani et al, 1997). Additionally, it is known that there exists some variation in expression of tRNA as a function of growth conditions in both bacteria (Dong et al, 1996) and unicellular eukaryotes (Heyman et al, 1994). Regardless of these caveats, tRNA gene number has been largely accepted as a means to estimate relative aa-tRNA concentrations in multiple organisms. It is important to note that correlations have indeed been found between tRNA gene number and the nonrandom use of synonymous codons in highly expressed genes in several unicellular organisms. This has led to the hypothesis that in organisms whose growth rates are largely dependent on the overall rate of protein production, the translation process has been accelerated, and thus optimized, by evolving codon usage in highly expressed genes to match the most abundant tRNAs (Ikemura, 1985). In other words, evolving highly

expressed genes to largely contain codons read by abundant tRNA would increase the rate of essential protein production and thus increase growth rates in these organisms. These codons were designated as “optimal codons” since they appeared to be favored over their synonymous counterparts in highly expressed genes. Conversely, codons rarely found in highly expressed genes were termed “non-optimal codons” because they were correlated with low abundance tRNAs, although to a lesser extent. Genes with low expression in these organisms, such as those encoding regulatory proteins, were found to be encoded by less biased usage of optimal and non-optimal codons. These results have led to the generalized assumption that frequently used codons are translated fast, and infrequently used codons are translated slowly across organisms, even though the inverse has been shown to occur for some codons (Bonekamp et al, 1989; Curran & Yarus, 1989). This is perhaps due to the fact that the correlation between codon usage frequency and tRNA availability is clearly not absolute (Figure 2, tabulated from the Genomic tRNA database http://gtrnadb.ucsc.edu/ (Chan & Lowe, 2009). For example, highest codon usage frequency and highest tRNA gene number agree only in 12 codons in human and 6 codons in *E. coli*. Furthermore, in most organisms, there are examples in which the most frequently used codon for a particular amino acid across the genome has zero cognate tRNA genes and thus must rely on a tRNA that decodes via non-Watson-Crick interactions, which, as mentioned above, is generally slower. For example, in *E. coli* and human, there are 9 and 4 cases, respectively, where the most frequently used codon for a particular amino acid has zero cognate tRNA genes (Figure 2). Furthermore, there are several instances where there are vastly more tRNA genes for a particular codon, but the

TABLE 1.1 DIFFERENCES IN TRNA GENE CONTENT FOR *E. COLI* AND *H. SAPIENS*
Codons boxed in blue denote tRNA genes often absent in bacteria *and* eukaryotes, while codons boxed in green denote genes mostly absent *only* in bacteria. Actual tRNA gene numbers and codon usage frequencies for humans and *E. coli* are provided as indicated. Numbers in red color denote most frequent codons for which there is no cognate tRNA gene in each organism. Data were obtained from (Chan & Lowe, 2009).



frequency with which that codon is used is only slightly higher (for example, the codons for Asn in humans, Figure 2). It is important to note here that there are different ways in which a codon can be designated as "frequent" or "rare". The original studies derived codon frequencies from *only highly expressed genes*, whereas modern databases (such as the one utilized to generate Figure 2) tabulate frequencies based on the total appearance of codons *across entire genomes*. There would undoubtedly be more agreement between high tRNA abundance and high usage frequency for *E. coli* if the codon usage data were restricted to highly expressed genes instead of considering all sequenced *E. coli* genes.

The correlation between tRNA abundance and codon usage is maintained for the previously discussed glutamate codons of *E. coli*, as GAA is more frequently used, has more cognate tRNA genes, and is translated faster than its synonymous glutamate encoding counterpart (Chan & Lowe, 2009; Sorensen & Pedersen, 1991). However, in the same study, the *in vivo* translation speeds of one frequent codon, CCG (Pro), and one rare codon, CGA (Arg), were translated at very similarly slow rates. This is likely due to the low availability of tRNAs to decode these codons (there are 1 and 0 cognate tRNA genes corresponding to these codons, respectively; Figure 2).

These findings and others of the time (Bennetzen & Hall, 1982; Grantham et al, 1980; Ikemura, 1985) cultivated an increased emphasis on biased codon usage frequencies in translation speed and evolution studies. In addition to the various datasets that can be utilized to measure codon frequencies, there are multiple formulas by which measures of codon frequency can be calculated, which have led to reports of significantly different usage frequency values (Perriere & Thioulouse, 2002) and thus variable correlations between "usage frequency" and "speed" (Stadler & Fire, 2011). Absolute codon frequency is the number of times a given codon is present in a given gene, set of genes, or an entire genome (Suzuki et al, 2008). The Genomic tRNA database (http://gtrnadb.ucsc.edu/) displays a value for absolute codon usage frequency as a percent of the occurrence of a particular codon throughout all coding sequences available for the organism listed, and does not take into account whether or not that codon is part of a synonymous codon block (Chan & Lowe, 2009; Nakamura et al, 2000). An important caveat of this method is that individual amino acids are not equally present in the coding

sequences and may introduce an amino acid-related bias in the observed codon usage frequency patterns. In order to represent codon usage bias independently of amino acid bias, relative frequencies can be calculated. Relative codon frequency is the ratio that results from dividing the absolute codon frequency of a particular codon by the sum of the absolute codon frequencies of all codons in a synonymous block (Perriere & Thioulouse, 2002). Another codon usage metric, Relative Synonymous Codon Usage (RSCU) (Sharp et al, 1986), takes the calculation one step further by normalizing equal codon usage frequencies within a synonymous block to 1.0 (by multiplying the relative codon frequency by the number of synonymous codons in that block). As stated above, highly expressed genes in bacteria and unicellular eukaryotes tend to be encoded by frequent codons. However, there is no evidence for such bias in the highly expressed genes of vertebrates (Ikemura, 1985; Stadler & Fire, 2011). Interestingly, in *C. elegans*, genes with high expression were found to be enriched for codons that the authors demonstrate to be translated faster by ribosomal occupancy times (Stadler & Fire, 2011). Therefore, the adequacy of codon bias for relative translation rate predictions is limited to highly expressed genes in some unicellular and simple multicellular organisms.

POLYPEPTIDE ELONGATION RATES AND PROTEIN FOLDING

To become biologically active, the great majority of proteins must fold into precise three-dimensional conformations. Invaluable insights regarding how protein chains acquire their so-called native states have come from *in vitro* refolding experiments (Anfinsen, 1973) and computational biology approaches (Bradley et al, 2005). These studies have demonstrated that the amino acid sequence of a protein encodes in its

entirety the necessary information to attain its native state *in vitro*. *De novo* protein folding in the cell differs from *in vitro* refolding in various fundamental aspects, which have just begun to be understood (Hartl & Hayer-Hartl, 2002; Kramer et al, 2009). *In vivo*, proteins emerge gradually from the ribosome as they are being synthesized. Thus, the full-length protein sequence is not available for folding all at once, as it is during *in vitro* refolding. Furthermore, the vectorial nature of ribosomal protein synthesis imparts additional constraints on the folding process. The N-terminus of the protein is always exposed to solvent before its more C-terminal elements, and the rate of appearance of the nascent chain is generally significantly slower (seconds to minutes) than observed rates of *in vitro* refolding (nanoseconds to seconds). Furthermore, in contrast to the optimal conditions prepared for refolding experiments, protein folding in the cell occurs under significant macromolecular crowding and at fixed temperature and ionic strength (Ellis & Minton, 2006). In order to allow efficient folding under these conditions, the cell has evolved proteins that assist during *de novo* folding. These proteins, known as "molecular chaperones", bind reversibly to emerging polypeptides and maintain them in an unfolded (or partially folded) state until sufficient sequence has been synthesized to form a native domain (Frydman, 2001; Hartl & Hayer-Hartl, 2009).

The ability to synthesize proteins recombinantly has shown that bacterial systems are often incapable of producing native proteins from human or other eukaryotic origins (Baneyx & Mujacic, 2004; Dingermann, 2008). The poor capacity of the bacterial cytosol to support efficient folding of certain model proteins has been exploited to investigate the mechanisms and molecules involved in this processes. It is possible that this inability

may be due to the presence of incompatible bacterial chaperones (Agashe et al, 2004; Kaiser et al, 2006) or the absence of specialized eukaryotic chaperones (Gautschi et al, 2002; Spiess et al, 2004). In addition to their distinct chaperone complements, a major difference between the protein biosynthetic machineries of bacteria and eukaryotes that has remained largely unexplored is the rate at which proteins are synthesized. In *E. coli*, polypeptide elongation rates vary from ~12 amino acids *per* second (aa/s) during slow growth to ~20 aa/s during fast growth (Bremer & Dennis, 1996). In contrast, elongation rates in eukaryotes are thought to be fairly constant and considerably slower (~5 aa/s) (Mathews et al, 2000). Thus, the folding pathways of nascent polypeptide chains in eukaryotes evolved in the context of synthesis rates slower than those of bacteria. Since translation is spatially and temporally coupled to protein folding, synthesis of certain eukaryotic proteins by bacterial ribosomes at abnormally fast speeds may be incompatible with their folding regimes.

Indeed, it has long been hypothesized that variations in mRNA translation rates could have significant impact on the folding of encoded polypeptides (Itano, 1968a; Purvis et al, 1987) and sequence-based manipulation constitutes a promising strategy to improve the folding of recombinant proteins in heterologous systems (Angov, 2011; Welch et al, 2009). The effect of globally altering translation speeds has been demonstrated by heterologous expression in an *E. coli* strain that has been mutated to produce slow-translating ribosomes (Siller et al, 2010). In this study, slow translation resulted in higher folding efficiency of the recombinant proteins compared to those that were translated by faster wild type ribosomes (Siller et al, 2010). The effects of regional

variations in translation rates on protein folding are generally addressed in two types of approaches: (1) computer-based searches for correlations between codon composition of mRNAs and structural features of the encoded polypeptides; and (2) biochemical investigations of the effects of silent substitutions on the activities of specific proteins (Table 1). These studies have found conflicting results on whether or not certain types of codons encode amino acid residues present in particular structures of the native protein, such as domain boundaries, regions of random coil, or certain secondary structural elements, *etc.* (Table 1). Similarly, there has been disagreement in the literature regarding the effect of "fast" or "slow" codons at certain positions on the solubility and activity of particular proteins (Table 1). These discrepancies are partially due to the fact that most of these studies base translation rate predictions on measures directly related to the above concept of biased codon usage (such as the Codon Adaptation Index (Sharp & Li, 1987) and %MinMax (Clarke & Clark, 2008)), which as stated above, may not accurately reflect polypeptide elongation rates.

How can subtle differences in polypeptide elongation rates impact the folding of the polypeptide emerging from the ribosome? Although 2-3 fold differences in the rates of ordinary reactions might not be generally considered significant from a chemical kinetics point of view, a 2-3 fold difference in the rate of synthesis of a protein may have profound biological consequences. For example, a subtle increase in the concentration of a partially folded, aggregation-prone polypeptide intermediate during translation may exceed the critical concentration of the intermediate and lead to its nucleation-dependent aggregation, thus forming intracellular aggregates. In essence, the fact that variation in

translation rates impact protein folding support the notion that not all proteins fold globally, but rather follow particular pathways throughout the available structural space, influenced by the speed at which they emerge vectorially from the ribosome. This idea may find applications in a variety of fields and settings, including improvements in the production of recalcitrant proteins for vaccine development, recombinant pharmaceuticals and structure-determination studies (Dingermann, 2008).

Knowledge of the determining factors of polypeptide elongation rates reviewed here should lead to more prudent speed designations for codons and thus more accurate predictions of variations in translation rates along mRNA. This information will help us to understand how this additional layer of information encoded in mRNA influences the resulting protein structure formation.

Year	Protein/Dataset	Methodology	Findings and Remarks	Ref.
1968	Human sickle cell hemoglobin	Theoretical	Proposed "the structure-rate hypothesis and the toll bridge analogy" to explain how a single codon change along the hemoglobin S molecule could result in misfolding.	a
1987	Feline pyruvate kinase	Theoretical	Correlated the occurrence of rare codons along the pyruvate kinase mRNA with its domain structure. Suggested controlled differential rates of translational elongation as a general mechanism for protein folding *in vivo*.	b
1989	Cytochromes; globins	Theoretical	Observed clusters of rare codons in the boundaries of segments encoding linkers connecting similar secondary structural elements. Suggested that the concentration of tRNA molecules allows sequential domain folding encoded in the mRNA	c
1994	Yeast TRP3	Experimental	Replacement of a segment of ten rare codons in a region predicted to lie between two folding units resulted in decreased specific activity. Removal of SSA (Hsp70) chaperones resulted in a further decrease in activity, supporting the notion of misfolding.	d
1996	37 *E. coli* proteins	Theoretical	Correlated codon frequency with protein domains and found that slow codons clustered around domain boundaries of multi-domain proteins. Utilized a combination of codon frequencies and codon adaptation index to predict translation rates.	e
1996	54 *E. coli* proteins	Theoretical	General trends found for helices to be encoded by codons predicted to be translated fast, and beta strands by codons predicted to be translated slowly. Utilized a combination of codon frequencies and codon adaptation index to predict translation rates.	f
1996	719 proteins	Theoretical	No correlations found between codons predicted to be translated slowly	g

	from bacteria and eukaryotes		and domain boundaries. Utilized codon adaptation index to predict translation rates.	
1996	109 mammalian sequences	Theoretical	Found that certain codons have a significantly different propensity for being located at the boundaries of secondary structural elements than the amino acids they encode.	h
1997	Human interferon	Experimental	Replacement of 11 rare Arg codons (AGG, AGA) with a frequent one (CGU) resulted in decreased specific activity upon recombinant production in *E. coli*. Supports idea that increased translation speed increases eukaryotic protein misfolding in *E. coli.*	i
1998	Yeast Ure2p	Experimental	Replacement of two rare Arg (AGA) codons by a more frequent one (CGU) resulted in a significant increment in the yield of biologically active protein upon production in *E. coli*. Does not support the idea that slower translation rates decrease misfolding of eukaryotic proteins in *E. coli*.	j
1999	Bacterial chloramphenicol acetyltransferase	Experimental	Replacement of a segment of 16 rare codons for frequent ones resulted in a 20% decrease in specific activity upon production in *E. coli*. Supports idea that increased translation speed increases protein misfolding.	k
2000	164 proteins from bacteria, yeast and humans	Theoretical	No species-invariant correlation between codon usage and secondary structural elements found, but significant differences for preferred codons found between helices and strands. Utilized synonymous codon usage as predictor of translation rates.	l
2002	cDNas from 21 bacterial species	Theoretical	The location of segments predicted to be translated slowest was mapped and found to be at codon ~155, proposed to correspond to the emergence of a “typical protein fold”. Translation rate prediction were based on codon frequency.	m
2003	200 proteins from SCOP dataset	Theoretical	Certain codons for Ile and Arg were found to be significantly enriched in folds composed of particular kinds of elements (*e.g.*, all alpha proteins). No correlations with predicted elongation rates were attempted.	n
2007	Human P-glycoprotein (MDR1)	Experimental	A silent single nucleotide polymorphism proposed to affect polypeptide elongation rates was found to result in a P-glycoprotein conformation with altered substrate characteristics.	o
2007	HIV *gag* p17	Experimental	A silent substitution in the gag p17 protein in virions incapable of seroconverting human hosts was found to interfere with viral assembly in cell culture models.	p
2009	*E. coli* SufI	Experimental	Correlated putative folding intermediates with regions along the mRNA predicted to be translated slowly. Translation rate predictions were based on a combination of codon frequency and tRNA concentrations.	q
2009	3636 proteins from *E. coli*, yeast, fly and mouse	Theoretical	“Translationally optimal codons” were found to associate with buried residues and with sites where mutations result in large changes in free energy. Translation efficiency was inferred from codon usage bias data.	r
2010	4406 proteins from bacteria and eukaryotes	Theoretical	No evidence found that domain boundaries are enriched in slow codons. However, translation rates predicted to decrease at the transitions into secondary structural elements. Found relative codon usage to be less informative than tRNA concentration for predicting translation rates	s
2010	Mammalian beta and gamma actins	Experimental	Differential arginylation of actin isoforms proposed to occur as a result of sequence-encoded differences in translation rates at the start of the mRNAs, which leads to differential degradation. Translation rate predictions were based on codon frequencies; translation rates were not experimentally determined.	t

TABLE 1.2 OVERVIEW OF STUDIES LINKING MRNA CODON COMPOSITION WITH PROTEIN FOLDING

References

a) Itano, H. A. (1968). The structure-rate hypothesis and the toll bridge analogy. In A. Rich & N. Davidson (Eds.), *Structural Chemistry and Molecular Biology* (pp. 275-280). San Francisco: W. H. Freeman and Company.

b) Purvis, I. J., Bettany, A. J., Santiago, T. C., Coggins, J. R., Duncan, K., Eason, R., & Brown, A. J. (1987). The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J Mol Biol, 193*(2), 413-417.

c) Krasheninnikov, I. A., Komar, A. A., & Adzhubei, I. A. (1989). [Role of the code redundancy in determining cotranslational protein folding]. *Biokhimiia, 54*(2), 187-200.

d) Crombie, T., Boyle, J. P., Coggins, J. R., & Brown, A. J. (1994). The folding of the bifunctional TRP3 protein in yeast is influenced by a translational pause which lies in a region of structural divergence with Escherichia coli indoleglycerol-phosphate synthase. *Eur J Biochem, 226*(2), 657-664.

e) Thanaraj, T. A., & Argos, P. (1996a). Protein secondary structural types are differentially coded on messenger RNA. *Protein Sci, 5*(10), 1973-1983. doi: 10.1002/pro.5560051003

f) Thanaraj, T. A., & Argos, P. (1996). Ribosome-mediated translational pause and protein domain organization. *Protein Sci, 5*(8), 1594-1612. doi: 10.1002/pro.5560050814.

g) Brunak, S., & Engelbrecht, J. (1996). Protein structure and the sequential structure of mRNA: alpha-helix and beta-sheet signals at the nucleotide level *Proteins* (Vol. 25, pp. 237-252). United States.

h) Adzhubei, A. A., Adzhubei, I. A., Krasheninnikov, I. A., & Neidle, S. (1996). Non-random usage of 'degenerate' codons is related to protein three-dimensional structure. *FEBS Lett, 399*(1-2), 78-82.

i) Ivanov, I. G., Saraffova, A. A., & Abouhaidar, M. G. (1997). Unusual effect of clusters of rare arginine (AGG) codons on the expression of human interferon alpha 1 gene in Escherichia coli *Int J Biochem Cell Biol* (Vol. 29, pp. 659-666). England.

j) Komar, A. A., Guillemet, E., Reiss, C., & Cullin, C. (1998). Enhanced expression of the yeast Ure2 protein in Escherichia coli: the effect of synonymous codon substitutions at a selected place in the gene. *Biol Chem, 379*(10), 1295-1300.

k) Komar, A. A., Lesnik, T., & Reiss, C. (1999). Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. *FEBS Lett, 462*(3), 387-391.

l) Gupta, S. K., Majumdar, S., Bhattacharya, T. K., & Ghosh, T. C. (2000). Studies on the relationships between the synonymous codon usage and protein secondary structural units *Biochem Biophys Res Commun* (Vol. 269, pp. 692-696). United States: 2000 Academic Press.

m) Makhoul, C. H., & Trifonov, E. N. (2002). Distribution of rare triplets along mRNA and their relation to protein folding *J Biomol Struct Dyn* (Vol. 20, pp. 413-420). United States.

n) Gu, W., Zhou, T., Ma, J., Sun, X., & Lu, Z. (2003). Folding type specific secondary structure propensities of synonymous codons. *IEEE Trans Nanobioscience, 2*(3), 150-157.

o) Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., & Gottesman, M. M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science, 315*(5811), 525-528. doi: 10.1126/science.1135308

p) Hamano, T., Matsuo, K., Hibi, Y., Victoriano, A. F., Takahashi, N., *et al.* (2007). A single-nucleotide synonymous mutation in the gag gene controlling human immunodeficiency virus type 1 virion production *J Virol* (Vol. 81, pp. 1528-1533). United States.

q) Zhang, G., Hubalewska, M., & Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat Struct Mol Biol, 16*(3), 274-280. doi: nsmb.1554

r) Zhou, T., Weems, M., & Wilke, C. O. (2009). Translationally optimal codons associate with structurally sensitive sites in proteins *Mol Biol Evol* (Vol. 26, pp. 1571-1580). United States.

s) Saunders, R., & Deane, C. M. (2010). Synonymous codon usage influences the local protein structure observed *Nucleic Acids Res* (Vol. 38, pp. 6719-6728). England.

t) Zhang, F., Saha, S., Shabalina, S. A., & Kashina, A. (2010). Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation *Science* (Vol. 329, pp. 1534-1537). United States.

CHAPTER 2: SILENT SUBSTITUTIONS PREDICTABLY ALTER TRANSLATION ELONGATION RATES AND PROTEIN FOLDING EFFIENCIES[3]

2.1 INTRODUCTION

In most organisms, 61 out of the 64 possible codon combinations are used to encode 20 different amino acids and thus, a single amino acid can be encoded by several (up to six) codons. The distribution of such *synonymous* codons along protein coding sequences is generally not uniform, suggesting that their properties are not entirely equivalent. This *biased codon usage* has been described, for example, in organisms where certain codons are more common than others within highly expressed genes (referred to as *frequent* or *optimal* codons) (Ikemura, 1985). Multiple theories have arisen to explain the biological significance of this biased codon selection, and most revolve around the notion that certain codons allow faster or more efficient translation while others result in slower rates (Deane & Saunders, 2011; Spencer & Barral, 2012). These different rates of polypeptide emergence from the ribosome are hypothesized to influence its folding properties (Deane & Saunders, 2011; Spencer & Barral, 2012). However, the factors that determine the rates at which different codons are translated have remained unclear, which has led to disagreements on whether or not changes in elongation rates have any influence on the properties of the encoded polypeptide (Deane & Saunders, 2011). tRNA selection has been determined to be rate limiting for translation elongation in various models (Gromadski & Rodnina, 2004; Johansson et al, 2008), and thus it is

[3] This chapter is reproduced from Journal of Molecular Biology, 2012, doi:10.1016/j.jmb.2012.06.010. Silent Substitutions Predictably Alter Translation Elongation Rates and Protein Folding Efficiencies. Spencer, P.S. *et al.* with permission.

likely that tRNA availability plays a critical role in determining translation elongation rates (Fluitt et al, 2007; Spencer & Barral, 2012; Varenne et al, 1984). Remarkably, in every organism examined to date, there are considerably fewer than 61 different tRNA species (Fig. 2.1), as certain tRNAs are capable of decoding more than one synonymous codon (Chan & Lowe, 2009).

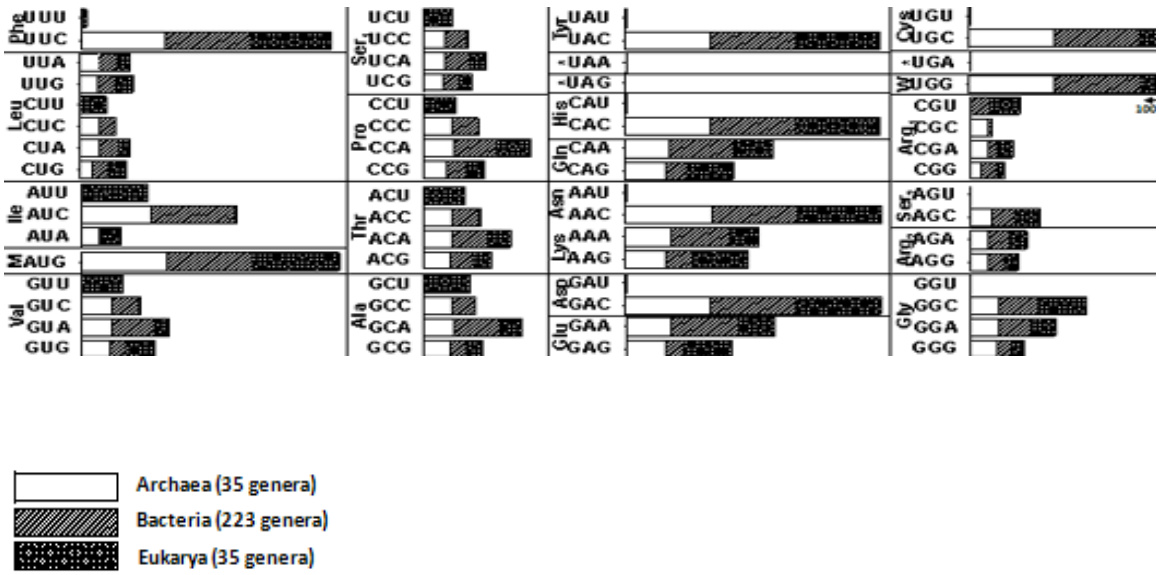


FIGURE 2.1 THE DISTRIBUTION OF GENES ENCODING TRNAS OF DIFFERENT DECODING CAPACITIES VARY AMONG ARCHAEA, BACTERIA, AND EUKARYA.
Predicted gene content for tRNAs capable of decoding the standard genetic code according to gtrnadb.ucsc.edu (Chan & Lowe, 2009) is plotted for each codon in histogram form (as indicated) by each domain of life in different colors (as indicated). The length of each box represents the extent to which genes for tRNAs capable of decoding the corresponding codon are present in a domain. For example, for Ala, no eukaryotic genera examined contain tRNA genes capable of decoding GCC, whereas ~60%, ~25% and ~15% of them contain tRNA genes to decode GCU, GCA and GCG, respectively. For Met or Trp, 100% of genera examined in each domain are predicted to contain a single species of tRNA genes to decode these codons (and thus the length of these bars corresponds to "100% exclusivity").

Thus, there are essentially two modes by which a particular tRNA molecule can decode a codon: (1) through strict Watson-Crick (WC) base pairing in all three positions of the codon:anticodon interaction and (2) through non-WC base pairing at the third position of the codon (referred to as a "wobble" interaction) (Crick, 1966). Previous studies have suggested that the speeds of decoding of these two mechanisms may be different with wobble-based decoding resulting in slower rates (Curran & Yarus, 1989; Sorensen & Pedersen, 1991; Stadler & Fire, 2011). Although the precise reasons for such rate differences are currently unclear, it is possible that they may reflect differences in dissociation rates between A-site tRNA and the mRNA after codon:anticodon binding, with wobble-type interactions displaying higher dissociation rates (Spencer & Barral, 2012). Importantly, direct determination of translation elongation rates based on these mechanisms for actual full-length polypeptides and their effect on protein folding are lacking.

2.2 RESULTS

We began by predicting relative codon translation speeds based on *E. coli* tRNA gene information (gtrnadb.ucsc.edu) (Chan & Lowe, 2009) and values derived from previously measured rates of select individual codons *in vivo* that allow the rate comparison between WC and wobble decoded codons (Curran & Yarus, 1989; Sorensen & Pedersen, 1991). We developed a formula that incorporated these parameters (see methods) and utilized it to generate predicted relative translation speed profiles of any mRNA in any organism of known tRNA gene content (Fig. 2.2).

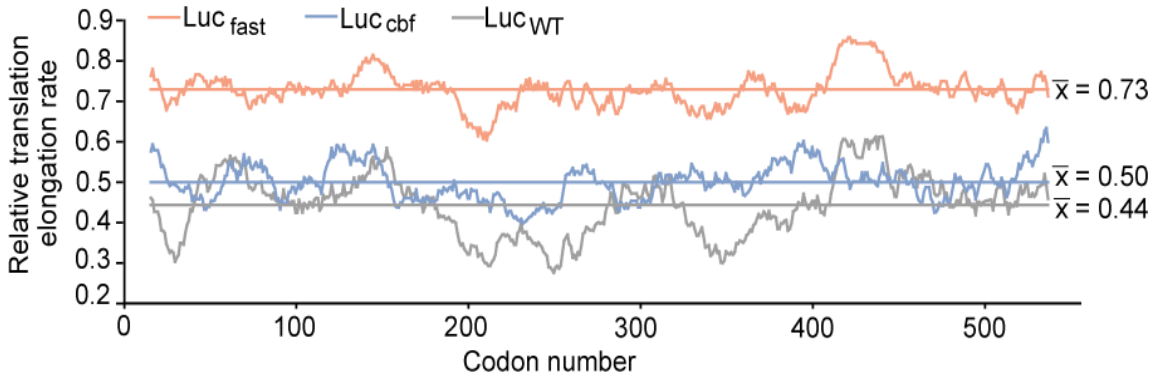


FIGURE 2.2 UTILIZATION OF TRNA GENE INFORMATION AND NATURE OF CODON:ANTICODON BASE PAIRING ALLOWS THE PREDICTION OF RELATIVE TRANSLATION ELONGATION RATES.
Predicted relative protein synthesis rates (see main text and Methods) in *E. coli* for Luc sequences lacking codons decoded by wobble-based tRNA interactions (Luc_{fast}, orange), containing the most frequent *E. coli* codons (Luc_{cbf}, blue) or the unmodified firefly coding sequence (Luc_{WT}, gray).

To determine whether WC-based decoding is indeed faster than that mediated by wobble interactions, we reasoned that complete elimination of wobble decoding along an mRNA molecule would result in a detectable enhancement of the translation rate of the encoded protein (Fig. 2.2). Thus, we employed DNA synthesis to engineer a bacterial expression construct for the model protein firefly luciferase (Luc) in which every amino acid is encoded by a synonymous codon read by a WC-pairing tRNA anticodon (Luc_{fast}) (Methods and Appendix), directly measured its translation rate by pulse-chase analysis in live *E. coli* cells (Sorensen & Pedersen, 1998), and compared it to that of the wild type sequence (Luc_{WT}) (Fig. 2.3), (their respective mRNAs accumulated to similar levels; Fig. 2.4). Interpretation of our pulse-chase experiments using a method that utilizes theoretical *constant* elongation rates to calculate protein synthesis rates (Sorensen & Pedersen, 1998) reported a speed of 9.8 amino acids *per* second (aa/s) for Luc_{WT} and 19.2 aa/s for Luc_{fast},

as judged by least sum of squares analysis.

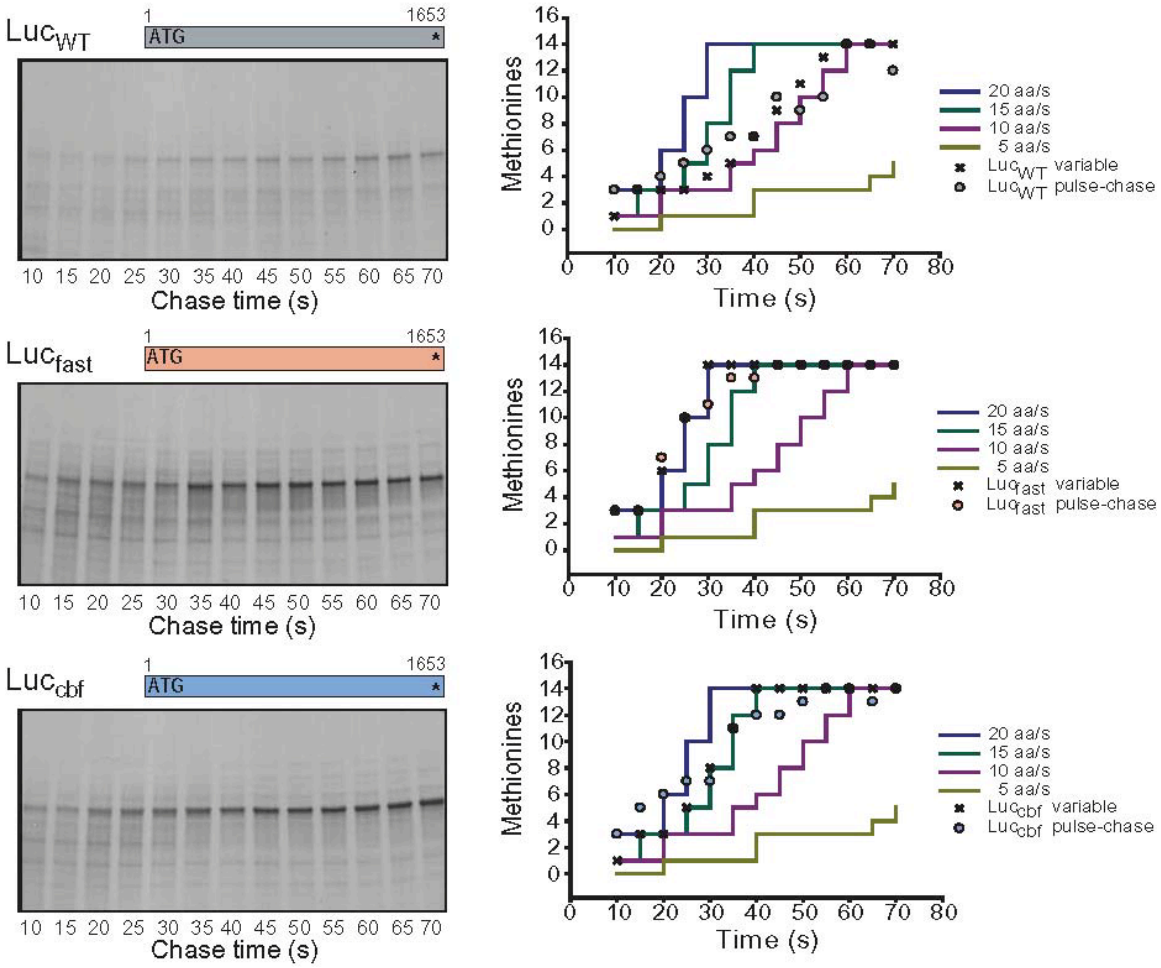


FIGURE 2.3 AVOIDANCE OF WOBBLE-BASED INTERACTIONS DURING MRNA DECODING RESULTS IN ACCELERATION OF TRANSLATION ELONGATION RATES *IN VIVO*.
Pulse-chase analyses (left panels) in live *E. coli* cells synthesizing recombinant Luc from the indicated constructs and plots (right panels) depicting the appearance of incorporated [^{35}S]methionine in full length Luc produced from the indicated constructs (colored dots), curves for the theoretical appearance of methionines with four calculated constant translation rates of the indicated constructs (colored lines) and calculated theoretical appearance of methionines according to our predicted variable rates (x symbols), which demonstrate that Luc$_{fast}$ is translated faster than Luc$_{WT}$ and Luc$_{cbf}$.

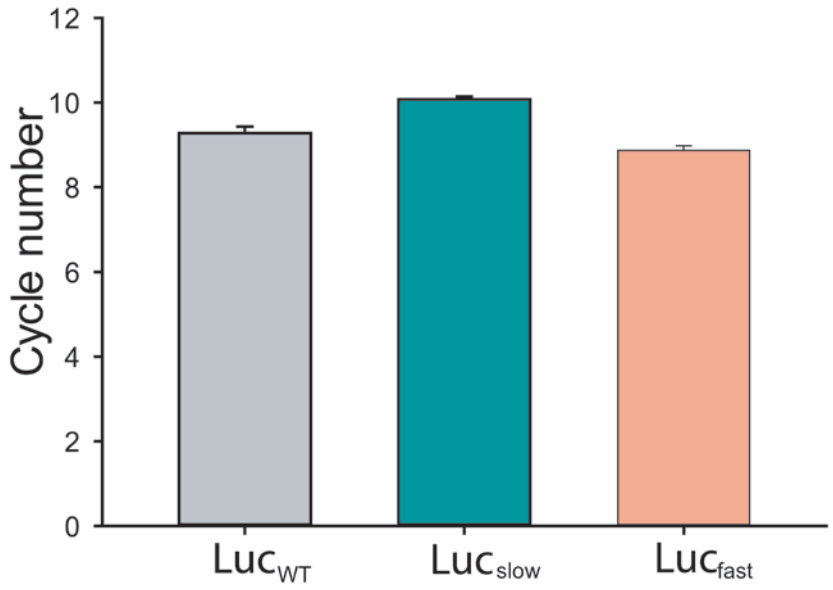


FIGURE 2.4 STEADY-STATE ACCUMULATION OF MRNA SYNTHESIZED FROM THE WILD TYPE AND SEQUENCE-ENGINEERED LUC CONSTRUCTS.
Bar graph depicting the results of a quantitative reverse transcriptase PCR reaction to evaluate the levels of accumulation of mRNA produced from each of the indicated constructs. Error bars represent standard errors of the mean.

However, our predictions (Fig. 2.2) suggest that elongation rates are not constant along the mRNAs of our Luc constructs. Thus, incorporation of theoretical *variable* rates in the interpretation of our pulse-chase data (see Methods) would be expected to yield a better fit. Indeed, for Luc_{WT}, calculations that utilized variable rates led to a least square value that was lower than any value that could be obtained using constant rates, reflecting a considerably better fit (Fig. 2.5). It is likely that the experimental methodologies utilized in this study may not be of sufficiently high resolution to reveal the finer details associated with regional variations in ribosome movement along mRNAs and thus the differences detected between constant and variable theoretical rates may actually be considerably greater than demonstrated here. For Luc_{fast}, we were unable to improve the

fit beyond the best-fit constant rate by using variable elongation rates. This was not unexpected since much of the variability that exists in elongation rate along the Luc_{WT} sequence was removed in the Luc_{fast} sequence by replacing slow with fast codons, which yield a more constant, fast speed profile (Fig. 2.2).

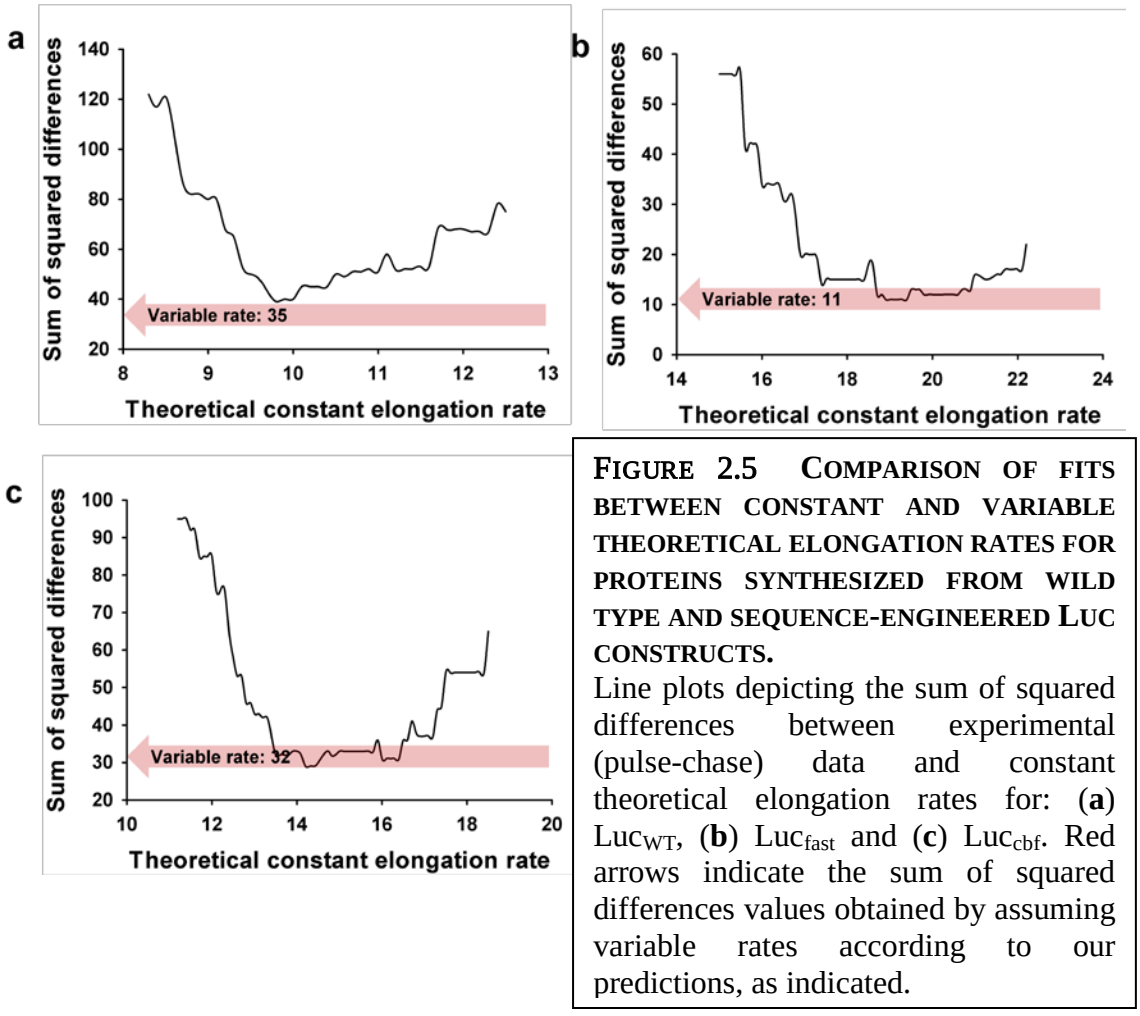


FIGURE 2.5 COMPARISON OF FITS BETWEEN CONSTANT AND VARIABLE THEORETICAL ELONGATION RATES FOR PROTEINS SYNTHESIZED FROM WILD TYPE AND SEQUENCE-ENGINEERED LUC CONSTRUCTS.
Line plots depicting the sum of squared differences between experimental (pulse-chase) data and constant theoretical elongation rates for: (**a**) Luc_{WT}, (**b**) Luc_{fast} and (**c**) Luc_{cbf}. Red arrows indicate the sum of squared differences values obtained by assuming variable rates according to our predictions, as indicated.

Regardless, the observed ~2-fold *average* increment in rate is of very similar magnitude to the one predicted by our metric (~1.7-fold; Fig. 2.2). Thus, WC-decoding appears to confer faster translation relative to wobble-based interactions. Frequent codons as determined by biased codon usage patterns have traditionally been considered “fast”,

aa	Codon	*E. coli* tRNA genes (#)	*E. coli* Codon usage (%)	*Drosophila* tRNA genes (#)	*Drosophila* Codon usage (%)	aa	Codon	*E. coli* tRNA genes (#)	*E. coli* Codon usage (%)	*Drosophila* tRNA genes (#)	*Drosophila* Codon usage (%)
Phe	UUU	0	2.22	0	1.32	Ser	UCU	0	0.8	8	0.7
	UUC	2	1.65	8	2.18		UCC	2	0.86	0	1.96
Leu	UUA	1	1.39	4	0.45		UCA	1	0.71	2	0.78
	UUG	1	1.36	4	1.61		UCG	1	0.89	4	1.66
	CUU	0	1.1	5	0.9	Pro	CUU	0	0.7	7	0.69
	CUC	1	1.11	0	1.38		CCC	1	0.55	0	1.81
	CUA	1	0.39	2	0.82		CCA	1	0.84	5	1.35
	CUG	4	5.29	8	3.82		CCG	1	2.32	5	1.58
Ile	AUU	0	3.04	9	1.66	Thr	ACU	0	0.89	9	0.95
	AUC	3	2.52	0	2.29		ACC	2	2.34	0	2.13
	AUA	0	0.43	2	0.95		ACA	1	0.7	6	1.1
Met	AUG	8	2.78	12	2.36		ACG	2	1.44	3	1.44
Val	GUU	0	1.83	6	1.1	Ala	GCU	0	1.53	12	1.44
	GUC	2	1.53	0	1.39		GCC	2	2.56	0	3.36
	GUA	5	1.09	2	0.64		GCA	3	2.02	2	1.28
	GUG	0	2.62	7	2.78		GCG	0	3.37	3	1.4
Tyr	UAU	0	1.6	0	1.08	Cys	UGU	0	0.5	0	0.54
	UAC	3	1.22	9	1.84		UGC	1	0.64	7	1.32
Stop	UAA		0.03		0.08	Trp Stop	UGA	1	0.09		0.05
	UAG		0.21		0.07	Trp	UGG	1	1.53	8	0.99
His	CAU	1	0.97	0	1.08	Arg	CGU	4	2.09	10	0.88
	CAC	0	1.29	5	1.62		CGC	0	2.2	0	1.8
Gln	CAA	2	1.54	4	1.56		CGA	0	0.35	10	0.84
	CAG	2	2.89	8	3.61		CGG	1	0.54	0	0.82
Asn	AAU	0	1.77	0	2.1	Ser	AGU	0	0.87	0	1.15
	AAC	4	2.16	12	2.62		AGC	1	1.6	6	2.04
Lys	AAA	6	3.37	6	1.7	Arg	AGA	1	0.2	3	0.51
	AAG	0	1.03	13	3.95		AGG	1	0.11	3	0.63
Asp	GAU	0	3.22	0	2.76	Gly	GGU	0	2.48	0	1.32
	GAC	3	1.91	14	2.46		GGC	4	2.97	14	2.67
Glu	GAA	4	3.96	6	2.11		GGA	1	0.79	6	1.8
	GAG	0	1.78	19	4.25		GGG	1	1.11	0	0.47

TABLE 2.6 tRNA GENE CONTENT AND BIASED CODON USAGE FREQUENCIES FOR *E. COLI* AND *D. MELANOGASTER*.

Table of data obtained from gtrnadb.ucsc.edu (Chan & Lowe, 2009) depicting the number of tRNAs capable of decoding each codon as well as the codon usage frequency of each codon of the bacterium *E. coli* and the fly *D. melanogaster*. Boxes shaded in green indicate instances where the most frequent codon coincides with the highest number of tRNA genes for that codon. Boxes shaded in red indicate instances where the most frequent codon has no tRNA genes available for strict WC decoding.

while rare ones have been predicted to be "slow" (Deane & Saunders, 2011; Ikemura, 1985). However, in every genome examined to date (gtrnadb.ucsc.edu) (Chan & Lowe, 2009), several of the most frequently utilized codons have no cognate tRNA genes and must rely on wobble-based decoding (Fig.2.1 and Fig. 2.6). To address this discrepancy, we designed a Luc construct composed exclusively of the most frequently utilized codons in *E. coli* regardless of the number of tRNA genes capable of decoding those codons (Luc_{cbf}) and compared its translation rate to that of Luc_{fast}. Calculations based on constant elongation rates determined that translation of Luc_{cbf} occurred at 14.3 aa/s (Fig. 2.3) and, probably because of reasons similar to Luc_{fast} (see above), there was no improvement in fit when variable rates were considered (Fig. 2.5). An intermediate rate between that of Luc_{fast} and Luc_{WT} is not unexpected, as a considerable fraction of the most frequent codons in *E. coli* correspond to codons decoded by WC tRNAs (Fig. 2.6) and thus Luc_{cbf} is indeed predicted to be translated at rates intermediate between Luc_{WT} and Luc_{fast} (Fig. 2.2). Attempts to determine the translation rate of a Luc sequence engineered to contain mostly wobble codons and thus be translated more slowly (Luc_{slow}) (Appendix) were unsuccessful because protein production was extremely limited, and precluded unambiguous identification of the full length Luc band (Fig. 2.7), probably as a result of marked ribosome sequestration along this recombinant mRNA (Endoh et al, 2012; Spencer & Barral, 2012; Stadler & Fire, 2011). These results show that WC-based codon:anticodon interactions lead to faster ribosome movement along an mRNA molecule *in vivo* and constitute a more accurate basis for predicting translation elongation rates than codon frequency *per se*.

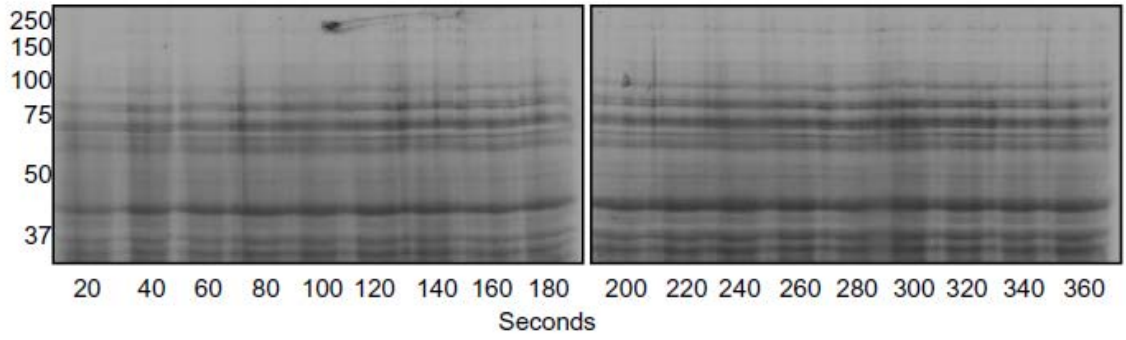


FIGURE 2.7 AUTORADIOGRAM OF AN SDS-PAGE FROM A PULSE-CHASE EXPERIMENT WITH THE LUC_{SLOW} CONSTRUCT.
Pulse-chase analysis on the protein synthesized from $pLuc_{slow}$ was carried out as described above, except that aliquots were taken for considerably longer times, as indicated. No clear band for full-length firefly Luc could be identified, which precluded calculation of polypeptide elongation rates.

In order to ensure that effects associated with translation initiation were not responsible for our observed effects on translation acceleration (Kozak, 2005; Kudla et al, 2009), we engineered Luc_{fast} and Luc_{cbf} sequences in which their first 50 nucleotides were identical to Luc_{WT}, to yield $Luc_{WT\text{-}fast}$ and $Luc_{WT\text{-}cbf}$ (Appendix and Fig. 2.8) and determined their translation rates to be 17.4 aa/s and 14.5 aa/s, very similar to those of their Luc_{fast} and Luc_{cbf} counterparts, respectively. Thus, we believe that the observed acceleration of translation is due to increased polypeptide elongation rates.

It has been previously demonstrated that decreased translation elongation rates enhance the folding efficiency of Luc upon expression in *E. coli*(Siller et al, 2010), and therefore we hypothesized that translation acceleration would result in the opposite effect. Thus, we measured enzymatic activity as an indication of acquisition of the native state and determined the fractional accumulation of the soluble (presumably folded) and

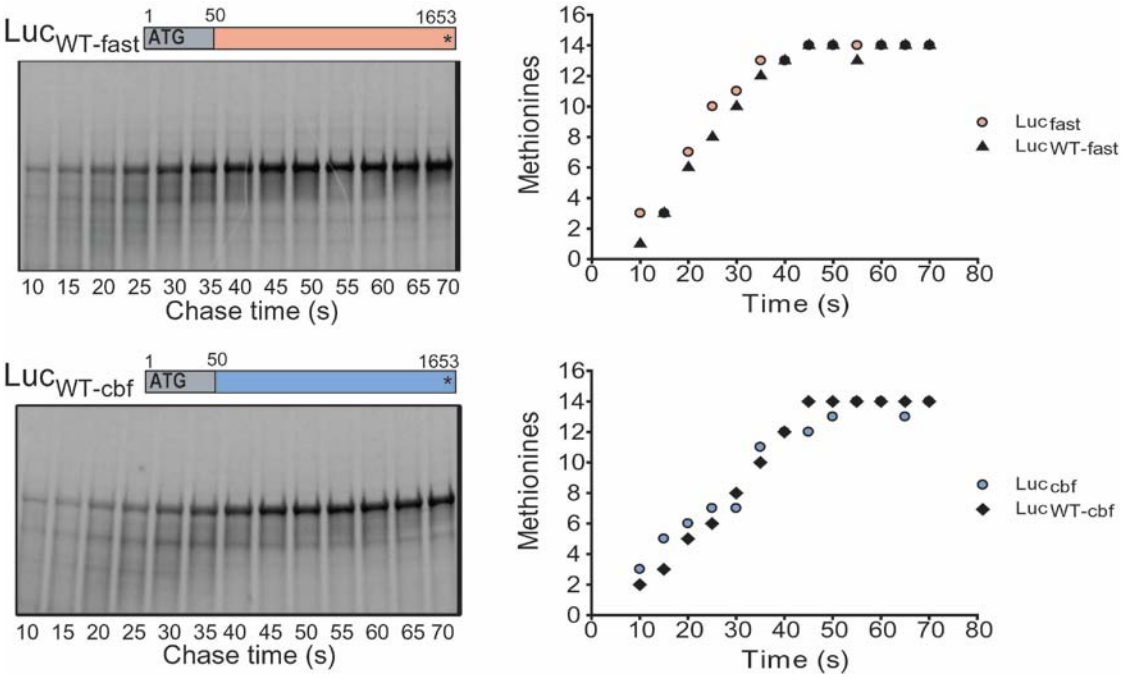


FIGURE 2.8 EFFECTS OF SILENT SUBSTITUTIONS ON OBSERVED TRANSLATION RATES ARE NOT DUE TO CHANGES IN TRANSLATION INITIATION.
Pulse-chase analyses (left panels) and plots (right panels) as in **Fig 2.3**, for the $Luc_{WT\text{-}fast}$ and $Luc_{WT\text{-}cbf}$ constructs, as indicated.

aggregated (misfolded) species of protein produced from the wild-type and engineered Luc sequences (which all contain identical amino acid sequences) (Appendix). At similar levels of total recombinant protein accumulation, the activity of the protein from the Luc_{fast} construct is less than half of that from Luc_{WT} (Fig. 2.9) and protein from Luc_{cbf} displays intermediate levels (Fig. 2.9). Consistently, a greater amount of protein partitioned into the aggregated fraction when translated from Luc_{fast}, with Luc_{cbf} yielding again intermediate levels (Fig. 2.9). Thus, it appears that, at least for Luc, increments in overall translation elongation rates correlate with decrements in folding efficiency.

tRNA gene content differs significantly between bacteria and eukaryotes (Grosjean et al, 2010) (Figs. 2.1 and 2.6). Thus, for a given mRNA sequence, the mode (WC- *vs.* wobble-

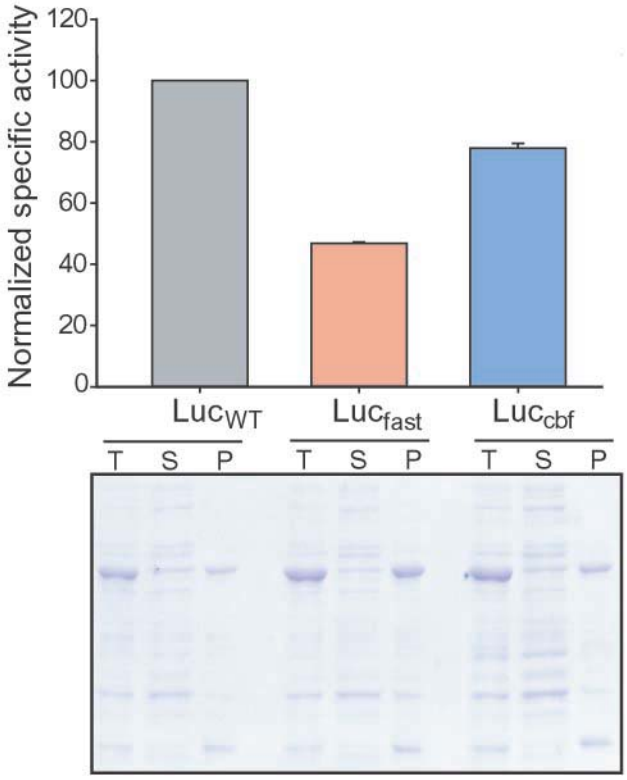


FIGURE 2.9 SYNONYMOUS SEQUENCE-BASED ACCELERATION INFLUENCES THE FOLDING OF THE ENCODED POLYPEPTIDE.
Specific activities of protein products identical in primary sequence produced from Luc_{WT}, Luc_{fast} and Luc_{cbf}, as indicated (top panel). The value of the protein from Luc_{WT} was set to 100. Error bars represent S.E.M. SDS-PAGE of total (T), soluble (S) and insoluble (P) recombinant protein produced in *E. coli* from the indicated sequence-engineered constructs (bottom panel).

based) by which a particular codon is decoded may differ depending on whether the mRNA is being translated in a eukaryotic or a bacterial cytosol. For example, in *E. coli* (Fig. 2.6), there are no tRNA genes that decode the GAG codon (glutamic acid) by strict WC base-pairing. This codon must rely on wobble-based decoding by tRNAs produced

from the four GAA tRNA genes present in that organism. In contrast, *D. melanogaster* contains six GAA tRNA genes and thus a GAA codon will be decoded by strict WC base-pairing tRNAs in addition to being decoded by wobble-based interactions from tRNAs produced by the 19 GAG tRNA genes. Thus, a GAA codon would be expected to be a "slow codon" in *E. coli* but a "fast codon" in *D. melanogaster*. If the relative translation elongation rates are calculated for the same mRNA sequence using our algorithm described above that takes into account these parameters (see methods), one would expect that the profiles would be considerably different depending on whether bacterial *vs.* eukaryotic tRNA gene contents were utilized. When such profiles are generated for Luc using tRNA gene data from *E. coli* or *D. melanogaster* (as the organism closest to the firefly available in the database; gtrnadb.ucsc.edu (Chan & Lowe, 2009)), we find that this is indeed the case (Fig. 2.10). We suggest that these profiles reflect differences in local rates of ribosome elongation along an mRNA in each particular organism, consistent with the finding that ribosome movement along natural mRNAs is likely not uniform (Stadler & Fire, 2011; Varenne et al, 1984), but rather punctuated by regions of acceleration and deceleration. It has been well established that translation elongation rates of eukaryotic ribosomes are generally slower than those of mesophilic bacteria (3-8 *vs.* 12-20 aa/s, respectively) (Liang et al, 2000; Mathews et al, 2000; Pedersen, 1984). By using mutant *E. coli* ribosomes that translate at slower rates (more similar to those of eukaryotes) we previously showed that a *general* reduction in elongation rates resulted in a reproducible, yet marginal increase in the folding efficiency of Luc (Siller et al, 2010). Here we show that the general increase in translation rate of

Luc_{fast} results in a converse decrease in its folding efficiency (Fig. 2.9). However, as mentioned above, we expect that in the insect, the ribosome will not move at a constant speed along the Luc mRNA, but rather increase and decrease its speed as it encounters

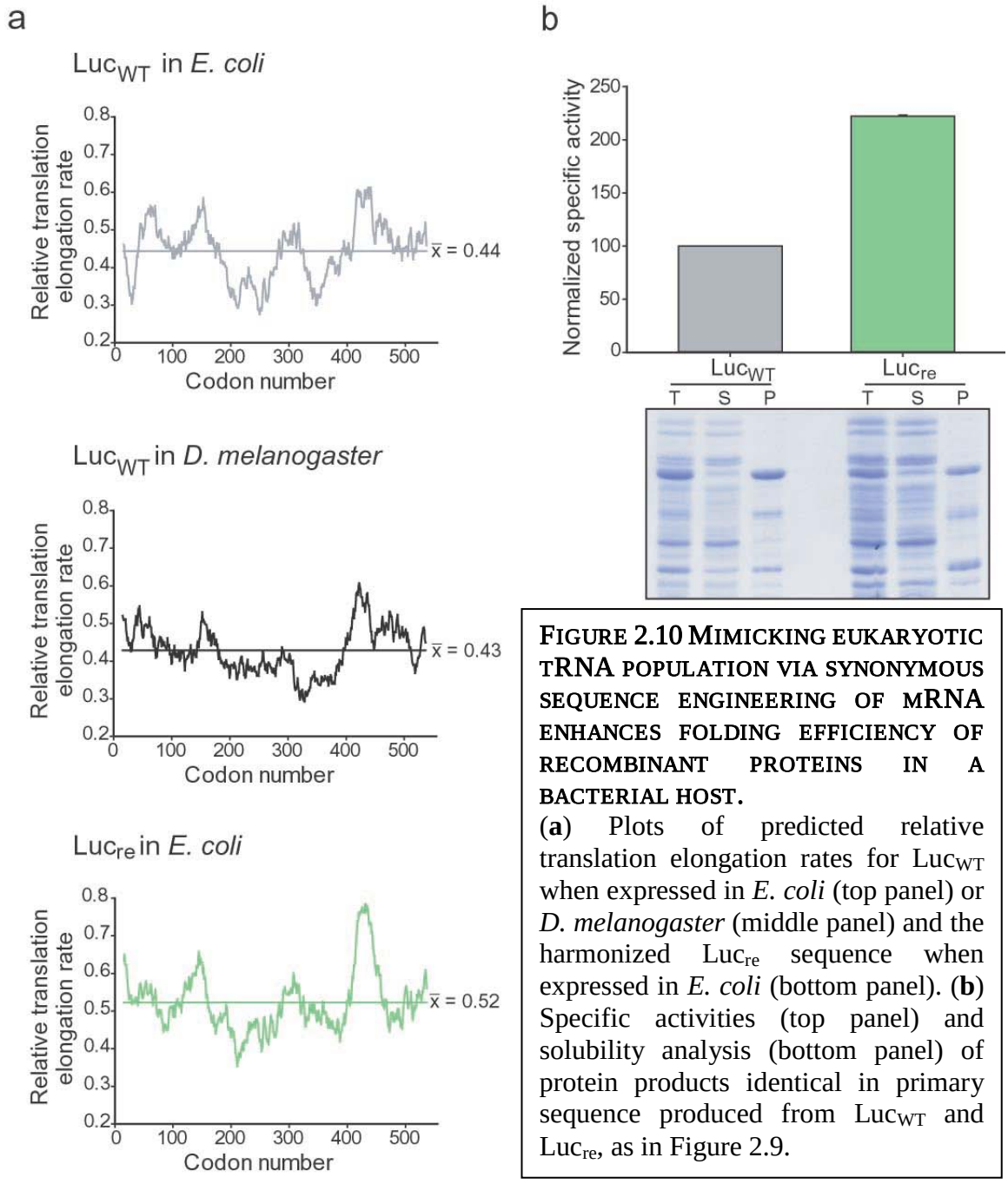


FIGURE 2.10 MIMICKING EUKARYOTIC TRNA POPULATION VIA SYNONYMOUS SEQUENCE ENGINEERING OF MRNA ENHANCES FOLDING EFFICIENCY OF RECOMBINANT PROTEINS IN A BACTERIAL HOST.

(**a**) Plots of predicted relative translation elongation rates for Luc_{WT} when expressed in *E. coli* (top panel) or *D. melanogaster* (middle panel) and the harmonized Luc_{re} sequence when expressed in *E. coli* (bottom panel). (**b**) Specific activities (top panel) and solubility analysis (bottom panel) of protein products identical in primary sequence produced from Luc_{WT} and Luc_{re}, as in Figure 2.9.

stretches of fast and slow codons (reflected in Fig. 2.10a, middle panel, by the profile's peaks and valleys, respectively). We propose that these variations in speed (a sort of ribosomal "rhythm") have been optimized throughout evolution to precisely orchestrate the emergence rates of each segment of the nascent polypeptide to fold or interact with molecular chaperones as it exits the ribosome. Thus, we reasoned that if we were capable of recreating these naturally occurring variations during expression of Luc in the heterologous *E. coli* cytosol, we might be able to mimic the natural rhythm that the ribosome follows in the insect, and thus increase its folding efficiency. Since the tRNA gene content of the firefly is not currently available, we utilized the tRNA gene content of *D. melanogaster* as the closest insect with a sequenced genome to conduct our Luc engineering. We created a Luc sequence (Luc_{re}; Appendix) in which fast codons in *D. melanogaster* (decoded by WC interactions; Fig. 2.6) were substituted with synonymous fast codons in *E. coli* (also translated by WC; Fig. 2.6) and similarly for slow (wobble-based) codons in each organism (Fig. 2.6; see methods). We expressed Luc_{re} in *E. coli* (which encodes an identical polypeptide to all our other Luc constructs; Appendix) and analyzed its folding efficiency (Fig. 2.10b). At similar levels of accumulation, the protein produced from Luc_{re} was more than twice as active and considerably more soluble than that from Luc_{WT} (Fig. 2.10b) although the predicted average (global) translation rates were very similar for both sequences (Fig. 2.10a). These results suggest that segmental variations in elongation rate can considerably influence the folding of the encoded polypeptide, even if these do not significantly alter the overall time that the ribosome spends along the mRNA. We thus propose that sequence engineering directed to mimic

the ribosome rhythm of the original host may constitute a valuable strategy for production of recombinant proteins in heterologous systems.

2.3 DISCUSSION

Our findings suggest that the genetic code has the capacity to regulate the rates of protein synthesis and folding. They support the notion that not all proteins fold *via* simple two-state mechanisms, but rather follow particular pathways throughout their available conformational space, influenced by the regional rates by which their nascent segments emerge unidirectionally from the ribosome. Although our predictions and experimental findings have captured principal features of the coupling between translation and folding, our model is likely oversimplified. For example, it is well known that post-transcriptional tRNA modifications can substantially influence codon:anticodon interactions (Agris, 2008) (particularly in eukaryotes (Grosjean et al, 2010)) and that codons neighboring the A-site may influence elongation rates (Boycheva et al, 2003), offering additional levels of speed modulation. Nevertheless, we believe that our study provides insight into how so-called *silent* polymorphisms may result in human disease (Kimchi-Sarfaty et al, 2007) and how variations in tRNA concentrations impact cellular proteostasis in a wide variety of developmental (Dittmar et al, 2006) and disease states (Pavon-Eternod et al, 2009).

2.4 MATERIALS AND METHODS

PREDICTION OF RELATIVE TRANSLATION ELONGATION RATES

In order to assign relative translation elongation values (v) to each of the 61 codons in a given organism, the following parameters were applied regarding the nature of the codon

($N_1N_2N_3$):anticodon($N_{34}N_{35}N_{36}$) interactions (where $N_1N_2N_3$ represents each trinucleotide along the 5' →3' direction in an mRNA and $N_{34}N_{35}N_{36}$ represents 5'→ 3'each trinucleotide of the anticodon loop of the decoding tRNA): (1) Watson-Crick (WC) interactions were allowed to occur between $N_1N_2G_3$:$C_{34}N_{35}N_{36}$, $N_1N_2C_3$:$G_{34}N_{35}N_{36}$, $N_1N_2A_3$:$U_{34}N_{35}N_{36}$, $N_1N_2U_3$:$A_{34}N_{35}N_{36}$ and $N_1N_2C_3$:$I_{34}N_{35}N_{36}$ (where I_{34} represents inosine, derived from post-transcriptional deamination of some A_{34}-bearing tRNAs); (2) non-Watson-Crick (wobble) interactions were allowed to occur between $N_1N_2G_3$:$U_{34}N_{35}N_{36}$, $N_1N_2U_3$:$G_{34}N_{35}N_{36}$, $N_1N_2U_3$:$I_{34}N_{35}N_{36}$, and $N_1N_2A_3$:$I_{34}N_{35}N_{36}$. Inosination was assumed to occur for all A_{34}-bearing tRNAs in eukaryotes and for A_{34}-bearing tRNAs that decode Arg codons in bacteria(Grosjean et al, 2010). Since a $U_{34}A_{35}C_{36}$-bearing species of tRNA is generally utilized to decode AUA codons in bacteria(Grosjean et al, 2010), it was assumed that a $U_{34}A_{35}C_{36}$-bearing tRNAs would partition equally for decoding AUG and AUA codons. In order to obtain normalized values for tRNA gene abundances across organisms for each codon, the number of tRNA genes for every codon documented in the Genomic tRNA database(Chan & Lowe, 2009) (gtrnadb.ucsc.edu) was divided by the total number of tRNA genes in the respective synonymous codon group for each organism. These values (termed $NNN_{\%}$ for each codon) were then utilized, according to the above parameters, to calculate a relative translation elongation value (v) for each codon (termed NNN_v) in a given organism according to the following formulas (where w is a "penalizing" factor for wobble interactions; in this study, w = 3 for all such interactions, as these have been experimentally shown to result in ~3-fold slower translation elongation rates(Curran &

Yarus, 1989; Pedersen, 1984)): For all bacterial codons (except those for Ile, Met and Arg): $NNU_v = NNU_{\%}+NNC_{\%}/w$; $NNC_v = NNC_{\%}$; $NNA_v = NNA_{\%}$; $NNG_v = NNA_{\%}/w$. (2) For bacterial Ile: $AUU_v = AUU_{\%} + AUC_{\%}/w$; $AUC_v = AUC_{\%}$ and $AUA_v = AUG_{\%}/w*2$. (3) For bacterial Met: $AUG_v = AUG_{\%}/2$. (4) For bacterial Arg: treat as a eukaryotic Arg. (5) For eukaryotic two-codon groups and both similar codons of six-codon groups: $NNU_v = NNU_{\%} + NNC_{\%}/w$; $NNC_v = NNC_{\%} + NNU_{\%}$; $NNA_v = NNA_{\%}$; $NNG_v = NNG_{\%} + NNA_{\%}/w$. (6) For eukaryotic four-codon groups, the four similar codons of six-codon groups and Ile: $NNU_v = NNU_{\%}/w + NNC_{\%}/w$; $NNC_v = NNC_{\%} + NNU_{\%}$; $NNA_v = NNA_{\%} + NNU_{\%}/w$; $NNG_v = NNG_{\%} + NNA_{\%}/w$. Values were then assigned to the corresponding codons of any protein coding sequence. From the start of the coding sequence, *v* values of 30 consecutive codons were added and the average value plotted at position number 15. The same operation was performed repeatedly by sliding the window of 30 values one codon position at a time, until the end of the coding sequence was reached. The resulting *v* values were plotted as a function of codon position.

CODING SEQUENCE ENGINEERING

Luc sequences predicted to be translated slowly (Luc_{slow}) were assembled by selecting codons for each amino acid that lack WC decoding tRNA genes in *E. coli* according to the Genomic tRNA database (gtrnadb.ucsc.edu) (Chan & Lowe, 2009) and thus necessitate decoding *via* non-strict WC-based interactions (with the exception of methionine and tryptophan). If genes for all the anticodons of a particular amino acid were present, the codon with the least amount of available anticodon interactions at the

wobble position was selected. Luc sequences predicted to be translated faster (Luc_{fast}), were assembled by selecting codons for each amino acid with the highest number of WC decoding tRNA genes in *E. coli* (and thus be decoded via WC-based interactions). In cases were more than one codon had the highest number of tRNA genes, the codon with the highest number of available anticodon interactions at the wobble position was selected. Similarly, Luc_{cbf} was designed to harbor codons that are the most frequently used in *E. coli*. Luc_{WT} contained the original, endogenous wild type sequence from the firefly *Photinus pyralis*. Sequences for $Luc_{WT\text{-}fast}$ and $Luc_{WT\text{-}cbf}$ contain nucleotides 1-50 from Luc_{WT} and the remaining nucleotides from Luc_{fast} and Luc_{cbf}, respectively. Luc_{re} was assembled by first determining the nature of the decoding interaction for each codon of Luc_{WT} with tRNA gene number data available for the evolutionarily related fly *Drosophila melanogaster* (gtrnadb.ucsc.edu) (Chan & Lowe, 2009) and subsequently selecting the codon that best matched that type of interaction for each codon from the tRNA gene number data of *E. coli*. All sequences contained a C-terminal c-myc-His_6 epitope tag that does not affect the activity of the protein (Agashe et al, 2004).

STRAINS AND GROWTH CONDITIONS

The *E. coli* strain utilized here was BL21 (New England Biolabs), transformed with one the following T7 promoter-based plasmids: $pLuc_{WT}$, $pLuc_{slow}$, $pLuc_{fast}$, $pLuc_{cbf}$, $pLuc_{WT\text{-}fast}$ and $pLuc_{WT\text{-}cbf}$ and $pLuc_{re}$, containing the sequences described above. For activity measurements, cells were grown in LB broth at 37 °C with 250 rpm orbital shaking in volumes that occupied at most one fourth of the total vessel volume, in the presence of

ampicillin (100 μg/ml). Volumes of cells containing equivalent A_{600} values were harvested by centrifugation and lysed by spheroplasting (Ausubel et al, 2003) under native conditions and separated by centrifugation into soluble and pellet fractions (Chang et al, 2005). For pulse-chase analysis, cells were grown in a methionine free defined medium (Teknova) at 37 °C with 250 rpm orbital shaking in volumes that occupied at most one fourth of the total vessel volume, in the presence of ampicillin (100 μg/ml).

PULSE-CHASE ANALYSIS AND CALCULATION OF POLYPEPTIDE ELONGATION RATES

Pulse-chase experiments were performed as described (Sorensen & Pedersen, 1998). Briefly, cells expressing the desired construct were grown and protein expression was induced as described above. At time 0, (30 minutes post-induction), ^{35}S-Met was added to the culture and 10 seconds later, excess unlabeled Met was added. Aliquots were taken every 5 seconds and placed in ice-cold tubes containing chloramphenicol (200 μg/ml final). Cells were harvested and lysates were run on SDS-PAGE followed by autoradiography and scintillation analyses. Polypeptide elongation rates were calculated essentially as described (Sorensen & Pedersen, 1998). In our constructs, there are 14 methionine residues at positions 92, 187, 189, 320, 336, 389, 421, 433, 467, 495, 518, 526, 555 and 584 from the C' terminus. The position-specific incorporation time for each methionine was calculated at various theoretical constant elongation speeds. Calculated times were then utilized to simulate the theoretical appearance of radiolabeled methionines for the various constant elongation speeds (5, 10, 15, 20 aa/s shown as representatives in Figure 2.3). The difference between simulated and measured data was

calculated at each time point and squared. The constant rate with the least sum of squared differences was determined to be the best fit to the data. To model the incorporation of methionine residues (Met) expected during variable (nonuniform) elongation, the average relative translation elongation rate of Luc mRNA sequences connecting neighboring Mets (*i.e.*, modeling the average rate of ribosomal movement from one Met to another) were calculated using the algorithm described above. Each regional relative translation elongation rate was then multiplied by the ratio of best fit constant rate to average relative translation elongation rate for the entire sequence, converting the relative rate (without units) to an elongation rate (aa/s). For example, for Luc_{WT}, the predicted relative rate from Met 92 to Met 187 is 0.475, which was multiplied by 9.8 (aa/s)/0.44 to yield a speed of 10.6 aa/s for that region. In this way, a rate can be predicted for each segment of sequence leading up to and including a particular Met. These rates were then used to calculate the time it would take for each Met to be incorporated as mentioned previously; however, in calculating appearance of Mets at variable rates along the sequence, the time of each Met appearance was calculated successively by taking into account the speed at which upstream Mets were translated. Since the most reliable parameter to assess overall elongation rate is the determination of the time at which no further increase in total radioactivity occurs (*e.g.*, the start of the plateau in the pulse-chase experiments) (Sorensen & Pedersen, 1998), variable rate simulations were normalized by allowing the first data point of each predicted plateau to occur no later than the experimental plateau. All simulated pulse-chase data of variable

rates were subjected to identical least squares analyses as to those of the constant theoretical rates (Fig. 2.5).

QUANTITATIVE PCR ASSAYS

We utilized the SYBR Green quantitative PCR technique to assess steady state accumulation of mRNA from our various constructs, as suggested by the manufacturer (Life technologies), from total RNA from *E. coli* harvested as outlined below for recombinant protein production after one hour of induction. The primer pairs utilized were: Luc_{WT}: forward 5' GCC AAG AAG GGC GGA AAG 3', reverse 5' GAA ATA AGT TTT TGT TCG GAT CGC 3'; Luc_{fast}: forward 5' AAA GGC GGC AAA TCC AAA C 3', reverse 5' GAA ATA AGT TTT TGT TCG GAT CGC 3'; Lucslow: forward 5' TAA GAA GGG TGG TAA GTC TAA GC 3', reverse 5' GAA ATA AGT TTT TGT TCG GAT CGC 3'.

RECOMBINANT PROTEIN PRODUCTION

Starter cultures were grown overnight as described above and diluted the next day. Protein expression was induced at A_{600} = 0.4 with 1 mM IPTG and harvested at 10 min intervals for activity measurements and at 5 second intervals for pulse-chase analysis. Total amounts of recombinant protein produced during each interval were assessed by examining equivalent amounts of cells (equal A_{600} values), which were subsequently lysed, ran on SDS-PAGE and either Coomassie brilliant blue-stained or Western blotted. Aliquots harvested at time points containing equivalent levels of each recombinant

protein produced were then lysed under native conditions as described (Chang et al, 2005) and their solubility and activity assessed as stated below.

DETERMINATION OF PROTEIN SOLUBILITY

Cells were harvested by centrifugation and spheroplasts were prepared as described (Ausubel et al, 2003). Spheroplasts were lysed by dilution into an equal volume of native lysis buffer (5 mM MgSO4, 0.2% (v/v) Triton X-100 (Sigma), Complete EDTA-free protease inhibitors (Roche), 100 units/ml Benzonase (Roche), 50 mM Tris-HCl, pH 7.5). Aliquots were centrifuged into supernatant and pellet fractions (20,000*g* for 10 min) and analyzed by SDS-PAGE followed by staining with Coomassie brilliant blue or Western blot analysis as described (Chang et al, 2005).

DETERMINATION OF LUCIFERASE ACTIVITY

Lysates from cells expressing Luc_{WT}, Luc_{fast}, Luc_{cbf} and Luc_{re} were prepared as above and equivalent dilutions to those used for solubility assessment were utilized. Luc activity was determined using the Luciferase Assay System (Promega) in a Sirius luminometer (Berthold) as described (Agashe et al, 2004). Specific activity values were calculated as described (Agashe et al, 2004). Briefly, total Luc activity values (in relative light units, R.L.U.) were divided by the densitometric value of the band corresponding to the full-length Luc on Coomassie brilliant blue-stained SDS-PAGE or Western blot of equivalent volumes of lysate. The ratio obtained for the protein from the Luc_{WT} construct was set to 100 in every experiment.

CHAPTER 3: POLYPEPTIDE ELONGATION RATES INFLUENCE CHAPERONE RECRUITMENT TO THE NASCENT POLYPEPTIDE

3.1 INTRODUCTION

Most proteins must fold into precise three-dimensional conformations in order to properly perform their respective functions. The information necessary for folding into this native conformation is included in the amino acid sequence of the protein (Anfinsen, 1973). However, for many proteins, amino acid sequence alone cannot account for their final conformation. Factors other than amino acid sequence have been shown to affect protein folding in the cell, including molecular chaperones and translation speed (Agashe et al, 2004; Siller et al, 2010; Spencer et al, 2012) (Chapter 2).

Molecular chaperones constitute a broad class of proteins that bind to other proteins and protect them from forming aberrant inter- and intra-molecular contacts that could lead to incorrect conformations, or misfolding, and aggregation (Bukau et al, 2006). So-called chaperone client proteins are able to fully fold only after they have been released from chaperones and can undergo multiple binding and release cycles before folding (Hartl & Hayer-Hartl, 2009; Young, 2010). The *E. coli* chaperone DnaK operates in this manner with the help of its co-chaperone, DnaJ, which delivers polypeptides to DnaK (Acebron et al, 2008; Liberek et al, 1991; Young, 2010). These chaperones interact with newly synthesized polypeptides both during and after translation to hold them in a folding competent state (Agashe et al, 2004; Hartl & Hayer-Hartl, 2009). However, in bacteria, the first chaperone to encounter the nascent polypeptide is

Trigger factor (TF), which is docked on the ribosome waiting to bind polypeptides as they emerge from the ribosomal exit tunnel (Ferbitz et al, 2004; Hesterkamp & Bukau, 1996; Kaiser et al, 2006). TF binds the emerging polypeptide after ~100 amino acids have emerged from the exit tunnel and detaches from the ribosome (Oh et al, 2011). TF can remain bound to the nascent chain for up to 35 seconds (Kaiser et al, 2006). Therefore, TF, DnaK, and DnaJ can all interact with a nascent polypeptide co-translationally, and this interaction has been shown to enhance yet delay folding relative to translation (Agashe et al, 2004). Another chaperone that the nascent polypeptide encounters is the GroEL/GroES chaperonin system. While TF, DnaK, and DnaJ can interact with the nascent polypeptide before synthesis is complete, GroEL only binds once the polypeptide has been released from the ribosome in wild type cells (Bukau & Horwich, 1998; Hartl & Hayer-Hartl, 2009).

Interestingly, the nonrandom use of redundant codons encoding the amino acid sequence of mammalian P-glycoprotein has been shown to influence polypeptide folding through proposed alterations in elongation rates (Kimchi-Sarfaty et al, 2007). Accelerating the translation speed of firefly Luc, without changing the identity of its amino acid sequence, increased the population of misfolded Luc and the formation of insoluble aggregates (Spencer et al, 2012) (Chapter 2). The precise mechanisms by which translation rates affect protein folding remain unclear. In addition, the possible contribution of molecular chaperones to the rate-induced changes in protein folding has not been explored. Since chaperones generally bind patches of hydrophobic amino acid residues along the protein sequence, alterations in the rate at which these hydrophobic

regions emerge from the ribosome may affect the ability of certain chaperones to bind the nascent chain cotranslationally. Here, we examine the effect of elongation rate on chaperone recruitment to nascent Luc synthesized in *E. coli*.

3.2 RESULTS

We have previously shown that Luc mRNA composed entirely of fast codons (Luc_{fast}), decoded by abundant tRNA via Watson-Crick interactions in the wobble position, is translated approximately two-fold faster than the synonymous wild type mRNA (Spencer et al, 2012) (Chapter 2). The twofold acceleration of elongation rate results in decreased folding efficiency of the encoded Luc compared to wild type rates yet both proteins contain identical amino acid sequences. The majority of Luc_{fast} protein was insoluble, reflected by a significantly greater fraction of recombinant protein present in the pellet fraction under mild centrifugation conditions (20,000*g*, 10 min., 4°C), compared to Luc_{WT}. As expected, the specific activity of the Luc_{fast} gene product was significantly reduced. Conversely, and consistent with these results, slowing bacterial translation of eukaryotic proteins results in increased folding efficiency of the encoded polypeptide (Siller et al, 2010). Thus, we predicted that the gene product of a synonymous Luc mRNA composed entirely of wobble-decoded codons (Luc_{slow}) would result in reduced elongation rates and increased folding efficiency. Attempts at pulse-chase determination of Luc_{slow} translation rates were unsuccessful owing to the lack of production of sufficient material to accurately quantify amounts of full length polypeptide synthesized (Spencer et al, 2012) (Chapter 2). However, enzymatic activity

determinations of equivalent amounts of synthesized protein revealed that Luc_{slow} exhibits higher specific activity and solubility compared to Luc_{WT} (Fig. 3.1). When the soluble fractions of Luc_{WT} and Luc_{slow} are compared, it can be observed that Luc_{slow} yields a considerably higher fraction of soluble and active Luc chains on a per mole basis. The fact that the soluble fraction of Luc_{WT} is not as active as that of Luc_{slow} implicates the existence of a population of improperly folded Luc in the soluble fraction of Luc_{WT} that escapes sedimentation under our centrifugation conditions. To obtain insight into the nature of the various species present in the soluble fractions of extracts from bacteria expressing proteins from these constructs, we analyzed them by size exclusion chromatography to obtain information on the apparent size and/or shapes of Luc-containing complexes. Native Luc predominantly elutes as a monomer, as reflected by maximal activity, in fraction 7 (Fig. 3.2). By comparing the elution patterns of Luc, it is clear that the rate at which Luc emerges from the ribosome affects its mobility on a gel filtration column: Luc_{slow} is mostly present in molecular complexes much larger than native Luc, while Luc_{fast} is present mostly in fractions that correspond to sizes of native or near-native molecules (Fig. 3.2). Remarkably, Luc_{WT}, naturally composed of both WC and wobble decoded codons and thus translated at a speed intermediate between Luc_{slow} and Luc_{fast}, exhibits an elution pattern that is also intermediate of Luc_{slow} and Luc_{fast}. These results indicate that the gradation of the rates at which Luc emerges from the ribosome can similarly affect its propensity to form what appear to be higher molecular weight complexes.

We next set out to determine whether the difference we observed in the fractionation pattern was due to the presence of a higher abundance of soluble aggregates of slowly translated Luc, and/or if it could be explained by preferential binding of molecular chaperones, known to actively participate in the folding of nascent Luc (Agashe et al, 2004). We analyzed the fractions obtained from size exclusion chromatography by performing immunoblots with antibodies for the nascent chain-interacting bacterial chaperones, TF, DnaK, DnaJ, and GroEL (Fig 3.3). As TF is the only ribosome-bound chaperone, and is thus the first chaperone to encounter the nascent

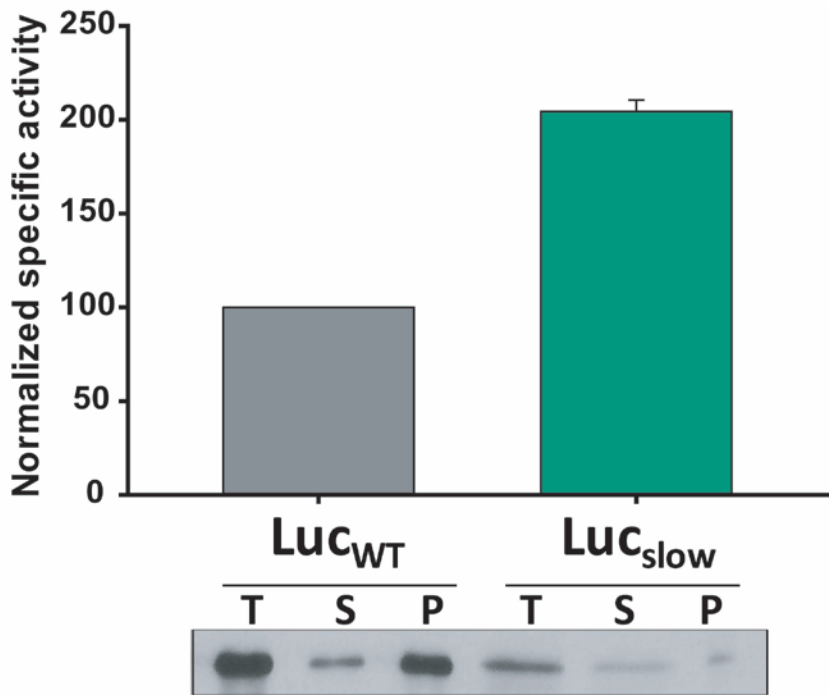


FIGURE 3.1 SYNONYMOUS SEQUENCE BASED DECELERATION INCREASES THE FOLDING EFFICIENCY OF THE ENCODED POLYPEPTIDE.
Specific activities of protein products identical in primary sequence produced from Luc_{WT} and Luc_{slow}, as indicated (top panel). The value of the protein from Luc_{WT} was set to 100. Error bars represent S.E.M. Western blot of total (T), soluble (S) and insoluble (P) recombinant protein produced in *E. coli* from the indicated sequence-engineered constructs (bottom panel).

polypeptide, we anticipated that differential TF binding might be responsible, at least partially, for the differences in elution profiles. Indeed, we observed that TF was the chaperone that displayed the most prominent differences in elution profiles among the

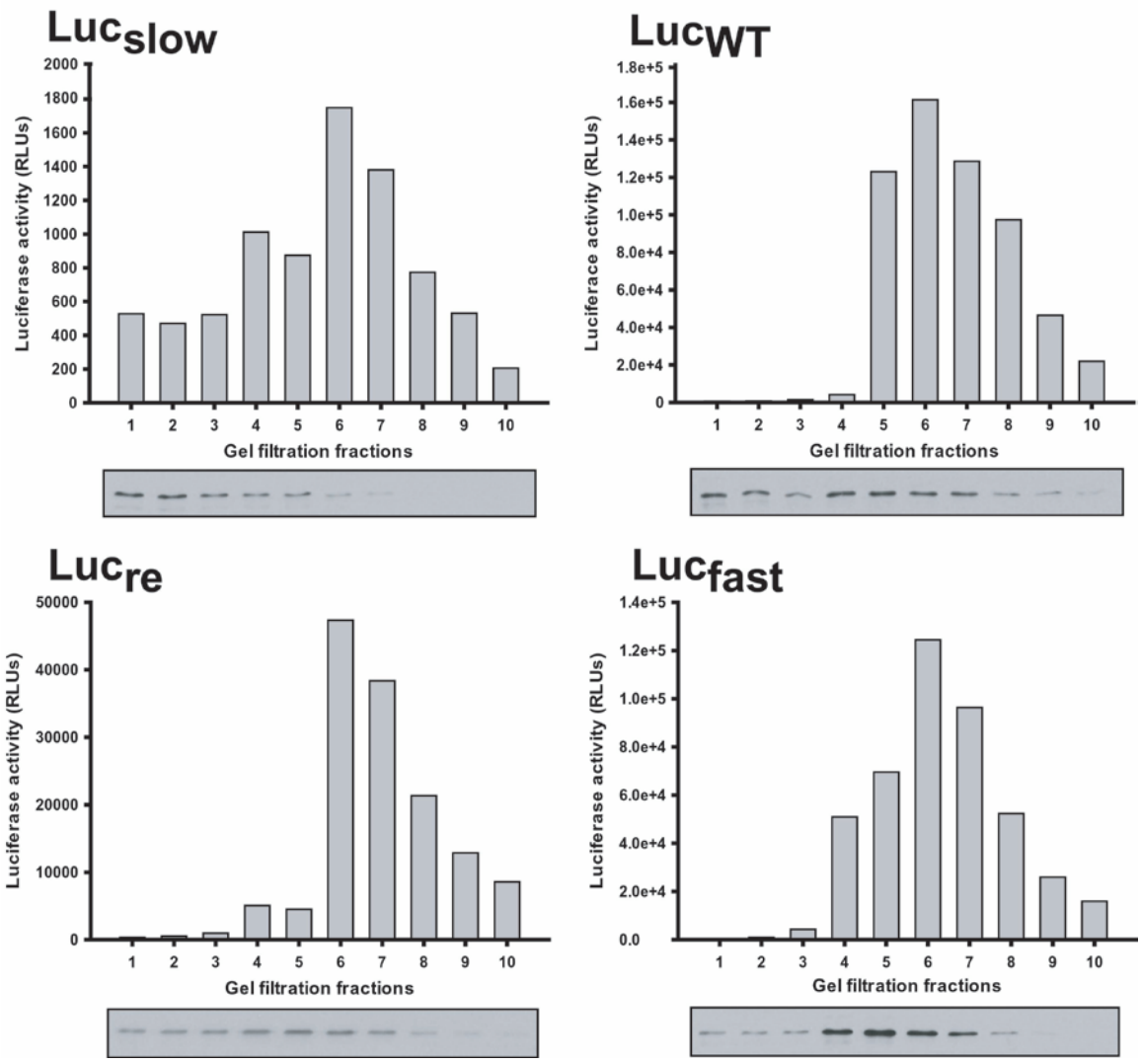


FIGURE 3.2 ELONGATION RATE DETERMINES LUCIFERASE MIGRATION PATTERN.
Lysates containing each sequence-engineered Luc construct were subjected to gel filtration chromatography. Fractions were collected beginning at the column void volume and their activities were measured (bar graphs). Fractions were also separated by SDS-PAGE and immunoblotted (panels below graphs).

clients of varying translation speeds (Fig. 3.3). This suggests that the rate of polypeptide chain emergence from the ribosome can alter the extent of TF-nascent chain interactions. Little variation was observed in elution profiles of the remaining chaperones (Fig. 3.3).

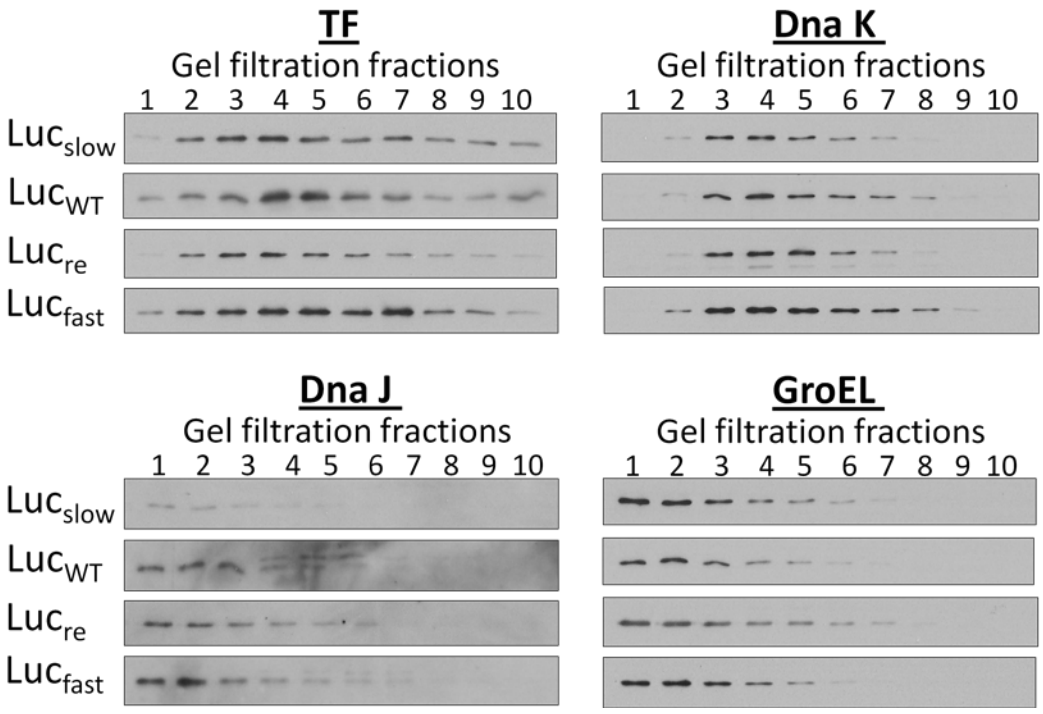


FIGURE 3.3 TEST OF MOLECULAR CHAPERONE CO-ELUTION PATTERN WITH LUC SYNTHESIZED FROM VARIOUS CONSTRUCTS.
Membranes from Fig. 3.2 were subjected to immunoblotting with antibodies for each of the listed chaperones.

However, these chaperones are likely bound to many client proteins other than Luc which could be masking any observable difference in their binding. In order to verify whether chaperone recruitment differs among the Luc speed variants, we performed pull down experiments from total lysates of cells expressing Luc_{WT}, Luc_{fast}, Luc_{re} and Luc_{slow} (Fig. 3.4)

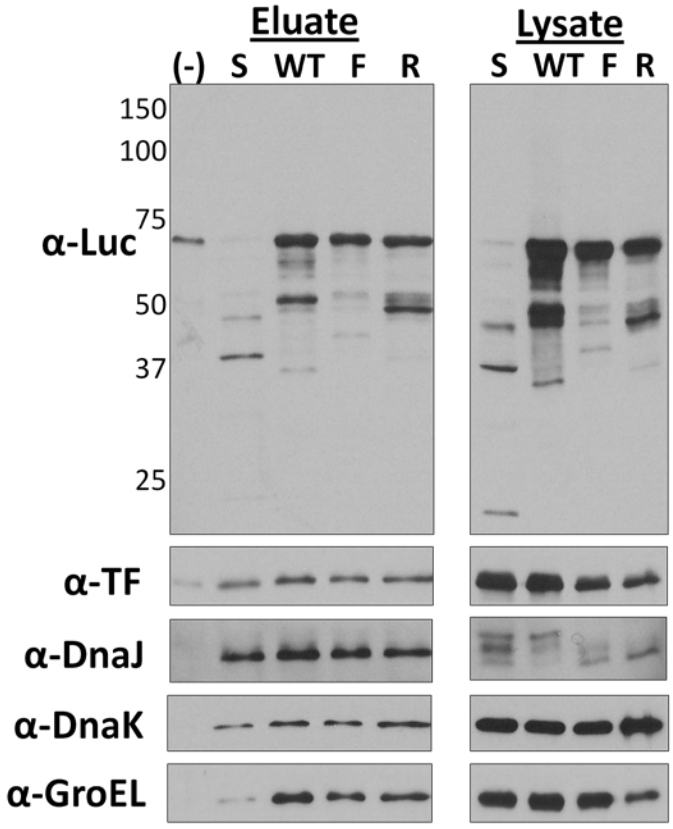


FIGURE 3.4 LUC PULLDOWN REVEALS INCREASED TF, DNAJ, AND DNAK BINDING TO LUC_{SLOW}.
Lysates containing each Luc construct (each containing a C-terminal c-myc-His_6 epitope tag) were precipitated on Ni^{2+}-charged beads. Eluates and lysates for each contruct were separated by SDS-PAGE and immunoblotted with antibodies targeting Luc, TF, DnaJ, DnaK, and GroEL. Negative control (-) refers to Luc-containing lysate incubated with uncharged beads.

Consistent with previous results (Spencer et al, 2012) (Chapter 2), very little Luc_{slow} protein was produced (Fig. 3.4). However, the small amount of Luc_{slow} that was pulled down, compared to the other constructs, was able to pull down TF, DnaJ, and DnaK at levels similar to those of other constructs, even though there is vastly more Luc present in those lysates (Fig. 3.4) This was not the case with GroEL, a chaperone that does not

interact with ribosome-bound nascent chains (Bukau & Horwich, 1998). These results show that TF, DnaK and DnaJ (but not GroEL) are bound to Luc_{slow} to a greater extent than Luc_{WT} and Luc_{fast}. This may also explain, in part, why Luc_{slow} elutes in higher molecular weight complexes than its speed variant counterparts. Taken together, these results indicate that slow translation increases recruitment of co-translationally relevant chaperones to nascent Luc.

It has been shown previously that chaperone binding of nascent polypeptide can delay folding relative to translation, which prevents co-translational misfolding and aggregation (Agashe et al, 2004). In the case of Luc, this chaperone-induced shift towards a more post-translational folding regime has been shown to increase its ability to fold to the native state. Since chaperone recruitment to Luc_{slow} is increased, we reasoned that its post-translational folding component might be enhanced. To explore this possibility, we utilized a previously described experimental paradigm that allows dissection of co-translational versus post-translational folding components in live *E. coli* cells (Agashe et al, 2004). Briefly, if translation is halted (with the antibiotic chloramphenicol) during Luc production, molecules that have completed their synthesis but have not yet folded (post-translational folding) will lead to a measurable increase in enzymatic activity after addition of the protein synthesis inhibitor. On the other hand, if molecules fold mostly co-translationally, no molecules would be expected to fold upon translation inhibition, and no increase in enzymatic activity would be observed. Luc_{slow} post-translational folding was enhanced compared to Luc_{WT} and Luc_{fast}, which both exhibited only a small signal of post-translational activity (Fig. 3.5) This result is

consistent with the roles of TF, DnaK and DnaJ in increasing folding efficiency of nascent polypeptides at the expense of folding time (Agashe et al, 2004).

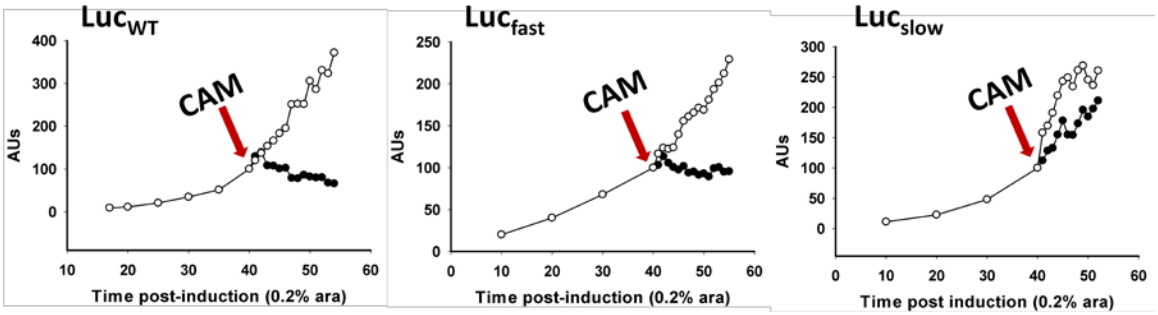


FIGURE 3.5 LUC$_{SLOW}$ POST-TRANSLATIONAL FOLDING REGIME IS ENHANCED RELATIVE TO FASTER CONSTRUCTS.
Bacterial cultures containing each Luc construct were induced (0.2% arabinose) and divided into equal volumes at t = 40 mins. The translation inhibitor chloramphenicol (CAM) was added to only one of the two cultures as indicated. Folding was monitoring by measuring Luc activities before and after CAM addition for each construct at the indicated time points.

3.3 DISCUSSION

Our results show that the rate at which a protein emerges from the ribosome can affect the recruitment of those chaperones which bind co-translationally. Chaperone recruitment to the nascent polypeptide was enhanced during slow elongation and likely plays a significant role in increasing the folding efficiency of Luc$_{slow}$ (Fig. 3.6). What remains unclear is how the elongation rate alters chaperone recruitment. We propose that slow elongation promotes chaperone recruitment either by affording the chaperone more time to bind to the nascent chain or by influencing the conformation of the emerging polypeptide in a way that enhances chaperone affinity. The majority of Luc produced

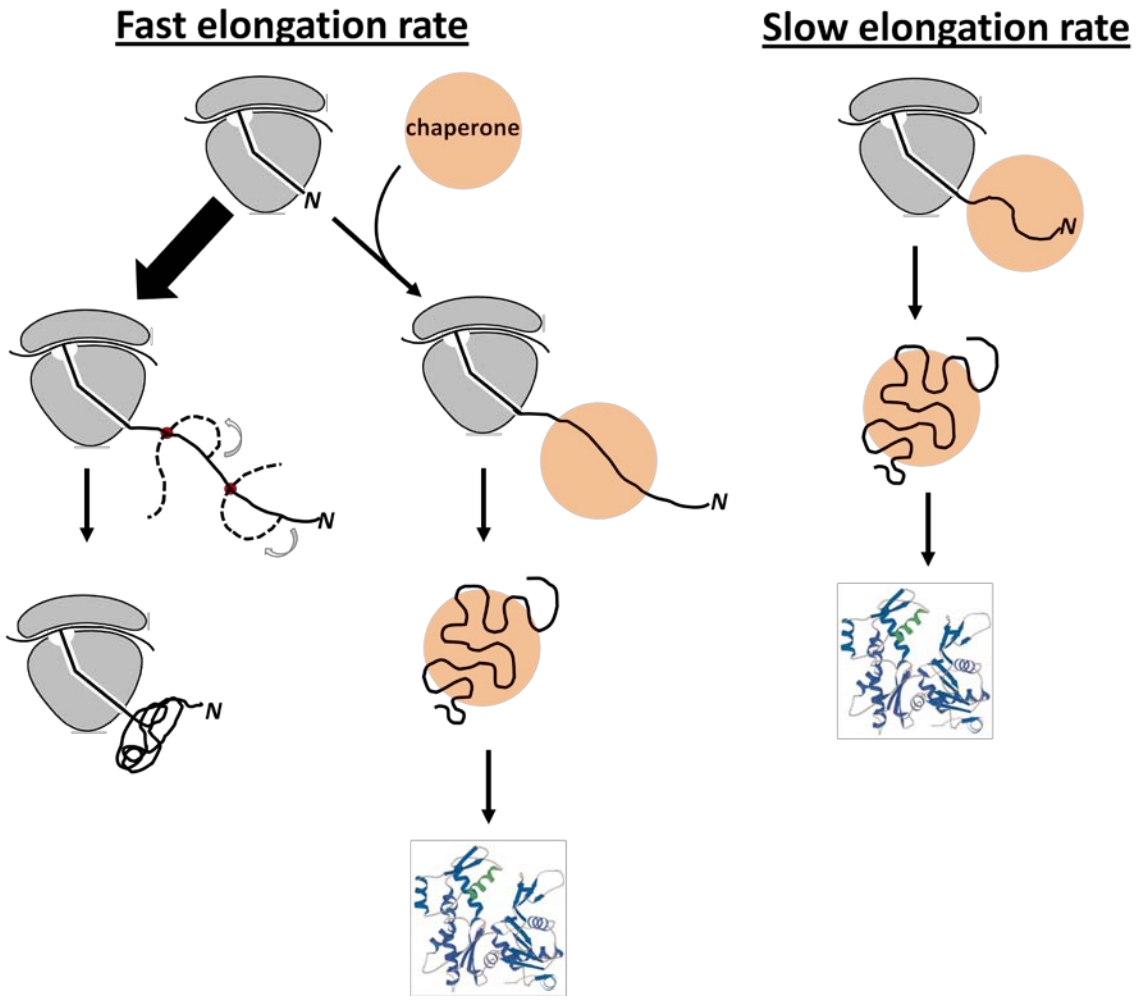


FIGURE 3.6 PROPOSED MECHANISM EXPLAINING CHAPERONE-ASSISTED INCREASED FOLDING EFFICIENCY.
Slow translation of Luc promotes increased co-translational chaperone binding that results in higher fractions of folded Luc. Chaperones are able to bind during fast translation to a lesser extent which is inadequate for folding the majority of fast-translated Luc.

during fast elongation appears to escape the extent of chaperone binding necessary for proper folding (Fig. 3.6). By the same reasoning as above, fast elongation may be too fast for adequate chaperone binding or may lead to the attainment of a conformation for

which the chaperones have low binding affinity. Nonetheless, these results provide, at least partially, a mechanism (Fig. 3.6) by which the rate of emergence from the ribosome affects polypeptide folding and provides further insight into how "silent" mutations may result in disease (Kimchi-Sarfaty et al, 2007).

3.4 MATERIALS AND METHODS

CODING SEQUENCE ENGINEERING

See Chapter 2 Materials and Methods (p. 50).

STRAINS AND GROWTH CONDITIONS

Plasmids used here (pBAD-Luc_{WT}, pBAD-Luc_{fast}, pBAD-Luc_{re}, pBAD-Luc_{slow}) contained pBAD (arabinose-inducible) promoters with Luc sequences identical to those utilized and described in Chapter 2 (p. 52). Growth conditions remained the same as in Chapter 2 (p. 52).

RECOMBINANT PROTEIN PRODUCTION

Protein expression was induced when bacterial cultures reached an A_{600} = 0.4 with 0.2% arabinose and harvested at 60 minutes post-induction for gel filtration, solubility assessment, and activity measurements as described in Chapter 2 Materials and Methods (p. 54).

GEL FILTRATION CHROMATOGRAPHY

Supernatent fractions (three loading loop volumes) of native lysates containing each Luc construct were applied to a Superdex 200 column (GE) pre-equilibrated in 1X phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4). Fractions were collected and equivalent volumes were immediately assayed for Luc activity (as above) or ran on SDS-PAGE and immunoblotted with the anti-Luc (monoclonal, Sigma), anti-TF (polyclonal, gift of Dr. Pierre Genevaux, Université Paul Sabatier, Toulouse, France), anti-DnaK (monoclonal, Stressgen), anti-DnaJ (polyclonal, Stressgen), and anti-GroEl (monoclonal, Enzo Life Sciences) antibodies.

PULL DOWN EXPERIMENTS

NTA agarose beads (Novagen) were charged with 100 mM $NiSO_4$ and equilibrated with 1X PBS. Lysates containing protein produced from our Luc constructs (each containing a C-terminal c-myc-His_6 epitope tag) were applied to the charged, PBS pre-equilibrated beads and washed with PBS. Luc was eluted with 1M imidazole in 1X PBS (pH 7.3). Eluates and Lysates were separated by SDS-PAGE and immunoblotted as mentioned above.

TRANSLATION INHIBITION EXPERIMENTS

Experiments were performed essentially as described (Agashe et al, 2004). Briefly, cultures containing the Luc_{WT}, Luc_{fast}, and Luc_{slow} were induced with 0.2% arabinose. Luc activity of each culture was measured at 10 min intervals post-induction. At 40 min post-induction, cultures were divided in halves with only one half treated with

chloramphenicol (CAM, 200 µg/mL). Cultures remained shaking at 37°C. 5µL samples were taken from each culture at one minute intervals post-CAM addition, diluted 100-fold into ice cold stopping buffer (25 mM Tris-Phosphate buffer [pH 7.4], 2 mM CDTA, 2 mM DTT, 1% Triton X-100, 1 mg/ml bovine serum albumin (BSA)), and placed on ice. Luc activities were measured as described previously in Chapter 2 (p. 55) (Spencer et al, 2012). Activities were normalized by setting t = 40 min activities to 100 arbitrary units (AU).

CHAPTER 4: ARE CERTAIN REGIONS OF A NASCENT POLYPEPTIDE CHAIN MORE MISFOLDING SENSITIVE TO ALTERATIONS IN TRANSLATION RATE?[4]

4.1 INTRODUCTION

Since the deciphering of the genetic code (Nirenberg et al, 1966), scientists have tried to determine its origin as well as investigate possible deeper connections between codons and the properties of the amino acids they encode. The search has even included examining affinities of amino acids for their corresponding anticodons, but this and other searches for physicochemical relationships between amino acids and their cognate codons and anticodons has shown weak correlations and has been largely inconclusive (Deane & Saunders, 2011; Koonin & Novozhilov, 2009). However, it has been shown that a correlation exists between the hydrophobic properties of amino acids and the nucleotide in the second position of the codon (Woese et al, 1966). This finding may explain, in part, the robustness of the genetic code in minimizing the effects of translation errors on protein structure, since the second position of the codon is found to be mistranslated significantly less than the first and third positions (Woese, 1965). A mutation in the first or third position, while changing the identity of the encoded amino acid, would largely conserve the polarity of the mutated residue. Another amino acid property, molecular weight, was found to share a negative correlation with the number of synonymous codons for each amino acid (Hasegawa & Miyata, 1980), a property our lab defines as "codability". More recently, upon the discovery that codons are translated at unequal

[4] This work was performed in collaboration with J. Wrabl and V. Hilser.

rates (Varenne et al, 1984), the focus has shifted to the search for relationships between codon translation rates and the folding properties of the amino acids that they encode. As discussed previously and summarized in Table 1.2, numerous correlations have been reported to exist between some measure of translation rate and some structural or folding property although these studies have not come to unequivocal conclusions (Deane & Saunders, 2011; Saunders & Deane, 2010; Spencer & Barral, 2012). This is perhaps due to a misunderstanding of codon translation rate determinants, which have led to inconsistencies among these studies in predictions of codon translation rates (Spencer & Barral, 2012). We have developed an algorithm that predicts relative codon translation rates that has been experimentally validated (Spencer et al, 2012) (Chapter 2). Furthermore, we have shown that altering translation rates, both globally and regionally, affects the folding of the encoded polypeptide (Spencer et al, 2012) (Chapter 2). These findings led us to wonder how the rate of polypeptide emergence from the ribosome might affect interactions between its amino acid residues which may influence its folding pathway to the native state.

Protein conformation is likely not a static structure but an ensemble of conformations. In the absence of energetic constraints, such as those provided by amino acid and solvent interactions, all possible conformations would be equally represented in the ensemble. However, some amino acid interactions are energetically lower and more favorable than others. The most populated conformation in the ensemble, and thus the most easily observed experimentally, is likely the one in which the all of the amino acid interactions in the protein chain collectively result in the lowest Gibbs free energy (dG)

conformation, referred to as native state (Anfinsen, 1973). The dG value describes the stability of a protein or a system. It is the pursuit of stability (*i.e.*, low dG) that drives atomic interactions in the protein.

Briefly, the dG is calculated from two other thermodynamic contributions (dG = dH – TdS). Enthalpy (dH) is energy absorbed or released into the system when a chemical bond is formed or broken. Minor conformational changes result in numerous changes in hydrogen bonding between amino acid residues and water molecules that are represented by dH and contribute to the overall stability of the system. Entropy (TdS) refers to the degree of disorder in the system. With no constraints, a system will become more disordered, which explains a simple observation of two soluble liquids mixing over time without the assistance of stirring. In the context of a protein, an amino acid region of little constraint, or high TdS, will be highly flexible in conformation and capable of exploring more conformational possibilities. Bond formation restricts the rotational freedom of the amino acid chain which reduces TdS. Therefore, dH and TdS are often opposing values which contribute to the overall dG of a system.

Most of the solved protein structures inform on amino acid placement in the context of the structure but not the thermodynamic parameters. The difficulty of predicting energies of entire proteins, and especially position specific energies of amino acids in the context of the protein, likely contributed to the lack of this type of information (Wrabl et al, 2002). However, this obstacle has been lessened by the COREX algorithm, which predicts global and local stabilities of proteins from known structures (Hilser & Freire, 1996). Briefly, COREX utilizes available x-ray

crystallography and nuclear magnetic resonance (NMR) data to generate an ensemble of unfolded and partially folded conformations by systematically folding and unfolding various regions of the protein. It then calculates the individual probabilities of all possible conformations in an ensemble which reveals the stability and likelihood of a given amino acid to be involved in a particular conformation. From here, the thermodynamic parameters described above can be calculated. The robustness of the algorithm has been corroborated with experimental data and is accurate at predicting structure formation (~77%) (Hilser & Freire, 1996). Furthermore, the development of the eScape (energetic landscape) algorithm, which is based on COREX, provides similarly robust predictions of position specific thermodynamic parameters from amino acid sequence alone (Gu & Hilser, 2008). This powerful tool permits the prediction of energetic profiles of proteins whose structures have yet to be determined and can be employed for a large dataset.

As an initial approach to elucidate the relationship between translation rate and protein folding properties of the polypeptide chain, we decided to search for testable correlations between codon translation rate, as predicted by our experimentally validated algorithm, and folding properties of the encoded polypeptide, namely the energetics of encoded amino acids interacting with one other in the context of the protein sequence. Here, we used eScape (Gu & Hilser, 2008) to computationally predict thermodynamic parameters along the polypeptide sequence to search for correlations with various codon properties including translation speed.

4.2 RESULTS

Based on known protein structure data, the program, eScape (Gu & Hilser, 2008), generates a predicted thermodynamic profile for a given protein from sequence alone, making it ideal for large-scale analysis of thermodynamic profiles. Thus, we examined a large dataset of 20,027 human proteins for correlations with codon usage bias and a parameter we have termed "codability" (the number synonymous codons which can code for one amino acid) along corresponding mRNA sequences. We found no general correlation between codon usage bias and any thermodynamic property (data not shown). However, a correlation (r (20,027) = 0.53, p <0.01) was found between codability and the conformational entropy of the denatured state (DTdSconf) of the protein sequences examined (Fig. 4.1) (Wrabl). Simply stated, DTdSconf is a statistical value describing the number of possible conformations a denatured protein can adopt. This proved to be a highly robust signal when compared to 1000 randomly shuffled variants of the actual codability, since no random shuffling of codabilities gave a higher correlation with DTdSconf than that of the actual codability (Fig. 4.1). These results show a relationship between denatured state conformational entropy and codability or, in other words, degrees of freedom in the denatured polypeptide conformation and the genetic code, respectively.

We next wished to determine whether a correlation exists between these properties of the amino acid sequence and their respective codon translation rates that might reveal a relationship that can be experimentally tested with a model protein. We chose green fluorescent protein (GFP, 27 kDa) from the eukaryote *Aequorea victoria* as a

model protein because: 1) its folding regime is not dependent on molecular chaperones (Chang et al, 2005) ; 2) its fluorescence emission is strictly dependent upon acquisition of the native state and is readily measurable (Tsien, 1998); 3) it folds only marginally upon

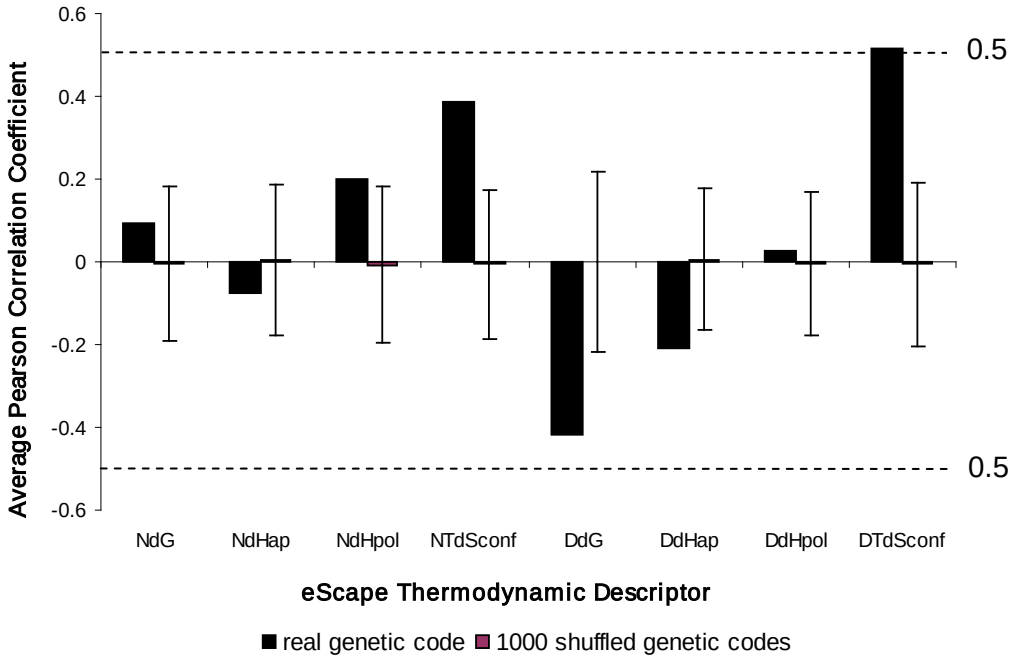


FIGURE 4.1 AVERAGE PEARSON CORRELATION COEFFICIENT (R) BETWEEN CODABILITY AND EACH THERMODYNAMIC PARAMETER FOR A LARGE SET OF HUMAN PROTEINS.

Codability values were obtained from the standard translation table of the NCBI site (www.ncbi.nlm.nih.gov) for 20,027 mRNA. Thermodynamic parameters were calculated for the corresponding protein sequences using eScape (Gu & Hilser, 2008). Black bars represent the average correlation coefficient between codability and each thermodynamic parameter for the 20,027 mRNA sequences and corresponding protein sequences, respectively. Magenta bars represent statistical significance of this average correlation which was computed by randomly shuffling the codability values for each of the 20 amino acids 1000 times, such that the result of each individual shuffle was not repeated. Then the average correlation coefficient over the 20,027 proteins was computed again so that a second average over the correlations was obtained from each of the 1,000 shuffles. DTdSconf was the only thermodynamic parameter to have an $r > 0.5$ as indicated by the dotted lines. (Analysis performed by James Wrabl)

production in *E. coli*, so that subtle folding enhancements can be detected (Chang et al, 2005); 4) altering the translation speed of the entire sequence affects its folding efficiency to a similar extent as what we have shown previously for Luc in Chapter 2 (Figs. 4.2 and 4.3); and 5) it appears to be particularly amenable to our segmental analysis of independently folding units, since it has been shown that some of its fragments can re-assemble *in trans* both *in vivo* and *in vitro* to yield a fluorescent protein (Kent et al, 2008; Pedelacq et al, 2006)

SEQUENCE MANIPULATION PREDICTABLY MODULATES GFP FOLDING

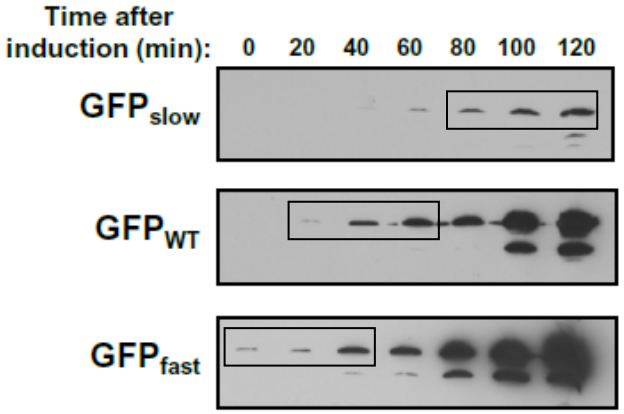


FIGURE 4.2 WESTERN BLOT OF GFP ACCUMULATION PRODUCED BY EACH RECODED CONSTRUCT.
Equal amounts of bacteria producing WT and sequence manipulated GFP were harvested at time points post-induction (0.2% arabinose) as indicated. GFP was detected on Western blot with anti-GFP antibodies (Clontech). Boxed bands indicate approximately equal amounts of protein accumulated to be utilized in the analysis of Fig. 4.3.

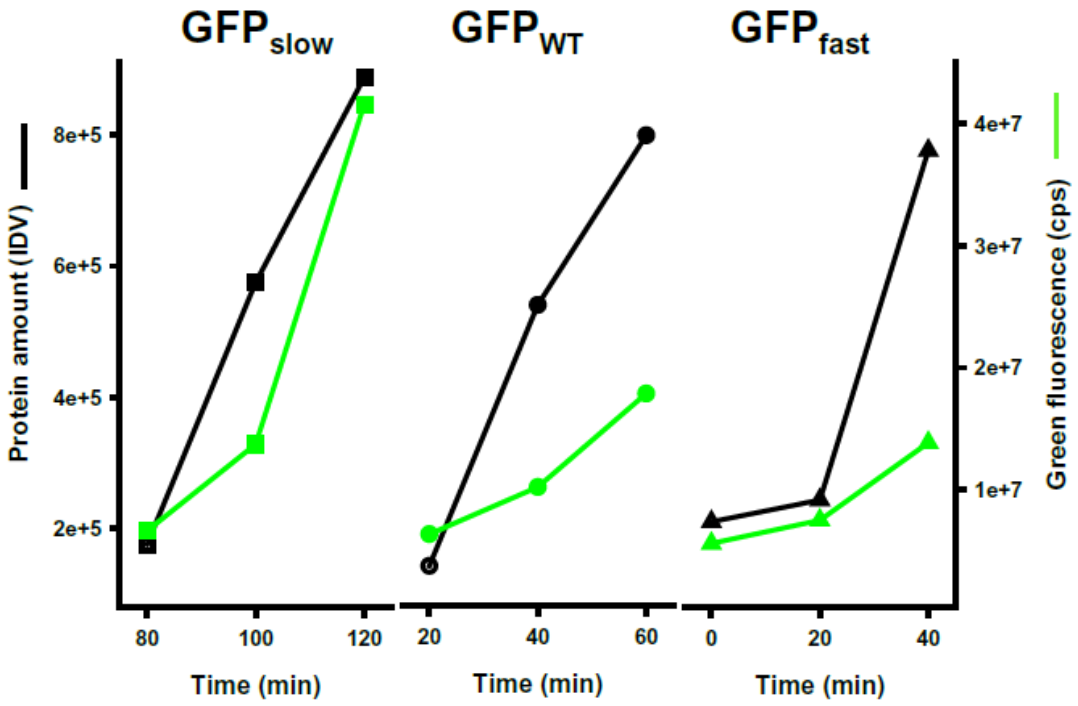


FIGURE 4.3 RELATIVE FLUORESCENCE EMISSION OF GFP VARIANTS.
Specific activities of sequence-manipulated and wild type GFP corresponding protein amounts measured by densitometry of Western blot from Fig. 4.2. Green fluorescence was measured at an excitation of wavelength 398 nm and emission wavelength 508 nm.

TESTABLE RELATIONSHIP EXISTS BETWEEN TRANSLATION RATE AND DTDSCONF

We began by plotting the codability, DTdSconf, and relative elongation rate profiles for GFP_{WT} (Fig. 4.4). The resulting plot shows some regions of significant divergence, which essentially display inverse correlations. In the first half of the protein there exists a region of high DTdSconf and codability and low elongation rate, whereas downstream in the sequence the opposite is observed. These divergent regions in the plot are of particular interest, since one testable hypothesis for the robust correlation between

DTdSconf and codability is a probable need for slower translation speeds in regions of the nascent chain where entropy values are high. Regions of amino acids with high DTdSconf values could explore more conformations that are presumably off the folding pathway and may benefit from slower translation rates. Indeed, the region corresponding approximately to residues 37-73 might appear to exhibit this behavior. Conversely, the regions corresponding approximately to residues 73-113 and 133-181 exhibit the opposite relationship, an area of low DTdSconf and codability that possibly would not benefit from translation speed regulation. Data from this plot indicate regions of GFP potentially sensitive to translation speed manipulation, which prompted us to perform an experimental test of the hypothesis that *regulation in the form of slower translation speeds in areas of high denatured state entropy and codability (region 37-73) is beneficial for protein folding*. If this hypothesis is correct, we would expect that slowing translation speed in region 37-73 would increase folding efficiency, whereas hastening translation speed in this same region would decrease folding efficiency.

DESIGN STRATEGY TO TEST TRANSLATION RATE AND DTDSCONF RELATIONSHIP

In order to test this experimentally, we recoded codons 1-77, corresponding to an area of high denatured state entropy/codability and low translation speed, to be slow by using wobble-based substitutions in that region (chimera 1, Fig. 4.5). We have recoded the same region to have a faster translation speed by using Watson-Crick-based substitutions (chimera 2, Fig. 4.5). It is important to note that the remaining residues, 78-240, are not recoded and, therefore, should retain WT speeds (Fig. 4.6, left panel). We

predict that these recoded regions will be the most sensitive to translation speed manipulation, as explained above. On the other hand, we have designed a pair of control chimeras (chimeras 3 and 4, Fig. 4.5), corresponding to the slow and fast recoding of region 113-190, respectively. The amino acid residues N- and C-terminal to this region

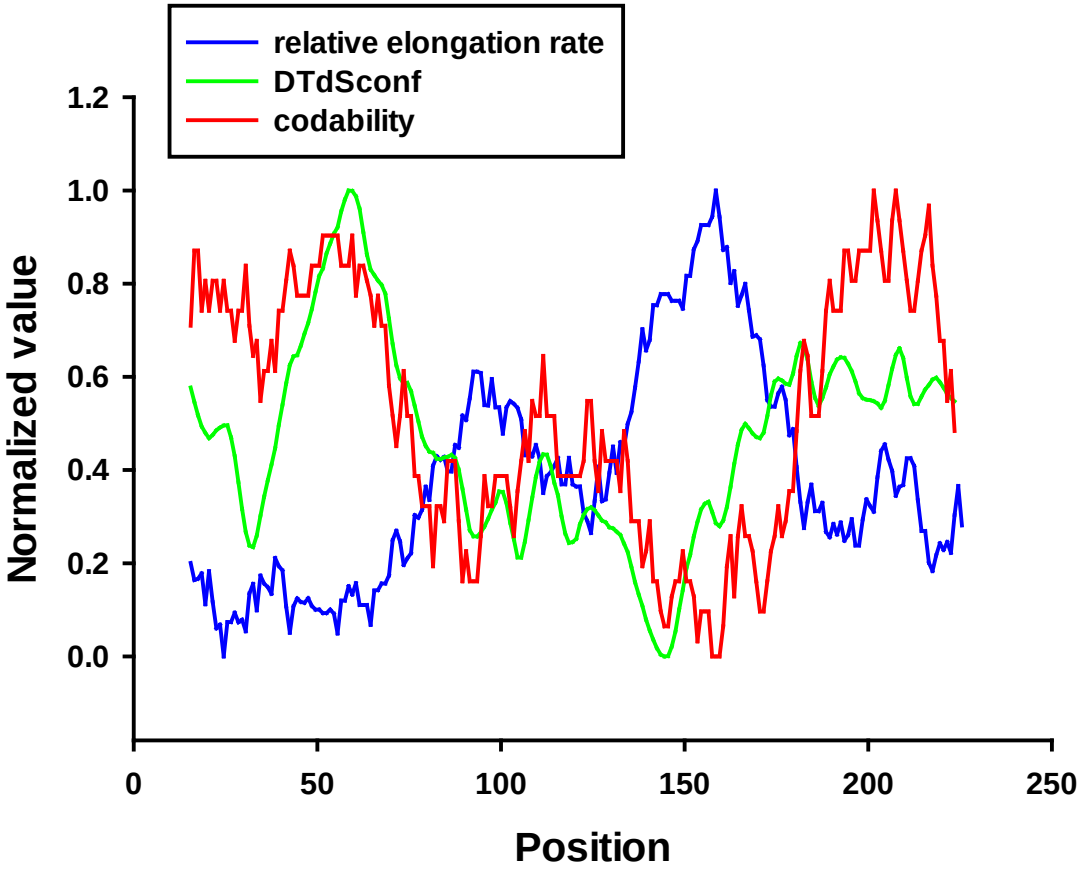


FIGURE 4.4 PLOT DISPLAYING DTDSCONF AND CODABILITY OF GFP PROTEIN AND RELATIVE ELONGATION RATE OF GFP MRNA EXPRESSED IN *E. COLI.*
The relative elongation rate profile (blue) (calculated using algorithm described Chapter 2) was plotted with DTdSconf (green) and codability (red) data generated for GFP as described in Materials and Methods. Each parameter was averaged over a window of 30 codons/amino acids and normalized such that minimum and maximum values were zero and one, respectively.

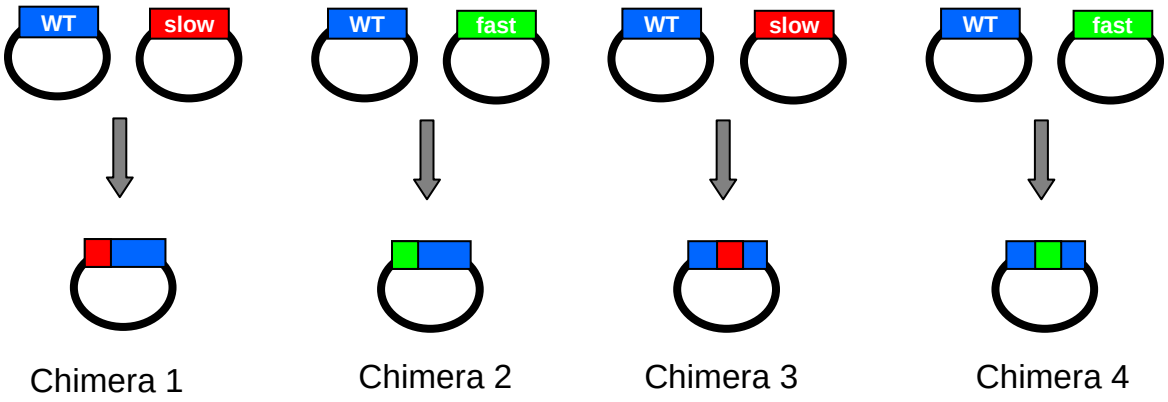


FIGURE 4.5 CONSTRUCTION STRATEGY FOR GFP CHIMERAS TO ASSESS THE IMPACT OF SPEED OF FOLDING OF SPECIFIC REGIONS.
For chimeras 1 and 2, a PCR-based approach was utilized to amplify the sequence corresponding to residues 78-240 from a GFP_{WT} vector, which was then recombined via blunt ligation and restriction endonuclease digestion to amplified sequences corresponding to residues 1-77 from GFP_{slow} and GFP_{fast} vectors, respectively. Chimeras 3 and 4 were constructed by blunt ligation of amplified sequences corresponding to residues 1-112 and 191-240 from GFP_{WT} vector to amplified sequences corresponding to residues 113-190 from GFP_{slow} and GFP_{fast}, respectively. Blue refers to WT mRNA sequence, while green and red refer to fast and slow recoded mRNA, respectively. This color scheme is also utilized in Fig. 4.6.

were not recoded and should be translated at WT speeds (Fig. 4.6, right panel). We have predicted that these chimeras will be relatively insensitive to these manipulations, given the low entropy/codability and high translation speed of this region. Interestingly, the region recoded in Ch 1 and Ch 2 includes the central alpha helix which is positioned inside the beta barrel and contains the chromophore, the formation of which is necessary for GFP activity (Fig. 4.7, left panel, highlighted in red). The sudden change in secondary structure from beta strand to alpha helix and the intricate placement of the helix inside the beta barrel could both pose a folding hurdle that may be aided by slow

translation. Since this region overlaps with the high DTdSconf region, an improvement in folding of the central alpha helix by slow translation would provide further basis for the expected increase in folding efficiency of Ch 1. On the other hand, the recoded region corresponding to Ch 3 and Ch 4 lack this potentially problematic change in folding (Fig. 4.7, right panel, highlighted in red).

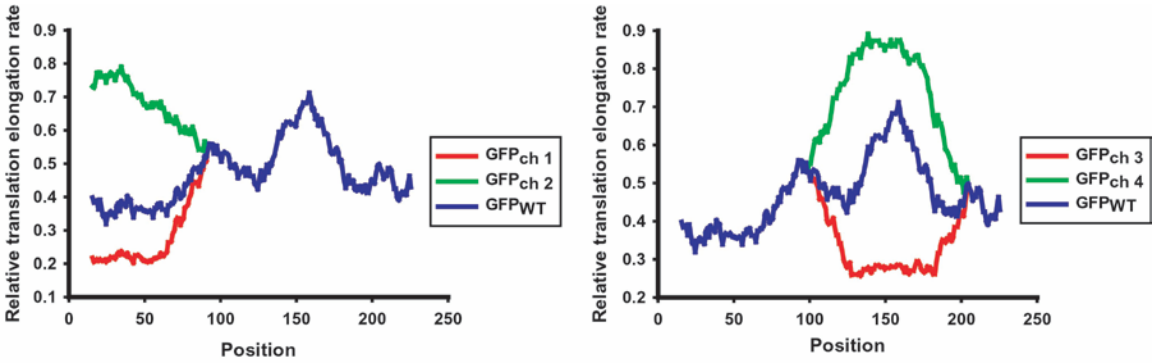


FIGURE 4.6 TRANSLATION SPEED PROFILES FOR GFP CHIMERAS.
Translation speed profiles were generated for chimeras 1 and 2 (left) and chimeras 3 and 4 (right) and overlaid with the wild type GFP speed profile to display areas in the sequence where rates were altered. Color scheme is identical to Fig. 4.5. Please note that red regions (as in Ch1 and Ch3) are predicted to be translated slower than the corresponding WT (blue) regions. Conversely, green regions (as in Ch2 and Ch4) are predicted to be translated faster than the corresponding WT regions.

GLOBAL AND LOCAL RATE DECELERATIONS INCREASE GFP FOLDING EFFICIENCY

Measurements of GFP fluorescence and protein abundance reveal that slowing translation in the high entropy region (GFP_{ch1}) results in increased specific activity relative to wild type, whereas accelerating translation in this same area (GFP_{ch2}) results in decreased specific activity relative to wild type (Fig. 4.8). Unexpectedly, similar results were

observed for speed alterations in the region corresponding to low entropy values (GFP_{ch3} and GFP_{ch4}, Fig. 4.8). Interestingly, we observe that altering the translation speed of only a fraction of the protein sequence resulted in folding efficiency changes of similar magnitude to full length sequence speed manipulations (GFP_{slow} and GFP_{fast}).

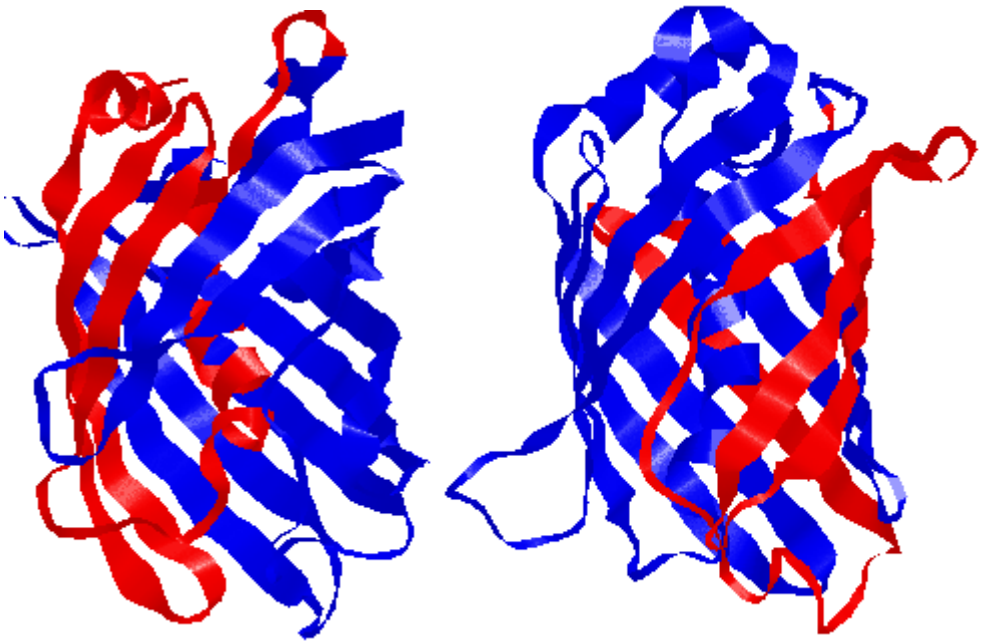


FIGURE 4.7 RECODED REGIONS HIGHLIGHTED IN GFP STRUCTURE.
Ribbon structure of GFP with structural regions highlighted in red corresponding to recoded regions in Chs 1 and 2 (left) and Chs 3 and 4 (right). The blue portions represent amino acids with WT mRNA. Note that the recoded region in Chs 1 and 2 contains the central alpha helix which includes the chromophore.

4.3 DISCUSSION

Based on these experiments, we are unable to conclude that one region of the protein is more sensitive to translation speed manipulation than any other. Nonetheless, the

hypothesis should not yet be rejected as the experimental design may be inadequate. For instance, it may be particularly important to shift our sequence manipulations to regions that are 30 residues C-terminal of the experimental regions. The rationale for this is that because the ~100 Å ribosomal exit tunnel can accommodate ~30 residues (up to 60 in alpha helical form) (Kramer et al, 2009), the speed at which the codon in the ribosomal A

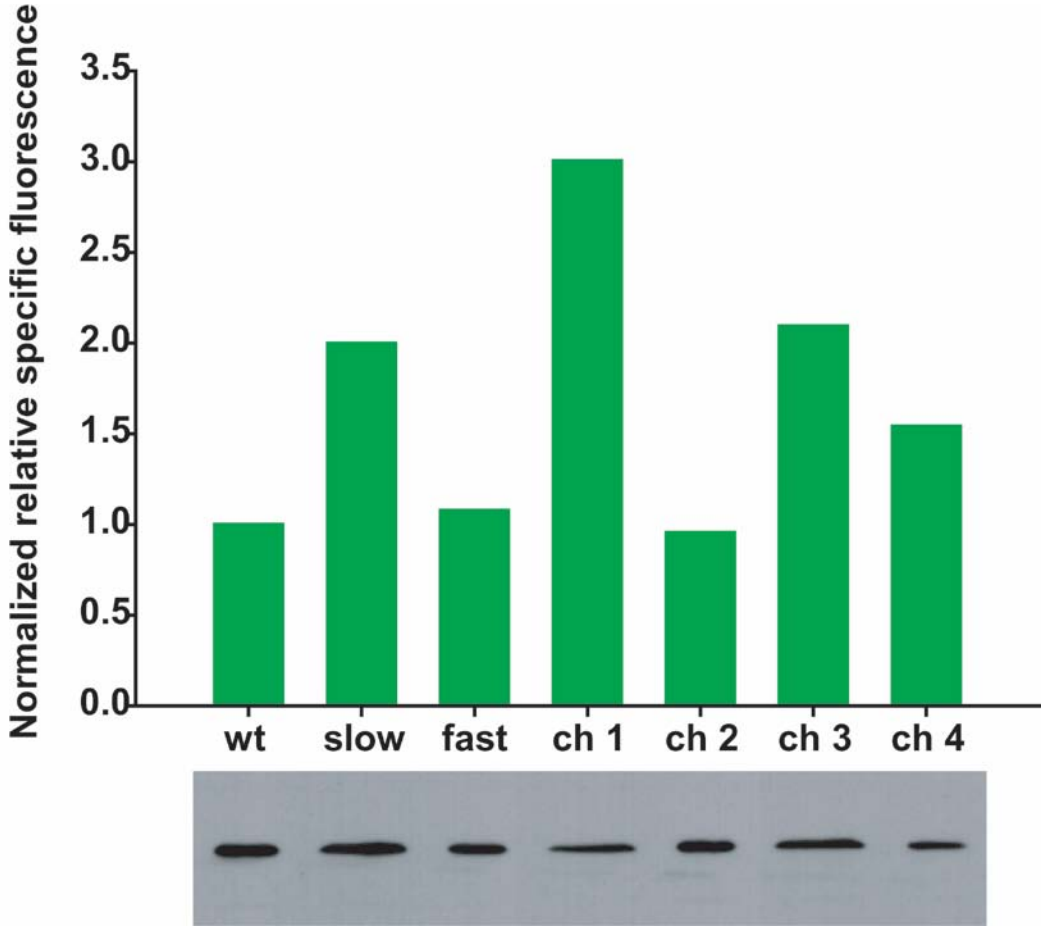


FIGURE 4.8 RECODING TRANSLATION SPEED AFFECTS FOLDING OF GFP CHIMERAS. Bar graph depicts specific GFP fluorescence relative to protein abundance of the corresponding total bands in lower panel, after normalization for loading (see Materials and Methods). Protein bands for high accumulation constructs (WT, fast, Ch 2, and Ch 4) represent one tenth of the total protein produced in order to maintain linearity in immunodectection signal. Relative specific fluorescence values were normalized such that GFP_{WT} fluorescence is set to 1.0.

site is translated will ultimately dictate how fast the amino acid approximately 30 residues away will emerge from the ribosomal exit tunnel (Kelkar et al, 2012). In addition, it is possible that our methods, particularly protein quantification by densitometry, are not of sufficient resolution to measure subtle differences in folding as a result of regional speed manipulation. Including these considerations in subsequent experimental designs may allow observable differences in folding efficiencies in the GFP regions tested that were not evident in these experiments.

4.4 MATERIALS AND METHODS

STRAIN AND GROWTH CONDITIONS

The *E. coli* strain utilized here was BL21 (New England Biolabs). For recombinant protein production, (see text below), this strain was transformed with arabinose-controlled promoter-based plasmids (Guzman et al, 1995) – pBAD-GFP_{uv} (encoding the Cycle3 variant of GFP (Crameri et al, 1996)), pBAD-GFP_{slow} and pBAD-GFP_{fast} (DNA 2.0), and pBAD-GFP_{ch1}, pBAD-GFP_{ch2}, pBAD-GFP_{ch3}, and pBAD-GFP_{ch4} (see text below). Cells were grown in LB broth at 37°C with 300 rpm of orbital shaking in volumes that occupied, at most, one fourth of the total vessel volume in the presence of ampicillin (100 µg/ml).

RECOMBINANT PROTEIN PRODUCTION

Starter cultures were grown as described above. Protein induction was carried out when the cell density had reached A_{600}=0.4 with 0.2% (wt/vol) arabinose. Volumes of

cells containing equivalent A_{600} (1 OD) values were harvested by centrifugation after a 30 min. incubation chloramphenicol (200 µg/ml) to halt any further protein production and to allow time for GFP chromophore formation. The samples were then centrifuged at 4°C and 20,000g for 1 min. The supernatant was removed, and the pellets were flash-frozen and stored at -20°C. Cell pellets were thawed on ice. Spheroplasts were prepared as described previously (Ausubel et al, 2003) and lysed under native conditions by dilution into an equal volume of native lysis buffer [5 mM $MgSO_4$, 0.2% (vol/vol) Triton X-100 (Sigma), Complete EDTA-free Protease Inhibitors (Roche), 100 U/ml Benzonase (Roche), and 50 mM Tris-HCl (pH 7.5)]. Samples were then run on SDS-PAGE and immunoblotted to assay their protein production, and fluorescence emission were assessed (see the text below).

RECODING GFP MRNA

GFP_{slow} and GFP_{fast} were designed in the same manner as before (Spencer et al, 2012) (Chapter 2) by recoding the entire gene using codons which rely only on wobble decoding or Watson-Crick decoding tRNA, respectively, synthesized with the appropriate flanking restriction sites by DNA 2.0, Inc. (Menlo Park, CA), and sub-cloned into the pBAD vector. tRNA gene information for *E. coli* was obtained from the genomic tRNA database, (www.gtrnadb.ucsc.edu)(Chan & Lowe, 2009). A PCR-based approach was utilized to amplify the sequence corresponding to residues 78-240 from pBAD-GFP_{WT}, which was recombined via blunt ligation and restriction endonuclease digestion by *Hind* III to amplified sequences corresponding to residues 1-77 from pBAD-GFP_{slow} and pBAD-GFP_{fast}, respectively, to form chimeras 1 (pBAD-GFP_{ch1}) and 2 (pBAD-GFP_{ch2}). Chimeras 3 (pBAD-GFP_{ch3}) and 4 (pBAD-GFP_{ch4}) were constructed by blunt ligation of

amplified sequences corresponding to residues 1-112 and 191-240 from pBAD-GFP_{WT} to amplified sequences corresponding to residues 113-190 from pBAD-GFP_{slow} and pBAD-GFP_{fast}, respectively. The design for chimeras 1-4 is illustrated in Fig. 4.5. The amino acid sequences remained unchanged.

DETERMINATION OF GREEN FLUORESCENCE

Green fluorescence was measured in a Fluorolog 3 fluorescence spectrometer (Horiba/Jobin Yvon) with excitation wavelength of 398 nm and emission wavelength of 508 nm, as described previously (Chang et al, 2005). Relative specific fluorescence was determined by dividing raw fluorescence by protein abundance (see below) of total GFP.

DETERMINATION OF GFP ACCUMULATION

Following native lysis as described above, total cell lysates were analyzed by SDS-PAGE (12% (wt/vol) acrylamide) followed by immunoblotting with the anti-GFP JL8 monoclonal antibody (Clontech). In Figure 4.8, all lysates containing GFP constructs with fast codons (WT, fast, Ch2, and Ch4) were diluted 10-fold in SDS loading dye in order to maintain linearity in immunodection signal among the GFP constructs.

AVERAGE PEARSON CORRELATION COEFFICIENT FOR CODABILITY AND THERMODYNAMIC PARAMETERS

20,027 human mRNA sequences and corresponding expressed protein sequences were taken from the NCBI Consensus Coding Regions Set (CCDS_nucleotide.20080430.fasta) and screened for translational accuracy between the

mRNA and corresponding protein (Pruitt et al, 2009). The program eScape (Gu & Hilser, 2008) was used to obtain predicted thermodynamic information for each protein sequence of the set. Codability values were obtained from the standard translation table of the NCBI site (www.ncbi.nlm.nih.gov). For each mRNA or protein sequence, codability values or denatured state conformational entropy values were averaged at each position over a window size of 30 residues. A Pearson correlation coefficient (Press, 1992) was computed for each protein between the averaged values of codability and thermodynamics. In these correlations, the first 15 and last 15 residues were neglected due to incomplete averaging. The average correlation coefficient over the 20,027 proteins in the set is reported as the black bars in Figure 4.1. To assess the statistical significance of this average correlation, the codability values for each of the 20 amino acids were randomly shuffled 1000 times such that the result of each individual shuffle was not repeated. For each shuffled codability, the average correlation coefficient over the 20,027 proteins was again computed, and a second average over the correlations obtained from each of the 1,000 shuffles is reported as the magenta bars in Figure 4.1.

CHAPTER 5: SUMMARY AND PERSPECTIVES

This dissertation was motivated by the desire to more fully understand the protein folding process as it occurs *in vivo*. Understanding the mechanisms by which a protein samples only a few of the copious conformational possibilities and ultimately adopts the conformation needed to be fully functional is important because, once they are attained, these folds can inform and empower proteins to perform specific functions. While it was once thought that amino acid sequence alone was entirely sufficient for proteins to fold properly (Anfinsen, 1973), auxiliary roles of molecular chaperones (Hartl & Hayer-Hartl, 2009) and translation speed (Deane & Saunders, 2011; Marin, 2008; Siller et al, 2010) in protein folding have been uncovered, albeit with far less experimental studies on the latter. The idea that translation speed might affect protein folding is not new. Indeed, it was not long after the discovery that protein synthesis is a unidirectional process (Naughton & Dintzis, 1962) that this was idea was postulated (Itano, 1968b). To date, however, most of the evidence generated in support of this notion has been theoretical or anecdotal findings. Here, we set out to strategically and experimentally investigate the effect of translation speed on protein folding. The data presented here reveal that the rate at which newly synthesized polypeptides emerge directionally from the ribosomal exit tunnel is genetically encoded and influences the folding fate of the encoded polypeptide.

We first determined the factors that dictate polypeptide elongation rate (nature of codon:anticodon interaction and tRNA concentration, see Chapter 1) since a consensus regarding the rate determining factors was lacking (Spencer & Barral, 2012). While many translation speed studies collectively pointed to codon:anticodon interaction and

tRNA concentration, the majority of those investigating translation rate and protein folding largely relied on codon bias (*i.e*., the biased usage of certain synonymous codons over others) as predictor of translation rate (Spencer & Barral, 2012) (Chapter 1). Using codon:anticodon interaction and tRNA concentration as rate determinants, we developed a predictive algorithm that computes a relative translation elongation rate for any codon in any mRNA from an organism of known tRNA gene content (Spencer et al, 2012) (Chapter 2). We showed experimentally that these determinants are more accurate predictors of translation elongation rate than the commonly used codon bias (Spencer et al, 2012) (Chapter 2). We then constructed synonymous Luc mRNAs that were globally fast or slow according to our predictions and used the pulse-chase method to measure polypeptide elongation rates of these constructs in live *E. coli* cells. To our knowledge, this is the first time actual translation rates of protein-encoding mRNA, strategically recoded to alter translation rates, have been measured in live E. coli cells. We found that the measured rate acceleration of Luc_{fast} was proportional to the acceleration of the overall average relative rate predicted by our algorithm. This finding validated the ability of our algorithm to predict average relative elongation rates for the entire length of an mRNA translated in *E. coli*. Regarding the validation of predicted *regional* rate variations, the pulse-chase method used here appears to lack the resolution necessary to adequately detect regional variations in elongation rate. However, when we modeled the pulse-chase data using variable rates based on our predictions, the agreement was either very similar or better than the data fitting using constant rates (Chapter 2). Methods with higher resolution would need to be employed to experimentally validate the *regional*

variations that our algorithm predicts. For example, ribosomal profiling of our Luc constructs could provide such resolution and may reveal higher ribosomal occupancy times for areas of predicted rate deceleration and shorter ribosomal occupancy times for areas of predicted rate acceleration. It is possible that it may also reveal ribosomal pausing in regions not predicted to be slow by our algorithm, such as those that contain the Shine-Dalgarno sequence (Li et al, 2012). Utilization of such methods would then provide a basis for the refinement of the algorithm, which we acknowledge is an oversimplification of a multifaceted biological process.

Synonymous recoding of Luc provided us with the unique opportunity to examine the role of various translation speeds in the folding of nascent polypeptides that are identical in amino acid sequence. Monitoring the activity and solubility of these proteins revealed an apparent increase in folding efficiency as the result of decreased translation elongation speed, which had been observed previously using mutant ribosomes (Siller et al, 2010). Luc_{slow} had a higher fraction of properly folded protein. We assume that Luc_{slow} was indeed translated more slowly, although we could not measure its translation rate by pulse-chase analysis due to low accumulation. We speculate that this is due to ribosomal stalling and possibly degradation, although this has not been validated. Further investigation revealed that increased co-translational chaperone binding (TF, DnaK, and DnaJ) during slow translation may explain the increased folding ability of Luc_{slow} (Chapter 3). We propose that altered chaperone binding is due to differences in accessibility to chaperone binding sites along the emerging polypeptide. Accessibility to chaperone binding sites could be altered in at least two ways. First, the rate of

polypeptide emergence from the ribosome may be incompatible with the rate of chaperone binding. Second, it is also possible that ribosomal emergence rates cause the nascent polypeptide to acquire a conformation that is incompatible with chaperone binding. Taken together, it may be that varying the rate at which hydrophobic residues exit the ribosomal tunnel, and thus the rate at which they encounter chaperones and neighboring hydrophobic residues, may result in the most significant changes in folding efficiency.

From a biotechnological perspective, we were interested in how our algorithm might aid in the improvement of recombinant protein production, since producing an adequate amount of properly folded protein is often a limiting step for structure determination studies, vaccine development, and protein recombinant pharmaceuticals (Dingermann, 2008). Attempts to improve recombinant protein production often include a speed harmonization technique whereby the codon usage of a given protein in its originating organism is recapitulated in the expression host by recoding the sequence to contain codons of equivalent usage bias by the host (Angov et al, 2008). Although this technique erroneously assumes codon bias to be the most adequate metric of translation speed, it often improves recombinant protein production. Since our algorithm is superior to codon bias at predicting relative translation rates, we reasoned that using it to recode Luc to recapitulate translation speed variations naturally occurring in the firefly when recombinantly produced in the expression host *E.coli,* would result in higher amounts of properly folded Luc. This is indeed what we observed with Luc_{re}, which accumulated to levels similar to that of wild type protein but displayed an improvement in folding that

was of equal magnitude to that observed by slowing translation speed. Interestingly, unlike Luc_{slow} the folding improvement of speed harmonized Luc did not appear to be a result of increased chaperone recruitment, possibly indicating two mechanisms by which folding pathways can be enhanced as a result of translation speed: increased chaperone recruitment or increased on-pathway intramolecular contacts. The next step would be to examine folding as a function of translation speed in the absence of co-translational chaperone assistance. This can be accomplished by expressing the speed variant messages in chaperone deleted (TF, DnaK/J) mutant *E. coli* cells to monitor folding efficiency as described in this work.

Aligned with this reasoning, we chose to recode the mRNA and monitor the folding of another bioluminescent protein, GFP (green fluorescent protein) whose folding regime is not dependent on molecular chaperones (Chang et al, 2005). We observed that GFP folding efficiency responded to global alteration of translation rate similarly to Luc speed variants. We did not perform sequence harmonization on GFP since its speed profiles in *E. coli* and *D. melanogaster* were quite similar already, and, therefore, harmonization would likely not result in a significant improvement in folding. Slow translation increased GFP folding efficiency, presumably without an increase in chaperone recruitment. A possible mechanism explaining this observation is that slow translation is compatible for native fold-promoting intramolecular interactions in GFP and other proteins which may or may not require chaperone assistance. If this is the case, then certain folding regions should be more sensitive than others to changes in translation speed. In search for such regions, we discovered that segments of GFP predicted to be

translated slowly corresponded to protein segments of high conformational entropy in the unfolded state (*i.e.* protein regions that, when denatured, have an innate conformational flexibility). We reasoned that the increased conformational freedom of these protein segments may benefit from slower translation. When we slowed translation rate only in these regions, we found an improvement in folding equal to what we observed with the globally slow message. However, the folding improvement was also similar to what we observed by slowing translation in regions corresponding to low entropy. As a result, we were unable to conclude that high entropic regions of GFP require slow translation speeds for optimal folding and suggest the refinement of the methods. For example, a better approach may involve recoding the translation rate *upstream* of the high entropy region so that the rate at which that region exits the ribosomal tunnel is affected. Also, it is possible that varying translation speed in different regions of GFP is changing its folding pathway but the differences are too subtle to be detected by our methodology, which only reports on folded or misfolded instead of more subtle conformational changes. An alternative approach would be to randomly, synonymously mutate GFP mRNA and examine the effects on folding. The silent mutations detrimental to GFP folding as well as those that improve or have no effect on folding could be mapped onto the speed and thermodynamic profiles of GFP to highlight regions that are sensitive or insensitive to translation rate modulation. Interestingly, GFP mRNA has already been randomly mutated in this manner by Kudla *et al.* who created a library of 154 chimeric GFP genes, although the authors only analyzed the data for expression effects as a result of mutations altering mRNA secondary structure (Kudla et al, 2009). The comparison of

these sequences with the folding efficiencies of the encoded proteins would nicely complement and clarify the results in Chapter 4, revealing protein folding regions that are sensitive to translation speed manipulation.

The degeneracy of the genetic code has provided coding options for most amino acids throughout evolution. While this redundancy may seem energetically wasteful, its significance to the encoded proteins and organisms themselves is becoming clearer, particularly by studying the nonrandom usage of synonymous codons. Many sequences contain clusters of slow codons and clusters of fast codons throughout their messages. Our results indicate that these variations are likely not due to chance and that these accelerations and decelerations have been maintained throughout evolution to aid in the folding of newly synthesized polypeptides. The folding requirements necessitating slow translation of one region of a protein and acceleration of another are not yet clear. However, understanding this relationship would represent a breakthrough in a largely unexplored area of science: examining the role of rate changing nucleotide substitutions (both "silent" and amino acid changing) in protein folding and disease. The most seminal finding in this area of study is that a silent SNP associated with the multi-drug resistance 1 gene (MDR1) resulted in altered function and conformation of its gene product, p-glycoprotein (Kimchi-Sarfaty et al, 2007). The authors hypothesized that the effect was due to a change in translation rate, although this has yet to be confirmed. We believe this effect may not be a singular event restricted only to the MDR1 gene but may be present in several SNP containing genes. Algorithms such as ours (to be made publicly available *via* the internet) can be utilized to generate speed profiles for numerous disease-

associated proteins, while mutations and single nucleotide polymorphisms (SNPs), available in the literature and in databases (http://www.ncbi.nlm.nih.gov/projects/SNP/), that exist in these proteins can then be mapped onto the their speed profiles to identify possible rate-changing mutations. This bioinformatic strategy could survey a large dataset of proteins and potentially identify numerous disease causing mutations that were previously overlooked because they did not alter amino acid identity. It could also be utilized to identify mutations that affect both rate and amino acid identity. To this end, strategies could be developed to dissect the rate contribution to folding from the contribution of the amino acid change. It is plausible that a mutation, which results in a minor amino acid identity change (one that maintains similar chemistry and size of the original amino acid) but drastically alters the translation speed in that position, may have a more pronounced effect on protein folding than the mutation which causes the same minor amino acid change but does not alter the speed. Also, just as a single amino acid change can cause a protein conformational change that is represented adequately enough in the ensemble to be observed by methods such as NMR (Alexander et al, 2009), one could reason that a silent, but rate changing, mutation could cause a conformational change (*e.g.*, a kinetically trapped intermediate) that could also be captured by these methods. Such a result would be revolutionary to the protein folding field and would provide a manner to experimentally examine the effects of speed altering mutations in the faulty folding of individual proteins associated with disease. It may also prompt the reexamination of so-called neutral substitutions throughout evolution. Ultimately, the

information gained by these studies could be utilized to identify new avenues for diagnosis and personalized treatment of these diseases.

Appendix

LUCIFERASE DNA AND PROTEIN SEQUENCES

Luc:	**Met**	**Glu**	**Asp**	**Ala**	**Lys**	**Asn**	**Ile**	**Lys**	**Lys**	**Gly**	**Pro**	**Ala**	**Pro**	**Phe**	**Tyr**	**Pro**	**Leu**	**Glu**	**Asp**	**Gly**
WT	ATG	GAA	GAC	GCC	AAA	AAC	ATA	AAG	AAA	GGC	CCG	GCG	CCA	TTC	TAT	CCT	CTA	GAG	GAT	GGA
WT_fast	ATG	GAA	GAC	GCC	AAA	AAC	ATA	AAG	AAA	GGC	CCG	GCG	CCA	TTC	TAT	CCT	CTG	GAA	GAC	GGC
WT_cbf	ATG	GAA	GAC	GCC	AAA	AAC	ATA	AAG	AAA	GGC	CCG	GCG	CCA	TTC	TAT	CCT	CTG	GAA	GAT	GGC
fast	ATG	GAA	GAC	GCA	AAA	AAC	ATC	AAA	AAA	GGC	CCG	GCA	CCG	TTC	TAC	CCG	CTG	GAA	GAC	GGC
cbf	ATG	GAA	GAT	GCG	AAA	AAC	ATT	AAA	AAA	GGC	CCG	GCG	CCG	TTT	TAT	CCG	CTG	GAA	GAT	GGC
re	ATG	GAA	GAC	GCT	AAA	AAC	ATC	AAA	AAA	GGC	CCG	GCC	CCA	TTC	TAT	CCA	CTC	GAA	GAT	GGC
slow	ATG	GAG	GAT	GCT	AAG	AAT	ATA	AAG	AAG	GGT	CCT	GCT	CCT	TTT	TAT	CCT	CTT	GAG	GAT	GGT

Luc:	**Thr**	**Ala**	**Gly**	**Glu**	**Gln**	**Leu**	**His**	**Lys**	**Ala**	**Met**	**Lys**	**Arg**	**Tyr**	**Ala**	**Leu**	**Val**	**Pro**	**Gly**	**Thr**	**Ile**
WT	ACC	GCT	GGA	GAG	CAA	CTG	CAT	AAG	GCT	ATG	AAG	AGA	TAC	GCC	CTG	GTT	CCT	GGA	ACA	ATT
WT_fast	ACC	GCA	GGC	GAA	CAG	CTG	CAC	AAA	GCA	ATG	AAA	CGT	TAC	GCA	CTG	GTA	CCG	GGC	ACC	ATC
WT_cbf	ACC	GCG	GGC	GAA	CAG	CTG	CAT	AAA	GCG	ATG	AAA	CGC	TAT	GCG	CTG	GTG	CCG	GGC	ACC	ATT
Fast	ACC	GCA	GGC	GAA	CAG	CTG	CAC	AAA	GCA	ATG	AAA	CGT	TAC	GCA	CTG	GTA	CCG	GGC	ACC	ATC
Cbf	ACC	GCG	GGC	GAA	CAG	CTG	CAT	AAA	GCG	ATG	AAA	CGC	TAT	GCG	CTG	GTG	CCG	GGC	ACC	ATT
Re	ACC	GCC	GGC	GAA	CAA	CTG	CAT	AAA	GCC	ATG	AAA	CGT	TAC	GCT	CTG	GTC	CCA	GGC	ACC	ATC
Slow	ACT	GCT	GGT	GAG	CAA	CTT	CAT	AAG	GCT	ATG	AAG	CGA	TAT	GCT	CTT	GTT	CCT	GGT	ACT	ATA

Luc:	**Ala**	**Phe**	**Thr**	**Asp**	**Ala**	**His**	**Ile**	**Glu**	**Val**	**Asn**	**Ile**	**Thr**	**Tyr**	**Ala**	**Glu**	**Tyr**	**Phe**	**Glu**	**Met**	**Ser**
WT	GCT	TTT	ACA	GAT	GCA	CAT	ATC	GAG	GTG	AAC	ATC	ACG	TAC	GCG	GAA	TAC	TTC	GAA	ATG	TCC
WT_fast	GCA	TTC	ACC	GAC	GCA	CAC	ATC	GAA	GTA	AAC	ATC	ACC	TAC	GCA	GAA	TAC	TTC	GAA	ATG	TCC
WT_cbf	GCG	TTT	ACC	GAT	GCG	CAT	ATT	GAA	GTG	AAC	ATT	ACC	TAT	GCG	GAA	TAT	TTT	GAA	ATG	AGC
Fast	GCA	TTC	ACC	GAC	GCA	CAC	ATC	GAA	GTA	AAC	ATC	ACC	TAC	GCA	GAA	TAC	TTC	GAA	ATG	TCC
Cbf	GCG	TTT	ACC	GAT	GCG	CAT	ATT	GAA	GTG	AAC	ATT	ACC	TAT	GCG	GAA	TAT	TTT	GAA	ATG	AGC
Re	GCC	TTT	ACC	GAT	GCC	CAT	ATT	GAA	GTG	AAC	ATT	ACT	TAC	GCC	GAA	TAC	TTC	GAA	ATG	TCT
Slow	GCT	TTT	ACT	GAT	GCT	CAT	ATA	GAG	GTT	AAT	ATA	ACT	TAT	GCT	GAG	TAT	TTT	GAG	ATG	TCT

Luc:	**Val**	**Arg**	**Leu**	**Ala**	**Glu**	**Ala**	**Met**	**Lys**	**Arg**	**Tyr**	**Gly**	**Leu**	**Asn**	**Thr**	**Asn**	**His**	**Arg**	**Ile**	**Val**	**Val**
WT	GTT	CGG	TTG	GCA	GAA	GCT	ATG	AAA	CGA	TAT	GGG	CTG	AAT	ACA	AAT	CAC	AGA	ATC	GTC	GTA
WT_fast	GTA	CGT	CTG	GCA	GAA	GCA	ATG	AAA	CGT	TAC	GGC	CTG	AAC	ACC	AAC	CAC	CGT	ATC	GTA	GTA
WT_cbf	GTG	CGC	CTG	GCG	GAA	GCG	ATG	AAA	CGC	TAT	GGC	CTG	AAC	ACC	AAC	CAT	CGC	ATT	GTG	GTG
Fast	GTA	CGT	CTG	GCA	GAA	GCA	ATG	AAA	CGT	TAC	GGC	CTG	AAC	ACC	AAC	CAC	CGT	ATC	GTA	GTA
Cbf	GTG	CGC	CTG	GCG	GAA	GCG	ATG	AAA	CGC	TAT	GGC	CTG	AAC	ACC	AAC	CAT	CGC	ATT	GTG	GTG
Re	GTC	CGC	CTG	GCC	GAA	GCC	ATG	AAA	CGT	TAT	GGT	CTG	AAT	ACC	AAT	CAC	CGT	ATT	GTT	GTC
Slow	GTT	CGA	CTT	GCT	GAG	GCT	ATG	AAG	CGA	TAT	GGT	CTT	AAT	ACT	AAT	CAT	CGA	ATA	GTT	GTT

Luc:	**Cys**	**Ser**	**Glu**	**Asn**	**Ser**	**Leu**	**Gln**	**Phe**	**Phe**	**Met**	**Pro**	**Val**	**Leu**	**Gly**	**Ala**	**Leu**	**Phe**	**Ile**	**Gly**	**Val**
WT	TGC	AGT	GAA	AAC	TCT	CTT	CAA	TTC	TTT	ATG	CCG	GTG	TTG	GGC	GCG	TTA	TTT	ATC	GGA	GTT
WT_fast	TGC	TCC	GAA	AAC	TCC	CTG	CAG	TTC	TTC	ATG	CCG	GTA	CTG	GGC	GCA	CTG	TTC	ATC	GGC	GTA
WT_cbf	TGC	AGC	GAA	AAC	AGC	CTG	CAG	TTT	TTT	ATG	CCG	GTG	CTG	GGC	GCG	CTG	TTT	ATT	GGC	GTG
Fast	TGC	TCC	GAA	AAC	TCC	CTG	CAG	TTC	TTC	ATG	CCG	GTA	CTG	GGC	GCA	CTG	TTC	ATC	GGC	GTA
Cbf	TGC	AGC	GAA	AAC	AGC	CTG	CAG	TTT	TTT	ATG	CCG	GTG	CTG	GGC	GCG	CTG	TTT	ATT	GGC	GTG
Re	TGC	TCT	GAA	AAC	TCC	CTC	CAA	TTC	TTT	ATG	CCG	GTG	CTG	GGC	GCC	CTC	TTT	ATT	GGC	GTC
Slow	TGT	TCT	GAG	AAT	TCT	CTT	CAA	TTT	TTT	ATG	CCT	GTT	CTT	GGT	GCT	CTT	TTT	ATA	GGT	GTT

Luc:	**Ala**	**Val**	**Ala**	**Pro**	**Ala**	**Asn**	**Asp**	**Ile**	**Tyr**	**Asn**	**Glu**	**Arg**	**Glu**	**Leu**	**Leu**	**Asn**	**Ser**	**Met**	**Asn**	**Ile**
WT	GCA	GTT	GCG	CCC	GCG	AAC	GAC	ATT	TAT	AAT	GAA	CGT	GAA	TTG	CTC	AAC	AGT	ATG	AAC	ATT
WT_fast	GCA	GTA	GCA	CCG	GCA	AAC	GAC	ATC	TAC	AAC	GAA	CGT	GAA	CTG	CTG	AAC	TCC	ATG	AAC	ATC
WT_cbf	GCG	GTG	GCG	CCG	GCG	AAC	GAT	ATT	TAT	AAC	GAA	CGC	GAA	CTG	CTG	AAC	AGC	ATG	AAC	ATT
Fast	GCA	GTA	GCA	CCG	GCA	AAC	GAC	ATC	TAC	AAC	GAA	CGT	GAA	CTG	CTG	AAC	TCC	ATG	AAC	ATC
Cbf	GCG	GTG	GCG	CCG	GCG	AAC	GAT	ATT	TAT	AAC	GAA	CGC	GAA	CTG	CTG	AAC	AGC	ATG	AAC	ATT
Re	GCC	GTC	GCC	CCT	GCC	AAC	GAC	ATC	TAT	AAT	GAA	CGT	GAA	CTG	CTT	AAC	TCT	ATG	AAC	ATC
Slow	GCT	GTT	GCT	CCT	GCT	AAT	GAT	ATA	TAT	AAT	GAG	CGA	GAG	CTT	CTT	AAT	TCT	ATG	AAT	ATA

Luc:	**Ser**	**Gln**	**Pro**	**Thr**	**Val**	**Val**	**Phe**	**Val**	**Ser**	**Lys**	**Lys**	**Gly**	**Leu**	**Gln**	**Lys**	**Ile**	**Leu**	**Asn**	**Val**	**Gln**
WT	TCG	CAG	CCT	ACC	GTA	GTG	TTT	GTT	TCC	AAA	AAG	GGG	TTG	CAA	AAA	ATT	TTG	AAC	GTG	CAA
WT_fast	TCC	CAG	CCG	ACC	GTA	GTA	TTC	GTA	TCC	AAA	AAA	GGC	CTG	CAG	AAA	ATC	CTG	AAC	GTA	CAG
WT_cbf	AGC	CAG	CCG	ACC	GTG	GTG	TTT	GTG	AGC	AAA	AAA	GGC	CTG	CAG	AAA	ATT	CTG	AAC	GTG	CAG
Fast	TCC	CAG	CCG	ACC	GTA	GTA	TTC	GTA	TCC	AAA	AAA	GGC	CTG	CAG	AAA	ATC	CTG	AAC	GTA	CAG
Cbf	AGC	CAG	CCG	ACC	GTG	GTG	TTT	GTG	AGC	AAA	AAA	GGC	CTG	CAG	AAA	ATT	CTG	AAC	GTG	CAG

Re	TCG	CAG	CCC	ACC	GTC	GTG	TTT	GTC	TCT	AAA	AAA	GGT	CTG	CAA	AAA	ATC	CTG	AAC	GTG	CAA
Slow	TCT	CAA	CCT	ACT	GTT	GTT	TTT	GTT	TCT	AAG	AAG	GGT	CTT	CAA	AAG	ATA	CTT	AAT	GTT	CAA

Luc:	**Lys**	**Lys**	**Leu**	**Pro**	**Ile**	**Ile**	**Gln**	**Lys**	**Ile**	**Ile**	**Ile**	**Met**	**Asp**	**Ser**	**Lys**	**Thr**	**Asp**	**Tyr**	**Gln**	**Gly**
WT	AAA	AAA	TTA	CCA	ATA	ATC	CAG	AAA	ATT	ATT	ATC	ATG	GAT	TCT	AAA	ACG	GAT	TAC	CAG	GGA
WT_fast	AAA	AAA	CTG	CCG	ATC	ATC	CAG	AAA	ATC	ATC	ATC	ATG	GAC	TCC	AAA	ACC	GAC	TAC	CAG	GGC
WT_cbf	AAA	AAA	CTG	CCG	ATT	ATT	CAG	AAA	ATT	ATT	ATT	ATG	GAT	AGC	AAA	ACC	GAT	TAT	CAG	GGC
Fast	AAA	AAA	CTG	CCG	ATC	ATC	CAG	AAA	ATC	ATC	ATC	ATG	GAC	TCC	AAA	ACC	GAC	TAC	CAG	GGC
Cbf	AAA	AAA	CTG	CCG	ATT	ATT	CAG	AAA	ATT	ATT	ATT	ATG	GAT	AGC	AAA	ACC	GAT	TAT	CAG	GGC
Re	AAA	AAA	CTC	CCC	ATC	ATT	CAG	AAA	ATC	ATC	ATT	ATG	GAT	TCC	AAA	ACT	GAT	TAC	CAG	GGC
Slow	AAG	AAG	CTT	CCT	ATA	ATA	CAA	AAG	ATA	ATA	ATA	ATG	GAT	TCT	AAG	ACT	GAT	TAT	CAA	GGT

Luc:	**Phe**	**Gln**	**Ser**	**Met**	**Tyr**	**Thr**	**Phe**	**Val**	**Thr**	**Ser**	**His**	**Leu**	**Pro**	**Pro**	**Gly**	**Phe**	**Asn**	**Glu**	**Tyr**	**Asp**
WT	TTT	CAG	TCG	ATG	TAC	ACG	TTC	GTC	ACA	TCT	CAT	CTA	CCT	CCC	GGT	TTT	AAT	GAA	TAC	GAT
WT_fast	TTC	CAG	TCC	ATG	TAC	ACC	TTC	GTA	ACC	TCC	CAC	CTG	CCG	CCG	GGC	TTC	AAC	GAA	TAC	GAC
WT_cbf	TTT	CAG	AGC	ATG	TAT	ACC	TTT	GTG	ACC	AGC	CAT	CTG	CCG	CCG	GGC	TTT	AAC	GAA	TAT	GAT
Fast	TTC	CAG	TCC	ATG	TAC	ACC	TTC	GTA	ACC	TCC	CAC	CTG	CCG	CCG	GGC	TTC	AAC	GAA	TAC	GAC
Cbf	TTT	CAG	AGC	ATG	TAT	ACC	TTT	GTG	ACC	AGC	CAT	CTG	CCG	CCG	GGC	TTT	AAC	GAA	TAT	GAT
Re	TTT	CAG	TCG	ATG	TAC	ACT	TTC	GTT	ACC	TCC	CAT	CTC	CCC	CCT	GGT	TTT	AAT	GAA	TAC	GAT
Slow	TTT	CAA	TCT	ATG	TAT	ACT	TTT	GTT	ACT	TCT	CAT	CTT	CCT	CCT	GGT	TTT	AAT	GAG	TAT	GAT

Luc:	**Phe**	**Val**	**Pro**	**Glu**	**Ser**	**Phe**	**Asp**	**Arg**	**Asp**	**Lys**	**Thr**	**Ile**	**Ala**	**Leu**	**Ile**	**Met**	**Asn**	**Ser**	**Ser**	**Gly**
WT	TTT	GTA	CCA	GAG	TCC	TTT	GAT	CGT	GAC	AAA	ACA	ATT	GCA	CTG	ATA	ATG	AAT	TCC	TCT	GGA
WT_fast	TTC	GTA	CCG	GAA	TCC	TTC	GAC	CGT	GAC	AAA	ACC	ATC	GCA	CTG	ATC	ATG	AAC	TCC	TCC	GGC
WT_cbf	TTT	GTG	CCG	GAA	AGC	TTT	GAT	CGC	GAT	AAA	ACC	ATT	GCG	CTG	ATT	ATG	AAC	AGC	AGC	GGC
Fast	TTC	GTA	CCG	GAA	TCC	TTC	GAC	CGT	GAC	AAA	ACC	ATC	GCA	CTG	ATC	ATG	AAC	TCC	TCC	GGC
Cbf	TTT	GTG	CCG	GAA	AGC	TTT	GAT	CGC	GAT	AAA	ACC	ATT	GCG	CTG	ATT	ATG	AAC	AGC	AGC	GGC
Re	TTT	GTC	CCC	GAA	TCT	TTT	GAT	CGT	GAC	AAA	ACC	ATC	GCC	CTG	ATC	ATG	AAT	TCT	TCC	GGC
Slow	TTT	GTT	CCT	GAG	TCT	TTT	GAT	CGA	GAT	AAG	ACT	ATA	GCT	CTT	ATA	ATG	AAT	TCT	TCT	GGT

Luc:	**Ser**	**Thr**	**Gly**	**Leu**	**Pro**	**Lys**	**Gly**	**Val**	**Ala**	**Leu**	**Pro**	**His**	**Arg**	**Thr**	**Ala**	**Cys**	**Val**	**Arg**	**Phe**	**Ser**
WT	TCT	ACT	GGG	TTA	CCT	AAG	GGT	GTG	GCC	CTT	CCG	CAT	AGA	ACT	GCC	TGC	GTC	AGA	TTC	TCG
WT_fast	TCC	ACC	GGC	CTG	CCG	AAA	GGC	GTA	GCA	CTG	CCG	CAC	CGT	ACC	GCA	TGC	GTA	CGT	TTC	TCC
WT_cbf	AGC	ACC	GGC	CTG	CCG	AAA	GGC	GTG	GCG	CTG	CCG	CAT	CGC	ACC	GCG	TGC	GTG	CGC	TTT	AGC
Fast	TCC	ACC	GGC	CTG	CCG	AAA	GGC	GTA	GCA	CTG	CCG	CAC	CGT	ACC	GCA	TGC	GTA	CGT	TTC	TCC
Cbf	AGC	ACC	GGC	CTG	CCG	AAA	GGC	GTG	GCG	CTG	CCG	CAT	CGC	ACC	GCG	TGC	GTG	CGC	TTT	AGC
Re	TCC	ACC	GGT	CTC	CCC	AAA	GGT	GTG	GCT	CTC	CCG	CAT	CGT	ACC	GCT	TGC	GTT	CGT	TTC	TCG
Slow	TCT	ACT	GGT	CTT	CCT	AAG	GGT	GTT	GCT	CTT	CCT	CAT	CGA	ACT	GCT	TGT	GTT	CGA	TTT	TCT

Luc:	**His**	**Ala**	**Arg**	**Asp**	**Pro**	**Ile**	**Phe**	**Gly**	**Asn**	**Gln**	**Ile**	**Ile**	**Pro**	**Asp**	**Thr**	**Ala**	**Ile**	**Leu**	**Ser**	**Val**
WT	CAT	GCC	AGA	GAT	CCT	ATT	TTT	GGC	AAT	CAA	ATC	ATT	CCG	GAT	ACT	GCG	ATT	TTA	AGT	GTT
WT_fast	CAC	GCA	CGT	GAC	CCG	ATC	TTC	GGC	AAC	CAG	ATC	ATC	CCG	GAC	ACC	GCA	ATC	CTG	TCC	GTA
WT_cbf	CAT	GCG	CGC	GAT	CCG	ATT	TTT	GGC	AAC	CAG	ATT	ATT	CCG	GAT	ACC	GCG	ATT	CTG	AGC	GTG
Fast	CAC	GCA	CGT	GAC	CCG	ATC	TTC	GGC	AAC	CAG	ATC	ATC	CCG	GAC	ACC	GCA	ATC	CTG	TCC	GTA
Cbf	CAT	GCG	CGC	GAT	CCG	ATT	TTT	GGC	AAC	CAG	ATT	ATT	CCG	GAT	ACC	GCG	ATT	CTG	AGC	GTG
Re	CAT	GCT	CGT	GAT	CCC	ATC	TTT	GGC	AAT	CAA	ATT	ATC	CCG	GAT	ACC	GCC	ATC	CTC	TCT	GTC
Slow	CAT	GCT	CGA	GAT	CCT	ATA	TTT	GGT	AAT	CAA	ATA	ATA	CCT	GAT	ACT	GCT	ATA	CTT	TCT	GTT

Luc:	**Val**	**Pro**	**Phe**	**His**	**His**	**Gly**	**Phe**	**Gly**	**Met**	**Phe**	**Thr**	**Thr**	**Leu**	**Gly**	**Tyr**	**Leu**	**Ile**	**Cys**	**Gly**	**Phe**
WT	GTT	CCA	TTC	CAT	CAC	GGT	TTT	GGA	ATG	TTT	ACT	ACA	CTC	GGA	TAT	TTG	ATA	TGT	GGA	TTT
WT_fast	GTA	CCG	TTC	CAC	CAC	GGC	TTC	GGC	ATG	TTC	ACC	ACC	CTG	GGC	TAC	CTG	ATC	TGC	GGC	TTC
WT_cbf	GTG	CCG	TTT	CAT	CAT	GGC	TTT	GGC	ATG	TTT	ACC	ACC	CTG	GGC	TAT	CTG	ATT	TGC	GGC	TTT
Fast	GTA	CCG	TTC	CAC	CAC	GGC	TTC	GGC	ATG	TTC	ACC	ACC	CTG	GGC	TAC	CTG	ATC	TGC	GGC	TTC
Cbf	GTG	CCG	TTT	CAT	CAT	GGC	TTT	GGC	ATG	TTT	ACC	ACC	CTG	GGC	TAT	CTG	ATT	TGC	GGC	TTT
Re	GTC	CCC	TTC	CAT	CAC	GGT	TTT	GGC	ATG	TTT	ACC	ACC	CTT	GGC	TAT	CTG	ATC	TGT	GGC	TTT
Slow	GTT	CCT	TTT	CAT	CAT	GGT	TTT	GGT	ATG	TTT	ACT	ACT	CTT	GGT	TAT	CTT	ATA	TGT	GGT	TTT

Luc:	**Arg**	**Val**	**Val**	**Leu**	**Met**	**Tyr**	**Arg**	**Phe**	**Glu**	**Glu**	**Glu**	**Leu**	**Phe**	**Leu**	**Arg**	**Ser**	**Leu**	**Gln**	**Asp**	**Tyr**
WT	CGA	GTC	GTC	TTA	ATG	TAT	AGA	TTT	GAA	GAA	GAG	CTG	TTT	TTA	CGA	TCC	CTT	CAG	GAT	TAC
WT_fast	CGT	GTA	GTA	CTG	ATG	TAC	CGT	TTC	GAA	GAA	GAA	CTG	TTC	CTG	CGT	TCC	CTG	CAG	GAC	TAC
WT_cbf	CGC	GTG	GTG	CTG	ATG	TAT	CGC	TTT	GAA	GAA	GAA	CTG	TTT	CTG	CGC	AGC	CTG	CAG	GAT	TAT
Fast	CGT	GTA	GTA	CTG	ATG	TAC	CGT	TTC	GAA	GAA	GAA	CTG	TTC	CTG	CGT	TCC	CTG	CAG	GAC	TAC
Cbf	CGC	GTG	GTG	CTG	ATG	TAT	CGC	TTT	GAA	GAA	GAA	CTG	TTT	CTG	CGC	AGC	CTG	CAG	GAT	TAT
Re	CGT	GTT	GTT	CTC	ATG	TAT	CGT	TTT	GAA	GAA	GAA	CTG	TTT	CTC	CGT	TCT	CTC	CAG	GAT	TAC
Slow	CGA	GTT	GTT	CTT	ATG	TAT	CGA	TTT	GAG	GAG	GAG	CTT	TTT	CTT	CGA	TCT	CTT	CAA	GAT	TAT

Luc:	**Lys**	**Ile**	**Gln**	**Ser**	**Ala**	**Leu**	**Leu**	**Val**	**Pro**	**Thr**	**Leu**	**Phe**	**Ser**	**Phe**	**Phe**	**Ala**	**Lys**	**Ser**	**Thr**	**Leu**
WT	AAA	ATT	CAA	AGT	GCG	TTG	CTA	GTA	CCA	ACC	CTA	TTT	TCA	TTC	TTC	GCC	AAA	AGC	ACT	CTG
WT_fast	AAA	ATC	CAG	TCC	GCA	CTG	CTG	GTA	CCG	ACC	CTG	TTC	TCC	TTC	TTC	GCA	AAA	TCC	ACC	CTG
WT_cbf	AAA	ATT	CAG	AGC	GCG	CTG	CTG	GTG	CCG	ACC	CTG	TTT	AGC	TTT	TTT	GCG	AAA	AGC	ACC	CTG
Fast	AAA	ATC	CAG	TCC	GCA	CTG	CTG	GTA	CCG	ACC	CTG	TTC	TCC	TTC	TTC	GCA	AAA	TCC	ACC	CTG
Cbf	AAA	ATT	CAG	AGC	GCG	CTG	CTG	GTG	CCG	ACC	CTG	TTT	AGC	TTT	TTT	GCG	AAA	AGC	ACC	CTG
Re	AAA	ATC	CAA	TCT	GCC	CTG	CTC	GTC	CCC	ACC	CTC	TTT	TCC	TTC	TTC	GCT	AAA	TCC	ACC	CTG
Slow	AAG	ATA	CAA	TCT	GCT	CTT	CTT	GTT	CCT	ACT	CTT	TTT	TCT	TTT	TTT	GCT	AAG	TCT	ACT	CTT

Luc:	**Ile**	**Asp**	**Lys**	**Tyr**	**Asp**	**Leu**	**Ser**	**Asn**	**Leu**	**His**	**Glu**	**Ile**	**Ala**	**Ser**	**Gly**	**Gly**	**Ala**	**Pro**	**Leu**	**Ser**
WT	ATT	GAC	AAA	TAC	GAT	TTA	TCT	AAT	TTA	CAC	GAA	ATT	GCT	TCT	GGG	GGC	GCA	CCT	CTT	TCG
WT_fast	ATC	GAC	AAA	TAC	GAC	CTG	TCC	AAC	CTG	CAC	GAA	ATC	GCA	TCC	GGC	GGC	GCA	CCG	CTG	TCC
WT_cbf	ATT	GAT	AAA	TAT	GAT	CTG	AGC	AAC	CTG	CAT	GAA	ATT	GCG	AGC	GGC	GGC	GCG	CCG	CTG	AGC
Fast	ATC	GAC	AAA	TAC	GAC	CTG	TCC	AAC	CTG	CAC	GAA	ATC	GCA	TCC	GGC	GGC	GCA	CCG	CTG	TCC
Cbf	ATT	GAT	AAA	TAT	GAT	CTG	AGC	AAC	CTG	CAT	GAA	ATT	GCG	AGC	GGC	GGC	GCG	CCG	CTG	AGC
Re	ATC	GAC	AAA	TAC	GAT	CTC	TCC	AAT	CTC	CAC	GAA	ATC	GCC	TCC	GGT	GGC	GCC	CCC	CTC	TCG
Slow	ATA	GAT	AAG	TAT	GAT	CTT	TCT	AAT	CTT	CAT	GAG	ATA	GCT	TCT	GGT	GGT	GCT	CCT	CTT	TCT

Luc:	**Lys**	**Glu**	**Val**	**Gly**	**Glu**	**Ala**	**Val**	**Ala**	**Lys**	**Arg**	**Phe**	**His**	**Leu**	**Pro**	**Gly**	**Ile**	**Arg**	**Gln**	**Gly**	**Tyr**
WT	AAA	GAA	GTC	GGG	GAA	GCG	GTT	GCA	AAA	CGC	TTC	CAT	CTT	CCA	GGG	ATA	CGA	CAA	GGA	TAT
WT_fast	AAA	GAA	GTA	GGC	GAA	GCA	GTA	GCA	AAA	CGT	TTC	CAC	CTG	CCG	GGC	ATC	CGT	CAG	GGC	TAC
WT_cbf	AAA	GAA	GTG	GGC	GAA	GCG	GTG	GCG	AAA	CGC	TTT	CAT	CTG	CCG	GGC	ATT	CGC	CAG	GGC	TAT
Fast	AAA	GAA	GTA	GGC	GAA	GCA	GTA	GCA	AAA	CGT	TTC	CAC	CTG	CCG	GGC	ATC	CGT	CAG	GGC	TAC
Cbf	AAA	GAA	GTG	GGC	GAA	GCG	GTG	GCG	AAA	CGC	TTT	CAT	CTG	CCG	GGC	ATT	CGC	CAG	GGC	TAT
Re	AAA	GAA	GTT	GGT	GAA	GCC	GTC	GCC	AAA	CGC	TTC	CAT	CTC	CCC	GGT	ATC	CGT	CAA	GGC	TAT
Slow	AAG	GAG	GTT	GGT	GAG	GCT	GTT	GCT	AAG	CGA	TTT	CAT	CTT	CCT	GGT	ATA	CGA	CAA	GGT	TAT

Luc:	**Gly**	**Leu**	**Thr**	**Glu**	**Thr**	**Thr**	**Ser**	**Ala**	**Ile**	**Leu**	**Ile**	**Thr**	**Pro**	**Glu**	**Gly**	**Asp**	**Asp**	**Lys**	**Pro**	**Gly**
WT	GGG	CTC	ACT	GAG	ACT	ACA	TCA	GCT	ATT	CTG	ATT	ACA	CCC	GAG	GGG	GAT	GAT	AAA	CCG	GGC
WT_fast	GGC	CTG	ACC	GAA	ACC	ACC	TCC	GCA	ATC	CTG	ATC	ACC	CCG	GAA	GGC	GAC	GAC	AAA	CCG	GGC
WT_cbf	GGC	CTG	ACC	GAA	ACC	ACC	AGC	GCG	ATT	CTG	ATT	ACC	CCG	GAA	GGC	GAT	GAT	AAA	CCG	GGC
Fast	GGC	CTG	ACC	GAA	ACC	ACC	TCC	GCA	ATC	CTG	ATC	ACC	CCG	GAA	GGC	GAC	GAC	AAA	CCG	GGC
Cbf	GGC	CTG	ACC	GAA	ACC	ACC	AGC	GCG	ATT	CTG	ATT	ACC	CCG	GAA	GGC	GAT	GAT	AAA	CCG	GGC
Re	GGT	CTT	ACC	GAA	ACC	ACC	TCC	GCC	ATC	CTG	ATC	ACC	CCT	GAA	GGT	GAT	GAT	AAA	CCG	GGC
Slow	GGT	CTT	ACT	GAG	ACT	ACT	TCT	GCT	ATA	CTT	ATA	ACT	CCT	GAG	GGT	GAT	GAT	AAG	CCT	GGT

Luc:	**Ala**	**Val**	**Gly**	**Lys**	**Val**	**Val**	**Pro**	**Phe**	**Phe**	**Glu**	**Ala**	**Lys**	**Val**	**Val**	**Asp**	**Leu**	**Asp**	**Thr**	**Gly**	**Lys**
WT	GCG	GTC	GGT	AAA	GTT	GTT	CCA	TTT	TTT	GAA	GCG	AAG	GTT	GTG	GAT	CTG	GAT	ACC	GGG	AAA
WT_fast	GCA	GTA	GGC	AAA	GTA	GTA	CCG	TTC	TTC	GAA	GCA	AAA	GTA	GTA	GAC	CTG	GAC	ACC	GGC	AAA
WT_cbf	GCG	GTG	GGC	AAA	GTG	GTG	CCG	TTT	TTT	GAA	GCG	AAA	GTG	GTG	GAT	CTG	GAT	ACC	GGC	AAA
Fast	GCA	GTA	GGC	AAA	GTA	GTA	CCG	TTC	TTC	GAA	GCA	AAA	GTA	GTA	GAC	CTG	GAC	ACC	GGC	AAA
Cbf	GCG	GTG	GGC	AAA	GTG	GTG	CCG	TTT	TTT	GAA	GCG	AAA	GTG	GTG	GAT	CTG	GAT	ACC	GGC	AAA
Re	GCC	GTT	GGT	AAA	GTC	GTC	CCC	TTT	TTT	GAA	GCC	AAA	GTC	GTG	GAT	CTG	GAT	ACC	GGT	AAA
Slow	GCT	GTT	GGT	AAG	GTT	GTT	CCT	TTT	TTT	GAG	GCT	AAG	GTT	GTT	GAT	CTT	GAT	ACT	GGT	AAG

Luc:	**Thr**	**Leu**	**Gly**	**Val**	**Asn**	**Gln**	**Arg**	**Gly**	**Glu**	**Leu**	**Cys**	**Val**	**Arg**	**Gly**	**Pro**	**Met**	**Ile**	**Met**	**Ser**	**Gly**
WT	ACG	CTG	GGC	GTT	AAT	CAG	AGA	GGC	GAA	TTA	TGT	GTC	AGA	GGA	CCT	ATG	ATT	ATG	TCC	GGT
WT_fast	ACC	CTG	GGC	GTA	AAC	CAG	CGT	GGC	GAA	CTG	TGC	GTA	CGT	GGC	CCG	ATG	ATC	ATG	TCC	GGC
WT_cbf	ACC	CTG	GGC	GTG	AAC	CAG	CGC	GGC	GAA	CTG	TGC	GTG	CGC	GGC	CCG	ATG	ATT	ATG	AGC	GGC
Fast	ACC	CTG	GGC	GTA	AAC	CAG	CGT	GGC	GAA	CTG	TGC	GTA	CGT	GGC	CCG	ATG	ATC	ATG	TCC	GGC
Cbf	ACC	CTG	GGC	GTG	AAC	CAG	CGC	GGC	GAA	CTG	TGC	GTG	CGC	GGC	CCG	ATG	ATT	ATG	AGC	GGC
Re	ACT	CTG	GGC	GTC	AAT	CAG	CGT	GGC	GAA	CTC	TGT	GTT	CGT	GGC	CCC	ATG	ATC	ATG	TCT	GGT
Slow	ACT	CTT	GGT	GTT	AAT	CAA	CGA	GGT	GAG	CTT	TGT	GTT	CGA	GGT	CCT	ATG	ATA	ATG	TCT	GGT

Luc:	**Tyr**	**Val**	**Asn**	**Asn**	**Pro**	**Glu**	**Ala**	**Thr**	**Asn**	**Ala**	**Leu**	**Ile**	**Asp**	**Lys**	**Asp**	**Gly**	**Trp**	**Leu**	**His**	**Ser**
WT	TAT	GTA	AAC	AAT	CCG	GAA	GCG	ACC	AAC	GCC	TTG	ATT	GAC	AAG	GAT	GGA	TGG	CTA	CAT	TCT
WT_fast	TAC	GTA	AAC	AAC	CCG	GAA	GCA	ACC	AAC	GCA	CTG	ATC	GAC	AAA	GAC	GGC	TGG	CTG	CAC	TCC
WT_cbf	TAT	GTG	AAC	AAC	CCG	GAA	GCG	ACC	AAC	GCG	CTG	ATT	GAT	AAA	GAT	GGC	TGG	CTG	CAT	AGC
Fast	TAC	GTA	AAC	AAC	CCG	GAA	GCA	ACC	AAC	GCA	CTG	ATC	GAC	AAA	GAC	GGC	TGG	CTG	CAC	TCC
Cbf	TAT	GTG	AAC	AAC	CCG	GAA	GCG	ACC	AAC	GCG	CTG	ATT	GAT	AAA	GAT	GGC	TGG	CTG	CAT	AGC
Re	TAT	GTC	AAC	AAT	CCG	GAA	GCC	ACC	AAC	GCT	CTG	ATC	GAC	AAA	GAT	GGC	TGG	CTC	CAT	TCC
Slow	TAT	GTT	AAT	AAT	CCT	GAG	GCT	ACT	AAT	GCT	CTT	ATA	GAT	AAG	GAT	GGT	TGG	CTT	CAT	TCT

Luc:	**Gly**	**Asp**	**Ile**	**Ala**	**Tyr**	**Trp**	**Asp**	**Glu**	**Asp**	**Glu**	**His**	**Phe**	**Phe**	**Ile**	**Val**	**Asp**	**Arg**	**Leu**	**Lys**	**Ser**
WT	GGA	GAC	ATA	GCT	TAC	TGG	GAC	GAA	GAC	GAA	CAC	TTC	TTC	ATA	GTT	GAC	CGC	TTG	AAG	TCT
WT_fast	GGC	GAC	ATC	GCA	TAC	TGG	GAC	GAA	GAC	GAA	CAC	TTC	TTC	ATC	GTA	GAC	CGT	CTG	AAA	TCC
Fast	GGC	GAC	ATC	GCA	TAC	TGG	GAC	GAA	GAC	GAA	CAC	TTC	TTC	ATC	GTA	GAC	CGT	CTG	AAA	TCC
WT_cbf	GGC	GAT	ATT	GCG	TAT	TGG	GAT	GAA	GAT	GAA	CAT	TTT	TTT	ATT	GTG	GAT	CGC	CTG	AAA	AGC
Cbf	GGC	GAT	ATT	GCG	TAT	TGG	GAT	GAA	GAT	GAA	CAT	TTT	TTT	ATT	GTG	GAT	CGC	CTG	AAA	AGC
Re	GGC	GAC	ATC	GCC	TAC	TGG	GAC	GAA	GAC	GAA	CAC	TTC	TTC	ATC	GTC	GAC	CGC	CTG	AAA	TCC
Slow	GGT	GAT	ATA	GCT	TAT	TGG	GAT	GAG	GAT	GAG	CAT	TTT	TTT	ATA	GTT	GAT	CGA	CTT	AAG	TCT

Luc:	**Leu**	**Ile**	**Lys**	**Tyr**	**Lys**	**Gly**	**Tyr**	**Gln**	**Val**	**Ala**	**Pro**	**Ala**	**Glu**	**Leu**	**Glu**	**Ser**	**Ile**	**Leu**	**Leu**	**Gln**
WT	TTA	ATT	AAA	TAC	AAA	GGA	TAT	CAG	GTG	GCC	CCC	GCT	GAA	TTG	GAA	TCG	ATA	TTG	TTA	CAA
WT_fast	CTG	ATC	AAA	TAC	AAA	GGC	TAC	CAG	GTA	GCA	CCG	GCA	GAA	CTG	GAA	TCC	ATC	CTG	CTG	CAG
WT_cbf	CTG	ATT	AAA	TAT	AAA	GGC	TAT	CAG	GTG	GCG	CCG	GCG	GAA	CTG	GAA	AGC	ATT	CTG	CTG	CAG
Fast	CTG	ATC	AAA	TAC	AAA	GGC	TAC	CAG	GTA	GCA	CCG	GCA	GAA	CTG	GAA	TCC	ATC	CTG	CTG	CAG
Cbf	CTG	ATT	AAA	TAT	AAA	GGC	TAT	CAG	GTG	GCG	CCG	GCG	GAA	CTG	GAA	AGC	ATT	CTG	CTG	CAG
Re	CTC	ATC	AAA	TAC	AAA	GGC	TAT	CAG	GTG	GCT	CCT	GCC	GAA	CTG	GAA	TCG	ATC	CTG	CTC	CAA
Slow	CTT	ATA	AAG	TAT	AAG	GGT	TAT	CAA	GTT	GCT	CCT	GCT	GAG	CTT	GAG	TCT	ATA	CTT	CTT	CAA

Luc:	**His**	**Pro**	**Asn**	**Ile**	**Phe**	**Asp**	**Ala**	**Gly**	**Val**	**Ala**	**Gly**	**Leu**	**Pro**	**Asp**	**Asp**	**Asp**	**Ala**	**Gly**	**Glu**	**Leu**
WT	CAC	CCC	AAC	ATC	TTC	GAC	GCG	GGC	GTG	GCA	GGT	CTT	CCC	GAC	GAT	GAC	GCC	GGT	GAA	CTT
WT_fast	CAC	CCG	AAC	ATC	TTC	GAC	GCA	GGC	GTA	GCA	GGC	CTG	CCG	GAC	GAC	GAC	GCA	GGC	GAA	CTG
WT_cbf	CAT	CCG	AAC	ATT	TTT	GAT	GCG	GGC	GTG	GCG	GGC	CTG	CCG	GAT	GAT	GAT	GCG	GGC	GAA	CTG
Fast	CAC	CCG	AAC	ATC	TTC	GAC	GCA	GGC	GTA	GCA	GGC	CTG	CCG	GAC	GAC	GAC	GCA	GGC	GAA	CTG
Cbf	CAT	CCG	AAC	ATT	TTT	GAT	GCG	GGC	GTG	GCG	GGC	CTG	CCG	GAT	GAT	GAT	GCG	GGC	GAA	CTG
Re	CAC	CCT	AAC	ATT	TTC	GAC	GCC	GGC	GTG	GCC	GGT	CTC	CCT	GAC	GAT	GAC	GCT	GGT	GAA	CTC
Slow	CAT	CCT	AAT	ATA	TTT	GAT	GCT	GGT	GTT	GCT	GGT	CTT	CCT	GAT	GAT	GAT	GCT	GGT	GAG	CTT

Luc:	**Pro**	**Ala**	**Ala**	**Val**	**Val**	**Val**	**Leu**	**Glu**	**His**	**Gly**	**Lys**	**Thr**	**Met**	**Thr**	**Glu**	**Lys**	**Glu**	**Ile**	**Val**	**Asp**
WT	CCC	GCC	GCC	GTT	GTT	GTT	TTG	GAG	CAC	GGA	AAG	ACG	ATG	ACG	GAA	AAA	GAG	ATC	GTG	GAT
WT_fast	CCG	GCA	GCA	GTA	GTA	GTA	CTG	GAA	CAC	GGC	AAA	ACC	ATG	ACC	GAA	AAA	GAA	ATC	GTA	GAC
WT_cbf	CCG	GCG	GCG	GTG	GTG	GTG	CTG	GAA	CAT	GGC	AAA	ACC	ATG	ACC	GAA	AAA	GAA	ATT	GTG	GAT
Fast	CCG	GCA	GCA	GTA	GTA	GTA	CTG	GAA	CAC	GGC	AAA	ACC	ATG	ACC	GAA	AAA	GAA	ATC	GTA	GAC
Cbf	CCG	GCG	GCG	GTG	GTG	GTG	CTG	GAA	CAT	GGC	AAA	ACC	ATG	ACC	GAA	AAA	GAA	ATT	GTG	GAT
Re	CCT	GCT	GCT	GTC	GTC	GTC	CTG	GAA	CAC	GGC	AAA	ACT	ATG	ACT	GAA	AAA	GAA	ATT	GTG	GAT
Slow	CCT	GCT	GCT	GTT	GTT	GTT	CTT	GAG	CAT	GGT	AAG	ACT	ATG	ACT	GAG	AAG	GAG	ATA	GTT	GAT

Luc:	**Tyr**	**Val**	**Ala**	**Ser**	**Gln**	**Val**	**Thr**	**Thr**	**Ala**	**Lys**	**Lys**	**Leu**	**Arg**	**Gly**	**Gly**	**Val**	**Val**	**Phe**	**Val**	**Asp**
WT	TAC	GTC	GCC	AGT	CAA	GTA	ACA	ACC	GCG	AAA	AAG	TTG	CGC	GGA	GGA	GTT	GTG	TTT	GTG	GAC
WT_fast	TAC	GTA	GCA	TCC	CAG	GTA	ACC	ACC	GCA	AAA	AAA	CTG	CGT	GGC	GGC	GTA	GTA	TTC	GTA	GAC
WT_cbf	TAT	GTG	GCG	AGC	CAG	GTG	ACC	ACC	GCG	AAA	AAA	CTG	CGC	GGC	GGC	GTG	GTG	TTT	GTG	GAT
Fast	TAC	GTA	GCA	TCC	CAG	GTA	ACC	ACC	GCA	AAA	AAA	CTG	CGT	GGC	GGC	GTA	GTA	TTC	GTA	GAC
Cbf	TAT	GTG	GCG	AGC	CAG	GTG	ACC	ACC	GCG	AAA	AAA	CTG	CGC	GGC	GGC	GTG	GTG	TTT	GTG	GAT
Re	TAC	GTT	GCT	TCT	CAA	GTC	ACC	ACC	GCC	AAA	AAA	CTG	CGC	GGC	GGC	GTC	GTG	TTT	GTG	GAC
Slow	TAT	GTT	GCT	TCT	CAA	GTT	ACT	ACT	GCT	AAG	AAG	CTT	CGA	GGT	GGT	GTT	GTT	TTT	GTT	GAT

Luc:	**Glu**	**Val**	**Pro**	**Lys**	**Gly**	**Leu**	**Thr**	**Gly**	**Lys**	**Leu**	**Asp**	**Ala**	**Arg**	**Lys**	**Ile**	**Arg**	**Glu**	**Ile**	**Leu**	**Ile**
WT	GAA	GTA	CCG	AAA	GGT	CTT	ACC	GGA	AAA	CTC	GAC	GCA	AGA	AAA	ATC	AGA	GAG	ATC	CTC	ATA
WT_fast	GAA	GTA	CCG	AAA	GGC	CTG	ACC	GGC	AAA	CTG	GAC	GCA	CGT	AAA	ATC	CGT	GAA	ATC	CTG	ATC
WT_cbf	GAA	GTG	CCG	AAA	GGC	CTG	ACC	GGC	AAA	CTG	GAT	GCG	CGC	AAA	ATT	CGC	GAA	ATT	CTG	ATT
Fast	GAA	GTA	CCG	AAA	GGC	CTG	ACC	GGC	AAA	CTG	GAC	GCA	CGT	AAA	ATC	CGT	GAA	ATC	CTG	ATC
Cbf	GAA	GTG	CCG	AAA	GGC	CTG	ACC	GGC	AAA	CTG	GAT	GCG	CGC	AAA	ATT	CGC	GAA	ATT	CTG	ATT
Re	GAA	GTC	CCG	AAA	GGT	CTC	ACC	GGC	AAA	CTT	GAC	GCC	CGT	AAA	ATT	CGT	GAA	ATT	CTT	ATC
Slow	GAG	GTT	CCT	AAG	GGT	CTT	ACT	GGT	AAG	CTT	GAT	GCT	CGA	AAG	ATA	CGA	GAG	ATA	CTT	ATA

Luc:	**Lys**	**Ala**	**Lys**	**Lys**	**Gly**	**Gly**	**Lys**	**Ser**	**Lys**	**Leu**	**Ile**	**Glu**	**Gly**	**Arg**	**Gly**	**Ser**	**Gly**	**Thr**	**Ser**	**Gly**
WT	AAG	GCC	AAG	AAG	GGC	GGA	AAG	TCC	AAA	TTG	ATC	GAA	GGC	CGC	GGA	TCT	GGT	ACT	AGT	GGC
WT_fast	AAA	GCA	AAA	AAA	GGC	GGC	AAA	TCC	AAA	CTG	ATC	GAA	GGC	CGC	GGA	TCT	GGT	ACT	AGT	GGC
WT_cbf	AAA	GCG	AAA	AAA	GGC	GGC	AAA	AGC	AAA	CTG	ATC	GAA	GGC	CGC	GGA	TCT	GGT	ACT	AGT	GGC
Fast	AAA	GCA	AAA	AAA	GGC	GGC	AAA	TCC	AAA	CTG	ATC	GAA	GGC	CGC	GGA	TCT	GGT	ACT	AGT	GGC
Cbf	AAA	GCG	AAA	AAA	GGC	GGC	AAA	AGC	AAA	CTG	ATC	GAA	GGC	CGC	GGA	TCT	GGT	ACT	AGT	GGC
Re	AAA	GCT	AAA	AAA	GGC	GGC	AAA	TCT	AAA	CTG	ATC	GAA	GGC	CGC	GGA	TCT	GGT	ACT	AGT	GGC
Slow	AAG	GCT	AAG	AAG	GGT	GGT	AAG	TCT	AAG	CTT	ATC	GAA	GGC	CGC	GGA	TCT	GGT	ACT	AGT	GGC

Luc:	**Gly**	**Ser**	**Gly**	**Gly**	**Ser**	**Gly**	**Arg**	**Ser**	**Glu**	**Gln**	**Lys**	**Leu**	**Ile**	**Ser**	**Glu**	**Glu**	**Asp**	**Leu**	**His**	**His**
WT	GGG	TCA	GGT	GGC	TCG	GGG	CGA	TCC	GAA	CAA	AAA	CTT	ATT	TCT	GAA	GAA	GAC	TTG	CAC	CAT
WT_fast	GGG	TCA	GGT	GGC	TCG	GGG	CGA	TCC	GAA	CAA	AAA	CTT	ATT	TCT	GAA	GAA	GAC	TTG	CAC	CAT
WT_cbf	GGG	TCA	GGT	GGC	TCG	GGG	CGA	TCC	GAA	CAA	AAA	CTT	ATT	TCT	GAA	GAA	GAC	TTG	CAC	CAT
Fast	GGG	TCA	GGT	GGC	TCG	GGG	CGA	TCC	GAA	CAA	AAA	CTT	ATT	TCT	GAA	GAA	GAC	TTG	CAC	CAT
Cbf	GGG	TCA	GGT	GGC	TCG	GGG	CGA	TCC	GAA	CAA	AAA	CTT	ATT	TCT	GAA	GAA	GAC	TTG	CAC	CAT
Re	GGG	TCA	GGT	GGC	TCG	GGG	CGA	TCC	GAA	CAA	AAA	CTT	ATT	TCT	GAA	GAA	GAC	TTG	CAC	CAT

Luc:	**His**	**His**	**His**	**His**	**---**
LucWT	CAC	CAT	CAC	CAT	TAA
WT_fast	CAC	CAT	CAC	CAT	TAA
WT_cbf	CAC	CAT	CAC	CAT	TAA
Fast	CAC	CAT	CAC	CAT	TAA
Cbf	CAC	CAT	CAC	CAT	TAA
Re	CAC	CAT	CAC	CAT	TAA
Slow	CAC	CAT	CAC	CAT	TAA

References:

Acebron SP, Fernandez-Saiz V, Taneva SG, Moro F, Muga A (2008) DnaJ recruits DnaK to protein aggregates. *J Biol Chem* **283:** 1381-1390

Agashe VR, Guha S, Chang HC, Genevaux P, Hayer-Hartl M, Stemp M, Georgopoulos C, Hartl FU, Barral JM (2004) Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* **117:** 199-209

Agris PF (2008) Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* **9:** 629-635

Alexander PA, He Y, Chen Y, Orban J, Bryan PN (2009) A minimal sequence code for switching protein structure and function. In *Proc Natl Acad Sci U S A* **106:** 21149-21154. United States

Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* **181:** 223-230

Angov E (2011) Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnol J* **6:** 650-659

Angov E, Hillier CJ, Kincaid RL, Lyon JA (2008) Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS One* **3:** e2189

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (2003) *Current Protocols in Molecular Biology*, New York: John Wiley & Sons, Inc.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289:** 905-920

Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in Escherichia coli. *Nat Biotechnol* **22:** 1399-1408

Bashan A, Yonath A (2008) Correlating ribosome function with high-resolution structures. *Trends Microbiol* **16:** 326-335

Bennetzen JL, Hall BD (1982) Codon selection in yeast. *J Biol Chem* **257:** 3026-3031

Bonekamp F, Dalboge H, Christensen T, Jensen KF (1989) Translation rates of individual codons are not correlated with tRNA abundances or with frequencies of utilization in Escherichia coli. *J Bacteriol* **171:** 5812-5816

Boycheva S, Chkodrov G, Ivanov I (2003) Codon pairs in the genome of Escherichia coli. *Bioinformatics* **19:** 987-998

Bradley P, Misura KM, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. *Science* **309:** 1868-1871

Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Neidhart FC (ed), pp 1553-1569. Washington, D. C.: ASM Press

Bukau B, Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* **92:** 351-366

Bukau B, Weissman J, Horwich A (2006) Molecular chaperones and protein quality control. *Cell* **125:** 443-451

Chan PP, Lowe TM (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37:** D93-97

Chang HC, Kaiser CM, Hartl FU, Barral JM (2005) De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria. *J Mol Biol* **353:** 397-409

Chu D, Barnes DJ, von der Haar T (2011) The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae. *Nucleic Acids Res* **39:** 6705-6714

Clarke TFt, Clark PL (2008) Rare codons cluster. *PLoS One* **3:** e3412

Crameri A, Whitehorn EA, Tate E, Stemmer WP (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* **14:** 315-319

Crick FH (1966) Codon--anticodon pairing: the wobble hypothesis. *J Mol Biol* **19:** 548-555

Curran JF, Yarus M (1989) Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J Mol Biol* **209:** 65-77

Deane CM, Saunders R (2011) The imprint of codons on protein structure. *Biotechnol J* **6:** 641-649

Dingermann T (2008) Recombinant therapeutic proteins: production platforms and challenges. *Biotechnol J* **3:** 90-97

Dittmar KA, Goodenbour JM, Pan T (2006) Tissue-specific differences in human transfer RNA expression. *PLoS Genet* **2:** e221

Dong H, Nilsson L, Kurland CG (1996) Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J Mol Biol* **260:** 649-663

Ehrenberg M, Kurland CG (1984) Costs of accuracy determined by a maximal growth rate constraint. *Q Rev Biophys* **17:** 45-82

Ellis RJ, Minton AP (2006) Protein aggregation in crowded environments. *Biol Chem* **387:** 485-497

Endoh T, Kawasaki Y, Sugimoto N (2012) Synchronized Translation for Detection of Temporal Stalling of Ribosome during Single-Turnover Translation. *Anal Chem* **84:** 857-861

Ferbitz L, Maier T, Patzelt H, Bukau B, Deuerling E, Ban N (2004) Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431:** 590-596

Fluitt A, Pienaar E, Viljoen H (2007) Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput Biol Chem* **31:** 335-346

Frydman J (2001) Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* **70:** 603-647

Gautschi M, Mun A, Ross S, Rospert S (2002) A functional chaperone triad on the yeast ribosome. *Proc Natl Acad Sci U S A* **99:** 4209-4214

Grantham R, Gautier C, Gouy M, Mercier R, Pave A (1980) Codon catalog usage and the genome hypothesis. *Nucleic Acids Res* **8:** r49-r62

Gromadski KB, Rodnina MV (2004) Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol Cell* **13:** 191-200

Grosjean H, de Crecy-Lagard V, Marck C (2010) Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett* **584:** 252-264

Gu J, Hilser VJ (2008) Predicting the energetics of conformational fluctuations in proteins from sequence: a strategy for profiling the proteome. *Structure* **16:** 1627-1637

Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177:** 4121-4130

Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295:** 1852-1858

Hartl FU, Hayer-Hartl M (2009) Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16:** 574-581

Hasegawa M, Miyata T (1980) On the antisymmetry of the amino acid code table. *Orig Life* **10:** 265-270

Hesterkamp T, Bukau B (1996) The Escherichia coli trigger factor. *FEBS Lett* **389:** 32-34

Heyman T, Agoutin B, Fix C, Dirheimer G, Keith G (1994) Yeast serine isoacceptor tRNAs: variations of their content as a function of growth conditions and primary structure of the minor tRNA(Ser)GCU. *FEBS Lett* **347:** 143-146

Hilser VJ, Freire E (1996) Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. *J Mol Biol* **262:** 756-772

Ikemura T (1981) Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* **146:** 1-21

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* **2:** 13-34

Itano HA (1968a) The structure-rate hypothesis and the toll bridge analogy. In *Structural Chemistry and Molecular Biology*, Rich A, Davidson N (eds), pp 275-280. San Francisco: W. H. Freeman and Company

Johansson M, Bouakaz E, Lovmar M, Ehrenberg M (2008) The kinetics of ribosomal peptidyl transfer revisited. *Mol Cell* **30:** 589-598

Kaiser CM, Chang HC, Agashe VR, Lakshmipathy SK, Etchells SA, Hayer-Hartl M, Hartl FU, Barral JM (2006) Real-time observation of trigger factor function on translating ribosomes. *Nature* **444:** 455-460

Kanaya S, Yamada Y, Kudo Y, Ikemura T (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene

expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238:** 143-155

Kelkar DA, Khushoo A, Yang Z, Skach WR (2012) Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate. In *J Biol Chem* **287:** 2568-2578. United States

Kent KP, Childs W, Boxer SG (2008) Deconstructing green fluorescent protein. *J Am Chem Soc* **130:** 9664-9665

Kimchi-Sarfaty C, Oh JM, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM (2007) A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* **315:** 525-528

Koonin EV, Novozhilov AS (2009) Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* **61:** 99-111

Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361:** 13-37

Kramer G, Boehringer D, Ban N, Bukau B (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol* **16:** 589-597

Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324:** 255-258

Kurland CG, Hughes D, Ehrenberg M (1996) Limitations of translational accuracy. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Neidhart FC (ed), pp 979-1004. Washington, D. C.: ASM Press

Lee TH, Blanchard SC, Kim HD, Puglisi JD, Chu S (2007) The role of fluctuations in tRNA selection by the ribosome. *Proc Natl Acad Sci U S A* **104:** 13661-13665

Li GW, Oh E, Weissman JS (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484:** 538-541

Liang ST, Xu YC, Dennis P, Bremer H (2000) mRNA composition and control of bacterial gene expression. *J Bacteriol* **182:** 3037-3044

Liberek K, Marszalek J, Ang D, Georgopoulos C, Zylicz M (1991) Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. *Proc Natl Acad Sci U S A* **88:** 2874-2878

Marin M (2008) Folding at the rhythm of the rare codon beat. *Biotechnol J* **3:** 1047-1057

Mathews MB, Sonenberg N, Hershey JWB (2000) Origins and principles of translational control. In *Translational control of gene expression*, Sonenberg N, Hershey JWB, Mathews MB (eds), pp 1-31. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28:** 292

Naughton MA, Dintzis HM (1962) Sequential biosynthesis of the peptide chains of hemoglobin. *Proc Natl Acad Sci U S A* **48:** 1822-1830

Nirenberg M, Caskey T, Marshall R, Brimacombe R, Kellogg D, Doctor B, Hatfield D, Levin J, Rottman F, Pestka S, Wilcox M, Anderson F (1966) The RNA code and protein synthesis. *Cold Spring Harb Symp Quant Biol* **31:** 11-24

Oh E, Becker AH, Sandikci A, Huber D, Chaba R, Gloge F, Nichols RJ, Typas A, Gross CA, Kramer G, Weissman JS, Bukau B (2011) Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. In *Cell* **147:** 1295-1308. United States.

Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T (2009) tRNA over-expression in breast cancer and functional consequences. *Nucleic Acids Res* **37:** 7268-7280

Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24:** 79-88

Pedersen S (1984) Escherichia coli ribosomes translate in vivo with variable rate. *Embo J* **3:** 2895-2898

Percudani R, Pavesi A, Ottonello S (1997) Transfer RNA gene redundancy and translational selection in Saccharomyces cerevisiae. *J Mol Biol* **268:** 322-330

Perriere G, Thioulouse J (2002) Use and misuse of correspondence analysis in codon usage studies. *Nucleic Acids Res* **30:** 4548-4555

Plant EP, Nguyen P, Russ JR, Pittman YR, Nguyen T, Quesinberry JT, Kinzy TG, Dinman JD (2007) Differentiating between near- and non-cognate codons in Saccharomyces cerevisiae. *PLoS One* **2:** e517

Press WHea. (1992) Numerical Recipes in C: the art of scientific computing. Cambridge University Press, New York.

Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland JE, Ruef BJ, Hart E, Suner MM, Landrum MJ, Aken B, Ayling S, Baertsch R, Fernandez-Banet J, Cherry JL, Curwen V, Dicuccio M, Kellis M, Lee J, Lin MF, Schuster M, Shkeda A, Amid C, Brown G, Dukhanina O, Frankish A, Hart J, Maidak BL, Mudge J, Murphy MR, Murphy T, Rajan J, Rajput B, Riddick LD, Snow C, Steward C, Webb D, Weber JA, Wilming L, Wu W, Birney E, Haussler D, Hubbard T, Ostell J, Durbin R, Lipman D (2009) The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. In *Genome Res* **19:** 1316-1323. United States

Purvis IJ, Bettany AJ, Santiago TC, Coggins JR, Duncan K, Eason R, Brown AJ (1987) The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *J Mol Biol* **193:** 413-417

Saunders R, Deane CM (2010) Synonymous codon usage influences the local protein structure observed. In *Nucleic Acids Res* **38:** 6719-6728. England

Schmeing TM, Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461:** 1234-1242

Sharp PM, Li WH (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15:** 1281-1295

Sharp PM, Tuohy TM, Mosurski KR (1986) Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* **14:** 5125-5143

Siller E, DeZwaan DC, Anderson JF, Freeman BC, Barral JM (2010) Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J Mol Biol* **396:** 1310-1318

Sorensen MA, Kurland CG, Pedersen S (1989) Codon usage determines translation rate in Escherichia coli. *J Mol Biol* **207:** 365-377

Sorensen MA, Pedersen S (1991) Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J Mol Biol* **222:** 265-280

Sorensen MA, Pedersen S (1998) Determination of the peptide elongation rate in vivo. *Methods Mol Biol* **77:** 129-142

Spencer PS, Barral JM (2012) Genetic code redundancy and its influence on the encoded polypeptides. *Computational and Structural Biotechnology Journal* **1:** e201204006. doi: http://dx.doi.org/201204010.201205936/csbj.201204006

Spencer PS, Siller E, Anderson JF, Barral JM (2012) Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *J Mol Biol* **422:** 328-335.

Spiess C, Meyer AS, Reissmann S, Frydman J (2004) Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol* **14:** 598-604

Stadler M, Fire A (2011) Wobble base-pairing slows in vivo translation elongation in metazoans. *Rna* **17:** 2063-2073

Steitz TA (2008) A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* **9:** 242-253

Suzuki H, Brown CJ, Forney LJ, Top EM (2008) Comparison of correspondence analysis methods for synonymous codon usage in bacteria. *DNA Res* **15:** 357-365

Takyar S, Hickerson RP, Noller HF (2005) mRNA helicase activity of the ribosome. *Cell* **120:** 49-58

Tsien RY (1998) The green fluorescent protein. *Annu Rev Biochem* **67:** 509-544

Uemura S, Aitken CE, Korlach J, Flusberg BA, Turner SW, Puglisi JD (2010) Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464:** 1012-1017

Varenne S, Buc J, Lloubes R, Lazdunski C (1984) Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J Mol Biol* **180:** 549-576

Welch M, Villalobos A, Gustafsson C, Minshull J (2009) You're one in a googol: optimizing genes for protein expression. *J R Soc Interface* **6 Suppl 4:** S467-476

Woese CR (1965) On the evolution of the genetic code. *Proc Natl Acad Sci U S A* **54:** 1546-1552

Woese CR, Dugre DH, Dugre SA, Kondo M, Saxinger WC (1966) On the fundamental nature and evolution of the genetic code. *Cold Spring Harb Symp Quant Biol* **31:** 723-736

Wrabl J, Spencer, P., Hilser, V., Barral, J. Unpublished data.

Wrabl JO, Larson SA, Hilser VJ (2002) Thermodynamic environments in proteins: fundamental determinants of fold specificity. *Protein Sci* **11:** 1945-1957

Young JC (2010) Mechanisms of the Hsp70 chaperone system. *Biochem Cell Biol* **88:** 291-300

VITA

Paige Sanders Spencer was born on August 8, 1985 in Sulphur, Louisiana to Jeffrey and Cheryl Sanders. She attended Hackberry High School from elementary school to high school until she graduated in 2003. She then attended McNeese State University where she earned a Bachelor of Science degree in Chemistry with a concentration in Biochemistry. Paige received several departmental awards for academic excellence. Additionally, she was awarded the Gladys Anderson Emerson Award, a nationally competitive award for female undergraduate students exemplifying excellence in Chemistry. The awarding society, Iota Sigma Pi, National Honor Society for Women in Chemistry, recognized Paige's academic, leadership, and research achievements. She conducted research under the tutelage of Dr. Mark E. Merchant at McNeese where she studied the mechanisms of the robust innate immune system of the American alligator, *Alligator mississippiensis*. Her work resulted in numerous presentations and four publications (listed below). While attending McNeese, Paige's teaching experience included tutoring and working as a teaching assistant for general chemistry lecture and labs and biochemistry lab. In addition, she served as president of the school's student affiliate chapter of the American Chemical Society for two years. Paige first joined UTMB in 2006 as a summer undergraduate research program (SURP) student where she worked with Dr. Todd Pappas in the Center for Biomedical Engineering. The goal of the project was to make nanoparticle thin films biologically compatible with neuroblastoma cells to ultimately be used as a novel mode of drug delivery. She presented this work at the UTMB SURP poster session where she was honored with an award for her outstanding poster presentation. In 2008, Paige joined UTMB Graduate School of Biomedical Science. Due to her interest in protein folding, she joined the lab of Dr. José M. Barral, as well as the Biochemistry and Molecular Biology program. Her research has

consisted of studying the effects of silent mutations on translation speed and protein folding. This work has resulted in two first author publications (listed below). Her research article, "Silent substitutions predictably alter translation elongation rates and protein folding efficiencies", published in *Journal of Molecular Biology*, has gained the interest of field experts and has prompted an expert commentary. Signifying the quality and impact of her findings, this article has been chosen to represent the cover of that journal issue. She has presented this work numerous times at various local conferences and once at an internationally recognized Cold Spring Harbor Meeting where audience enthusiasm was readily evident. She has been recognized by UTMB Neuroscience and Cell Biology (NCB) Department for an outstanding podium presentation of her research. Paige's outstanding research contributions in the area of Biochemistry and Molecular Biology were recognized by the graduate school with the awarding of the Marianne Blum PhD endowed scholarship. At UTMB, she has taught Introduction to Biological Systems, a course in the graduate school's Basic Biomedical Science Curriculum (BBSC). She has also lectured on the topic of Protein Biogenesis for the NCB Advanced Cell Biology course and has facilitated small group discussions for the BBSC Biochemistry course. Paige has also served as Secretary and Vice-Chair of UTMB Biological Chemistry Student Organization (BCSO).

HONORS

2004	**Chemistry Excellence Award**. MSU, Lake Charles, Louisiana.
2005	**Merck/AAAS Undergraduate Research Scholarship**. MSU, Lake Charles, Louisiana
2005	**Chemistry Excellence Award**. MSU, Lake Charles, Louisiana.
2006	**Chemistry Excellence Award**. MSU, Lake Charles, Louisiana.
2006	**Gladys Anderson Emerson Award** for excellence in Chemistry and Biochemistry. Iota Sigma Pi, National Honor Society for Women in Chemistry.
2006	**Summer Undergraduate Research Program**.UTMB, Galveston, Texas
2006	**Outstanding Poster Presentation Award**. Center for Addiction Research, UTMB, Galveston, Texas
2011	**Marianne Blum PhD Endowed Scholarship** for outstanding research in the area of Biochemistry and Molecular Biology. UTMB, Galveston, Texas.
2012	**Outstanding Student Podium Presentation Award.** Neuroscience and Cell Biology Department, UTMB, Galveston, TX

ABSTRACTS

"A Rapid and Inexpensive Method for the Spectroscopic Determination of Innate Immune Activity of Crocodilians". 80th Annual Louisiana Academy of Sciences: Poster Presentation. March 10, 2006. University of Louisiana-Lafayette, Lafayette, Louisiana.

"The role of integrin alpha 1 beta 1 in the biocompatibility of semiconductor nanoparticle thin films" 15th Annual Summer Undergraduate Research Program: Poster Presentation. August 11, 2006. University of Texas Medical Branch, Galveston, Texas.

"Iron withholding as an innate immune mechanism in the American alligator (*Alligator mississippienis*)". 62nd Annual Southwest Regional Meeting of the American Chemical Society: Poster Presentation. October 19-22, 2006. Houston, Texas.

"Iron withholding as an innate immune mechanism in the American alligator (*Alligator mississippienis*)". 81st Annual Louisiana Academy of Sciences: Podium Presentation. March 16, 2007. Southern University and A&M College, Baton Rouge, Louisiana.

"Onset of serum complement activity of American alligator hatchlings in different environments". 82nd Annual Louisiana Academy of Sciences: Podium Presentation.
March 14, 2008. Northwestern State University of Louisiana, Natchitoches, Louisiana.
"Translation Speed and Protein Folding". 2nd Annual Neuroscience and Cell Biology Retreat: Poster Presentation. October 18, 2010. University of Texas Medical Branch, Galveston, Texas.

"Investigating the role of naturally occurring variations in ribosomal elongation rates". 19th Annual Texas Protein Folders Meeting: Poster Presentation. March 26, 2011. Camp Allen, Navasota, Texas.

"Investigating the role of naturally occurring variations in ribosomal elongation rates". 16th Annual Structural Biology Symposium: Poster Presentation. April 8, 2011. Galveston, Texas.

"Silent substitutions predictably alter polypeptide elongation rates and protein folding efficiencies". 20th Annual Texas Protein Folders Meeting: Poster Presentation. April 14, 2012. Camp Allen, Navasota, Texas.

"Silent substitutions predictably alter polypeptide elongation rates and protein folding efficiencies". 17th Annual Structural Biology Symposium: Poster Presentation. April 27, 2012. Galveston, Texas.

"Silent substitutions predictably alter polypeptide elongation rates and protein folding efficiencies". Molecular Chaperones and Stress Responses: Poster Presentation. May 1-5, 2012. Cold Spring Harbor Laboratory, Syosset, New York.

"Silent nucleotide substitutions affect the encoded polypeptide". Annual Neuroscience and Cell Biology Research Retreat: Podium Presentation. June 22, 2012. University of Texas Medical Branch, Galveston, Texas.

PUBLICATIONS

Merchant, M.; Hammock,T.; Sanders, P. and Dronette, J. A rapid and inexpensive method for the spectroscopic determination of innate immune activity of Crocodilians. ***Spectroscopy Letters***. 39:337-343; 2006.

Merchant, M.; Sanders, P.; Dronette, J.; Mills, K. and Berkin, J. Iron withholding as an innate immune mechanism in the American alligator (*Alligator mississippienis*). ***J. Exp. Zool. A. Ecol. Genet. Physiol.*** 307:406-410; 2007.

Merchant, M.; Kinney, C. and Sanders, P. Differential protein expression in alligator leukocytes in response in bacterial lipopolysaccharide injection. ***Comp Biochem Physiol Part D Genomics Proteomics.*** 4:300-304; 2009.

Maccha, V.; Sanders, P. and Merchant, M. Effects of Leukocyte Extract from the American Alligator (*Alligator mississippiensis*) on Antibiotic-Resistant Bacteria. ***The Open Zoology Journal.*** 3: 60-64; 2010.

Spencer, P.S. and Barral, JM. Genetic code redundancy and its influence on the encoded polypeptides. ***Computational and Structural Biotechnology Journal.*** 1 (1): e201204006. doi: http://dx.doi.org/10.5936/csbj.201204006; 2012

Spencer, P.S.; Siller, E.; Anderson, J. and Barral, J.M. Silent substitutions predictably alter polypeptide elongation rates and protein folding efficiencies. ***Journal of Molecular Biology*** (2012), doi: 10.1016/j.jmb.2012.06.010

IN PREPARATION:

Spencer, P.S. and Barral, J.M. Polypeptide elongation rates influence chaperone recruitment to the nascent polypeptide.

Spencer, P.S. and Barral, J.M. "Silent nucleotide substitutions affect the encoded polypeptide" New York: Springer Science